Cynthia V. Rider · Jane Ellen Simmons
Editors

Chemical Mixtures and Combined Chemical and Nonchemical Stressors

Exposure, Toxicity, Analysis, and Risk

Springer

Chemical Mixtures and Combined Chemical and Nonchemical Stressors

Cynthia V. Rider · Jane Ellen Simmons
Editors

Chemical Mixtures and Combined Chemical and Nonchemical Stressors

Exposure, Toxicity, Analysis, and Risk

Foreword by Linda S. Birnbaum

Springer

Editors
Cynthia V. Rider
National Toxicology Program
National Institute of Environmental
Health Sciences
Research Triangle Park
NC, USA

Jane Ellen Simmons
National Health and Environmental Effects
Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park
NC, USA

ISBN 978-3-319-56232-2 ISBN 978-3-319-56234-6 (eBook)
https://doi.org/10.1007/978-3-319-56234-6

Library of Congress Control Number: 2018930751

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To all who came before, paving the way, to those currently engaged in the active practice of mixtures research – from unraveling complex exposures to performing toxicological evaluations of mixtures and assessing cumulative risk. In particular, we recognize George Alexeeff, who dedicated himself to advancing public health and understanding cumulative health impacts, with particular attention to vulnerable populations overburdened by multiple sources of pollution.

Foreword

Mixtures are our everyday reality. We are exposed to numerous chemicals throughout our lifetimes from various sources in our environment – personal care products, food and water contaminants, occupational exposures, traffic pollution, molds and allergens, pesticides, pharmaceuticals, and too many others to list. These external exposures are influenced by our internal milieu, which reflects background genetics and acquired epigenetic changes, as well as a host of nonchemical environmental factors (e.g., microbiome, psychosocial stressors, disease states, nutritional status). Considering this complex and dynamic exposure scenario, it has long been recognized that evaluating exposures and their effects on a chemical-by-chemical basis is not adequate for protecting public health. However, there has not been a clear path forward for changing the paradigm, and the complexities involved in mixtures research and risk assessment have often been used as justification for perpetuating the standard approach of assessing one chemical or one exposure at a time. Despite the challenges, many researchers across diverse fields of science have been actively engaged in the study of mixtures. Through these efforts, we have gained a significant understanding of the key issues in mixtures science and developed many approaches for addressing these issues. In this book, insights from leading-edge researchers and analysts have been pulled together to present a comprehensive picture of the current state of mixtures science and provide tools for practitioners engaged in assessment of risk from exposure to mixtures.

My interest in mixtures has spanned the breadth of my career. Through my own research program and as director of the National Institute of Environmental Health Sciences (NIEHS) and the National Toxicology Program (NTP), I have had the opportunity to be a part of the mixtures research story. One of my early interests was in mixtures of dioxin-like chemicals. Work from my lab and others contributed to some of the first efforts to account for the cumulative risk associated with exposure to mixtures. Dioxins, typified by the reference contaminant 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), represent persistent organic pollutants that are highly toxic and exist in complex mixtures. Dioxins accumulate in the food chain, and people are exposed today mainly through food consumption. The toxic

equivalency factor approach, developed to sum the total burden of dioxin-like chemicals, was a pioneering effort to move beyond single chemical analyses of risk and account for the cumulative effects of dioxin-containing mixtures. Lessons learned through that effort have been applied to many other classes of environmental contaminants. I have since been involved in mixtures research touching upon endocrine disruptors, flame retardants, organophosphates, and perfluorinated compounds, among others.

During my tenure as director of NIEHS, I have had the pleasure of seeing mixtures research elevated through numerous workshops and NIEHS-wide research efforts. In 2011, NIEHS hosted a meeting titled "Advancing Research on Mixtures: New Perspectives and Approaches for Predicting Adverse Human Health Effects." This workshop brought together mixtures experts from exposure science, toxicology, epidemiology, statistics, and risk assessment to outline challenges in mixtures research and discuss approaches to address those challenges. More recently, NIEHS organized a workshop on "Statistical Approaches for Assessing Health Effects of Environmental Chemical Mixtures in Epidemiology Studies." Development of statistical methods for analysis of mixtures in epidemiological studies was an area specifically identified in the 2011 workshop as requiring research attention. During this innovative workshop, participants were given multiple epidemiological datasets and asked to apply their analysis methods, which were then compared. In addition to workshops, the NIEHS has also demonstrated a commitment to mixtures research by including "Understand how combined environmental exposures affect disease pathogenesis" as Goal #4 in the 2012–2017 NIEHS Strategic Plan, which identifies priority areas of research. This goal includes assessing the joint action of multiple environmental factors, including both chemical and nonchemical stressors. Finally, numerous projects led by NIEHS scientists and grantees are dedicated to better understanding the potential health effects of exposure to mixtures. Projects supported by NIEHS range widely from defining the totality of human exposure through research into the exposome to targeted projects that address the toxicity of specific complex mixtures, such as toxicity testing of botanical dietary supplements at the NTP.

The future of mixtures research is bright. Mixtures research has moved beyond simply combining chemicals to look for greater than additive interactions. Instead, we are using the latest understanding of biological systems to predict how combinations of chemicals and nonchemical factors might interact by targeting an adverse outcome pathway. We are developing hypotheses of combined effects and using every tool available to test these hypotheses. Combinations of in silico, in vitro, alternative animal and traditional toxicity studies are being employed to prioritize mixtures for study and to routinely assess both defined and complex mixtures. We are developing more sophisticated methods to analyze "big data" resulting from high-content assays and refining methods to predict mixture effects. All of these efforts inform risk assessment efforts that are increasingly expanding beyond single chemicals to address cumulative and community-specific risks. As we move forward with mixtures research, we are critically evaluating findings in the context of our historical knowledge of mixtures.

This book is an excellent example of the type of thoughtful collaboration that is required to understand the human health consequences of a life lived, beginning before conception, in a soup of chemical and nonchemical stressors. The editors and the authors collaborated on producing a book that stands in sharp contrast to most multiauthored books. The authors agreed to use a common set of definitions and terminology, greatly enhancing the ability of the reader to move between chapters and sections. Conference calls were held at the request of various writing teams with other writing teams. Further, the authors shared draft chapters within and between sections to ensure continuity and lack of duplication. The book and the reader directly benefit from the intense effort required to accomplish this level of integration. The book loosely follows the risk assessment paradigm (exposure, hazard identification, risk characterization), providing the reader with essential information and tools. It also highlights some recent advances in predicting co-occurrence, using the new adverse outcome pathway concept to group chemicals and to identify the appropriate risk assessment strategy, and environment-wide association studies (EWAS) for identification of the effect drivers within complex exposures. The book concludes with suggested approaches for incorporating nonchemical stressors into cumulative risk assessment. This book provides a sound foundation for anyone engaging in some aspect of consideration of chemical mixtures and chemical and nonchemical stressors and their impact on human health.

National Institute of Environmental Health Sciences and National Toxicology Program,
Research Triangle Park, NC, USA

Linda S. Birnbaum

Acknowledgments

We sincerely thank all of the authors for their time and energy in contributing to this book. We would also like to thank the Springer editorial team. Jane Ellen appreciates the support of Roel and Johanna during this process. Cynthia thanks Marcus, Vaughn, and Laidy for their encouragement along the way.

Contents

Chapter 1
Introduction

Cynthia V. Rider and Jane Ellen Simmons

Abstract All people are exposed to complex and dynamic mixtures of chemicals and nonchemical stressors throughout their lives. Understanding how these combined exposures impact human health is an active area of research spanning exposure science, toxicology, epidemiology, statistics, and risk analysis. Mixtures under study range from simple combinations of chemicals to the complete exposure profile known as the exposome. Research efforts to explore mixtures have used individual chemical data to estimate mixture effects in a bottom-up approach and have evaluated the effects of whole mixtures in a top-down approach. Considering the numerous perspectives and approaches, mixture terminology has been particularly challenging. In this introductory chapter, we lay the groundwork for the book by providing a rationale for the study of mixtures, defining important terms, and describing the flow of the book.

Keywords Mixtures · Nonchemical stressors · Combined exposures · Mixture terminology

Disclaimer The views expressed in this chapter are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

C. V. Rider (✉)
National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA
e-mail: cynthia.rider@nih.gov

J. E. Simmons
National Health and Environmental Effects Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA
e-mail: Simmons.Jane@epa.gov

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_1

1.1 To Study Mixtures or Not to Study Mixtures, That Is the Question

The study of mixtures is often paradoxically referred to as essential and as prohibitively complicated. Essential, because it is widely accepted that humans are not exposed to a single chemical throughout their lifetimes, but to a complex milieu of chemicals that differ circumstantially and across time (Pohl and Abadin 2008). Therefore, efforts to better understand how public health and ecological systems are affected by their ambient conditions must include consideration of mixtures (Sexton 2012). Prohibitively complicated, because as researchers move toward addressing more environmentally relevant mixtures, uncertainty increases in terms of study design, interpretation of results, and translation into risk estimates (Fig. 1.1). This sentiment is captured in a 2004 paper titled "Chemical Mixtures: An Unsolvable Riddle?" which details the increased uncertainty in assessing the risk of mixtures as compared to single chemicals (Borgert 2004). These two opposing viewpoints (i.e., essentiality versus prohibitive complexity) have provided the backdrop for decisions on whether to continue on a path of assessing chemicals individually or to move toward more standard consideration of mixtures. Fortunately, many investigators have braved the various obstacles and contributed to the significant body of mixtures research that now exists. It is important to note that the study of mixtures has included effort directed at building on knowledge from the study of individual chemicals and simple mixtures and developing new ways to interpret information from the study of complex mixtures. In other words, mixture studies have included bottom-up approaches to understand how components of a mixture contribute to mixture effects and top-down approaches involving assessment of complex mixtures and developing methods to interpret and apply information gained.

1.1.1 Mixtures Are Reality

According to the Environmental Protection Agency's Toxic Substances Control Act (TSCA) Inventory, there are approximately 85,000 chemical substances manufactured or processed in the United States (U.S. EPA 2017a). People are constantly exposed to dynamic mixtures of chemicals, through the air we breathe; our diets; use of personal care and household products; pharmaceutical intake;

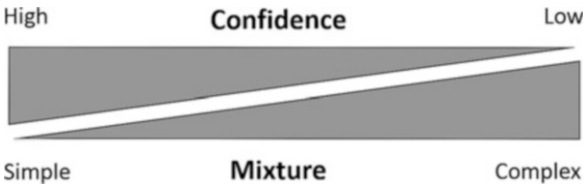

Fig. 1.1 As the complexity of the study mixture increases, confidence in associated data and interpretation decreases

occupational, recreational, accidental, and intentional exposures; etc. It is clear from biomonitoring efforts such as the National Health and Nutrition Examination Survey (NHANES), which measures over 200 chemicals in human samples, that chemicals are making their way from the environment into our bodies (CDC 2015). Although chemicals have been the focus of the majority of mixtures research, many of the concepts also apply to nonchemical factors. There are many pathways for nonchemical factors to influence toxicity outcomes.

Consideration of the many nonchemical factors that can potentially affect human health or modify our response to chemical exposures moves us closer to the real-world scenario. People are exposed to various nonchemical factors that can include physical stressors (e.g., heat, radiation, allergens, and noise) as well as psychosocial stressors (e.g., circumstances or events that elicit an acute or chronic stress response). Furthermore, there are some factors that do not fit neatly into categories. For example, over-consumption of an essential trace mineral (e.g., manganese) could be considered as a chemical exposure with the potential to elicit toxicity, while a trace mineral deficiency could be considered as a nonchemical factor that could also negatively affect health but with a different constellation of effects.

People are exposed to different mixtures throughout their lifetime, and exposure profiles can change drastically based on behavior as well as age. For example, crawling and hand-to-mouth behaviors in an infant may result in relatively high exposure to house dust and chemicals attached to dust, as compared to adults (Pohl and Abadin 2008). Product use and diet can change over time, as can surroundings. Whereas young children may spend a significant portion of their time at home, outside, or in a daycare setting, adults may be exposed to very different mixtures depending on their occupation. In addition to changing exposures over time, there are also differences in the responses of individuals to exposure that can depend on genetic background, life stage, or disease state.

1.1.2 What Mixtures Are Being Studied?

Accepting that there are an infinite number of potential combined exposures for study, it is important to define the term "mixture" before discussing the associated science. Although the majority of this book addresses mixtures of chemicals, combined exposures to any factors (chemical, physical, or psychosocial) to which a human, animal, or cell are exposed, either concurrently or separated in time, may also be relevant determinants of outcome. Mixtures span a wide range of complexity. The following is a presentation of common terms and definitions used to describe mixtures. It is not meant to be exhaustive, but to provide illustrative examples of the types of mixtures that are the focus of research attention.

Mixture Types The term *binary mixtures* refers to combinations of two factors and represent the simplest possible mixtures. Good examples of binary mixtures can be found in the pharmacology field, where binary combinations of drugs are either recommended for increased efficacy/selectivity (e.g., combination

chemotherapy) or discouraged based on the potential for increased toxicity (e.g., many drugs are counter-indicated for people taking the anticoagulant warfarin). A specific example of a well-studied binary mixture can be found in combined use of alcohol and acetaminophen – a combination that can lead to increased hepatotoxicity as compared to use of one or the other alone. Typically, binary mixtures are the subject of focused studies to assess interactions among select factors. Ternary and quaternary mixtures can be considered as incrementally more complex versions of this simple binary mixture category.

Following binary mixtures, the next level of complexity can be captured by the term *defined mixtures* (which subsumes binary, ternary, and quaternary combinations). In defined mixtures, all of the constituents and their concentrations are known. Many studies with defined mixtures have been conducted to assess the performance of predictive models of mixture toxicity based on single chemical data (i.e., component-based approaches) (Kortenkamp 2007; Howdeshell et al. 2017). Defined mixtures are also studied under the assumption that complex mixture toxicity can be estimated by evaluating a subset of known active constituents. For example, research efforts can focus on a subset of dioxin-like chemicals that have demonstrated binding to the aryl hydrocarbon receptor, while the actual mixture to which people are exposed could contain a more complex suite of structurally diverse chemicals. Alternatively, exposure data can be used to identify a subset of chemicals from which to build defined mixtures for study. Although defined mixtures do not recapitulate the complexity of real-world exposures, they offer an intermediate step in understanding the behavior of chemicals acting jointly.

Moving beyond defined mixtures are different types of *whole mixtures* or *complex mixtures*. Recently, the more descriptive term to categorize these mixtures, "chemical substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials," abbreviated as UVCB, has been used by the EPA in the Toxic Substances Control Act (TSCA) Chemical Substance Inventory (U.S. EPA 2017b) and by the European Chemicals Agency (ECHA) (ECHA 2017). Complex mixtures can be further subcategorized into *formulations* and *environmental mixtures*. Formulations are whole mixtures that contain one or more purported active chemical(s) and a number of potentially inert ingredients. Although it is possible for formulations to be defined mixtures, they are categorized as whole mixtures here, because many of their ingredients can be proprietary or can be variable in terms of their concentration. Consider personal care products, which may list "fragrance" as an ingredient, or commercial formulations of flame retardants, such as Firemaster 550, which include a proprietary mixture of different active classes of chemicals (e.g., brominated and non-halogen flame retardants). Although the active constituent(s) are often identified in formulations, they are often unknown in environmental mixtures. These mixtures include samples taken from hazardous waste sites (e.g., Superfund sites), spills or catastrophic events (e.g., Elk River chemical spill, Gulf oil spill, Fukushima disaster), or those created by intentional processes (e.g., drinking water disinfection, gas extraction via hydraulic fracturing).

Finally, the most complex and challenging type of mixture to study is that encompassed by the *exposome* concept – all exposures over a lifetime. This concept includes all factors (chemical, physical, biological, psychosocial) that an individual is exposed to from conception to death (Wild 2005). Essentially, the exposome represents the real-world scenario. The difficulties in assessing the exposome are in measuring the totality of exposures and deconvoluting how these exposures collectively affect health. It is important to note that to develop effective interventions, the drivers of harm must be identified.

In addition to defining the types of mixtures that are typically the focus of research, there is also a need to define the relevant terms that have been used in mixture science. As discussed below, terminology has been particularly confusing in the field of mixtures, with multiple meanings associated with a single term and meanings differing between disciplines. The next section does not offer an official consensus on terminology, but instead attempts to clarify meanings for the terms as they are used in this volume.

1.1.3 Additional Mixture Terminology

The topic of mixtures has been of interest to many diverse fields, including ecology, epidemiology, exposure science, pharmacology, risk analysis science, statistics, and toxicology. Scientists in these fields approach mixtures from perspectives informed by their expertise and apply specialized terminology to address mixture questions most relevant to their discipline. While this multidisciplinary treatment is critical to understanding environmental and human health effects resulting from exposure to mixtures, it has also led to confusion over terminology.

A list of common terms and definitions is provided below to clearly articulate the interpretation of each term applied within this text. Throughout the book, the terms will be used as defined below with exceptions explicitly noted by chapter authors.

1.1.3.1 Terms to Identify the Combined Exposure or Mixture of Interest

Combined Exposure This represents a broad term that encompasses any combination of two or more exposures or factors of interest. Exposures can include chemical and/or nonchemical stressors and can be concurrent or separated in time. Often, this term is used interchangeably with the term "mixture." However, while "combined exposure" can refer to chemical and/or nonchemical exposures, the term "mixture" is limited to chemical exposures in this text.

Mixture A mixture is any combination of two or more chemicals of interest present concurrently or separated in time. See additional definitions of mixture types above. A *simple mixture* is one where the effects of the mixture can be

reasonably estimated based on knowledge of its components. Although the upper limit of a simple mixture has not yet been defined and may depend on the nature of the component chemicals, it is estimated that mixtures between 2 and 25 components would qualify as simple mixtures. A *complex mixture* contains too many components to allow for a reasonable estimate of mixture effects based on component chemicals and usually contains some *unidentified fraction* (portion of a mixture that has not been chemically characterized). In contrast, a *defined mixture* contains chemicals that are identified and present in known quantities (see above for more detailed description).

Chemical and Nonchemical Factors The term "chemical" refers to any natural or anthropogenic substance to which a person or animal can be exposed. Unless noted, this does not include endogenous chemicals (e.g., natural hormones produced in the body). Chemicals are differentiated from nonchemical factors, which can be subdivided into *physical* and *psychosocial stressors*. Physical stressors are defined here as biological agents (e.g., viruses) or external forces (e.g., radiation, noise) that can modify exposure or elicit a physiological response from the exposed organism. Psychosocial stressors are defined as factors in the external environment that are perceived to be harmful (e.g., fear of violence), which can result in physiological changes (e.g., increased production of cortisol).

1.1.3.2 Terms to Describe the Joint Action of Exposures

Dose Addition Dose addition and *concentration addition* are essentially the same concept, with the term "dose" applying where appropriate (e.g., oral gavage to rodents, ingestion of pharmaceuticals) and the term "concentration" applying where appropriate (e.g., aquatic exposures, cell-based experiments). Under dose addition, the effect (response) of the mixture is predicted by summing the exposure doses of the component chemicals. A key concept is that the doses of the component chemicals are weighted by their toxic potency. In the idealized situation, the component chemicals behave as concentrations or dilutions of each other. Dose addition is thought to be best applied to those chemicals that share a common or similar mode of action or similarity of target organ. Thus, the behavior of a chemical mixture is considered dose additive if the effects of the combined components (i.e., the effect of the mixture) can be estimated from the sum of the scaled doses of the individual components.

Independent Action Under independent action (also called *independent joint action* or *response addition*), the effect (response) of the mixture is predicted by summing the effects (responses) of the component chemicals. A key concept is that the mixture response is predictable by the sum of the responses of the components using the formula for the sum of the probabilities of independent events. Independent action is thought to be best applied to mixtures of chemicals that have dissimilar modes of action; these chemicals are toxicologically independent (i.e., the biological response to each chemical is the same whether or not the other

chemical(s) is present). Thus, the behavior of a chemical mixture is considered to be consistent with independent action if the effects of the combined components (the effect of the mixture) can be estimated from combining the responses of the individual chemicals using an equation to describe the probability of independent events co-occurring. *Effect summation*, which represents a simple summation of component effects, can be viewed as a special case of independent action. Although effect summation is commonly used in the mixture literature, its application should be limited (see Chap. 9).

Interaction The term "interaction" is common in both toxicology and epidemiology and has notably different implications depending on the context (see Chap. 10). Unless otherwise stated, "interaction" will be used to describe a joint action among combined exposures that differs significantly from the clearly stated expectation of additivity (e.g., predicted effects based on dose addition or independent action).

Greater-than-Additive and Less-than-Additive It is recommended that conclusions regarding interactions be drawn as to whether the response of the mixture in question is consistent with a specific definition of additivity as in "no detectable deviation from additivity" or inconsistent with the specific definition, showing "greater-than-additive" or "less-than-additive" responses. The definition of additivity should be specific as to dose addition or independent action with appropriate reference to the underlying literature. It is highly recommended that the use of the terms "synergy" and "antagonism" be avoided due to the vast confusion that has plagued chemical mixture toxicology and risk assessment. In effect, they are problematic because of the many differing definitions of these terms and their widespread use without articulation of the meaning ascribed by the user. To avoid confusion "synergy" is replaced with "greater-than-additive," and "antagonism" is replaced with "less-than-additive." When the term "synergy" cannot be avoided, it should be defined within the context of the definition of additivity being used.

1.1.3.3 Terms to Describe Exposure or Risk

Aggregate Exposure, Aggregate Risk The term "aggregate" is used here to indicate the summing of exposure for an individual chemical across all relevant routes, so that the total dose to the person/animal model can be used to estimate the aggregate risk. For example, in the case of bromodichloromethane, multiple routes of exposure (i.e., oral, inhalation, and dermal) make significant contributions to internal dose and contribute to the aggregate exposure and aggregate risk (Haddad et al. 2006).

Cumulative Exposure, Cumulative Risk The term "cumulative" is used here to indicate consideration of more than one stressor (chemical or nonchemical) in an exposure or risk assessment. This is a general term that can be applied to any exposure characterization or risk assessment that includes multiple factors (e.g.,

cumulative assessment of organophosphate pesticides, community-based assessment involving chemical and nonchemical exposures in a select population) and can be contrasted against single chemical exposure or risk evaluations. *Cumulative is notably distinct from aggregate and should not be used interchangeably. However, an exposure characterization or risk assessment can be both aggregate and cumulative*. It is an umbrella term that does not dictate the specific model used to assess cumulative risk, and concepts of either dose addition or independent action can be used as a basis for the calculation of cumulative risk. It is important to note the distinction between the concepts used to describe joint action (dose addition and independent action) and the methods available for calculation of risk (e.g., hazard index, relative potency factors) that are built upon those concepts.

Exposure/Response Modifier Due to inconsistent definitions of the terms susceptibility, vulnerability, and sensitivity, as well as interchangeable use in the literature, the single term "exposure/response modifier" has been suggested as an umbrella term to capture any condition or state that could alter exposure or response to a chemical or nonchemical stressor or buffer. Although the use of this more generic term is recommended, the terms *susceptibility, vulnerability, and sensitivity* have a long history of use, particularly in epidemiology, and do appear in this volume. Throughout the book, these three terms are used interchangeably and defined as any factor or set of factors that increases the likelihood of harm from an exposure. Examples include low socioeconomic status and other psychosocial stressors (e.g., exposure to violence, lack of access to healthcare) within a population that contribute to decreasing the resiliency reservoir required to maintain health. For more information, see Gee and Payne-Sturges (2004). The traditional use of "genetic susceptibility" refers specifically to genetic variations that increase the likelihood of harm from an exposure. For example, inherited mutations in the tumor suppressor genes *BRCA1* and *BRCA2* predispose women to development of breast cancer (Valencia et al. 2017).

1.2 Challenges in Mixtures Research

As discussed previously, mixtures can be approached from a reductionist (i.e., bottom-up) or holistic (i.e., top-down) perspective. The challenges associated with mixtures research differ depending on which approach is used. In the reductionist approach, the challenges can lie in relating findings from carefully controlled experiments with defined mixtures to the real-world scenario. On the other hand, the holistic approach presents a different set of challenges in understanding how to interpret data and develop targeted interventions based on findings from complex mixtures.

Specific challenges associated with a reductionist approach include deciding on chemicals to incorporate in the defined mixture for study and understanding how the defined mixture fits into the bigger exposure picture. Prioritizing chemicals for

inclusion in mixtures research or cumulative risk assessment can be driven by exposure information or biological considerations (e.g., common toxicity target). The NHANES database has been an excellent source of information for better understanding of co-exposures. In terms of using biological information to prioritize mixtures for study, there has been movement from focusing on isolated targets (e.g., estrogen receptor, liver) toward a systems biology view that considers the complex network of signaling pathways involved in disease manifestation. This evolution has expanded our view on which chemicals and/or factors to include in defined mixture studies.

Challenges in the study of whole mixtures include identifying which chemicals are present in the mixture of concern and which chemicals are responsible for eliciting the observed effects. Identifying the active constituents within a complex mixture can require significant investment in chemical analysis without guaranteed success. Furthermore, efforts to confirm the biological activity of possible active constituents by isolating, identifying, and testing individual chemicals are complicated by the possibility of interactions among constituents. In effect, it is often very difficult to disentangle the problem when the mixture, the biological target, and the interaction of the mixture with the biological target are all complex.

1.3 Progress to Date and Future Directions

Despite the many challenges associated with the study of mixtures, there have been significant advancements over the past century. The first major step was the development of modeling approaches that used single chemical dose-response data to predict mixture effects in the mid-1900s (detailed in Chap. 9). Almost half a century later, these models were further advanced and applied to different types of mixtures. There followed a period defined by the search for unexpected interactions, thereby beginning the controversy over use of the term "synergy." Perhaps a lack of discovery of the "holy grail" (i.e., chemicals that produced thousandfold greater effects when tested in combination versus alone) shifted focus away from the search for such remarkable deviations from additivity toward more sophisticated modeling of predicted mixture responses.

This advancement in modeling of predicted mixture toxicity continues today. Work in this area represents a joint effort between statisticians and toxicologists to design more efficient and appropriate studies to assess mixtures, to develop methods to better fit individual chemical data, and to utilize appropriate statistical methods for comparing predicted responses to observed responses. Increasingly, there has been a recognition that methods to assess mixtures should be fit for purpose, not one size fits all.

Another more recent evolution is from mixtures research focusing exclusively on simple mixtures to an expansion into complex mixtures. Although there is a long history of assessing the toxicity and risk associated with complex mixtures (e.g., diesel exhaust, tobacco smoke) as a single entity, attempts to better understand

complex mixtures and compare across them represent a relatively recent development. There are still only a handful of studies that attempt to assess whether or not complex mixtures are *sufficiently similar*, meaning that data from a well-characterized reference mixture can be applied to a related, untested mixture of interest. This represents an area where more research is needed.

Finally, incorporation of nonchemical stressors, including both physical and psychosocial, into mixtures research and cumulative risk assessments remains an area that requires significant attention. The default has been to apply methods from chemical mixtures to these nonchemical stressors, but more work is needed to validate this application. The development of case studies that incorporate both chemical and nonchemical stressors will offer opportunities to identify data needs and areas that require refinement.

1.4 Flow of Book Sections

In this volume, the current state of the science on mixtures is presented. Beginning with exposure, the measurement of chemicals is addressed, including both internal measures of chemicals in human samples (biomonitoring) (Chap. 2) and external measures of chemicals in the environment (Chap. 3). These chapters on exposure measurement are complemented with a chapter covering modeling approaches to describe exposure (Chap. 4). The set of three chapters provides insight into the characterization of mixtures in the environment, people, and populations.

Following identification of chemicals, a prioritization step is needed to focus attention on mixtures that require research attention. Prioritization can be based on the presence and concentration of chemicals (e.g., high levels measured in an NHANES population), or it can be driven by a particular goal (e.g., occupational risk assessment). The section on prioritization of chemicals focuses on newer approaches that have been developed to intelligently design mixtures for study. These include using statistical tools for determining association of exposure to effect markers in Environment-Wide Association Studies (EWAS; Chap. 5), methods adapted from the field of biogeography to examine chemical co-occurrence (Chap. 6), and building adverse outcome networks using mechanistic data to inform mixture decisions (Chap. 7).

Having identified the mixture(s) of interest, it is then critical to understand how to go about evaluating the mixture. In the section on Mixture Toxicology, there is a thorough coverage of evaluating defined mixtures using component-based approaches. A chapter on dose-response evaluation addresses single chemical data that are used to understand mixtures (Chap. 8). Predicting mixture effects is addressed in multiple chapters, which cover models of additivity conceptually (Chap. 9), contrasting toxicological and epidemiological approaches (Chap. 10), and statistical comparison of observed and predicted mixture responses (Chap. 11). The challenging topic of physiologically based pharmacokinetic modeling of

mixture toxicity is also addressed (Chap. 12). Finally, elements of experimental design in the study of mixtures are examined (Chap. 13).

The risk assessment section covers both component-based (Chap. 14) and whole mixture (Chap. 15) approaches. There are many well-established component-based methods for estimating cumulative risk. Whole mixtures are often handled like single chemicals; however, there is a need for methods to determine sufficient similarity of whole mixtures. Throughout these chapters, areas of uncertainty are highlighted.

In the last section of the book, nonchemical stressors are brought into the picture. Physical stressors (e.g., heat, radiation) are discussed in terms of how to begin to quantify the effects of these stressors and bring them into cumulative risk assessments (Chap. 16). Psychosocial stressors (e.g., fear of violence) represent a challenging area of combined exposure research and cumulative risk evaluation that include the unique aspect of perception (Chap. 17). Finally, many of the concepts presented throughout the book are brought together by providing an example of community-based cumulative impact assessment (Chap. 18).

Acknowledgement This work was supported in part by the NIH, National Institute of Environmental Health Sciences.

References

Borgert, C.J. 2004. Chemical mixtures: An unsolvable riddle? *Human and Ecological Risk Assessment* 10: 619–629.

CDC. 2015. Fourth National Report on Human Exposure to Environmental Chemicals. In*:* CDC (ed.).

ECHA. 2017. *What is a substance?* [Online]. Available: https://echa.europa.eu/support/substance-identification/what-is-a-substance. Accessed 15 Dec 2017.

Gee, G.C., and D.C. Payne-Sturges. 2004. Environmental health disparities: A framework integrating psychosocial and environmental concepts. *Environmental Health Perspectives* 112: 1645–1653.

Haddad, S., G.C. Tardif, and R. Tardif. 2006. Development of physiologically based toxicokinetic models for improving the human indoor exposure assessment to water contaminants: Trichloroethylene and trihalomethanes. *Journal of Toxicology and Environmental Health-Part A-Current Issues* 69: 2095–2136.

Howdeshell, K.L., A.K. Hotchkiss, and L.E. Gray. 2017. Cumulative effects of antiandrogenic chemical mixtures and their relevance to human health risk assessment. *International Journal of Hygiene and Environmental Health* 220: 179–188.

Kortenkamp, A. 2007. Ten years of mixing cocktails: A review of combination effects of endocrine-disrupting chemicals. *Environmental Health Perspectives* 115: 98–105.

Pohl, H.R., and H.G. Abadin. 2008. Chemical mixtures: Evaluation of risk for child-specific exposures in a mufti-stressor environment. *Toxicology and Applied Pharmacology* 233: 116–125.

Sexton, K. 2012. Cumulative risk assessment: An overview of methodological approaches for evaluating combined health effects from exposure to multiple environmental stressors. *International Journal of Environmental Research and Public Health* 9: 370–390.

U.S. EPA. 2017a. *About the TSCA chemical substance inventory* [Online]. U.S. EPA. Accessed 15 Dec 2017.

———. 2017b. *Chemical substances of unknown or variable composition, complex reaction products and biological materials (UVCB Substance) on the TSCA Inventory* [Online]. U.S. EPA. Available: https://www.epa.gov/tsca-inventory/chemical-substances-unknown-or-variable-composition-complex-reaction-products-and. Accessed 15 Dec 2017.

Valencia, O.M., S.E. Samuel, R.K. Viscusi, T.S. Riall, L.A. Neumayer, and H. Aziz. 2017. The role of genetic testing in patients with breast cancer a review. *JAMA Surgery* 152 (6): 589–594.

Wild, C.P. 2005. Complementing the genome with an exposome: The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology, Biomarkers & Prevention* 14: 1847–1850.

Part I
Combined Exposures

Chapter 2
Biomonitoring to Assess Exposures to Mixtures of Environmental Chemicals

Antonia M. Calafat

Abstract In modern societies, humans may be exposed to a wide spectrum of environmental stressors, including mixtures of anthropogenic chemicals. Furthermore, because human exposure does not occur under controlled conditions of dose-response evaluations in animal studies, exposure assessment is complex. Three main tools have been used to assess human exposures: history/questionnaire information, environmental monitoring, and biomonitoring (i.e., measuring concentrations of the chemicals or their metabolites or adducts in human specimens). In this chapter, we will discuss the suitability of biomonitoring data for evaluating exposures to mixtures of environmental chemicals.

Keywords Biomonitoring · Exposure · NHANES · Endocrine disruptors · Environmental chemicals

2.1 Introduction

In the course of their daily routines, humans are exposed to a large number and variety of physical, biological, psychosocial, and chemical stressors. All of these stressors, their timing, and duration along with each person's genetic makeup, diet, and lifestyle can affect human health (Needham et al. 2005a; Birnbaum 2010). Because of the complexity of such exposures and their interactions, understanding the potential effects of the exposures on health requires a multidisciplinary approach—a topic of interest to several scientific fields including, among others, chemistry, ecology, epidemiology, exposure science, pharmacology, risk analysis, statistics, and toxicology (Carlin et al. 2013).

Disclaimer The findings and conclusions in this report are those of the author and do not necessarily represent the official position of the Centers for Disease Control and Prevention (CDC).

A. M. Calafat (✉)
Centers for Disease Control and Prevention, Atlanta, GA, USA
e-mail: aic7@cdc.gov

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_2

Three main tools have been used to assess chemical exposures: history/questionnaire information, environmental monitoring, and biomonitoring (i.e., measuring concentrations of the chemicals or their metabolites or adducts in human specimens) (Calafat et al. 2006; Sexton et al. 2004; Needham et al. 2005b). Exposure models, covered in the last chapter of this section, usually incorporate information from the three approaches. The use of history/questionnaire data to assess human exposure to environmental chemicals falls within the purview of environmental epidemiology. Indirect measures of exposure (e.g., environmental monitoring) are the subject of the second chapter of this section. In this chapter, we will cover the assessment of internal exposures using biomonitoring.

2.2 Biomonitoring Overview

Biomonitoring is the assessment of internal dose (i.e., body burden) by measuring the parent chemical (or its metabolite or reaction product) in human samples. Biomonitoring, a "gold standard for assessing exposure to chemicals" (Sexton et al. 2004), has many potential uses in the public health context of preventing disease related to people's exposure to chemicals. Biomonitoring can be used to detect and monitor chemical exposures, to assess people's health risk as a result of such exposures, to develop and implement interventions to reduce exposures, and to evaluate the effectiveness of those interventions (CDC 2009; National Research Council 2012).

In some cases, evidence of chemical exposures and their human health effects (e.g., lead poisoning) have been known since antiquity (Waldron 1973; U.S. EPA 1985), although the use of biomonitoring to track lead poisoning did not start until the late 1890s with the screening of factory workers' blood and urine (Sexton et al. 2004). Since then and thanks, in part, to access to and availability of sophisticated analytical chemistry techniques, trace levels of lead and many other chemicals in a person's body can be routinely measured with high precision and accuracy (Angerer et al. 2007; Pirkle et al. 1995).

These scientific and technologic advances along with the increase in global production of chemicals and their use in a myriad of industrial and consumer products starting in the twentieth century (UNEP 2013) have contributed to the remarkable growth of human biomonitoring research in the last few decades (Angerer et al. 2006; National Research Council 2006; Needham et al. 2007). For example, biomonitoring concentrations are increasingly used to categorize exposures (e.g., low, medium, high) within populations to assess internal exposure to environmental chemicals (National Research Council 2012). However, the scenario of chemical human exposures is complex (Table 2.1). First, controlled conditions, as in traditional animal studies based on the administration of a single chemical and identification of potential target organs (Carlin et al. 2013), do not generally apply. Second, intensity, duration, and frequency of the exposures are normally unknown and often changing. Third, the timing of the exposure is seldom known. Fourth,

Table 2.1 Typical scenarios of human vs animal exposures to environmental chemicals

Chemical-dependent Variable	Human	Animal
Dose	Low? (known?)	High (controlled/ known)
Intensity	Unknown	Known
Timing	Variable (known?)	Fixed
Frequency	Unknown, likely episodic yet chronic	Known
Pathway	Multiple (known?)	Single and identifiable
Chemicals	Many (known?) Metabolites?	Single (mixtures)
Target organ	Accessible? (known?)	Accessible

exposure routes and sources are numerous and, at times, even unknown. Finally, in a world where more than 80,000 chemicals are used in commerce (Bell and Edwards 2015), people are exposed to "cocktails" (multiple/mixtures) of chemicals. Therefore, mixtures encompass the large majority of environmental or background chemical exposures, even in situations when other exposures to mainly single chemicals or chemical classes may occur (e.g., accidental exposures). The fact that all of the above considerations would apply to each of the individual components of the mixtures may further complicate the interpretation of human biomonitoring data. Nevertheless, because biomonitoring per its nature provides an aggregate measure of exposure, biomonitoring has the potential to provide invaluable information for the exposure assessment of chemical mixtures.

2.3 Analytical Aspects of Biomonitoring

Biomonitoring relies on a targeted analysis to provide a quantitative measure of the amount of a chemical or chemicals present in the human body. These chemical biomarkers can be markers of exposure, effect, or susceptibility (National Research Council 2006). As defined by the World Health Organization (WHO), a biomarker of exposure is a "chemical or its metabolite or the product of an interaction between a chemical and some target molecule or cell that is measured in a compartment in an organism," a biomarker of effect is "a measurable biochemical, physiologic, behavioral, or other alteration in an organism that, depending on the magnitude can be recognized as associated with an established or possible health impairment or disease" (e.g., DNA adduct), and a biomarker of susceptibility is "an indicator of an inherent or acquired ability of an organism to respond to the challenge of exposure to a specific chemical substance" (e.g., glucose-6-phosphate dehydrogenase deficiency) (National Research Council 2006). In this chapter, we will focus on biomarkers of exposure.

The success of a biomonitoring approach greatly depends on the adequate selection of the exposure biomarkers, the matrix, and the analytical method (Needham et al. 2007). Knowledge of the physicochemical properties of the target chemicals is important in the choice of exposure biomarkers (e.g., parent compound vs metabolite) and biomonitoring matrix. In general, persistent compounds are commonly measured in blood or blood products, while metabolites of nonpersistent compounds (chemicals with half-lives of the order of hours) are measured in urine (Needham et al. 2007). Measuring blood concentrations of nonpersistent chemicals may be possible when the blood is collected soon after the exposure and if the analytical method is sensitive enough to detect the much lower blood concentrations of the chemical than of its metabolites in urine (Needham and Sexton 2000).

Because exogenous chemicals are present in the biological matrix at much lower concentrations than other endogenous compounds, biomonitoring methods generally include steps to (a) preconcentrate the target analytes and eliminate unwanted matrix components and (b) separate the target analytes from each other and from residual matrix constituents before quantification. The chemical nature of the biomarker and availability of instrumentation can impact the choice of preconcentration (e.g., liquid-liquid extraction, solid-phase extraction) and separation (e.g., chromatography) techniques. For example, separation of volatile organic chemicals is generally achieved by gas chromatography, while liquid chromatography separates nonvolatile organic compounds; nonvolatile organic compounds may also be amenable to gas chromatography after suitable chemical derivatization (Needham et al. 2005c). For human biomonitoring of organic chemicals, isotope-dilution mass spectrometry is generally considered the gold standard quantification technique (Needham et al. 2005c; WHO 2011). Other techniques (e.g., enzyme-linked immunosorbent assay [ELISA], fluorescence) may have the required sensitivity but generally lack adequate analytical selectivity and specificity (WHO 2011).

Biomonitoring methods, rooted in their analytical chemistry foundation, must be sensitive (i.e., capable of accurately measuring small amounts of a given substance in a sample (Saah and Hoover 1997)), selective and specific (i.e., able to measure one particular substance, rather than others in a sample (Saah and Hoover 1997)), and accurate and precise at trace levels (Needham et al. 2005c; Calafat and Needham 2009). Thanks to the scientific and technical advances in robotics and analytical instrumentation in the past few decades, such methods have become increasingly common in many laboratories (Angerer et al. 2007). In addition, biomonitoring methods should preferably use minimal matrix volume, be high throughput, show sustained reproducibility, and concurrently quantify multiple biomarkers. To achieve such characteristics, biomonitoring requires highly trained staff, top quality analytical standards (often custom-synthesized), and the use and maintenance of state-of-the-art instrumentation and facilities. Because of the uniqueness of these resources, biomonitoring is relatively expensive. Nevertheless, despite cost and other challenges associated with the interpretation of biomonitoring data (e.g., study design, communication of results), the use of biomonitoring in environmental public health is on the rise (National Research

Council 2012, 2006; Albertini et al. 2006; Morello-Frosch et al. 2015). Specifically, the possibility of measuring multiple chemicals simultaneously in a small amount of biospecimen makes biomonitoring uniquely suited to study human exposures to chemical mixtures and the potential effects of such exposures on health.

2.4 Interpretation of Biomonitoring Data

Biomonitoring provides information on the concentrations (i.e., amount) of select chemicals that were absorbed into the body after contact between the chemicals—or their precursors—and the body (Needham et al. 2005b). Noteworthy, such concentrations are in the range of trace levels (compared to the generally higher concentrations of the chemicals in the environment) and integrate all environmental pathways and routes (e.g., food, water, air, dust, product use). Furthermore, using biomonitoring concentrations to estimate exposure can pose study design challenges related both to the nature of the biomarker (e.g., specificity, temporality) as well as to the adequacy of the sampling process.

Specificity of the Biomarker Interpreting biomonitoring data requires a good understanding of the toxicokinetics of the target biomarkers. In general, relying on the concentrations of the most abundant biomarker for a given chemical will likely minimize exposure misclassification. For example, di-isononyl phthalate (DINP) metabolizes into mono-isononyl phthalate (MINP) before forming several oxidative metabolites which are the major DINP metabolites in urine (Koch and Angerer 2007). In general population settings, using urinary concentrations of MINP may underestimate exposure to DINP because MINP represents only a minor fraction (~2%) of the DINP excreted in urine compared to the oxidative metabolites (~44%). In fact, exposure to DINP in approximately 82% of Americans would have been misclassified based on the concentrations of the insensitive biomarker MINP, highlighting the importance of selecting a priori the best biomarkers for the intended purposes of the study (Calafat et al. 2011). An additional benefit of measuring phthalate oxidative metabolites is that they cannot be formed as a result of external contamination with the parent phthalate (Koch and Calafat 2009), thus also increasing the specificity of the measurement. In general, when a compound is converted to multiple metabolites, the quantification of all metabolites provides the best biomonitoring approach for exposure assessment. Depending on the aims of the study, exposure categorization can then be based on the concentrations of the individual biomarkers (e.g., four di (2-ethylhexyl) phthalate [DEHP] metabolites) and/or their sum (e.g., sum of all measured DEHP metabolites) (Dales et al. 2018; Kasper-Sonnenberg et al. 2017; Sathyanarayana et al. 2016; Huang et al. 2016; Axelsson et al. 2015; Ferguson et al. 2014; Guo et al. 2014; Kim and Hong 2014; Larsson et al. 2014; Mervish et al. 2014; Watkins et al. 2014; Zhang et al. 2014; Kim et al. 2013; Park et al. 2013; Tellez-Rojo et al. 2013; Upson et al.

2013; Braun et al. 2012; James-Todd et al. 2012; Kasper-Sonnenberg et al. 2012; Teitelbaum et al. 2012; Frederiksen et al. 2011; Romero-Franco et al. 2011).

The fact that several chemicals can metabolize into the same end product may also complicate the interpretation of biomonitoring data. For example, a number of synthetic pyrethroid insecticides are converted to 3-phenoxybenzoic acid (Leng et al. 2003). Therefore, the presence of 3-phenoxybenzoic acid in urine suggests exposure to pyrethroids, but 3-phenoxybenzoic acid concentrations per se cannot pinpoint the specific pyrethroid(s) to which exposure occurred. Similarly, certain chemicals (e.g., organophosphate insecticides) may degrade in the environment, and exposure could be to both the parent compound and the preformed degradate (e.g., dialkylphosphates) (Needham et al. 2005a; Barr et al. 2004). Yet, biomonitoring concentrations of degradates will reflect exposure to both the parent chemical and the preformed metabolites. In the above scenarios, interpreting the concentrations of the biomarkers may be challenging, particularly when the bioactivity of the precursors/parent compounds and their corresponding degradates or non-specific metabolites differ (Duggan et al. 2003). Nonetheless, use of these "non-specific" urinary biomarkers can still provide useful information about cumulative exposure to the parent class of compounds (e.g., pyrethroids, organophosphates).

Temporality of the Biomarker Exposure biomarkers should reflect a person's exposure to the target chemicals or their precursors within a specific time period (e.g., pregnancy) (Calafat et al. 2015). However, with a few exceptions of defined patterns of exposure, such as scheduled tasks in occupational settings (Arnold et al. 2013), the timing, duration, and intensity of chemical exposures are generally unknown and likely different for each chemical in a mixture. As a result, even though biomarker concentrations can accurately rank a person's exposure at a single time point, to evaluate exposure over weeks, months, or years may require different approaches.

In general, for persistent compounds, the timing of the exposure relative to sample collection is not critical. Regardless of the nature of the exposure (i.e., constant vs episodic), its duration, intensity, or timing, a single biomarker concentration at a given time point adequately represents exposure over an extended time (e.g., years) because persistent compounds have relative long elimination half-lives (Needham et al. 2005a; Meeker et al. 2009; Makey et al. 2014). Nonetheless, in certain situations, more than one sample may be needed. For example, chemical-specific toxicokinetics, including transplacental transfer or distribution into breast milk, can result in biomarker concentration changes which, in turn, could impact long-term exposure estimates obtained using a single sample collected during pregnancy or lactation (Adetona et al. 2013; Hooper et al. 2007; Glynn et al. 2012; Kato et al. 2014).

Variability in concentrations is much more pronounced for nonpersistent than for persistent chemicals because concentrations of the target biomarkers for non-persistent chemicals increase and decay rapidly in blood and urine after exposure

(Needham et al. 2005a, 2008; Needham and Sexton 2000; Aylward et al. 2014). Therefore, the intrinsic short half-life (e.g., hours) of nonpersistent chemicals, the nature of the exposure (i.e., intensity, duration, recurrence), and the time passed between exposure and collection of the biomonitoring matrix will impact the reproducibility of the biomarker concentrations (Needham et al. 2005a; Needham and Sexton 2000; Aylward et al. 2014). The timing, duration, and intensity of recurrent activities including use of personal care products or diet—likely pathways of exposure to environmental chemicals—generally change not only between days but even within a day and result in individual variability in concentrations of nonpersistent chemical biomarkers (Aylward et al. 2017; Koch and Angerer 2007; Preau et al. 2010; Teeguarden et al. 2011; Ye et al. 2011; Koch et al. 2004, 2005, 2012, 2013, 2014; Frederiksen et al. 2013; Li et al. 2010; Lassen et al. 2013; Fromme et al. 2007). Assuming a recurrent exposure of constant intensity and duration, this variability will increase as the half-life of the target chemical becomes shorter and the exposure becomes more infrequent (Aylward et al. 2014). The real situation, though, is not as simple: exposures likely recur but at unknown times, and with variable duration and intensity. Moreover, the variability of concentrations will differ for the individual components of chemical mixtures. Nevertheless, because background chemical exposures arise from everyday activities (e.g., food consumption, product use), acceptable biomarker variability over time exists although a person's concentration of a nonpersistent chemical will vary considerably (Calafat et al. 2015; Koch et al. 2014; Preau et al. 2010; Ye et al. 2011; Frederiksen et al. 2013; Li et al. 2010; Lassen et al. 2013; Bradman et al. 2013; Wielgomas 2013). Therefore, biomarker concentrations of a single sample—representing each person's variable exposure scenario—obtained from a large number of individuals can adequately characterize a population's average concentration over extended time because every person's exposure (represented by the biomarker concentration in his/her single sample) would contribute to the mean and extent of exposure (categorized from the mean and range of biomarker concentration, respectively) of the overall population (Calafat et al. 2015; Aylward et al. 2014). This principle is at the foundation of population-based or nation-based biomonitoring programs discussed later in this chapter.

In environmental epidemiology, the intrinsic variability in biomarker urinary concentrations may result in considerable exposure misclassification and bias associations between exposures and health outcomes toward the null hypothesis. Therefore, optimizing the design of exposure assessment in epidemiological studies will require information on the temporal variability of concentrations of the target compounds, particularly for nonpersistent compounds. Determining exposure over weeks, months, or years may require multiple measurements or use of composite (i.e., pooled) specimens to minimize this variability (Ferguson et al. 2014; Adibi et al. 2008; Bertelsen et al. 2014; Braun et al. 2011; Braun et al. 2012; Cantonwine et al. 2014; Engel et al. 2014; Hauser et al. 2004; Irvin et al. 2010; Lewis et al. 2014, 2015; Mahalingaiah et al. 2008; Meeker et al. 2012; Peck et al. 2010; Philippat et al. 2013; Quiros-Alcala et al. 2013; Smith et al. 2012; Teitelbaum et al. 2008; Watkins et al. 2014; Whyatt et al. 2012; Baird et al. 2010; Valvi et al. 2015; Hoppin et al.

2002; Townsend et al. 2013; Heffernan et al. 2014; Jusko et al. 2014; Nepomnaschy et al. 2009; Meeker et al. 2005; Perrier et al. 2016; Weiss et al. 2015; Morgan et al. 2015; Geens et al. 2015; Spaan et al. 2015; Reeves et al. 2014; Fisher et al. 2015; Ackerman et al. 2014; Guidry et al. 2015; Pollack et al. 2016; Romano et al. 2017); the biomarker concentrations obtained from multiple measures (preferably from collections at various times of the day and on multiple days during the study period) or pooled specimens would "average" the concentrations of the target chemicals during the study period.

Of interest, the reproducibility of such measurements will differ, depending on the chemical, the nature of the exposure, and the study population. For example, for nonpersistent compounds for which exposure is largely driven by use of personal care products (e.g., certain phthalates, parabens), the largest variability contributor will be interindividual, while for nonpersistent chemicals with mainly dietary sources that can change substantively from day to day and even within a day (e.g., bisphenol A [BPA], polycyclic aromatic hydrocarbons [PAHs]), large intraindividual differences are expected (Koch et al. 2014; Preau et al. 2010; Teeguarden et al. 2011; Ye et al. 2011; Frederiksen et al. 2013; Li et al. 2010; Lassen et al. 2013; Fromme et al. 2007; Bradman et al. 2013; Wielgomas 2013). For example, the intraclass correlation coefficient (ICC) describes the reliability of repeated measures over time and is defined as the ratio of between-subject variance to total (between-subject plus within-subject) variance. ICC ranges from zero (no reproducibility) to one (perfect reproducibility). Interestingly, ICCs of biomarker concentrations associated with episodic dietary exposures tend to be lower than for other exposures (e.g., use of personal care products), regardless of the study population (e.g., children, pregnant women, other adults) and the timeframe evaluated (days, months, years) (Ferguson et al. 2014; Adibi et al. 2008; Bertelsen et al. 2014; Braun et al. 2011, 2012; Cantonwine et al. 2014; Engel et al. 2014; Hauser et al. 2004; Irvin et al. 2010; Lewis et al. 2014, 2015; Mahalingaiah et al. 2008; Meeker et al. 2012; Peck et al. 2010; Philippat et al. 2013; Quiros-Alcala et al. 2013; Smith et al. 2012; Teitelbaum et al. 2008; Watkins et al. 2014; Whyatt et al. 2012; Baird et al. 2010; Valvi et al. 2015; Hoppin et al. 2002; Townsend et al. 2013; Heffernan et al. 2014; Jusko et al. 2014; Nepomnaschy et al. 2009; Meeker et al. 2005; Perrier et al. 2016; Weiss et al. 2015; Morgan et al. 2015; Geens et al. 2015; Spaan et al. 2015; Reeves et al. 2014; Fisher et al. 2015; Ackerman et al. 2014; Guidry et al. 2015; Dewalque et al. 2015). These findings suggest that short-term variability in biomarker concentrations may also apply to long-term variability because the associated exposures related to lifestyle habits do not change considerably over time (Calafat et al. 2015), particularly if commercial formulations of the chemicals do not change within the study timeframe and geographic location.

Sampling Strategies The use of spot urine samples (i.e., single, untimed urine specimen, voided spontaneously) is common in environmental epidemiology despite the well-recognized fact that spot concentrations for short-lived chemicals can show considerable inter- and intraindividual temporal variability, particularly for episodic exposures (Calafat et al. 2015; Aylward et al. 2014). Collecting spot

samples, including first morning voids, is easier than collecting 24-h samples and may facilitate compliance in epidemiologic studies.

In recent years, several investigations evaluated the suitability of spot samples, including first morning voids, and 24-h collections to categorize exposures to nonpersistent compounds such as plastic components (phthalates, BPA), personal care product chemicals (e.g., parabens, triclosan), pesticides, and PAHs from the biomarker urinary concentrations (Koch et al. 2014; Preau et al. 2010; Ye et al. 2011; Frederiksen et al. 2013; Li et al. 2010; Lassen et al. 2013; Bradman et al. 2013; Wielgomas 2013; Aylward et al. 2017). Interestingly, regardless of the collection strategy (spot, first morning, 24-h), urinary concentrations of the target biomarkers varied considerably during the 6–7 consecutive days of study, and the main contributors to the total variance differed. Specifically, biomarker urinary concentrations of chemicals with mainly dietary sources showed higher inter-day variability than those of chemicals with mostly non-dietary sources (e.g., use of personal care products) for which inter-person variability was most pronounced. These findings and the fact that the concentrations of biomarkers in first morning voids or spot samples correlated moderately to highly with those from 24-h composites suggest that, for exposure assessment, collecting 24-h voids may not be advantageous compared to multiple spot collections.

The nature of the exposures (diet vs other lifestyle factors) and timing of urine sampling to evaluate chemical exposures should be considered. This critical aspect is particularly relevant when evaluating exposures to chemical mixtures because the sources and timing of the exposures are likely to vary depending on the chemical. Changing the time of collection of spot samples and recording the time of urine collection and time since last void would provide useful information to interpret biomonitoring data for chemical mixtures.

Because not all sources of the more than 80,000 chemicals commercially used are known, understanding the details related to collecting and processing of biomonitoring specimens is also critical. Sampling must ensure that the biomarker concentrations reflect contact with the chemicals or their precursor(s) from a person's usual exposures over time (e.g., during pregnancy) and not from recent contact, such as from use of medical equipment or medical intervention or from specimen contamination (Calafat and Needham 2009; Calafat et al. 2015). Field or travel blanks are advantageous to identify potential external contamination during sample collection (Calafat and Needham 2009) but are inadequate to highlight contact with the chemical shortly before sampling. For example, during medical interventions people may be exposed to chemicals used in plastics such as DEHP and BPA (Huygh et al. 2015; Su et al. 2012; Weuve et al. 2006; Green et al. 2005; Calafat et al. 2004; Duty et al. 2013; Calafat et al. 2009). Concentrations of these chemicals or their metabolites in specimens collected soon after medical treatment would reflect true exposures (Calafat et al. 2015; Vandentorren et al. 2011; Yan et al. 2009), but would not be representative of typical daily exposures and, therefore, likely would be irrelevant in environmental epidemiology studies.

2.5 Population-Based Biomonitoring Surveys

Human biomonitoring data per se do not provide information on health effects (CDC 2009), but general population human biomonitoring programs are useful for investigating human exposure to chemical mixtures and are important tools for integrating environment and health (Angerer et al. 2006; National Research Council 2006; Bell and Edwards 2015; Sobus et al. 2015).

For example, in the United States, the Centers for Disease Control and Prevention (CDC) conducts the National Health and Nutrition Examination Survey (NHANES) (CDC 2011, 2017). Beginning in 1971 and then yearly since 1999, NHANES collects detailed medical history and performs physical examinations on all survey participants, approximately 5000 per year—selected through a complex statistical process using the most current census information—and includes a biomonitoring component from participants 1 year of age and older (CDC 2017; Calafat 2012). Most NHANES biomonitoring data are based on the analysis of samples from representative subsets of participants (e.g., one-third sample); subset distributions often change with survey cycles. Of note, biomonitoring measurements in blood or blood serum are only available from participants 12 years and older, with the exception of lead, cadmium, mercury, and, since 2011, manganese and selenium (minimum age of 1 year), and cotinine (minimum age of 3 years); biomonitoring measurements in urine are available for persons 6 years (3 years starting in 2015) of age and older (CDC 2011; Calafat 2012). Since 1999, the number of chemicals measured in people's blood has increased from 27 to more than 300 (CDC 2017). The NHANES biomonitoring program provides the most comprehensive assessment of the U.S. general population's exposure to hundreds of environmental chemicals (CDC 2011, 2017). These data, which are representative of the U.S. population, have been used to establish population reference ranges, track exposure trends, identify populations with higher than background exposures, prioritize research needs, and, sometimes, inform chemical risk assessments (National Research Council 2012; Sobus et al. 2015). For example, several federal agencies in the United States have used NHANES biomonitoring data to support various research and regulatory activities, including, among others, the U.S. Environmental Protection Agency (EPA) decision to remove lead from gasoline and the reregistration of triclosan and the U.S. Consumer Products Safety Commission (CPSC) decision to prohibit the use of certain phthalates in children's toys and child care articles (Sobus et al. 2015; U.S. EPA 2008; CPSC 2017). More recently, EPA also used NHANES data for lead, mercury, perchlorate, and several organic compounds—polyfluoroalkyl and perfluoroalkyl substances, polychlorinated biphenyls, polybrominated diphenyl ethers, phthalates, and BPA—in its third edition of America's Children and the Environment (U.S. EPA 2013), a report which, among other goals, can help policymakers and the public track and understand the potential impacts of environmental contaminants on children's health and, ultimately, to identify and evaluate ways to minimize these impacts.

Nationwide human biomonitoring initiatives also exist in other countries, including Austria, Belgium (Flanders), Canada, Czech Republic, France, Germany, Israel, Japan, Korea, Slovenia, and Spain (Health Canada 2017; Kolossa-Gehring et al. 2012; Kawamoto et al. 2014; Perez-Gomez et al. 2013; Cerna et al. 2012; Frery et al. 2012; Schoeters et al. 2012; Perharic and Vracko 2012; Berman et al. 2012; Casteleyn et al. 2015; La Corte and Wuttke 2012; Hohenblum et al. 2012; Ha et al. 2014). Biological samples collected as part of these programs have been used to provide the most comprehensive assessment of these countries' general populations' exposure to select environmental chemicals (CDC 2017; Becker et al. 2009; Koch et al. 2007; Schulz et al. 2007; Haines and Murray 2012; Jeong et al. 2014; Geens et al. 2014; Saoudi et al. 2014; Puklova et al. 2010; Bartolome et al. 2015; Levine et al. 2015; Černá et al. 2017) and also have the potential to inform chemical risk assessments (Albertini et al. 2006).

In addition to nationwide general population programs, biomonitoring has been increasingly used in environmental epidemiology such as birth cohort studies as well as cohorts of specific population groups. The list of programs relying on biomonitoring measurements is so extensive that it can't be covered in this short chapter. The examples below are meant to illustrate the breadth and versatility of populations for which biomonitoring can provide useful information in environmental epidemiology research.

Birth cohort studies include, among many others, the U.S. National Institute of Environmental Health Sciences/EPA's Children's Centers for Environmental Health and Disease Prevention (https://www.epa.gov/research-grants/niehsepa-childrens-environmental-health-and-disease-prevention-research-centers), Project VIVA, The Infant Development and the Environment Study, the Mothers and Children's Environmental Health Study, the Norwegian Mother and Child Cohort Study, the Danish National Birth Cohort, and the Northern Norway Mother-and-Child Contaminant Cohort (Kim et al. 2009; Zeman et al. 2013; Ronningen et al. 2006; Guxens et al. 2012; Olsen 2012; Veyhe et al. 2015; Oken et al. 2015; Barrett et al. 2014). Cohorts of specific population groups include, among many, the Framingham Heart Study, the Nurses' Health Study, the Women's Health Initiative, the Childhood Autism Risks from Genetics and the Environment Study, the Shanghai Women's Health Study, and the Shanghai Men's Health Study (Zheng et al. 2005; Cupples et al. 2007; Colditz and Hankinson 2005; Hays et al. 2003; Hertz-Picciotto et al. 2006; Cai et al. 2007). Lactating women may be considered a special category within these specific population groups. Human milk is regarded as infants' best nutrition source (Landrigan et al. 2002). Nonetheless, environmental chemicals can also partition into breast milk (Lehmann et al. 2014), and breast milk monitoring programs can provide useful information to study early life exposures to environmental chemicals (Lehmann et al. 2014; Solomon and Weiss 2002; Berlin et al. 2005; Fenton et al. 2005; LaKind et al. 2005; Pan et al. 2010; LaKind et al. 2001; Colles et al. 2008; Li et al. 2009; Croes et al. 2012). Breast milk programs, for instance, albeit restricted to one specific sector of the population, have been useful to assess temporal trends to select persistent organic compounds (Solomon and Weiss 2002; Konishi et al. 2001; Kunisue et al. 2006; Raab et al. 2007; Abballe

et al. 2008; Polder et al. 2008; Lignell et al. 2009; Sundstrom et al. 2011; Fang et al. 2013; Vukavic et al. 2013; Darnerud et al. 2015).

For these population-based initiatives, biomonitoring can provide important exposure information as well as data to understand the potential impact of the environment, including chemical mixtures, among other factors, on various aspects of human health such as reproductive outcomes, development, behavior, and respiratory health. Age; diet and lifestyle; route, frequency, and intensity of the exposures; potential interactions among chemicals; and genetic factors, among others, are critical in determining health outcomes from exposures to chemical mixtures. In the context of environmental public health, the recent increase in studies linking biomonitoring exposure data and human health would benefit from the definition and implementation of best practices for analysis, interpretation, and communication of biomonitoring data (Sobus et al. 2015; LaKind et al. 2014).

2.6 Conclusions

Biomonitoring provides an integrated measure of exposure to chemicals from all sources and routes and can be reliably used to estimate internal doses. Because of technological advances, a vast amount of biomonitoring exposure data already exist, and the volume of studies including a biomonitoring component for exposure assessment and, most recently, for relating internal exposure to potential health effects continue to increase. The proper use and interpretation of biomonitoring depend in large part on the study objectives which, in turn, dictate the study design. Adequate selection of the study population, procurement and type of biospecimens, and choice of analytical methods are critical to a successful biomonitoring initiative.

References

Abballe, A., T.J. Ballard, E. Dellatte, et al. 2008. Persistent environmental contaminants in human milk: Concentrations and time trends in Italy. *Chemosphere* 73: S220–S227.

Ackerman, J.M., R.E. Dodson, C.L. Engel, et al. 2014. Temporal variability of urinary di (2-ethylhexyl) phthalate metabolites during a dietary intervention study. *Journal of Exposure Science and Environmental Epidemiology* 24: 595–601.

Adetona, O., K. Horton, A. Sjodin, et al. 2013. Concentrations of select persistent organic pollutants across pregnancy trimesters in maternal and in cord serum in Trujillo, Peru. *Chemosphere* 91: 1426–1433.

Adibi, J.J., R.M. Whyatt, P.L. Williams, et al. 2008. Characterization of phthalate exposure among pregnant women assessed by repeat air and urine samples. *Environmental Health Perspectives* 116: 467–473.

Albertini, R., M. Bird, N. Doerrer, et al. 2006. The use of biomonitoring data in exposure and human health risk assessments. *Environmental Health Perspectives* 114: 1755–1762.

Angerer, J., M.G. Bird, T.A. Burke, et al. 2006. Strategic biomonitoring initiatives: Moving the science forward. *Toxicological Sciences* 93: 3–10.

Angerer, J., U. Ewers, and M. Wilhelm. 2007. Human biomonitoring: State of the art. *International Journal of Hygiene and Environmental Health* 210: 201–228.

Arnold, S.M., J. Angerer, P.J. Boogaard, et al. 2013. The use of biomonitoring data in exposure and human health risk assessment: Benzene case study. *Critical Reviews in Toxicology* 43: 119–153.

Axelsson, J., L. Rylander, A. Rignell-Hydbom, et al. 2015. Phthalate exposure and reproductive parameters in young men from the general Swedish population. *Environment International* 85: 54–60.

Aylward, L.L., S.M. Hays, R. Smolders, et al. 2014. Sources of variability in biomarker concentrations. *Journal of Toxicology and Environmental Health, Part B* 17: 45–61.

Aylward, L.L., S.M. Hays, and A. Zidek. 2017. Variation in urinary spot sample, 24 h samples, and longer-term average urinary concentrations of shortlived environmental chemicals: implications for exposure assessment and reverse dosimetry. *Journal of Exposure Science and Environmental Epidemiology* 27 (6): 582–590.

Baird, D.D., T.M. Saldana, P.A. Nepomnaschy, et al. 2010. Within-person variability in urinary phthalate metabolite concentrations: Measurements from specimens after long-term frozen storage. *Journal of Exposure Science and Environmental Epidemiology* 20: 169–175.

Barr, D.B., R. Bravo, G. Weerasekera, et al. 2004. Concentrations of dialkyl phosphate metabolites of organophosphorus pesticides in the US population. *Environmental Health Perspectives* 112: 186–200.

Barrett, E.S., S. Sathyanarayana, S. Janssen, et al. 2014. Environmental health attitudes and behaviors: Findings from a large pregnancy cohort study. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 176: 119–125.

Bartolome, M., J.J. Ramos, F. Cutanda, et al. 2015. Urinary polycyclic aromatic hydrocarbon metabolites levels in a representative sample of the Spanish adult population: The BIOAMBIENT.ES project. *Chemosphere* 135: 436–446.

Becker, K., T. Goen, M. Seiwert, et al. 2009. GerES IV: Phthalate metabolites and bisphenol A in urine of German children. *International Journal of Hygiene and Environmental Health* 212: 685–692.

Bell, S.M., and S.W. Edwards. 2015. Identification and prioritization of relationships between environmental stressors and adverse human health impacts. *Environmental Health Perspectives* 123: 1193–1199.

Berlin, C.M., B.L. Crase, P. Furst, et al. 2005. Methodologic considerations for improving and facilitating human milk research. *Journal of Toxicology and Environmental Health, Part A* 68: 1803–1823.

Berman, T., Y. Amitai, S. Almog, et al. 2012. Human biomonitoring in Israel: Past, present, future. *International Journal of Hygiene and Environmental Health* 215: 138–141.

Bertelsen, R.J., S.M. Engel, T.A. Jusko, et al. 2014. Reliability of triclosan measures in repeated urine samples from Norwegian pregnant women. *Journal of Exposure Science and Environmental Epidemiology* 24: 517–521.

Birnbaum, L.S. 2010. Applying research to public health questions: Biologically relevant exposures. *Environmental Health Perspectives* 118: A152.

Bradman, A., K. Kogut, E.A. Eisen, et al. 2013. Variability of organophosphorous pesticide metabolite levels in spot and 24-hr urine samples collected from young children during 1 week. *Environmental Health Perspectives* 121: 118–124.

Braun, J.M., A.E. Kalkbrenner, A.M. Calafat, et al. 2011. Variability and predictors of urinary bisphenol A concentrations during pregnancy. *Environmental Health Perspectives* 119: 131–137.

Braun, J.M., K.W. Smith, P.L. Williams, et al. 2012. Variability of urinary phthalate metabolite and bisphenol A concentrations before and during pregnancy. *Environmental Health Perspectives* 120: 739–745.

Cai, H., W. Zheng, Y.B. Xiang, et al. 2007. Dietary patterns and their correlates among middle-aged and elderly Chinese men: A report from the Shanghai Men's Health Study. *British Journal of Nutrition* 98: 1006–1013.

Calafat, A.M. 2012. The U.S. National Health and Nutrition Examination Survey and human exposure to environmental chemicals. *International Journal of Hygiene and Environmental Health* 215: 99–101.

Calafat, A.M., and L.L. Needham. 2009. What additional factors beyond state-of-the-art analytical methods are needed for optimal generation and interpretation of biomonitoring data? *Environmental Health Perspectives* 117: 1481–1485.

Calafat, A.M., L.L. Needham, M.J. Silva, et al. 2004. Exposure to di-(2-ethylhexyl) phthalate among premature neonates in a neonatal intensive care unit. *Pediatrics* 113: e429–e434.

Calafat, A.M., X. Ye, M.J. Silva, et al. 2006. Human exposure assessment to environmental chemicals using biomonitoring. *International Journal of Andrology* 29: 166–171.

Calafat, A.M., J. Weuve, X.Y. Ye, et al. 2009. Exposure to bisphenol A and other phenols in neonatal intensive care unit premature infants. *Environmental Health Perspectives* 117: 639–644.

Calafat, A.M., L.Y. Wong, M.J. Silva, et al. 2011. Selecting adequate exposure biomarkers of diisononyl and diisodecyl phthalates: Data from the 2005-2006 National Health and Nutrition Examination Survey. *Environmental Health Perspectives* 119: 50–55.

Calafat, A.M., M.P. Longnecker, H.M. Koch, et al. 2015. Optimal exposure biomarkers for nonpersistent chemicals in environmental epidemiology. *Environmental Health Perspectives* 123: A166–A168.

Cantonwine, D.E., J.F. Cordero, L.O. Rivera-Gonzalez, et al. 2014. Urinary phthalate metabolite concentrations among pregnant women in Northern Puerto Rico: Distribution, temporal variability, and predictors. *Environment International* 62: 1–11.

Carlin, D.J., C.V. Rider, R. Woychik, et al. 2013. Unraveling the health effects of environmental mixtures: An NIEHS priority. *Environmental Health Perspectives* 121: A6–A8.

Casteleyn, L., B. Dumez, K. Becker, et al. 2015. A pilot study on the feasibility of European harmonized Human Biomonitoring: Strategies towards a common approach, challenges and opportunities. *Environmental Research* 141: 3–14.

CDC. 2009. Fourth National Report on Human Exposure to Environmental Chemicals, Centers for Disease Control and Prevention; National Center for Environmental Health; Division of Laboratory Sciences, Atlanta. https://www.cdc.gov/exposurereport/pdf/fourthreport.pdf. Accessed 9 Dec 2017.

———. 2011. NHANES environmental chemical data tutorial. National Center for Health Statistics. http://www.cdc.gov/nchs/tutorials/environmental/index.htm. Accessed 9 Dec 2017.

———. 2017. Fourth National Report on Human Exposure to Environmental Chemicals. Updated Tables, January 2017, Centers for Disease Control and Prevention; National Center for Environmental Health; Division of Laboratory Sciences, Atlanta. https://www.cdc.gov/exposurereport/. Accessed 9 Dec 2017.

Cerna, M., A. Krskova, M. Cejchanova, et al. 2012. Human biomonitoring in the Czech Republic: An overview. *International Journal of Hygiene and Environmental Health* 215: 109–119.

Černá, M., V. Puklová, L. Hanzlíková, et al. 2017. 25 years of HBM in the Czech Republic. *International Journal of Hygiene and Environmental Health* 220 (2): 3–5.

Colditz, G.A., and S.E. Hankinson. 2005. The Nurses' Health Study: Lifestyle and health among women. *Nature Reviews Cancer* 5: 388–396.

Colles, A., G. Koppen, V. Hanot, et al. 2008. Fourth WHO-coordinated survey of human milk for persistent organic pollutants (POPs): Belgian results. *Chemosphere* 73: 907–914.

CPSC. 2017. Prohibition of children's toys and child care articles containing specified phthalates. *Federal Register* 82: 49938–49982.

Croes, K., A. Colles, G. Koppen, et al. 2012. Persistent organic pollutants (POPs) in human milk: A biomonitoring study in rural areas of Flanders (Belgium). *Chemosphere* 89: 988–994.

Cupples, L.A., H.T. Arruda, E.J. Benjamin, et al. 2007. The Framingham Heart Study 100K SNP genome-wide association study resource: Overview of 17 phenotype working group reports. *BMC Medical Genetics* 8 (Suppl 1): S1.

Dales, R.E., L.M. Kauri, and S. Cakmak. 2018. The associations between phthalate exposure and insulin resistance, beta-cell function and blood glucose control in a population-based sample. *Science of The Total Environment* 612: 1287–1292.

Darnerud, P.O., S. Lignell, M. Aune, et al. 2015. Time trends of polybrominated diphenylether (PBDE) congeners in serum of Swedish mothers and comparisons to breast milk data. *Environmental Research* 138: 352–360.

Dewalque, L., C. Pirard, S. Vandepaer, et al. 2015. Temporal variability of urinary concentrations of phthalate metabolites, parabens and benzophenone-3 in a Belgian adult population. *Environmental Research* 142: 414–423.

Duggan, A., G. Charnley, W. Chen, et al. 2003. Di-alkyl phosphate biomonitoring data: Assessing cumulative exposure to organophosphate pesticides. *Regulatory Toxicology and Pharmacology* 37: 382–395.

Duty, S.M., K. Mendonca, R. Hauser, et al. 2013. Potential sources of bisphenol A in the neonatal intensive care unit. *Pediatrics* 131: 483–489.

Engel, L.S., J.P. Buckley, G. Yang, et al. 2014. Predictors and variability of repeat measurements of urinary phenols and parabens in a cohort of Shanghai women and men. *Environmental Health Perspectives* 122: 733–740.

Fang, J., E. Nyberg, A. Bignert, et al. 2013. Temporal trends of polychlorinated dibenzo-p-dioxins and dibenzofurans and dioxin-like polychlorinated biphenyls in mothers' milk from Sweden, 1972-2011. *Environment International* 60: 224–231.

Fenton, S.E., M. Condon, A.S. Ettinger, et al. 2005. Collection and use of exposure data from human milk biomonitoring in the United States. *Journal of Toxicology and Environmental Health, Part A* 68: 1691–1712.

Ferguson, K.K., T.F. McElrath, Y.A. Ko, et al. 2014. Variability in urinary phthalate metabolite levels across pregnancy and sensitive windows of exposure for the risk of preterm birth. *Environment International* 70: 118–124.

Fisher, M., T.E. Arbuckle, R. Mallick, et al. 2015. Bisphenol A and phthalate metabolite urinary concentrations: Daily and across pregnancy variability. *Journal of Exposure Science and Environmental Epidemiology* 25: 231–239.

Frederiksen, H., L. Aksglaede, K. Sorensen, et al. 2011. Urinary excretion of phthalate metabolites in 129 healthy Danish children and adolescents: Estimation of daily phthalate intake. *Environmental Research* 111: 656–663.

Frederiksen, H., S.K. Kranich, N. Jorgensen, et al. 2013. Temporal variability in urinary phthalate metabolite excretion based on spot, morning, and 24-h urine samples: Considerations for epidemiological studies. *Environmental Science & Technology* 47: 958–967.

Frery, N., S. Vandentorren, A. Etchevers, et al. 2012. Highlights of recent studies and future plans for the French human biomonitoring (HBM) programme. *International Journal of Hygiene and Environmental Health* 215: 127–132.

Fromme, H., G. Bolte, H.M. Koch, et al. 2007. Occurrence and daily variation of phthalate metabolites in the urine of an adult population. *International Journal of Hygiene and Environmental Health* 210: 21–33.

Geens, T., L. Bruckers, A. Covaci, et al. 2014. Determinants of bisphenol A and phthalate metabolites in urine of Flemish adolescents. *Environmental Research* 134: 110–117.

Geens, T., A.C. Dirtu, E. Dirinck, et al. 2015. Daily intake of bisphenol A and triclosan and their association with anthropometric data, thyroid hormones and weight loss in overweight and obese individuals. *Environment International* 76: 98–105.

Glynn, A., U. Berger, A. Bignert, et al. 2012. Perfluorinated alkyl acids in blood serum from primiparous women in Sweden: Serial sampling during pregnancy and nursing, and temporal trends 1996-2010. *Environmental Science & Technology* 46: 9071–9079.

Green, R., R. Hauser, A.M. Calafat, et al. 2005. Use of di(2-ethylhexyl) phthalate-containing medical products and urinary levels of mono(2-ethylhexyl) phthalate in neonatal intensive care unit infants. *Environmental Health Perspectives* 113: 1222–1225.

Guidry, V.T., M.P. Longnecker, H. Aase, et al. 2015. Measurement of total and free urinary phenol and paraben concentrations over the course of pregnancy: Assessing reliability and contamination of specimens in the norwegian mother and child cohort study. *Environmental Health Perspectives* 123: 705–711.

Guo, Y., J. Weck, R. Sundaram, et al. 2014. Urinary concentrations of phthalates in couples planning pregnancy and its association with 8-hydroxy-2'-deoxyguanosine, a biomarker of oxidative stress: Longitudinal Investigation of Fertility and the Environment Study. *Environmental Science & Technology* 48: 9804–9811.

Guxens, M., F. Ballester, M. Espada, et al. 2012. Cohort profile: The INMA--INfancia y Medio Ambiente--(environment and childhood) project. *International Journal of Epidemiology* 41: 930–940.

Ha, M., H.J. Kwon, J.H. Leem, et al. 2014. Korean Environmental Health Survey in Children and Adolescents (KorEHS-C): Survey design and pilot study results on selected exposure biomarkers. *International Journal of Hygiene and Environmental Health* 217: 260–270.

Haines, D.A., and J. Murray. 2012. Human biomonitoring of environmental chemicals--early results of the 2007-2009 Canadian Health Measures Survey for males and females. *International Journal of Hygiene and Environmental Health* 215: 133–137.

Hauser, R., J.D. Meeker, S. Park, et al. 2004. Temporal variability of urinary phthalate metabolite levels in men of reproductive age. *Environmental Health Perspectives* 112: 1734–1740.

Hays, J., J.R. Hunt, F.A. Hubbell, et al. 2003. The Women's Health Initiative recruitment methods and results. *Annals of Epidemiology* 13: S18–S77.

Health Canada. 2017. Fourth report on human biomonitoring of environmental chemicals in Canada: results of the Canadian Health Measures Survey Cycle 4 (2014–2015), Health Canada, Ottawa. https://www.canada.ca/en/health-canada/services/environmental-workplacehealth/reports-publications/environmental-contaminants/fourth-report-human-biomonitoring-environmental-chemicals-canada.html. Accessed 9 Dec 2017.

Heffernan, A.L., L.L. Aylward, A.J. Samidurai, et al. 2014. Short term variability in urinary bisphenol A in Australian children. *Environment International* 68: 139–143.

Hertz-Picciotto, I., L.A. Croen, R. Hansen, et al. 2006. The CHARGE study: An epidemiologic investigation of genetic and environmental factors contributing to autism. *Environmental Health Perspectives* 114: 1119–1125.

Hohenblum, P., P. Steinbichl, W. Raffesberg, et al. 2012. Pollution gets personal! A first population-based human biomonitoring study in Austria. *International Journal of Hygiene and Environmental Health* 215: 176–179.

Hooper, K., J. She, M. Sharp, et al. 2007. Depuration of polybrominated diphenyl ethers (PBDEs) and polychlorinated biphenyls (PCBs) in breast milk from California first-time mothers (primiparae). *Environmental Health Perspectives* 115: 1271–1275.

Hoppin, J.A., J.W. Brock, B.J. Davis, et al. 2002. Reproducibility of urinary phthalate metabolites in first morning urine samples. *Environmental Health Perspectives* 110: 515–518.

Huang, P.-C., C.-H. Tsai, W.-Y. Liang, et al. 2016. Early phthalates exposure in pregnant women is associated with alteration of thyroid hormones. *Plos One* 11 (7): e0159398.

Huygh, J., K. Clotman, G. Malarvannan, et al. 2015. Considerable exposure to the endocrine disrupting chemicals phthalates and bisphenol-A in intensive care unit (ICU) patients. *Environment International* 81: 64–72.

Irvin, E.A., A.M. Calafat, M.J. Silva, et al. 2010. An estimate of phthalate exposure among pregnant women living in Trujillo, Peru. *Chemosphere* 80: 1301–1307.

James-Todd, T., R. Stahlhut, J.D. Meeker, et al. 2012. Urinary phthalate metabolite concentrations and diabetes among women in the National Health and Nutrition Examination Survey (NHANES) 2001-2008. *Environmental Health Perspectives* 120: 1307–1313.

Jeong, S.W., C.K. Lee, C.H. Suh, et al. 2014. Blood lead concentration and related factors in Korea from the 2008 National Survey for Environmental Pollutants in the Human Body. *International Journal of Hygiene and Environmental Health* 217: 871–877.

Jusko, T.A., P.A. Shaw, C.A. Snijder, et al. 2014. Reproducibility of urinary bisphenol A concentrations measured during pregnancy in the Generation R Study. *Journal of Exposure Science and Environmental Epidemiology* 24: 532–536.

Kasper-Sonnenberg, M., H.M. Koch, J. Wittsiepe, et al. 2012. Levels of phthalate metabolites in urine among mother-child-pairs – results from the Duisburg birth cohort study, Germany. *International Journal of Hygiene and Environmental Health* 215: 373–382.

Kasper-Sonnenberg, M., J. Wittsiepe, K. Wald, et al. 2017. Pre-pubertal exposure with phthalates and bisphenol A and pubertal development. *Plos One* 12 (11): e0187922.

Kato, K., L.Y. Wong, A.M. Chen, et al. 2014. Changes in serum concentrations of maternal poly- and perfluoroalkyl substances over the course of pregnancy and predictors of exposure in a multiethnic cohort of Cincinnati, Ohio pregnant women during 2003-2006. *Environmental Science & Technology* 48: 9600–9608.

Kawamoto, T., H. Nitta, K. Murata, et al. 2014. Rationale and study design of the Japan environment and children's study (JECS). *BMC Public Health* 14: 25.

Kim, J.H., and Y.C. Hong. 2014. HSP70-hom gene polymorphisms modify the association of diethylhexyl phthalates with insulin resistance. *Environmental and Molecular Mutagenesis* 55: 727–734.

Kim, B.M., M. Ha, H.S. Park, et al. 2009. The mothers and children's environmental health (MOCEH) study. *European Journal of Epidemiology* 24: 573–583.

Kim, J.H., H.Y. Park, S. Bae, et al. 2013. Diethylhexyl phthalates is associated with insulin resistance via oxidative stress in the elderly: A panel study. *Plos One* 8: e71392.

Koch, H.M., and J. Angerer. 2007. Di-iso-nonylphthalate (DINP) metabolites in human urine after a single oral dose of deuterium-labelled DINP. *International Journal of Hygiene and Environmental Health* 210: 9–19.

Koch, H.M., and A.M. Calafat. 2009. Human body burdens of chemicals used in plastic manufacture. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364: 2063–2078.

Koch, H.M., H.M. Bolt, and J. Angerer. 2004. Di(2-ethylhexyl)phthalate (DEHP) metabolites in human urine and serum after a single oral dose of deuterium-labelled DEHP. *Archives of Toxicology* 78: 123–130.

Koch, H.M., H.M. Bolt, R. Preuss, et al. 2005. New metabolites of di(2-ethylhexyl)phthalate (DEHP) in human urine and serum after single oral doses of deuterium-labelled DEHP. *Archives of Toxicology* 79: 367–376.

Koch, H.M., K. Becker, M. Wittassek, et al. 2007. Di-n-butylphthalate and butylbenzylphthalate – urinary metabolite levels and estimated daily intakes: Pilot study for the German Environmental Survey on children. *Journal of Exposure Science and Environmental Epidemiology* 17: 378–387.

Koch, H.M., K.L.Y. Christensen, V. Harth, et al. 2012. Di-n-butyl phthalate (DnBP) and diisobutyl phthalate (DiBP) metabolism in a human volunteer after single oral doses. *Archives of Toxicology* 86: 1829–1839.

Koch, H.M., A. Schutze, C. Palmke, et al. 2013. Metabolism of the plasticizer and phthalate substitute diisononyl-cyclohexane-1,2-dicarboxylate (DINCH) in humans after single oral doses. *Archives of Toxicology* 87: 799–806.

Koch, H.M., L.L. Aylward, S.M. Hays, et al. 2014. Inter- and intra-individual variation in urinary biomarker concentrations over a 6-day sampling period. Part 2: Personal care product ingredients. *Toxicology Letters* 231: 261–269.

Kolossa-Gehring, M., K. Becker, A. Conrad, et al. 2012. Environmental surveys, specimen bank and health related environmental monitoring in Germany. *International Journal of Hygiene and Environmental Health* 215: 120–126.

Konishi, Y., K. Kuwabara, and S. Hori. 2001. Continuous surveillance of organochlorine compounds in human breast milk from 1972 to 1998 in Osaka, Japan. *Archives of Environmental Contamination and Toxicology* 40: 571–578.

Kunisue, T., M. Muraoka, M. Ohtake, et al. 2006. Contamination status of persistent organochlorines in human breast milk from Japan: Recent levels and temporal trend. *Chemosphere* 64: 1601–1608.

La Corte, E., and S. Wuttke. 2012. The First Nations Biomonitoring Initiative-FNBI. *International Journal of Hygiene and Environmental Health* 215: 168–171.

LaKind, J.S., C.M. Berlin, and D.Q. Naiman. 2001. Infant exposure to chemicals in breast milk in the United States: What we need to learn from a breast milk monitoring program. *Environmental Health Perspectives* 109: 75–88.

LaKind, J.S., R.L. Brent, M.L. Dourson, et al. 2005. Human milk biomonitoring data: Interpretation and risk assessment issues. *Journal of Toxicology and Environmental Health, Part A* 68: 1713–1769.

LaKind, J.S., J.R. Sobus, M. Goodman, et al. 2014. A proposal for assessing study quality: Biomonitoring, Environmental Epidemiology, and Short-lived Chemicals (BEES-C) instrument. *Environment International* 73: 195–207.

Landrigan, P.J., B. Sonawane, D. Mattison, et al. 2002. Chemical contaminants in breast milk and their impacts on children's health: An overview. *Environmental Health Perspectives* 110: A313–A315.

Larsson, K., K.L. Bjorklund, B. Palm, et al. 2014. Exposure determinants of phthalates, parabens, bisphenol A and triclosan in Swedish mothers and their children. *Environment International* 73: 323–333.

Lassen, T.H., H. Frederiksen, T.K. Jensen, et al. 2013. Temporal variability in urinary excretion of bisphenol A and seven other phenols in spot, morning, and 24-h urine samples. *Environmental Research* 126: 164–170.

Lehmann, G.M., M.A. Verner, B. Luukinen, et al. 2014. Improving the risk assessment of lipophilic persistent environmental chemicals in breast milk. *Critical Reviews in Toxicology* 44: 600–617.

Leng, G., U. Ranft, D. Sugiri, et al. 2003. Pyrethroids used indoors--biological monitoring of exposure to pyrethroids following an indoor pest control operation. *International Journal of Hygiene and Environmental Health* 206: 85–92.

Levine, H., T. Berman, R. Goldsmith, et al. 2015. Urinary concentrations of polycyclic aromatic hydrocarbons in Israeli adults: Demographic and life-style predictors. *International Journal of Hygiene and Environmental Health* 218: 123–131.

Lewis, R.C., D.E. Cantonwine, L.V. Anzalota Del Toro, et al. 2014. Urinary biomarkers of exposure to insecticides, herbicides, and one insect repellent among pregnant women in Puerto Rico. *Environmental Health* 13: 97.

———. 2015. Distribution and determinants of urinary biomarkers of exposure to organophosphate insecticides in Puerto Rican pregnant women. *Science of The Total Environment* 15: 337–344.

Li, J., L. Zhang, Y. Wu, et al. 2009. A national survey of polychlorinated dioxins, furans (PCDD/Fs) and dioxin-like polychlorinated biphenyls (dl-PCBs) in human milk in China. *Chemosphere* 75: 1236–1242.

Li, Z., L.C. Romanoff, M.D. Lewin, et al. 2010. Variability of urinary concentrations of polycyclic aromatic hydrocarbon metabolite in general population and comparison of spot, first-morning, and 24-h void sampling. *Journal of Exposure Science and Environmental Epidemiology* 20: 526–535.

Lignell, S., M. Aune, P.O. Darnerud, et al. 2009. Persistent organochlorine and organobromine compounds in mother's milk from Sweden 1996-2006: Compound-specific temporal trends. *Environmental Research* 109: 760–767.

Mahalingaiah, S., J.D. Meeker, K.R. Pearson, et al. 2008. Temporal variability and predictors of urinary bisphenol a concentrations in men and women. *Environmental Health Perspectives* 116: 173–178.

Makey, C.M., M.D. McClean, A. Sjodin, et al. 2014. Temporal variability of polybrominated diphenyl ether (PBDE) serum concentrations over one year. *Environmental Science & Technology* 48: 14642–14649.

Meeker, J.D., D.B. Barr, L. Ryan, et al. 2005. Temporal variability of urinary levels of nonpersistent insecticides in adult men. *Journal of Exposure Science and Environmental Epidemiology* 15: 271–281.

Meeker, J.D., S.A. Missmer, L. Altshul, et al. 2009. Serum and follicular fluid organochlorine concentrations among women undergoing assisted reproduction technologies. *Environmental Health* 8: 32.

Meeker, J.D., A.M. Calafat, and R. Hauser. 2012. Urinary phthalate metabolites and their biotransformation products: Predictors and temporal variability among men and women. *Journal of Exposure Science and Environmental Epidemiology* 22: 376–385.

Mervish, N., K.J. McGovern, S.L. Teitelbaum, et al. 2014. Dietary predictors of urinary environmental biomarkers in young girls, BCERP, 2004-7. *Environmental Research* 133: 12–19.

Morello-Frosch, R., J. Varshavsky, M. Liboiron, et al. 2015. Communicating results in post-Belmont era biomonitoring studies: Lessons from genetics and neuroimaging research. *Environmental Research* 136: 363–372.

Morgan, M., P. Jones, and J. Sobus. 2015. Short-term variability and predictors of urinary pentachlorophenol levels in Ohio preschool children. *International Journal of Environmental Research and Public Health* 12: 800–815.

National Research Council. 2006. *Human biomonitoring for environmental chemicals*. Washington, DC: The National Academies Press. https://doi.org/10.17226/11700. Accessed 9 Dec 2017.

———. 2012. Exposure science in the 21st century: A vision and a strategy. National Academies Press (US), Washington (DC). http://www.nap.edu/catalog/13507/exposure-science-in-the-21st-century-a-vision-and-a. Accessed 9 Dec 2017.

Needham, L.L., and K. Sexton. 2000. Assessing children's exposure to hazardous environmental chemicals: An overview of selected research challenges and complexities. *Journal of Exposure Analysis & Environmental Epidemiology* 10: 611–629.

Needham, L.L., D.B. Barr, and A.M. Calafat. 2005a. Characterizing children's exposures: Beyond NHANES. *Neurotoxicology* 26: 547–553.

Needham, L.L., H. Ozkaynak, R.M. Whyatt, et al. 2005b. Exposure assessment in the National Children's Study: Introduction. *Environmental Health Perspectives* 113: 1076–1082.

Needham, L.L., D.G. Patterson, D.B. Barr, et al. 2005c. Uses of speciation techniques in biomonitoring for assessing human exposure to organic environmental chemicals. *Analytical and Bioanalytical Chemistry* 381: 397–404.

Needham, L.L., A.M. Calafat, and D.B. Barr. 2007. Uses and issues of biomonitoring. *International Journal of Hygiene and Environmental Health* 210: 229–238.

———. 2008. Assessing developmental toxicant exposures via biomonitoring. *Basic & Clinical Pharmacology & Toxicology* 102: 100–108.

Nepomnaschy, P.A., D.D. Baird, C.R. Weinberg, et al. 2009. Within-person variability in urinary bisphenol A concentrations: Measurements from specimens after long-term frozen storage. *Environmental Research* 109: 734–737.

Oken, E., A.A. Baccarelli, D.R. Gold, et al. 2015. Cohort profile: Project viva. *International Journal of Epidemiology* 44: 37–48.

Olsen, J. 2012. Nine months that last a lifetime. Experience from the Danish National Birth Cohort and lessons learned. *International Journal of Hygiene and Environmental Health* 215: 142–144.

Pan, I.J., J.L. Daniels, A.H. Herring, et al. 2010. Lactational exposure to polychlorinated biphenyls, dichlorodiphenyltrichloroethane, and dichlorodiphenyldichloroethylene and infant

growth: An analysis of the Pregnancy, Infection, and Nutrition Babies Study. *Paediatric and Perinatal Epidemiology* 24: 262–271.

Park, H.Y., J.H. Kim, Y.H. Lim, et al. 2013. Influence of genetic polymorphisms on the association between phthalate exposure and pulmonary function in the elderly. *Environmental Research* 122: 18–24.

Peck, J.D., A.M. Sweeney, E. Symanski, et al. 2010. Intra- and inter-individual variability of urinary phthalate metabolite concentrations in Hmong women of reproductive age. *Journal of Exposure Science and Environmental Epidemiology* 20: 90–100.

Perez-Gomez, B., R. Pastor-Barriuso, M. Cervantes-Amat, et al. 2013. BIOAMBIENT.ES study protocol: Rationale and design of a cross-sectional human biomonitoring survey in Spain. *Environmental Science and Pollution Research* 20: 1193–1202.

Perharic, L., and P. Vracko. 2012. Development of national human biomonitoring programme in Slovenia. *International Journal of Hygiene and Environmental Health* 215: 180–184.

Perrier, F., L. Giorgis-Allemand, R. Slama, et al. 2016. Within-subject pooling of biological samples to reduce exposure misclassification in biomarker-based studies. *Epidemiology* 27: 378–388.

Philippat, C., M.S. Wolff, A.M. Calafat, et al. 2013. Prenatal exposure to environmental phenols: Concentrations in amniotic fluid and variability in urinary concentrations during pregnancy. *Environmental Health Perspectives* 121: 1225–1231.

Pirkle, J.L., L.L. Needham, and K. Sexton. 1995. Improving exposure assessment by monitoring human tissues for toxic chemicals. *Journal of Exposure Analysis and Environmental Epidemiology* 5: 405–424.

Pollack, A.Z., N.J. Perkins, L. Sjaarda, et al. 2016. Variability and exposure classification of urinary phenol and paraben metabolite concentrations in reproductive-aged women. *Environmental Research* 151: 513–520.

Polder, A., C. Thomsen, G. Lindstrom, et al. 2008. Levels and temporal trends of chlorinated pesticides, polychlorinated biphenyls and brominated flame retardants in individual human breast milk samples from Northern and Southern Norway. *Chemosphere* 73: 14–23.

Preau, J.L., L.Y. Wong, M.J. Silva, et al. 2010. Variability over 1 week in the urinary concentrations of metabolites of diethyl phthalate and di(2-ethylhexyl) phthalate among eight adults: An observational study. *Environmental Health Perspectives* 118: 1748–1754.

Puklova, V., A. Krskova, M. Cerna, et al. 2010. The mercury burden of the Czech population: An integrated approach. *International Journal of Hygiene and Environmental Health* 213: 243–251.

Quiros-Alcala, L., B. Eskenazi, A. Bradman, et al. 2013. Determinants of urinary bisphenol A concentrations in Mexican/Mexican-American pregnant women. *Environment International* 59: 152–160.

Raab, U., U. Schwegler, U. Preiss, et al. 2007. Bavarian breast milk survey--pilot study and future developments. *International Journal of Hygiene and Environmental Health* 210: 341–344.

Reeves, K.W., J. Luo, S.E. Hankinson, et al. 2014. Within-person variability of urinary bisphenol-A in postmenopausal women. *Environmental Research* 135: 285–288.

Romano, M.E., N.L. Hawley, M. Eliot, et al. 2017. Variability and predictors of urinary concentrations of organophosphate flame retardant metabolites among pregnant women in Rhode Island. *Environmental Health* 16 (1).

Romero-Franco, M., R.U. Hernandez-Ramirez, A.M. Calafat, et al. 2011. Personal care product use and urinary levels of phthalate metabolites in Mexican women. *Environment International* 37: 867–871.

Ronningen, K.S., L. Paltiel, H.M. Meltzer, et al. 2006. The biobank of the Norwegian Mother and Child Cohort Study: A resource for the next 100 years. *European Journal of Epidemiology* 21: 619–625.

Saah, A.J., and D.R. Hoover. 1997. “Sensitivity” and “specificity” reconsidered: The meaning of these terms in analytical and diagnostic settings. *Annals of Internal Medicine* 126: 91–94.

Saoudi, A., N. Frery, A. Zeghnoun, et al. 2014. Serum levels of organochlorine pesticides in the French adult population: The French National Nutrition and Health Study (ENNS), 2006-2007. *Science of The Total Environment* 472: 1089–1099.

Sathyanarayana, S., R. Grady, E.S. Barrett, et al. 2016. First trimester phthalate exposure and male newborn genital anomalies. *Environmental Research* 151: 777–782.

Schoeters, G., H.E. Den, A. Colles, et al. 2012. Concept of the Flemish human biomonitoring programme. *International Journal of Hygiene and Environmental Health* 215: 102–108.

Schulz, C., A. Conrad, K. Becker, et al. 2007. Twenty years of the German Environmental Survey (GerES): Human biomonitoring – temporal and spatial (West Germany/East Germany) differences in population exposure. *International Journal of Hygiene and Environmental Health* 210: 271–297.

Sexton, K., L.L. Needham, and J.L. Pirkle. 2004. Human biomonitoring of environmental chemicals. *Am Sci* 92: 38–45.

Smith, K.W., J.M. Braun, P. Williams, et al. 2012. Predictors and variability of urinary paraben concentrations in men and women, including before and during pregnancy. *Environmental Health Perspectives* 120: 1538–1543.

Sobus, J.R., R.S. DeWoskin, Y.M. Tan, et al. 2015. Uses of NHANES biomarker data for chemical risk assessment: Trends, challenges, and opportunities. *Environmental Health Perspectives* 123: 919–927.

Solomon, G.M., and P.M. Weiss. 2002. Chemical contaminants in breast milk: Time trends and regional variability. *Environmental Health Perspectives* 110: A339–A347.

Spaan, S., A. Pronk, H.M. Koch, et al. 2015. Reliability of concentrations of organophosphate pesticide metabolites in serial urine specimens from pregnancy in the Generation R Study. *Journal of Exposure Science and Environmental Epidemiology* 25: 286–294.

Su, P.H., Y.Z. Chang, H.P. Chang, et al. 2012. Exposure to di(2-ethylhexyl) phthalate in premature neonates in a neonatal intensive care unit in Taiwan. *Pediatric Critical Care Medicine* 13: 671–677.

Sundstrom, M., D.J. Ehresman, A. Bignert, et al. 2011. A temporal trend study (1972-2008) of perfluorooctanesulfonate, perfluorohexanesulfonate, and perfluorooctanoate in pooled human milk samples from Stockholm, Sweden. *Environment International* 37: 178–183.

Teeguarden, J.G., A.M. Calafat, X.Y. Ye, et al. 2011. Twenty-four hour human urine and serum profiles of bisphenol A during high-dietary exposure. *Toxicological Sciences* 123: 48–57.

Teitelbaum, S.L., J.A. Britton, A.M. Calafat, et al. 2008. Temporal variability in urinary concentrations of phthalate metabolites, phytoestrogens and phenols among minority children in the United States. *Environmental Research* 106: 257–269.

Teitelbaum, S.L., N. Mervish, E.L. Moshier, et al. 2012. Associations between phthalate metabolite urinary concentrations and body size measures in New York City children. *Environmental Research* 112: 186–193.

Tellez-Rojo, M.M., A. Cantoral, D.E. Cantonwine, et al. 2013. Prenatal urinary phthalate metabolites levels and neurodevelopment in children at two and three years of age. *Science of The Total Environment* 461: 386–390.

Townsend, M.K., A.A. Franke, X.N. Li, et al. 2013. Within-person reproducibility of urinary bisphenol A and phthalate metabolites over a 1 to 3 year period among women in the Nurses' Health Studies: A prospective cohort study. *Environmental Health* 12: 80.

UNEP. 2013. Global chemicals outlook – Towards sound management of chemicals. https://sustainabledevelopment.un.org/content/documents/1966Global%20Chemical.pdf. Accessed 9 Dec 2017.

Upson, K., S. Sathyanarayana, A.J. De Roos, et al. 2013. Phthalates and risk of endometriosis. *Environmental Research* 126: 91–97.

U.S. EPA. 1985. Lead poisoning: A historical perspective. https://archive.epa.gov/epa/aboutepa/lead-poisoning-historical-perspective.html. Accessed 9 Dec 2017.

———. 2008. 5-Chloro-2-(2,4-dichlorophenoxy) phenol (Triclosan): Risk Assessment for the Reregistration Eligibility Decision (RED) Document. Case No 2340. DP Barcode 343544. PC

Code: 054901, U.S. Environmental Protection Agency. https://archive.epa.gov/pesticides/reregistration/web/html/index-316.html. Accessed 9 Dec 2017.

———. 2013. America's children and the environment, Third Edition (ACE3). http://www.epa.gov/ace/index.html. Accessed 9 Dec 2017.

Valvi, D., N. Monfort, R. Ventura, et al. 2015. Variability and predictors of urinary phthalate metabolites in Spanish pregnant women. *International Journal of Hygiene and Environmental Health* 218: 220–231.

Vandentorren, S., F. Zeman, L. Morin, et al. 2011. Bisphenol-A and phthalates contamination of urine samples by catheters in the Elfe pilot study: Implications for large-scale biomonitoring studies. *Environmental Research* 111: 761–764.

Veyhe, A.S., D. Hofoss, S. Hansen, et al. 2015. The Northern Norway Mother-and-Child Contaminant Cohort (MISA) Study: PCA analyses of environmental contaminants in maternal sera and dietary intake in early pregnancy. *International Journal of Hygiene and Environmental Health* 218: 254–264.

Vukavic, T., M.M. Vojinovic, I. Mihajlovic, et al. 2013. Human milk POPs and neonatal risk trend from 1982 to 2009 in the same geographic region in Serbia. *Environment International* 54: 45–49.

Waldron, H.A. 1973. Lead poisoning in the ancient world. *Medical History* 17: 391–399.

Watkins, D.J., M. Eliot, S. Sathyanarayana, et al. 2014. Variability and predictors of urinary concentrations of phthalate metabolites during early childhood. *Environmental Science & Technology* 48: 8881–8890.

Weiss, L., T.E. Arbuckle, M. Fisher, et al. 2015. Temporal variability and sources of triclosan exposure in pregnancy. *International Journal of Hygiene and Environmental Health* 218: 507–513.

Weuve, J., B.N. Sanchez, A.M. Calafat, et al. 2006. Exposure to phthalates in neonatal intensive care unit infants: Urinary concentrations of monoesters and oxidative metabolites. *Environmental Health Perspectives* 114: 1424–1431.

WHO. 2011. Toxicological and health aspects of bisphenol A. Report of Joint FAO/WHO Expert Meeting 2–5 November 2010 and Report of Stakeholder Meeting on Bisphenol A 1 November 2010. http://whqlibdoc.who.int/publications/2011/97892141564274_eng.pdf. Accessed 9 Dec 2017.

Whyatt, R.M., X.H. Liu, V.A. Rauh, et al. 2012. Maternal prenatal urinary phthalate metabolite concentrations and child mental, psychomotor, and behavioral development at 3 years of age. *Environmental Health Perspectives* 120: 290–295.

Wielgomas, B. 2013. Variability of urinary excretion of pyrethroid metabolites in seven persons over seven consecutive days-implications for observational studies. *Toxicology Letters* 221: 15–22.

Yan, X., A. Calafat, S. Lashley, et al. 2009. Phthalates biomarker identification and exposure estimates in a population of pregnant women. *Human and Ecological Risk Assessment* 15: 565–578.

Ye, X.Y., L.Y. Wong, A.M. Bishop, et al. 2011. Variability of urinary concentrations of bisphenol A in spot samples, first morning voids, and 24-hour collections. *Environmental Health Perspectives* 119: 983–988.

Zeman, F.A., C. Boudet, K. Tack, et al. 2013. Exposure assessment of phthalates in French pregnant women: Results of the ELFE pilot study. *International Journal of Hygiene and Environmental Health* 216: 271–279.

Zhang, Y.H., X.Z. Meng, L. Chen, et al. 2014. Age and sex-specific relationships between phthalate exposures and obesity in Chinese children at puberty. *Plos One* 9: e104852.

Zheng, W., W.H. Chow, G. Yang, et al. 2005. The Shanghai Women's Health Study: Rationale, study design, and baseline characteristics. *American Journal of Epidemiology* 162: 1123–1131.

Chapter 3
Considerations for Measuring Exposure to Chemical Mixtures

L. Blair Paulik and Kim A. Anderson

Abstract Exposure to chemical mixtures contributes to human disease risk. While analytical capability is continuing to increase, many chemicals remain understudied with regard to environmental occurrence and to toxicity. If a commercial chemical standard does not exist for a given chemical, then little to no quantitative data likely exists for that chemical. This chapter discusses exposures to organic chemical mixtures, lists necessary considerations for studying those exposures, and highlights research needs to continue to advance mixtures exposure science. When planning or reviewing studies that focus on exposure to chemical mixtures, important considerations include: spatial orientation of sampling, temporality of sampling, bioavailability of measured chemicals, measuring enough of the appropriate chemicals, potential for chemical transformations, and mixture effects. Importantly, relatively little is known about how exposure to mixtures of chemicals differs from exposure to individual chemicals. The majority of toxicity studies are performed using individual chemicals, so characterizing the toxicity of chemical mixtures should be a priority for the scientific community.

Keywords Exposure assessment · External exposure · Environmental chemistry · Environmental toxicology · Chemical mixtures · Sampling strategies · Passive sampling · Effect-directed analysis

3.1 Introduction: Measuring Chemical Stressors in the Environment

Nonscientists are often surprised to learn that science is incapable of measuring *all* of the chemicals people are exposed to on a daily basis. In reality, data about environmental chemical exposures only exists for chemicals for which detection

L. B. Paulik · K. A. Anderson (✉)
Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR, USA
e-mail: kim.anderson@oregonstate.edu

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_3

methods have been developed. Those methods only exist for chemicals that (a) have existing analytical standards and (b) are stable enough to reasonably be measured (Baird 2015). Additionally, the chemicals that meet these criteria are not always the ones that are the most relevant to environmental or human health, either with regard to toxicity or frequency of exposure (Brack 2003). In fact, we are exposed to many chemicals on a daily basis for which data about toxicity and/or environmental fate are sparse or nonexistent. To add another layer of complexity to this problem, people are constantly exposed to *mixtures* of chemicals. While we know that chemicals in mixtures do not necessarily behave the same as they do individually, we have limited information about what those differences are (Chap. 13, Cedergreen et al. 2008). Regardless researchers, risk assessors, and other decision-makers must regularly determine whether exposures to measured levels of pollutants are concerning or not. In the face of these uncertainties and data gaps, these can be difficult decisions to make.

This chapter includes considerations for assessing exposure to mixtures of organic chemicals in the environment. These considerations should be addressed both when designing studies to answer questions about chemical exposures and when interpreting results of such studies. Specifically, characterization of exposures to common chemical mixtures (Sect. 3.2), considerations for assessing exposure to these mixtures (Sect. 3.3), case studies to illustrate these considerations (Sect. 3.4), and priority research needs to improve the assessment of exposure to chemical mixtures (Sect. 3.5) are discussed.

3.2 Exposure to Common Chemical Mixtures

3.2.1 Why Characterize Exposure to Chemical Mixtures?

Accurately assessing a person's chemical exposures is essential for determining the impact of environmental exposures on human health. It is well established that exposure to many environmental pollutants is linked with disease (Wild 2012). While the science of externally measuring chemical exposures has come a long way, there is still much work to be done. The main principle of toxicology is that the dose determines the effects. It follows that a toxic substance does not increase disease risk if no exposure occurs. Consequently, there can be no useful assessment of toxicological risks without appropriately and quantitatively assessing chemical exposure. Therefore, it is important to continue to learn both about what chemicals should be prioritized for study and about how, where, and when chemical exposures happen. The importance of this area of study was emphasized in the new Strategic Plan from the U.S. National Institute of Environmental Health Sciences (NIEHS). This plan aims to characterize human exposure to chemical mixtures over a lifetime and to learn how those exposures may affect health and disease risk (Birnbaum 2012).

Although the risk of developing disease is attributed to both genetic and environmental factors, it has been suggested that differences in environmental factors may substantially affect disease risk (Willett 2002; Rappaport and Smith 2010; Wu et al. 2016). While the exact proportion of disease risk that can be attributed to environmental factors is under debate (Tomasetti and Vogelstein 2015; Wu et al. 2016), it is established that environmental factors, including environmental pollution, play an important role in disease risk (Lichtenstein et al. 2000; Wu et al. 2016). Although recent definitions of a person's "environment" have encompassed the internal, external, and psychological environment, and it has been suggested many of these varied factors may contribute to disease risk (Wild 2012), the relationship between exposure to environmental pollutants and health remains a key area of study. Additionally, there are many chemical exposures that are still undercharacterized (Wild 2012). To move exposure science forward, the scientific community needs more accurate and comprehensive assessments of chemical exposure. It is also important that these studies take into account how exposures vary with space and time.

3.2.2 How to Characterize Chemical Mixtures in the Environment: What Are the Options?

3.2.2.1 Existing Technologies for Measuring Chemical Exposures in the Field

3.2.2.1.1 Traditional Methods

Epidemiologists frequently evaluate the effect of environmental exposures on disease using data from questionnaires. Inferences gleaned from these questionnaires often require extrapolating from a participants' response to a chemical exposure and can therefore be uncertain (Rappaport 2011). For instance, a questionnaire may ask participants whether they grilled meat during the course of a study, and a "yes" response may be interpreted as exposure to carcinogenic chemicals that are products of incomplete combustion. However, this response would not provide information about which chemicals participants were exposed to, what concentrations they were exposed to, or the duration or frequency of this exposure.

All of this extra information could affect how this exposure is interpreted by researchers, but none of it would be captured by the questionnaire. Situations like this likely increase uncertainty in data from questionnaires. Thus, much care should be taken when either developing exposure assessment questionnaires or interpreting data from such questionnaires (Nieuwenhuijsen 2005). Additionally, epidemiologists rarely have access to baseline or historical exposure data. The lack of low-cost, easy-to-use sampling technology for directly measuring chemicals in the environment, or for directly measuring exposure to those chemicals, hinders

epidemiological studies. In rare cases when exposure assessments are based on data from environmental sampling, they often are limited in sampling time points or sampling locations, are limited to readily accessible tissues, and are limited in subject size due to challenges of cost and compliance.

Exposures to chemical mixtures can be measured either internally (e.g., with biomonitoring) or externally. As described in the previous chapter (Chap. 2), biomonitoring has both substantial strengths and limitations. Biomonitoring has the advantage of directly measuring concentrations in participants' bodies. When a sufficiently specific and sensitive biomarker exists for a contaminant of concern, the direct nature of this tool can greatly reduce uncertainty about whether participants were exposed to that contaminant. However, biomonitoring studies can be hindered by the lack of specificity and sensitivity of many biomarkers, temporal challenges related to timing of exposure, and a relatively short list of measureable exposures. Another challenge is that biomonitoring estimates are transient snapshots of exposure, which can be difficult to interpret (Schwartz et al. 2005). Collecting biological samples from study participants also presents additional challenges for participants and can therefore have a negative impact on participant compliance compared to less invasive methods.

Exposure to environmental pollutants happens via three main exposure routes: inhalation, ingestion, and dermal. Additionally, characterizing chemical pollution for health assessments may entail measuring chemical mixtures in four broad types of media: sediment/soil, water, air, and food products. It is important to keep in mind that measuring chemicals in these media is only useful where it addresses exposures that may occur through at least one of the three main exposure routes (inhalation, ingestion, and dermal). While the last 50–60 years saw significant advances in analytical chemistry (e.g., enabling chemists to detect as little as a few femtograms of thousands of chemicals), much less progress has been made in how we collect samples from the environment. Additionally, while it is encouraging that analytical chemists can now detect a few thousand chemicals, this does not mean that methods exist to detect all of these chemicals in all types of samples (e.g., soil, water, air, food). And there are still numerous chemicals present in the environment for which no analytical methods exist at all.

Below we discuss existing methodologies for externally *measuring* chemicals in the environment, for use in estimating human health risks associated with exposure to those chemicals:

There are a number of existing technologies for measuring exposures externally. Some traditional measurement techniques are unique to one of the three main exposure routes, while others bridge multiple pathways. In general, people can be exposed to chemicals measured in the sediment or soil through dermal exposure, and potentially through ingestion (if a child ingests the contaminated soil or sediment directly, or if that chemical moves from the soil or sediment into a plant or shellfish that develops in the soil or sediment). People are generally exposed to chemicals measured in water dermally, through ingestion if a chemical moves from the water into a fish that a person eats, through ingestion if that chemical is not

filtered out before the water is used as drinking water, or through inhalation of volatiles/semi-volatiles vaporized from water. Exposures to chemicals measured in the air happen predominantly through inhalation. Finally, exposures to chemicals measured in foods occur through ingestion.

There are multiple techniques for measuring chemicals in each of the media listed above. To measure chemicals in the sediment/soil, traditional methods include taking what is known as a "grab sample" of sediment/soil, extracting the whole sediment/soil, and reporting the chemicals measured in the whole sample. Traditional methods of measuring chemicals in water include taking grab samples of water, extracting chemicals from that water, and reporting the totality of what is measured. Traditional methods of measuring chemicals in air involve setting up an active air sampler which pumps air over a filter or other sorbent, and then extracting that filter. To measure exposure via ingestion of food grown at a contaminated site, traditional methods include directly sampling the food of concern and then performing an extraction and measuring chemicals in that food.

Extraction techniques are also often specific to the type of sample (e.g., air or sediment) and to a certain chemical or class of chemicals (e.g., only for mercury or only for hydrocarbons). The near-infinite list of combinations of chemicals and sample types necessitates developing many different extraction methods, and this further complicates the task of efficiently measuring chemicals in the environment.

When coupled with appropriate analytical techniques, each of the traditional methods described above yields chemical concentrations that are representative of the specific location and time of sampling. Given that exposures to pollutants in the environment are dynamic, it is necessary to extrapolate from environmental concentrations to estimate human exposure. However, data from traditional sampling methods are commonly used to assess pollution levels that an individual, or a population, is exposed to. This is done using exposure factors to approximate how much of a chemical a person might be exposed to on a regular basis. Depending on the exposure route of interest, different exposure factors are required to make this extrapolation. Exposure factors range from estimating the number of hours individuals in a population spend in their houses each day to the average number of grams of crayfish individuals eat each day, to the number of years individuals live adjacent to a source of pollution, and everything in between. Specific examples of many of these exposure factors that are commonly used in the U.S. Environmental Protection Agency (U.S. EPA) risk assessment are given in the U.S. EPA's 2011 Exposure Factors Handbook (U.S. EPA 2011).

Other chemical sampling techniques aim to measure chemical exposures in an individual's personal environment, by placing samplers directly on individuals. These methods are attractive because they eliminate (or greatly reduce) the need to extrapolate between a measured chemical concentration and an external exposure. Traditionally, techniques for measuring personal exposure to organic contaminants have included putting air-sampling backpacks on study participants (Perera et al. 2003) or putting active samplers on participants' lapels (Tsai and Vincente 2001). However, these tend to be costly, monitor relatively few environmental

contaminants, and require committed participants because the devices can be bulky or noisy during use. Monitors exist to measure other types of contaminants, such as particulate matter in the air, in real time (Chung et al. 2001). Real-time monitors for freely dissolved organic contaminants, however, are still in development. While not a focus of this chapter, other techniques exist for measuring personal exposure to inorganic contaminants, such as metals. One such technique uses an acidic hand wipe to assess personal exposure to metals (Lidén et al. 2008). Existing tools for measuring personal organic chemical exposure all require some in-lab analysis after deployment. Sampling devices for individuals should ideally be easily worn, adaptable, capable of measuring many chemicals, easy to use by research staff, integrated with many simultaneous measures (e.g., location, chemical exposures, health outcomes like lung function), rugged, and noninvasive and have the capacity to store and transmit data.

3.2.2.1.2 Passive Sampling

Thinking outside the "sampling jar" is required to accurately and efficiently characterize chemical exposures. Since the early 1990s, passive sampling has been gaining momentum as an effective tool for measuring trace levels of contaminants of concern (Petty et al. 2000; O'Connell et al. 2014). Passive samplers measure time-integrated concentrations of the freely dissolved concentration (C_{free}) of contaminants. Passive samplers are relatively low cost, and they do not require energy or maintenance while deployed. The ability to infuse passive samplers with performance reference compounds (PRCs) before deployment further improves this tool's ability to accurately assess contaminant levels (Huckins et al. 2006).

Numerous polymer materials and membrane technologies for passive sampling of water, air, and the personal environment have been explored. Low-density polyethylene (LDPE) and silicone are two widely used materials to make passive samplers for measuring organic contaminants (Anderson et al. 2008; O'Connell et al. 2014). When deployed in air, water, or sediment porewater, these polymers absorb hydrophobic organic contaminants via simple diffusion from the environment into the hydrophobic membrane. This process is analogous to uptake across a phospholipid membrane into an organism, making passive samplers well-suited to serve as surrogates for contamination in organisms (Booij et al. 2006; Allan et al. 2011, Fernandez and Gschwend 2015; Paulik et al. 2016a; Forsberg et al. 2014). Other samplers, such as Diffusive Gradient Thin Films, or DGT, passively absorb inorganic contaminants (Pérez and Anderson 2009a; Vrana et al. 2005; Gimpel et al. 2003).

Another attractive aspect of passive sampling is that some passive samplers can be paired with high-throughput bioassay systems, such as the embryonic zebra fish assay. This allows chemical mixtures measured in the environment (by a passive sampler) to be tested in toxicity bioassays (Allan et al. 2012). The zebra fish bioassay is also attractive because it requires relatively small amounts of sample. Traditional animal bioassays are much more expensive and time-consuming and

require much larger amounts of sample. By pairing a passive sampler extract with the zebra fish assay, the bioassay helps reveal the toxicity of the whole mixture. More can be learned when bioassays are combined with fractionation techniques to separate components of the mixture based on physical or chemical parameters (e.g., polarity, size) (Burgess et al. 2013). This will be discussed further below. While there has been significant progress with passive technologies, the nexus of the passive sampling platform technology is yet to be fully exploited.

Recently new technology has emerged which passively samples individuals' chemical exposures. The passive wristband sampler is a new, wearable passive sampler that directly measures the chemicals a person is exposed to (O'Connell et al. 2014). These wristbands accumulate thousands of common environmental chemicals (O'Connell et al. 2014), including many used in commerce and organic compounds formed during natural and industrial processes. The wristband can also capture emerging chemicals of concern. For example, it quantifies the concentrations of oxygenated polycyclic aromatic hydrocarbons (OPAHs), toxic compounds in asphalt fumes (O'Connell et al. 2014). Another promising aspect of the wristband sampler is that it is easily used by citizen scientists, making it easier to engage community members in research studies.

3.2.2.1.3 Limitations of Passive Sampling:

As with any technology, passive sampling has its limitations. One challenge is accurately calculating an environmental (or a personal exposure) concentration of a chemical from what is measured in a passive sampling device. This requires adding labeled chemicals to the passive sampler before deployment to act as PRCs and measuring the loss of those PRCs during the deployment. The amount of PRC lost is used to estimate whether the sampler had reached equilibrium with the system for that chemical. This information is then used to correct the concentration measured in the sampler, to more accurately reflect the concentration in the environment. While the scientific community accepts this approach (Huckins et al. 2006), it does introduce some uncertainty into reported concentrations of environmental chemicals that are measured using passive samplers (Khairy and Lohmann 2012; Allan et al. 2009). Another aspect to consider is that passive samplers produce time-integrated concentrations of chemicals (Greenberg et al. 2014; Petty et al. 1993). This is a strength or weakness of the technology depending on the goals of the study. It is a strength if the goal is to assess a person's total individual chemical exposure over a period of time or to measure an average concentration of a chemical over a given time period (Tidwell et al. 2016; Khairy and Lohmann 2012). It is a weakness, however, if the researcher wants to capture the elevated concentration at a specific moment during a pulse of contaminants (e.g., during a brief emission of air pollution).

3.2.2.2 Existing Technologies for Measuring Chemical Exposures in the Lab

3.2.2.2.1 Traditional Methods: Which Chemicals Are Measured?

Traditional analytical methods only include the relatively few chemicals that are feasible to measure. These are not necessarily the chemicals that should be prioritized due to being frequently encountered in the environment or due to being the most toxic (Brack 2003). In some cases, chemicals have been prioritized for study based solely on availability of analytical standards (Baird 2015). Additionally, if a standard is mistakenly misidentified, this can lead to years of incorrect conclusions about that chemical (Baird 2015). There are many chemicals, therefore, for which environmental occurrence data, toxicity data, or both are sparse. There are other chemicals for which these data do not exist at all. While this chapter focuses on human exposures to chemical mixtures, the contaminants we discuss also affect the environment as a whole.

For many chemicals found in the environment, toxicity information is limited, or regulatory limits do not exist. Therefore, even when these chemicals are included in analytical methods and detected in the environment, they are often excluded from risk assessments. If there is no toxicity information for a measured chemical, it is impossible to include it in a risk assessment. Several approaches exist to close some of these toxicity data gaps. Many of these approaches involve predicting toxicity of understudied chemicals (one such set of predictive tools is quantitative structure-activity relationships, or QSARs). The U.S. EPA's ToxCast is a promising initiative, which aims to efficiently predict and characterize the toxicity of thousands of chemicals (U.S. EPA 2015). In this program, EPA scientists screen thousands of chemicals using high-throughput toxicity bioassays. This enables researchers to prioritize which chemicals to study based on toxicity.

The gap between environmental measurements and toxicity is further widened as most dose-response studies (Wright and Welbourne 2001) are not conducted at environmentally relevant doses and are not performed with mixtures that are realistically found at contaminated sites. Conventional sampling methods may allow pollutants with unknown toxicological relevance to be overlooked, but passive samplers begin to address this problem. Although passive samplers are not exhaustive tools for chemical extraction from the environment, they can be designed to extract a wide range of contaminants. PAHs (Baussant et al. 2001; Lohmann et al. 2001; Sun et al. 2008), polychlorinated biphenyls (Anderson and Johnson 2001; Sethajintanin and Anderson 2006), pesticides (Sethajintanin and Anderson 2006), flame retardants (Booij et al. 2002), dioxins (Lohmann et al. 2001), and metals (Pérez and Anderson 2009b; Pérez and Anderson 2009a; Zhang et al. 1995; Zhang and Davison 1995) are all examples of contaminants that can be captured by passive samplers. Additionally, passive samplers provide real-world mixtures of chemicals that can be directly integrated with bioassays for toxicity testing. This is useful when studying contaminants for which toxicity data

are sparse. Oxygenated PAHs (OPAHs) are an example of a group of chemicals that is often encountered in the environment but for which relatively little toxicity information exists. OPAHs, also known as PAH ketones or quinones, have one or more oxygen-containing functional groups attached to the aromatic ring structure and may also contain other chemical groups (Wischmann and Steinhart 1997). It is possible to use passive samplers to simultaneously sample PAHs (which are well-studied) and OPAHs, allowing the researcher to answer multiple research questions with the same sampling campaign.

3.2.2.2.2 Unmonitored and Infrequently Monitored Chemicals: Approaches for Identification and Toxicity Exploration

Many commonly encountered chemicals (which may or may not contribute substantially to exposures affecting human health) are not currently included in analytical methods. In some cases, it is possible to glean more data from existing analytical techniques. For instance, it is possible to look more closely at a chromatogram obtained from high-resolution GC-MS analysis of an environmental sample and to identify unmonitored chemicals in that sample. Additional analytical approaches that can be used to identify nontarget chemicals include nuclear magnetic resonance (NMR) and two-dimensional gas chromatography (GC x GC). If unmonitored chemicals are identified in this way, then their toxicities can be explored. These approaches involve searching through analytical data and trying to identify individual chemicals that are not traditionally monitored, using chemical parameters (e.g., structure, molecular weight, charge). These techniques are often very time intensive (requiring days, weeks, or even months), and there are often no analytical standards available to confirm chemical identification. This creates the potential to misidentify chemicals (potentially causing errors in databases) due to multiple chemicals having the same chemical formula, but differing in structure. Another challenge is that, the methods and software for identifying unmonitored chemicals are often written by individual investigators in individual research labs. This can make it difficult to reproduce or compare data among different laboratories.

Another method for identifying nontarget chemicals is known as "effect-directed analysis," or EDA. In this approach, environmental chemical mixtures are separated through a series of chemical and/or physical separations, creating different chemical fractions of the mixture (Burgess et al. 2013; Brack 2003) (see Fig. 3.1). Fractionation techniques include normal-phase chromatography, size exclusion chromatography, and many others. After fractionation, EDA can help researchers learn which fractions (and which chemicals) are causing toxicity in the mixture. EDA is also useful to help researchers identify which chemicals in mixtures are *not* causing observed toxicity. Which fractionation techniques are most appropriate, and which fractions are of most interest, depends on the specific research question being asked. EDA enables the researcher to use these fractions in toxicity bioassays, to explore the toxicities of each fraction, and ultimately to elucidate which

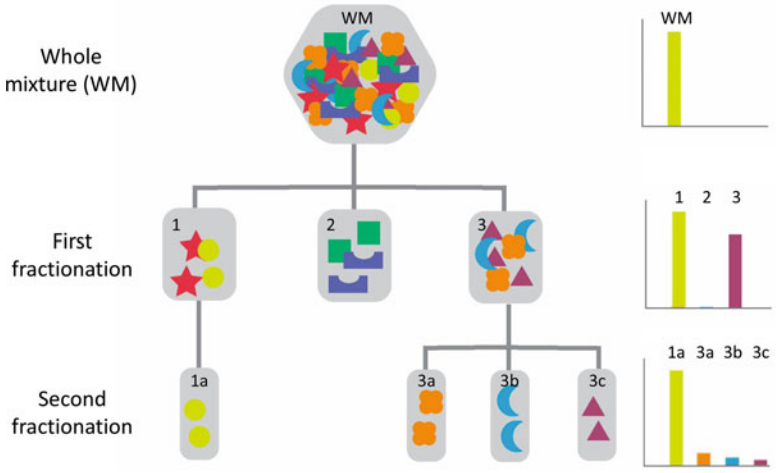

Fig. 3.1 A whole chemical mixture is separated into fractions using effect-directed analysis (EDA). The toxicity in the whole mixture is depicted by the yellow bar in the top graph. In the first fractionation, the whole mixture (WM) was separated into three fractions (1–3). Both fraction 1 and 3 contained some toxicity, but no toxicity was observed in fraction 2. The toxicity of fraction 1 appears to be the same as the toxicity of the whole mixture. However, there is additional toxicity in fraction 3. This suggests that there was less than additive toxicity in the whole mixture before fractionation. This illustrates the power of EDA to help researchers identify both toxic and inert fractions within chemical mixtures. In the second level of fractionation, fractions 1 and 3 were separated into fractions 1a and 3a–c. In these fractions, we see that the toxicity of fraction 1a is the same as the toxicity of fraction 1, and of the whole mixture. This could be interpreted to mean that the majority of toxicities observed in the original mixture and in fraction 1 were coming from the yellow circles in the mixtures. When these circles were completely isolated in fraction 1a, this fraction retained the same level of toxicity as in the whole mixture and in fraction 1. However, in this final fractionation, we also see that the sum of the toxicities of fractions 3a–c is less than the toxicity of fraction 3. This suggests that there was the potential for greater than additive toxicity in fraction 3 before the second fractionation. This example illustrates the utility of EDA to help researchers identify: components that are causing toxicity within a mixture, mixtures (and fractions of mixtures) where greater or less than additive toxicity is observed, and fractions of mixtures that do not appear to cause toxicity but may influence the toxicity of the whole mixture

components of the chemical mixture are causing toxicity. This is a useful tool for identifying both environmental occurrence and toxicity of chemicals that are rarely or infrequently included in environmental monitoring or sampling. In a recent review, Brack suggested that EDA is especially useful for measuring nontarget compounds at sites that are known to be heavily polluted (Brack 2003).

3.3 Considerations for Assessing Exposure to Chemical Mixtures

3.3.1 Field Methods: Appropriate Sampling Techniques

3.3.1.1 Spatial Considerations: Stationary (or Population-Based) vs. Individual

In many cases, chemical levels measured at stationary sampling sites in air, water, sediment/soil, or food are used to estimate exposure levels for people living nearby or for larger populations (e.g., cities or regions). To use a concentration measured at a specific location in the environment to estimate a personal exposure level, researchers must make assumptions about how often, and for how long, individuals in a population are exposed to the contaminant in that environmental compartment through inhalation (if measured in air), dermally (if measured in water or sediment/soil), or through ingestion (if measured in food). A good example of this is when data from immobile air-monitoring stations are used to estimate human health risks based on inhaling those contaminants. A few limitations with this approach are that fixed monitoring stations may be many miles away from the location(s) of exposure, they are not always continuously operated, they often only monitor a few chemicals, and they are likely not reflective of indoor air. Thus, there is a lot of uncertainty inherent in this approach that reduces its utility as an accurate surrogate for an individual's inhalation exposure.

Examples of typical exposures occurring through the three main exposure routes, in different locations, are shown in Fig. 3.2. Each exposure route requires extrapolation with different exposure factors. For instance, to estimate inhalation exposure, the researcher must estimate how many hours per day, and days per year, individuals would be exposed to contaminants at the levels measured at the stationary air-sampling site. For ingestion exposure, on the other hand, the researcher would need to estimate how often the consumer ate the food that was sampled and how much of it they ate each time. These estimated exposure levels are used in risk assessment to determine whether the exposure exceeds acceptable regulatory levels of pollution. Estimating these exposure factors can increase uncertainty in exposure estimates, because not all members of a population have the same behaviors. The benefit of stationary or population-based measurements is that they can reduce costs, as a small number of samples can be used to estimate exposure levels for large numbers of people.

More recently, methods have been evolving to sample chemical exposures directly on individuals. One such tool is the passive wristband sampler, introduced in Sect. 3.2.2 above. When used in risk assessment, personal sampling techniques require much less extrapolation than stationary environmental sampling techniques. When the sampler is on a participant constantly, there is no need to make assumptions regarding the frequency of exposure, duration of exposure, or other factors that are required when making exposure inferences based on chemical

Fig. 3.2 People are exposed to chemical mixtures, for example PAHs, through a variety of scenarios every day. Here, snapshots of a typical day depict four different exposure scenarios. This progression illustrates the spatial and temporal components of exposure and provides examples of common exposures occurring through the three main exposure routes. On the left, an ingestion exposure is shown in the woman's kitchen. This could include ingesting chemicals like PAHs in cereals. Next, as the woman commutes to work in heavy traffic, an inhalation exposure is shown. She may be exposed to elevated levels of PAHs and other chemicals emitted from car exhaust while driving to work. Next, additional examples of inhalation and ingestion exposure are shown as she walks in the city. She may be exposed to PAHs and chemical mixtures by inhaling chemicals emitted from the myriad anthropogenic pollution sources present in a congested city. On the right, examples of inhalation, ingestion, and dermal exposure are shown in this recreational scene. While enjoying a campfire, there is the potential to inhale PAHs and other chemical mixtures from smoke, to ingest chemicals present in food that has been roasted over fire, and to experience dermal exposure. Taken together, these vignettes represent a realistic suite of PAH and chemical mixture exposures that could occur in the daily lives of many people. All of the exposure scenarios described, as well as others that are not pictured in these four snapshots, would combine to yield this person's cumulative daily PAH and chemical mixture exposures

concentrations at stationary monitors. Therefore, using personal sampling tools may provide more accurate individual exposure information, reducing uncertainty in exposure estimates used in risk assessments. For example, if a personal sampler were put on the woman in Fig. 3.2, it would accumulate exposure information from all four of those exposure scenarios, as well as others throughout her day. However, if the goal is to estimate exposure for a population, extrapolation would still be required from the exposures of the individuals measured to the larger population. Additionally, depending on which sampling techniques are used, it may be prohibitively costly to sample enough individuals to characterize a population.

Moving forward, successful strategies for assessing exposures to mixtures of environmental chemicals will likely incorporate both stationary/population-based and individual sampling techniques. When choosing a sampling technique for a study assessing chemical exposures, it is important to consider which of these techniques best addresses the question being asked. In some cases, the best option may be to use a combination of sampling tools.

3.3.1.2 Temporal Considerations: Temporality of Sampling Should Match Temporality of Exposure

Understanding how temporality of sampling compares to temporality of exposure is another challenging aspect of assessing exposure to chemical mixtures. These two should be as well matched as possible, and temporality should be considered both within the day and within or among the year(s). For instance, if the goal is to study an exposure to emissions from a factory, and those emissions vary throughout the day, it would be important to sample for at least one full day to account for that variability. At the same time, if exposures are most likely to occur at one time of the day, then it may be appropriate to emphasize this time in sampling. Examples of exposures occurring sequentially throughout a day are visualized in Fig. 3.2. There is also seasonal variation in levels of some chemicals in the environment (Sower and Anderson 2008; Brun et al. 2004). This seasonal variation can come from emissions varying with changing practices in different seasons or from climactic conditions changing available concentrations. If the goal is to assess an exposure that would occur over many years, best practice would be to sample at multiple points throughout the year to gain as much information as possible about the potential exposures. If an exposure would only occur at one time of year, over a lifetime, best practice may be to sample at the same time in multiple years. Additionally, exposures at different stages of life may lead to different toxic responses. An exposure to a developing fetus may have a much more substantial impact on health risk than an exposure to a healthy adult (Wild et al. 2013; Wild 2012). It is therefore important to capture the effects of exposures at various stages of life, especially during life stages where people may be more susceptible (e.g., during development, or during old age). The importance of assessing exposures at susceptible life stages, to more accurately assess exposure to chemical mixtures over the entire life span, was identified by the NIEHS in their most recent strategic plan (Birnbaum 2012).

3.3.1.2.1 Time-Integrated Concentration vs. Grab Sample

As mentioned above, traditional "grab samples," only allow assessment of contaminant levels in the environment at one moment in time. In contrast, passive samplers sequester chemicals over time, yielding time-integrated concentrations. This is useful because passive samplers can capture less frequent, acute episodes of exposure. It is also useful because passive samplers absorb contaminants over time, making them good surrogates for the fraction of contaminants sequestered by an organism over time in the same environment. However this makes it important to consider the length of deployment, especially if the goal is to sample a brief pulse of contamination, as in episodic or catastrophic events (e.g., after spills or hurricanes). It may be most appropriate in these cases to use short deployment times.

3.3.1.3 Bioavailability: Fraction of Chemicals Sampled Should Match the Fraction to Which the Individual Is Exposed

It is generally accepted that measuring the total amount of an individual chemical concentration in the environment is not enough to predict biological effects. Conventional methods for human exposure assessment involve measuring total contaminant concentrations in the ambient environment (e.g., water, sediment) and extrapolating to toxicological endpoints; however this approach has proven ineffective (Alexander 2000; Petty et al. 2004; Mayer et al. 2014). Measuring total ambient contaminant concentrations yields at best rough estimates of exposure (Schwartz et al. 2005) and in many cases does not reflect the fraction of contaminants to which people or organisms are actually exposed (Allan et al. 2006; Spacie et al. 1995). Instead, measuring the bioavailable fraction is thought to be most relevant to human and ecological health. This is the contaminant portion to which an organism is directly exposed or that a person is exposed to when eating a contaminated organism. Scientists call this the "freely dissolved fraction," or "C_{free}," in water or sediment porewater. Measuring the freely dissolved fraction of a contaminant is imperative when assessing chemical bioavailability, toxicity, mobility, and degradation (Alexander 2000; Escher and Hermens 2004; Mayer et al. 2014). For most routes of exposure and health endpoints, it is the freely dissolved, or unassociated, form of hydrophobic contaminants that is transported across biological membranes of organisms and may exert toxic effects (Suffet et al. 1994; Escher and Hermens 2004). A decrease in freely dissolved contaminants directly reduces bioavailability and vice versa. It is therefore this bioavailable fraction that is thought to be the most relevant fraction to measure to understand exposures.

The bioavailable fraction of contaminants can be quantified using various analytical approaches. Passive sampling is well-suited to measure the bioavailable fraction of chemicals in water and sediment/soil, as passive samplers mimic the uptake of a cell or organism via both chemical and physical processes. Passive samplers may be used to assess contaminant levels, and subsequently exposure, in water (Ke et al. 2007; Utvik et al. 1999), air (Bartkow et al. 2004), and sediment/soils (Tao et al. 2008; Huckins et al. 1990, 2004; Gimpel et al. 2003; Wennrich et al. 2002; Anderson and Johnson 2001; Lohmann et al. 2001; Harner et al. 2003; Wania et al. 2003; Martin et al. 2003; Kingston et al. 2000; Wells and Lanno 2000; Burgess et al. 2017). A major advantage of passive samplers is the ability to distinguish between dissolved and bound molecules, rather than assessing the mere presence or absence of chemicals (Mayer et al. 2003). One important consideration, however, is that when passive samplers are used to measure contaminants in air, they only absorb the fraction of contaminants in the vapor phase. Given that both the vapor phase and particulate-bound phase of contaminants may be inhaled, this means that passive sampler-generated assessments of air contamination may be underrepresentative of total contaminants available for inhalation exposure. One definition of the bioavailable fraction (when measuring mixtures of chemicals in

water) is the portion of contaminants that can be taken up by an organism. It has been observed that this fraction can be accurately measured using passive samplers (Booij et al. 2006; Huckins et al. 2006; Forsberg et al. 2014; Paulik et al. 2016a).

3.3.1.4 Measuring the External Aspect of Chemical Exposure

Timely, high-quality data are needed to bridge the gap between environmental monitoring data and quantitative data about individual chemical exposures and to identify which of these exposures are most relevant to human health. In recent years there has been a movement toward measuring complete, lifelong exposures (Wild 2005, 2012; Wild et al. 2013). The goal of this work is to learn as much as possible about the relationship between environmental exposures and disease risks and to use these findings to improve public health decision-making (Wild 2012). Similarly, one of the goals of the U.S. NIEHS' recent Strategic Plan is to "transform exposure science by enabling consideration of the totality of human exposures," (Birnbaum 2012). There is evidence that environmental factors contribute heavily to disease risk and that environmental chemical exposures are an important piece of the environment that may affect that risk (Wild 2012; Wild et al. 2013). It has been suggested that interdisciplinary teams of scientists should work together to tackle this challenge with innovative technologies (Wild et al. 2013). Effectively assessing cumulative chemical exposure will likely require collaboration among chemists, toxicologists, immunologists, public health specialists, epidemiologists, and others. From the perspective of measuring external chemical exposures, it will be important to choose environmental sampling techniques that measure as much of a person's chemical exposures as possible, as accurately as possible. Personal sampling devices, such as the passive wristband sampler, may be some of the best existing tools to address this piece of the exposure assessment puzzle.

3.3.2 Lab Methods

3.3.2.1 Which Chemicals Should Be Measured?

The chemicals that are most heavily studied are not necessarily those that people are most commonly exposed to, or that are most toxic, but rather that existing methods can detect. While the goal is to measure chemicals that are the most relevant to environmental and human health, there is also preference toward measuring chemicals for which regulatory guidelines exist (Rappaport 2011; Baird 2015). In many cases this is appropriate (i.e., a potentially hazardous chemical has been previously identified and regulated, and so it receives attention). The challenge is that this can make it difficult to study certain chemicals (e.g., a new chemical that does not fall into an existing regulatory category). Developing methods to detect chemicals requires analytical standards, and each analytical standard must be

created in response to demand for that specific standard. So, there are many chemicals for which no standards exist and thus for which there is little to no data about their environmental occurrence or toxicity.

Most analytical methods are only used to quantify a few dozen chemicals. However, in most cases, there is additional information that could be harnessed from existing chromatographic analyses. Hundreds of additional chemicals could be quantified from many existing chromatographic methods, if the methods were further developed. By the time the sample is ready for analysis, much of the expense of the sample collection and processing has already been incurred. Often, precious research funds could produce more results if methods were further optimized. Generally only a few chemicals are quantified, when the same sample could be used to quantify hundreds or thousands of additional chemicals in the mixture.

Chemicals that share certain properties are often lumped into a group, or "class," as if all chemicals in a class have the same mode of action and toxicity. In reality, not all chemicals in a class behave the same. Chemicals within a given class can follow wildly different environmental pathways after emission and also can have different toxic modes of action and potencies.

3.3.2.1.1 Example Chemical Class: Polycyclic Aromatic Hydrocarbons (PAHs)

One class of environmental pollutants that has diverse physicochemical properties and modes of toxic action is polycyclic aromatic hydrocarbons, or PAHs. PAHs are pervasive environmental pollutants of concern, associated with hydrocarbon extraction and adverse health impacts (U.S. EPA 2010; Ravindra et al. 2008). The main categories of health concerns associated with exposure to PAH mixtures are cancer risk and respiratory distress. Some PAHs are pro-carcinogens, meaning they can be metabolically activated to create biologically active intermediates which can form DNA adducts (Baird et al. 2005). Thus, research has focused primarily on PAHs' carcinogenic risk (IARC and Monographs 2005; IARC and Monographs 2009). PAH-related cancer risk has been studied in relation to oil spills, traffic exhaust, wood smoke, and cooking. However, exposure to PAHs also increases the risk of cardiovascular disease (Lee et al. 2011) and the risk of mortality from heart attack (Burstyn et al. 2005; Lee et al. 2011). Animal studies indicate that PAHs can increase blood pressure and heart rate and accelerate the progression of atherosclerosis (Gentner and Weber 2011; Knaapen et al. 2007; Penn and Snyder 1988, Wang et al. 2009). Mechanistic evidence and epidemiological evidence associate PAHs with airway inflammation and asthma (Al-Daghri et al. 2013; Lee et al. 2005), and exposures are associated with developmental and behavioral deficits (Dejmek et al. 1999; Perera et al. 1999, 2009; Dejmek et al. 2000; Wu et al. 2010; Perera et al. 2006). There are also multiple biological effects and targets for PAHs that remain unknown or understudied. Additional toxic endpoints that have been studied in relation to PAHs include adverse developmental, reproductive, respiratory, and neurological effects (Herbstman et al. 2012; Perera et al. 2006; Miller et al. 2010;

Miller et al. 2004; Rosa et al. 2011). All of this evidence suggests that assuming the potencies and modes of action of all PAHs are the same is an oversimplification.

Benzo[a]pyrene, or BaP, has been extensively studied in relation to its carcinogenicity. It is used as a model carcinogenic PAH in many studies and in regulatory guidance (U.S. EPA 2010). While there is substantial evidence that BaP is indeed a carcinogenic PAH, it is often incorrectly assumed that it is so widely studied because it is *the most* carcinogenic PAH. However, research suggests that other PAHs are as or more carcinogenic than BaP. Analytical methods are lacking for some PAHs. For others, measurement methods are available, but they are rarely if ever encountered in the environment, and so they are not practical to study with regard to human exposures. However, there are some highly carcinogenic PAHs that are both measurable and found in the environment and that receive much less research attention than BaP. One such PAH is dibenzo[a,l]pyrene (also known as dibenzo[def,p]chrysene, or DBC) (Baird 2015). DBC was included in the U.S. EPA's 2010 list of carcinogenic PAHs (U.S. EPA 2010). In this document, relative potency factors (RPFs) were assigned to 26 unsubstituted PAHs, to scale their carcinogenic potencies relative to that of BaP. BaP was assigned an RPF of 1, and the rest of the compounds' RPFs were scaled relative to BaP. DBC was given an RPF of 30, suggesting that it is 30 times as carcinogenic as BaP at the same concentration. While there are uncertainties inherent in this assessment, it is worth noting that DBC is still not included in regular environmental monitoring regimes. This may simply be the natural progression of the identification of hazardous chemicals. Momentum must gain behind a chemical before it can truly be well-characterized. For comparison, the study of the carcinogenicity of BaP began in the late eighteenth century, when English surgeon Percivall Pott observed that chimney sweeps in London had higher rates of scrotal cancer. In the 1930s, what is now known as BaP was directly isolated from 2 tons of coal tar, and it was identified as a cancer-causing agent (Baird 2015). Since that time, the carcinogenicity of BaP has been demonstrated in numerous studies (U.S. EPA 2010). Given the advances that science has made since the initial identification of BaP as a carcinogen, identifying and prioritizing toxic chemicals should be much more efficient now.

It is similarly incorrect to assume that all PAHs behave the same in the environment. For instance, four examples of PAHs commonly measured in the environment are naphthalene, phenanthrene, pyrene, and benzo[e]pyrene. These four PAHs have pure water solubility values of 32, 1.0, 0.1, and 0.004 mg/L, respectively, ranging about five orders of magnitude. This means that the fate of each of these PAHs is very different once it is released into the environment. Additionally, while water solubility values are reported for pure water above, few environmental waters are even close to pure. As the amount of dissolved organic carbon (DOC) in water increases, the solubility of organic contaminants increases as well. For instance, Johnson-Logan et al. demonstrated the solubility of the pesticide chlordane in groundwater with a mere 34 mg/L DOC increased 500% (Johnson-Logan et al. 1992). The enhanced solubility is due to partitioning of hydrophobic organic contaminants onto the dissolved organic carbon within the water column. An increase in DOC can increase solubility, but it may or may not

increase bioavailability. However, it most certainly affects fate and transport of the chemical.

PAH exposures occur to complex mixtures of PAHs, and the composition of these mixtures can differ dramatically depending on the source(s) (Tidwell et al 2016; Allan et al. 2011; Tidwell et al. 2016). For instance, the vapor-phase PAH profile from woodburning is different from diesel exhaust, which is different from petroleum. Stout et al. recently observed that relying only on the U.S. EPA's 16 priority pollutant PAHs can inhibit the researchers' ability to determine the source of the PAHs (Stout et al. 2015). Diagnostic isomer ratios and alkylation patterns of PAHs are therefore commonly used to identify sources of PAHs (Yunker et al. 2002; Tobiszewski and Namieśnik 2012; Stout et al. 2015). Different sources emit PAH mixtures with differing magnitudes of individual PAHs and with different ratios of PAH isomers. Including more alkylated PAHs and more isomers in analysis allows for more robust source identification. It is also important to understand exposure to individual PAHs, because they have different modes of action. For instance, Jung et al. found that childhood asthma was associated with pyrene but not as strongly associated with five other measured PAHs (Jung et al. 2012).

Another interesting facet of PAHs is that exposures occur through all three of the main exposure routes (inhalation, ingestion, and dermal contact, with inhalation and ingestion typically being the primary routes). Respiratory PAH burden includes exposure to both PAHs in the vapor-phase and the particulate-bound fraction. Air sampling is often focused on determining the concentration of particulate-bound chemicals, but exposure to PAHs in the vapor phase has also been shown to contribute to the cancer risk from inhalation exposure (Ramírez et al. 2011; Tsai et al. 2002; Liu et al. 2007). Hassan et al. demonstrated that 67% of inhalable PAHs were in the vapor phase at a study site in Giza, Egypt (Hassan and Khoder 2012). Significant effort has been put toward clarifying the association between PAH inhalation and increased incidence of respiratory syndromes, especially asthma and lung cancer (Karimi et al. 2015, Kuo et al. 1998), so accurately understanding PAH levels in air is important.

3.3.2.2 Accounting for Transformations of Chemicals in the Environment, Between the Source and the Exposure

The importance of chemical transformations in the environment is often underrepresented in study designs to measure chemicals in the environment and interpretation of environmental chemical occurrence data. Chemical exposure estimates are often calculated based on emission reports from point sources, from large-volume chemical use reports, or from environmental monitoring programs. However, many of these reporting techniques only focus on short lists of chemicals and do not account for chemical transformations that occur in the environment. Degradation, adsorption, transport, and other chemical fate processes can transform the chemical composition of a point source emission.

Chemical transformations may lead to less or more toxic chemical mixtures than what is measured at an emission source. In some cases, transformations that alter the parent compound after emission may make a product that is less toxic than the parent (Baird 2015). In other cases, transformative processes can create chemicals that are more toxic than what was in the original mixture. These transformation products are often not included in analytical methods, and so they often cannot be detected, and thus are entirely missed from the exposure discussion. For many degradation products, toxicity information is sparse. In some cases, specific transformation products that lead to increased toxicity after transformation may not have been identified. Interestingly, EDA has been identified as a useful tool for identifying specific transformation products that are causing toxicity in transformed mixtures (Brack 2003).

Regardless, transformations often make the exposure experienced by an individual different from the original source of emissions. The magnitude of this difference depends on the chemicals in the original mixture (and their potential for transformations) and on the time and distance from the emission source to the exposure. Exposure to sunlight, microbial activity, changing temperature, and precipitation are all factors that could transform chemicals in the environment, after they leave their sources and before exposure occurs.

Additionally, it is important to consider that PAH mixtures redistribute in the environment. For instance, if a certain mixture of PAHs is emitted from a source into the air, that mixture will be different at a sampling device 10 miles away, even without chemical transformations. This is partly due to differences in physical and chemical parameters of chemicals, which dictate the environmental fate of each individual chemical. For instance, if a mixture of PAHs is emitted into the air, more of the higher molecular weight PAHs may partition preferentially into the soil, while more of the lower molecular weight PAHs may remain in the air. Thus, the mixture measured in those two matrices would not look the same as the mixture at the source.

One example of compounds that are commonly formed through environmental transformations, may have greater toxicity than their parent compounds, and are relatively understudied are oxygenated PAHs or OPAHs (Lundstedt et al. 2007; O'Connell et al. 2014). OPAHs can be formed from parent PAHs in the environment, through chemical oxidation, photo-oxidation, or biological transformation (Lundstedt et al. 2007). There is evidence that some OPAH compounds are more toxic than the unsubstituted parent PAHs (Lampi et al. 2006; Lundstedt et al. 2007; Yu 2002, Bamforth and Singleton 2005; Knecht et al. 2013). However, they are rarely included in environmental monitoring. When designing a study, it is important to consider how much transformation may have occurred in the environment for the chemicals of interest. If there is the potential for transformations, then both the degree of transformation of the parent compounds and the creation of new transformation products should be considered. It is worth considering how these transformations might alter the toxicity of the mixture a receptor is exposed to compared with the toxicity of the mixture at the original source. An exposure experienced by an individual near an emission source might be quite different

from an exposure experienced by an individual that is distant from the source. This example illustrates that personal monitoring can provide more accurate estimates of exposure than stationary monitoring.

Transformation products that are formed through environmental degradation processes are sometimes the same products that are created during in vivo metabolism. This can mean that the same chemical could enter a person's body both exogenously (as from a source where it was used in the environment) and endogenously (through metabolism of a different chemical). This can lead to errors in assessment of chemical exposure. One example of this is that mammalian metabolism of organophosphate pesticides (OPs) results in formation of a series of dialkyl phosphates (DAPs) that are excreted in the urine. It has often been assumed that the concentration of DAPs in the urine is directly related to dietary exposure to parent OPs. Improved analytical ability to measure DAPs in urine has resulted in an increase in studies using this biomarker of OP exposure. However, the same enzyme-mediated oxidation and hydrolysis reactions that produce DAPs in humans are also responsible for the transformation of OPs in the environment. Thus, DAPs are formed in the environment and are present on many foods before they are ingested. When DAPs are consumed, they do not change in vivo and are excreted (Forsberg et al. 2011). Traditionally, studies of the environmental fate of OPs have not analyzed for DAPs. This means that if all urinary DAPs are assumed to be from OP exposures, then this is likely a significant overestimate of exposure. Given that OPs have known health effects, and DAPs are generally considered nontoxic, it is important to distinguish between exposure to OPs and exposure to DAPs. However, this has not traditionally always been done. This is an example of the importance of considering whether transformations in the environment are possible with the contaminants of interest and how those transformations may impact the data.

3.3.2.3 Considerations for Assessing Exposures to Mixtures of Environmental Chemicals

3.3.2.3.1 Limited Data on Mixture Interactions

The health community has traditionally evaluated the toxicity of one compound at a time, identified its health effects, and determined an acceptable exposure level for regulatory use (e.g., a permissible exposure limit in the case of occupational exposures regulated by OSHA in the United States). However, we are constantly exposed to *mixtures* of chemicals in our daily lives. Limited data exists concerning how exposure to these mixtures may differ from exposure to the individual components of the mixture. In mixtures such as car exhaust fumes, the toxicity of PAHs may be additive, greater than additive, or less than additive (Siddens et al. 2012). Siddens et al. demonstrated that PAH mixtures are more potent skin carcinogens compared to benzo[a]pyrene than the EPA's Relative Potency Factor approach would suggest (Siddens et al. 2012). Novel approaches to assess the health impacts of mixtures are needed.

3.3.2.3.2 Risk Assessment of Mixtures: Appropriateness of Assuming Additivity?

Current methods of estimating risk from exposure to chemical mixtures often assume that the toxicities of individual chemicals from the same class (e.g., PAHs) are dose-additive. This assumption requires presuming that the mode of action of each PAH is the same. However, it has been observed that interactions between individual compounds in a mixture can lead to a greater or less than additive toxic responses (Chap. 13, Cedergreen et al. 2008). One way to improve risk estimates would be to directly test the mixtures of chemicals measured in the environment in toxicity assays. While testing the endless list of potential mixtures would be prohibitively challenging, a starting place may be to assess the combined toxicities of mixtures that commonly occur in the environment. One example of this could be the mixture of PAHs measured in water or an organism at a site contaminated by a common profile of pollution sources (Paulik et al. 2016a). This would require close communication between exposure scientists and mixture toxicologists. For instance, a recent study measured an almost identical mixture of 10 carcinogenic PAHs in crayfish tissues, in a Superfund site 10 years apart (Paulik et al. 2016a). Given that the same mixture was measured in the crayfish collected 10 years apart in this study, it may be a useful mixture to test in bioassays to learn more about mixture effects.

EDA is a promising framework that can be used to assess the toxicity of chemical mixtures measured in the environment. It is especially attractive because it allows the researcher to break down the chemical mixture piece by piece and explore the toxicity of each of these fractions (Fig. 3.1) (Brack 2003; Burgess et al. 2013). This technique has potential to help researchers characterize which components of a chemical mixture are contributing which aspects to the observed toxicity. Importantly, EDA may also eliminate broad groups or classes of chemicals that are not contributing to toxicity. It can also help elucidate how various chemicals behave in the presence of the whole mixture, relative to just in the presence of a chemically similar fraction of that mixture, relative to individually. With new low-volume bioassays, the fractionation process can be scaled accordingly, and this allows for rapid turnaround of fractionations and bioassay assessments on the order of days.

Historically EDA would employ strong extraction techniques, such as Soxhlet extraction, of samples before fractionation. However these extracts yielded fractions that bore little connection to the chemical exposures. Soxhlet extraction involves using a strong solvent, high heat, and/or elevated pressure to extract as much contaminant as possible from an environmental sample (e.g., sediment). Thus, the concentration of contaminants measured using Soxhlet extraction may be much higher than what would truly be bioavailable in the environment. Often chemicals may be toxic at high levels, but are not bioavailable in situ at concentrations sufficient to cause toxicity. EDA aims to characterize the toxicity of mixtures. So, if it is performed using chemical mixtures that are not representative of true environmental concentrations, this defeats the purpose. One way to avoid this problem is to use passive sampling to collect mixtures for use in EDA. Passive

samplers absorb bioavailable fraction of contaminants when deployed in the environment. Passive samplers are extracted using simple solvent extractions, without elevating temperature or pressure. This yields more realistic estimates of the bioavailable fraction of contaminants in the environment. These estimates are more relevant for use in EDA than artificially heightened concentrations that may be obtained using other techniques. After extraction, EDA employs multiple rounds of fractionation using various techniques (e.g., normal-phase chromatography, size exclusion chromatography, etc.) to separate the various chemicals within the mixture. These fractions can then be individually used in toxicity bioassays, allowing the researcher to further elucidate which parts of the chemical mixture may cause toxicity and which parts are not causing toxicity.

All of these tools and more need to be employed for exposure science to begin to address the substantial data gaps surrounding exposures to mixtures of chemicals in the environment.

3.3.2.3.3 Importance of Looking for Risk-Driving Chemicals

In the absence of comprehensive data describing exposures to all chemical mixtures, the majority of chemical risk assessments assume that the toxicities of chemicals in a mixture are additive (i.e., conform to an assumption of either dose addition or independent action). This approach requires assuming that there are no pharmacokinetic or pharmacodynamic interactions among chemicals. As long as this is the paradigm, it will be especially important to measure the chemicals that are most impactful to health risks. The importance of measuring the more toxic PAHs in the mixture (or, at least, the PAHs with the highest known toxicities) for use in risk assessment was illustrated in a recent study measuring PAH mixtures in crayfish tissue (Paulik et al. 2016a). In this work, it was observed that risk estimates were slightly higher than estimates presented in a previous public health assessment for the study area. One factor likely increasing the estimates in the study was that the study's analysis used the EPA's 2010 RPF approach to scale the carcinogenic potency of the PAH mixture relative to that of benzo[a]pyrene (U.S. EPA 2010). When this was combined with an analytical method that quantifies 23 of the 26 PAHs that were given RPFs by the EPA 2010 (U.S. EPA 2010) document, risk estimates increased relative to previous methods. This increase is due both to quantifying more PAHs (e.g., relative to looking for the EPA's 16 priority pollutant PAHs) and to using the extended list of potency values presented by the EPA in 2010 (compared to a shorter list, such as the EPA's 1993 list of 7 RPFs (U.S. EPA 1993)). Additionally, the crayfish tissue contained a few of the PAHs from the EPA 2010 document that have RPFs higher than benzo[a]pyrene. Even a small quantity of a highly potent chemical can change the risk assessment picture. For instance, benzo[c]fluorene was measured in some of the crayfish tissues presented in this study (Paulik et al. 2016a). This compound has an RPF of 20, suggesting that if a sample had the same concentration of benzo[a]pyrene and of benzo[c]fluorene, benzo[c]fluorene would contribute *20 times* more carcinogenic risk. However, this

compound is not included in traditional monitoring programs and is not in the EPA's previous priority pollutant list (Keith 2014). As the EPA's 16 priority pollutant list is still often used as the standard PAHs to measure, benzo[c]fluorene or other carcinogenic PAHs such as DBC (discussed above) could often be present in the environment without being detected. Assuming that these newer estimates are accurate would suggest that previous risk assessments, which only included the EPA's shorter list of priority pollutants, may have been inadvertently underrepresenting the potential for risk.

3.3.3 Summary of Considerations for Interpreting Data from Exposure Assessment Studies

It is important to consider the factors mentioned in this section when interpreting data from studies assessing environmental exposures. *Below is a list summarizing the main considerations outlined above:*

- *Spatial orientation and choice of sampling technique*: Does the chosen sampling technique fit the research question? For instance, does the choice between a personal or stationary sampler make sense? If one was used and not the other, consider how the choice of sampling technique may affect the results. It is also important to consider sources of uncertainty related to each sampling technique.
- *Temporality of sampling technique:* Was the sampling method timed appropriately to capture the exposure scenario of interest? How does temporality of the sampling event compare to expected temporality of exposure? If they are different, what impact might that have on the results?
- *Is the appropriate fraction of contaminants measured?* Did the sampling strategy measure the fraction of contaminants that a person or other biological receptor would be exposed to, via the exposure route of interest? Did the sampling design measure the bioavailable fraction of contaminants? If not, what impact could that have on results?
- *What chemicals are measured?* Are they the appropriate ones to answer the question or enough of the appropriate ones to answer the question? Can the toxicity of the entire mixture be explored (e.g., through EDA)? Could more chemical data be gleaned from the analytical techniques that were used?
- *Transformations of contaminants in the environment*: Is there potential for transformations to have occurred between the chemicals' source and the sampling site or the exposure site? If so, was this addressed? If not, what impact might that have on the results?
- *Mixture effects*: Was the challenge of assessing exposure to mixtures addressed? Was a form of dose-additivity assumed? If so, is that consistent with EPA guidance for mixtures risk assessment (e.g., (U.S. EPA 2000))? How might any assumptions, or data gaps, about exposures to the mixtures being studied affect results?

3.4 Case Study: Ingestion and Inhalation/Personal

Here a two-part case study is presented, illustrating many of the concepts discussed in the preceding sections. This case study focuses on measuring PAHs associated with unconventional natural gas extraction (NGE) and addressing potential exposures associated with these emissions. In this exposure scenario, there is potential for people living or working nearby to be exposed to PAHs through two of the dominant exposure routes (ingestion and inhalation). There may also be potential for dermal PAH exposure, predominantly for individuals working near NGE well pads, but this will not be a focus of this case study. The case study is broken into two pieces, with one outlining a potential pathway for ingestion exposure (Sect. 4.2) and another describing an application of personal passive samplers to assess individual PAH exposures, predominantly through inhalation (Sect. 4.3). The sidebars in this section draw attention to how the considerations described in Sect. 3.2.2 of this chapter are relevant to the case study presented here.

3.4.1 Introduction: PAHs and NGE

Natural gas extraction (NGE) has expanded rapidly in the USA in the last 15 years. This has been largely due to improvements to the technologies of hydraulic fracturing and horizontal drilling, which liberate previously inaccessible gas reserves from shale (EIA 2011). Due to the influence of these techniques, this activity has broadly been referred to as "fracking" and has begun to receive more attention from both the public and regulators. Despite this increase in NGE in the U.S. interior, there has been relatively little investigation of the impacts NGE activity may have on health. A few studies have measured impacts to air quality, predominantly concluding that NGE contributes contaminants to the environment (Colborn et al. 2014; McKenzie et al. 2012, McKenzie et al. 2014; Paulik et al. 2016b).

Air emissions have been identified as one of the main potential exposures pathways through which NGE may impact the health of nearby communities or workers. There is evidence that NGE emits volatile organic compounds (VOCs) and semi-volatile organic compounds (SVOCs) at levels that may impact human health. However, there is still limited information about levels of these contaminants in the environment. One class of SVOCs that has recently been measured in the air associated with NGE emissions is polycyclic aromatic hydrocarbons (PAHs) (Paulik et al. 2016b; Colborn et al. 2014).

3.4.2 *Ingestion Case Study: Potential for Natural Gas Extraction to Affect Nearby Wheat Crops and Local Wheat Consumers*

The potential for crop plants to sequester PAHs has not been considered in relation to NGE. It is established that PAHs enter plants through dry deposition from air (Simonich and Hites 1994a, b). This phenomenon has been specifically shown in agricultural crops such as wheat (Jones et al. 1989; Tuominen et al. 1988; Kobayashi et al. 2008). Wheat is a major U.S. crop (behind only corn and soybeans) and a substantial component of diets worldwide (USDA 2014). Thus, there is significant potential for changes in contaminant levels in wheat to cause changes in PAH-related health risk for wheat consumers. Here we consider the potential for PAHs emitted from NGE to be sequestered by wheat plants and to impact the PAH exposures of wheat consumers.

Ingestion is a major route of human exposure to PAHs (Phillips 1999). Additionally, cereals are one of the main constituents of ingestion-related PAH exposure for the average person in the USA (Menzie et al. 1992; Phillips 1999). This is due both to the ability of cereals such as wheat to sequester SVOCs including PAHs and to the large amount of cereals that are consumed as part of the Western diet. In contrast, smoked salmon can contain higher PAH levels per unit mass than cereal, yet eating smoked salmon presents relatively little risk for most people, due to low-average ingestion rates (Forsberg et al. 2012).

It is worth considering how PAH exposure from eating wheat may compare to an individual's overall PAH exposure. This has been given some thought previously by Kobayashi et al., comparing risk associated with consumption of PAHs in wheat to that of inhalation from living in the region where the wheat was collected (Kobayashi et al. 2008). The authors conclude that risk associated with eating wheat sampled in the study ranged from 0.05% to 76% of the inhalation risk for someone living in the study area (Kobayashi et al. 2008). While seasonal variation and differences in risk assessment parameters add considerable variability to this estimate, it is striking that PAHs from one food source could potentially contribute as much as 76% of a person's risk due to inhalation.

Considerations from Section 3, Applied to Case Study 4.2

- *Spatial orientation and choice of sampling technique*: Because relatively little is known about emissions from NGE, this is an instance where it would be important to carefully choose the appropriate sampling technique. For instance, it may be most relevant to sample wheat, but it may also be relevant to sample the air near the NGE wells. It would also be useful to consider the spatial component of sampling in this project and ideally to sample at various locations a range of distances from NGE. If,

(continued)

Considerations from Section 3, Applied to Case Study 4.2 (continued)

for instance, PAH levels in wheat are only affected when wheat plants are within a short distance from NGE wells, then this could potentially be mitigated by planting wheat a certain distance away from NGE activity. Both the sampling matrix and the spatial considerations would help the researcher assess whether PAHs were in fact moving from the NGE wells through the air and into wheat plants. If the concentration data suggested such movement (e.g., PAH levels elevated in air, soil, and wheat plants near NGE relative to levels measured farther from NGE), this would make a stronger case for NGE activity elevating PAH levels in wheat than if only wheat plants were sampled or if wheat were only sampled at one distance from NGE.

Another consideration with regard to spatial sampling is dilution of wheat grown near NGE, between the point when wheat leaves the farm and when it reaches the consumer's plate. This would have a big impact on the actual PAH levels to which wheat consumers were exposed. While the Western diet includes a large amount of wheat, determining definitively what fraction of the wheat consumed by an individual was grown near NGE would be difficult. Very few, if any, wheat consumers likely eat products containing wheat exclusively grown near NGE. Many processed foods likely contain wheat from a variety of locations. While there are many locations within the USA where wheat growing and NGE co-occur (NAWG 2015; EIA 2011), there would still likely be more locations where wheat was not grown near NGE. This could be addressed partially with careful spatial sampling. In addition to sampling at a range of distances close to NGE, researchers should also measure PAH concentrations in wheat grown far from NGE. Even if PAH levels in wheat grain were substantially elevated in wheat grown near NGE, this would have no impact on health risk to a consumer who only ate wheat grown far from NGE. On the contrary, if 100% of a consumer's wheat were grown near NGE, that consumer would experience an increase in health risk due to the increased PAH levels in wheat grain grown near NGE. This disparity in risk estimates could be mitigated by performing risk assessment using both the "best-case scenario" and "worst-case scenario"—assuming that 0% or 100% of a consumer's wheat was grown near NGE. This would bracket the range of potential effects dilution of wheat may have on health risk to wheat consumers. It may also be worth considering estimates assuming that 50% of a consumer's wheat was grown near NGE, or other ratios. None of that risk assessment flexibility would be possible if wheat were only sampled near NGE. The importance of this dilution potential would depend on the goals of study. If the study simply aimed to determine

(continued)

Considerations from Section 3, Applied to Case Study 4.2 (continued)

whether PAHs could feasibly travel from an NGE well pad into a wheat grain, then it would be appropriate only to sample wheat from fields at various distances from NGE well pads, including some with no NGE well pads nearby. However, if the goal was to directly assess potential human exposures to elevated PAH levels related to NGE emissions, then a much more sophisticated spatial sampling campaign would be needed, to account for transport of wheat around the country and the world, and the associated dilution of wheat grown close to NGE wells.

Additionally, if the goal of the study is to assess impacts of NGE on PAH exposure to wheat consumers, it should also be considered whether PAHs should be directly measured in consumables made from the wheat that came from the various sampling locations. This may be the most relevant sample to collect, depending on the specific goals of the study. Processed foods, for instance, may contain very little wheat grown near NGE and would likely have very diluted PAH levels. On the other extreme, if a consumer made bread with locally sourced wheat that was grown near NGE, this would represent minimal dilution of PAH levels in wheat. Thus, directly measuring PAHs in those consumer products would give researchers more information about what actual risks wheat consumers may face.

- *Temporality of sampling technique:* It would be important to consider temporality of sampling in this project. Specifically, it would be preferable to sample wheat in different years. PAH movement throughout the environment can change with environmental conditions (e.g., temperature, humidity, and precipitation), as well as with changing activity levels at the emission source. The concentration, as well as profile, of PAHs reaching the target site (in this case the wheat plant) can therefore vary year to year. Best practice would be to collect samples in various years, to most accurately estimate the range of potential exposure concentrations. For wheat samples, the grain would need to be thoroughly sampled at the time of harvest. This would ideally include sampling in multiple years. It would be especially important to consider temporal aspects of sampling design if the study aimed to approximate lifetime ingestion risk associated with exposure to measured concentrations. It could be misleading, for instance, if a concentration were only measured during 1 year, and then that concentration was extrapolated to represent the concentration of PAHs in wheat over a lifetime when assessing ingestion risks. The resulting estimate could inadvertently over- or underestimate risk.

  However, it would also be important to keep in mind that there is much uncertainty inherent in risk assessment. So, another approach could be to

(continued)

Considerations from Section 3, Applied to Case Study 4.2 (continued)

use a few samples to characterize the temporal variation expected in the samples and assess how this uncertainty compared to the overall uncertainty is inherent in the process. If it were estimated to only increase uncertainty slightly above existing levels of uncertainty, then an argument could be made for a simpler sampling design.

- *Is the appropriate fraction of contaminants measured?* When assessing the human health risk of ingesting PAHs in wheat, it would be important to consider the bioavailability of PAHs in wheat. This would require knowing whether the PAHs measured in wheat are capable of crossing the GI tract, thus reaching target sites where they may be metabolized and may potentially lead to DNA adducts. There is limited information in the scientific literature about the bioavailability of PAHs in wheat for uptake in humans. If PAHs are somehow inextricably bound to the wheat particles and not able to cross the GI tract, then no level of PAHs found in wheat would contribute additional human health risk. Fully answering this question would require further research. It would be an important consideration for the researcher to keep in mind and to communicate when communicating results of the project.
- *What chemicals are measured?* Because only a few studies have measured SVOCs emitted from NGE, and even less have measured PAHs in this context, the researcher should take care to measure PAHs that are most likely to both be emitted by NGE and be sequestered by wheat. Additionally, many studies traditionally only measure a small subset of PAHs (e.g., the EPA's 16 priority pollutants). Given that these are not necessarily the most toxic PAHs, it is important to measure as many as possible of the more toxic PAHs as well.
- *Transformations of contaminants in the environment*: This scenario has the potential for PAHs to be transformed between being emitted from the NGE well and deposited in the wheat plant. Given that the proposed mechanism of NGE altering PAH levels in wheat is via transport through the air, it would be important to consider transformations that might occur to parent PAHs emitted from NGE and how these may affect resulting contaminants measured in wheat. For instance, the presence of UV light may transform some parent PAHs to OPAHs. So, it may also be useful to analyze the air, soil, or wheat for OPAHs as well as PAHs. It would also be useful to consider how the composition of PAH mixtures might change between the emission source and being sampled in wheat, air, or soil. All of this would help the researcher learn more about how PAHs are moving through the system and about what is measured in wheat compares to what is emitted from NGE. Additionally, given that some OPAHs may be more

(continued)

Considerations from Section 3, Applied to Case Study 4.2 (continued)

toxic than the parent PAHs, sampling transformation products in all samples would inform the researcher more about toxicity of chemical mixtures in the system.

There could also be transformations that occur to PAHs after they are deposited on the plants' surface, after they have been absorbed into the plant, or during processing of wheat to make consumables. All of these would warrant consideration during study design and data interpretation.

- *Mixture effects*: If the PAH data measured in this study were used in risk assessment, it would be important to consider that mixtures of PAHs may or may not have toxicity equal to the toxicity expected under an assumption of dose addition. Given that present regulatory guidance (e.g., from the U.S. EPA (U.S. EPA 2010)) assumes that the carcinogenic risk of a PAH mixture is dose-additive, this is a challenge facing all PAH risk assessors and is not unique to this project. However, it would merit consideration.

3.4.3 Personal/Inhalation Exposure Case Study: Using a Personal Passive Sampler to Assess Individual Exposure to Natural Gas Extraction

The second part of this case study illustrates a real-world example of using an innovative personal passive sampling technology to assess individual exposure to air emissions from an understudied potential source of pollution. Instances where steps were taken to address sampling considerations raised in this chapter are highlighted in the sidebar.

3.4.3.1 Introduction to Individual Inhalation Exposure to NGE

While there is only a small amount of recent data regarding contaminant levels in the environment related to NGE, there is even less data about personal exposure to contaminants emitted from NGE. Some studies have used data from stationary monitors to estimate community-level health impacts (Paulik et al. 2016b; McKenzie et al. 2012; Bunch et al. 2014), while others have used results from questionnaires to approximate individual health impacts of NGE (Brasier et al. 2011; Bamberger and Oswald 2014; Rabinowitz et al. 2015). The majority of these studies have concluded that NGE has the potential to impact human health. Still other studies have concluded that much research is still needed to assess the public health impact of NGE. Several recent studies have addressed the need to directly measure the impact of NGE on environmental and human health (Goldstein et al. 2014; Penning et al. 2014; Adgate et al. 2014).

Some studies have used stationary monitors to measure VOCs or SVOCs in the air near NGE (Paulik et al. 2016b; Colborn et al. 2014; McKenzie et al. 2012). However, no study has directly measured the individual exposures of people living or working near NGE. This information would reduce the uncertainty in assessing environmental impacts of NGE. Knowing what contaminants an individual was exposed to would greatly improve the ability to assess risk(s) associated with living or working near NGE. Individualized exposure data yields much more accurate risk estimates, compared to approximating exposure from questionnaires or extrapolating exposure from stationary monitoring data. Estimates from questionnaire data are fraught with potential for miscommunication between the researcher and the respondent, while estimates from stationary monitors necessitate numerous assumptions about timing, frequency, and duration of exposure. Personal monitoring bypasses all of these uncertainties, by directly measuring the chemicals to which an individual is exposed.

A pilot study was conducted in a rural Ohio community that has been heavily affected by the U.S. natural gas boom. At the time of the study, this area was one of the most densely affected counties in Ohio, with more than one natural gas well pad per square mile. Because this area was historically rural (and thus had limited pre-existing anthropogenic sources of pollution, compared to an industrial area or a city), this community presented researchers with a good opportunity to measure any potential pollutant increases related to NGE. Volunteers were identified through collaboration with a concerned citizens group in the area.

3.4.3.2 Methods for Assessing Individual Inhalation/Personal Exposure to NGE

This study employed the use of a novel personal passive sampler, the passive wristband sampler (O'Connell et al. 2014). While the participant is wearing the wristband, it absorbs contaminants from the vapor phase in the surrounding air. The fraction of contaminants sequestered by the wristband is similar to the fraction that is inhaled by the participant, making the wristband a good surrogate for individual inhalation exposure. Compared to other personal sampling technologies, such as wearing a heavy backpack with a noisy active sampler or giving blood or urine samples for biomonitoring, the wristband is noninvasive and easy to use. This makes it easier for researchers to achieve high rates of compliance with the wristband, than with some more cumbersome or invasive traditional techniques. Thus, the wristband was selected as an ideal tool to estimate personal exposure to NGE.

To ensure that the wristband only sampled the contaminants each participant was exposed to, wristbands were transported between the lab and the study area and back in airtight Teflon bags. This study also engaged the participants as citizen scientists. At the time when the research team gave the participant the wristband,

each participant was trained in how to properly wear the wristband and in how to mail it back to the lab for chemical analysis. Importantly, this included education about how the technology worked, so that participants could be mindful not to accidentally bias their wristbands. Overall, this pilot study had over 91% compliance with the wristband, which is very promising.

Considerations from Section 3, Applied to Case Study 4.3

- *Spatial orientation and choice of sampling technique*: While the wristband has limitations, it is important to put it in the context of existing technologies for assessing personal exposure to chemicals. As described above, other techniques that have been used to address personal exposure to PAH emissions are questionnaires and stationary monitors, which both include substantial uncertainty about accurately assessing personal exposures to pollutants. Thus, as long as the wristband's limitations are considered in study design and data interpretation, the wristband is a good tool to address this research question. It is easy to use, and thus participants are more likely to use it and to use it correctly. Combined with the increased accuracy compared to questionnaires, tools like the wristband will be a useful addition to the science of measuring environmental exposures.

  From a spatial perspective, the wristband is a far more appropriate tool for assessing personal exposures than other existing tools. Simply, the wristband goes everywhere the participant goes and thus directly represents the person's exposure. Eliminating the spatial extrapolation that has historically been required to estimate exposure is a major strength of this technology.
- *Temporality of sampling technique:* Temporality of sampling would be important to consider in this project. Specifically, it would be preferable to sample the personal environment at different times of the year. As described above, PAH movement throughout a system can change with environmental conditions, as well as with changing activity levels at the emission source, and thus can vary with time of day and time of year. Thus, best practice would be to collect samples at various times throughout the year, to more accurately estimate the totality of potential exposure concentrations. Personal samples, collected using the wristband, have the advantage of sampling continuously over the course of multiple days. So, variation in PAH concentration throughout the day would automatically be taken into account by the sampling technology. It is especially important to consider temporal aspects of sampling design when the study aims to approximate lifetime risk (likely predominantly due to inhalation in this case) associated with exposure to measured concentrations. It could be misleading, for instance, if a concentration were only measured during one

(continued)

Considerations from Section 3, Applied to Case Study 4.3 (continued)

time of year, and then that concentration was extrapolated to represent the concentration of PAHs over a lifetime when calculating risk.

- *Is the appropriate fraction of contaminants measured?* The passive wristband sampler would measure the majority of what the individual is exposed to via inhalation. However, the wristband only samples vapor-phase contaminants in air and does not capture PAHs that are particulate-bound. Many of the more carcinogenic PAHs are higher molecular weight. Given that the particulate-bound phase typically contains a greater fraction of higher-molecular-weight PAHs, while the vapor phase typically contains a higher fraction of lower-molecular-weight PAHs, excluding the particulate phase means missing a fraction of the carcinogenic PAHs that are available for inhalation (Kameda et al. 2005, Hassan and Khoder 2012). Thus, the wristband would yield a concentration of PAHs in the air that is lower than the total. This would need to be acknowledged in any risk assessment using concentrations measured by the wristband.
- *What chemicals are measured?* As described in the first piece of this case study, it would be best to measure PAHs that are most likely to be emitted from NGE. However, as data are currently limited, this could be challenging to determine. If an aim was to perform human health risk assessment, it would be useful to measure as many of the risk-driving PAHs as possible. Additionally, after a sample has been collected and analyzed for PAHs in the lab, many more chemicals can often still be measured in that sample. This could be achieved by combining compatible analytical methods and by exploiting data that is already collected in analyses. This would allow researchers to further characterize chemical mixtures and learn more about chemical exposures.
- *Transformations of contaminants in the environment*: The considerations here would not differ markedly from the considerations described for transformations for the previous case study. In short, there would be potential for the PAHs emitted from NGE to transform before being absorbed by the wristband. Also, the profile of PAHs would likely change between the emission source and the wristband. This should be taken into account during data collection and analysis. Additionally, including OPAHs, or other chemicals that can be formed through transformations of parent PAHs in the environment, could help inform researchers about any potential transformation that occurred between PAH emissions and compounds sequestering into the wristband.
- *Mixture effects*: As in Case Study 4.2, if the PAH data measured in this study are to be used in risk assessment, it would be important to consider that mixtures of PAHs may or may not have toxicity expected under an

(continued)

Considerations from Section 3, Applied to Case Study 4.3 (continued)

assumption of dose-additivity of the individual PAHs in the mixture. Given that present regulatory guidance (e.g., from the U.S. EPA (U.S. EPA 2010)) assumes that risk associated with exposure to PAH mixtures is dose-additive, this is a challenge facing all PAH risk assessors and is not unique to this project. However, it would merit consideration.

The wristband generates chemical mixtures that represent personal exposures in the real world. This type of study could therefore be a useful piece of a larger project, by investigating the toxicity of those mixtures to learn more about potential nonadditive toxicity in mixtures.

3.5 Priority Research Needs for Assessing Exposure to Chemical Mixtures

3.5.1 Field

3.5.1.1 Develop Sampling Techniques that Accurately Assess Exposures, While Minimizing Cost and Maximizing Compliance

Optimal sampling techniques for any study ideally consider spatial and temporal factors of exposure, and whether personal or stationary/population-based sampling is most appropriate. As identified by the NIEHS's strategic plan, this must include considering how exposures may differ at more susceptible life stages (Birnbaum 2012). In many cases, a personal sampler may be the best tool for assessing exposures. To make this feasible, there is a need to develop more sampling tools that are cost-effective and have the ability to measure a wide range of chemicals. By continually measuring chemicals on an individual, these tools allow researchers to more accurately estimate what chemicals an individual is exposed to, over a longer period of time. These samplers can also be paired with toxicity bioassays and incorporated into pre-existing or new public health studies. The passive wristband sampler is an example of such a tool.

3.5.1.2 Move Toward Personal, in situ Sampling—More Representative of True Exposure

To best assess total chemical exposures, we must develop interdisciplinary teams of scientists measuring many aspects of exposures (Wild et al. 2013). From the environmental chemistry perspective, it seems appropriate to focus efforts on measuring *personal* chemical exposures, in situ.

3.5.1.3 Integrate Personal Sampling with Other Technologies to Learn Even More

For example, a recent study demonstrated the utility of combining a personal sampling device with a GPS-tracking device, questionnaires, and a spirometer to track lung function (Rohlman et al. 2015). This is an excellent example of an interdisciplinary project, where multiple forms of data are used to assess the relationship between exposures and adverse health outcomes.

3.5.1.4 Sampling Technologies That More Accurately Reflect Temporal Changes

Technology that continually samples chemical concentrations, as opposed to only measuring concentrations at distinct time points, enables estimations of chemical exposures to be normalized over time. This is more representative of the entirety of a person's exposure over a given unit of time. These technologies are therefore promising for use in epidemiological studies or for use in human health risk assessment.

3.5.2 Lab

3.5.2.1 Which Chemicals to Measure? How to Prioritize Them?

The EPA's ToxCast work is one example of how chemicals are being prioritized for further study (U.S. EPA 2015). ToxCast rapidly screens and predicts the toxicity of thousands of chemicals, in order to prioritize which to study. Another useful tool is the EPA's ExpoCast (U.S. EPA 2016). ExpoCast uses high-throughput approaches to rapidly estimate exposure for thousands of chemicals. Taken together, ToxCast and ExpoCast should help exposure scientists prioritize which chemicals are most relevant for study, based both on toxicity and exposure. Additionally, EDA is another tool that can help identify nontarget chemicals that are eliciting adverse effects. One way to identify currently unmonitored chemicals is to look more closely at and reexamine existing analytical data. There is often additional information in analytical results that just needs to be identified. This is an argument for collecting and sharing all analytical data possible. Even if some of the data are outside the main goals of a study, sharing all of it allows other researchers to use it to make new discoveries and to make observations and connections the primary researchers may not have had resources to exploit.

3.5.2.2 Assess Risk of Exposure to Chemical Mixtures Using Various Approaches (And Importance of Using the Right Metric to Estimate Risk)

3.5.2.2.1 Screen Commonly Found Mixtures

The exposure science community needs more studies directly measuring the bioavailable fraction of chemicals in the environment. This will add to our knowledge of the chemicals where human exposure is common or frequent, and therefore which to prioritize for study. This will also improve our ability to make public health decisions that have the greatest possible reduction in public health risks associated with environmental exposures. Additionally, mixtures that people are commonly exposed to should be prioritized. A recent study used ecological niche theory to identify that certain pesticides co-occur much more often than others in U.S. childcare centers (Chap. 6, Tornero-Velez et al. 2012). Characterizing mixture effects in these commonly occurring pesticide mixtures would be much more useful than in mixtures that rarely occur in the environment. Continuing to identify commonly occurring chemical mixtures would allow toxicologists and risk assessors to better prioritize which mixtures to study and to more accurately assess risk associated with exposure to chemical mixtures.

3.5.2.2.2 Interface with Bioassays

Rarely can currently used technologies for measuring mixtures of chemicals in the environment integrate directly with in vitro or in vivo toxicity assays. The environmental health science community needs more comprehensive measures of exposure and how they vary as a function of space and time. An example of this is testing a chemical extract from an environmentally deployed passive sampler, representing a mixture of chemicals from the environment, in a bioassay such as the embryonic zebra fish model. The utility of this technique has been demonstrated previously (Hillwalker et al. 2010; Allan et al. 2012). Combining techniques like these allows the researcher to directly observe health effects of real-world mixtures.

3.5.2.2.3 Holism vs. Reductionism

Holism and reductionism represent two different approaches to revealing the links between chemicals and our health. Holism attempts to understand the properties of a whole system by studying all of its parts together. The basis of the approach is that some system properties cannot be found by studying the individual components separately due to the complexity and integration of the system. Reductionism attempts to reveal the properties of the system by separating the components (e.g., measuring the toxicity of chemicals individually). While this approach is

the foundation of centuries of successful scientific exploration, the method has shortcomings when it comes to assessing the effects of chemical exposures on human health. In the same way that an organism would not be well represented by its isolated cells, a chemical mixture should not be described exclusively by the properties (environmental occurrence, toxicity, etc.) of its components. One could argue we need both holistic and reductionist approaches. Because it is much easier to apply the reductionist method to analytical methods, this has been the predominant approach for decades. However, the exposure science community should be striving to include more holistic approaches, both for analytical methods and exposure estimation.

3.5.2.2.4 Effect-Directed Analysis (EDA)

The goal of EDA is to determine which chemical (or chemicals) in a complex sample may be causing toxicity, by manipulating the sample to simplify the analysis. Equally importantly, it may be possible to use EDA to eliminate large numbers of chemicals that do not elicit toxicity. However, to conduct an accurate effect-directed analysis, it is necessary to consider bioavailability. An additional area of useful research would be developing other methods of extracting the bioavailable fraction of contaminants from the environment, for exploration through EDA.

Acknowledgments The authors would like to acknowledge Darlene Veenhuizen for illustrating Figs. 3.1 and 3.2 and Dr. William M. Baird, Dr. Alan J. Bergmann, Dr. Carey E. Donald, and Holly M. Dixon for insightful conversations.

References

Adgate, J.L., B.D. Goldstein, and L.M. Mckenzie. 2014. Potential public health hazards, exposures and health effects from unconventional natural gas development. *Environmental Science & Technology* 48: 8307–8320.

Al-Daghri, N.M., M.S. Alokail, S.H. Abd-Alrahman, H.M. Draz, S.M. Yakout, and M. Clerici. 2013. Polycyclic aromatic hydrocarbon exposure and pediatric asthma in children: A case-control study. *Environmental Health* 12 (1): 89.

Alexander, M. 2000. Aging, bioavailability, and overestimation of risk from environmental pollutants. *Environmental Science and Technology* 34: 4259–4265.

Allan, I.J., K. Booij, A. Paschke, B. Vrana, G.A. Mills, and R. Greenwood. 2009. Field performance of seven passive sampling devices for monitoring of hydrophobic substances. *Environmental Science & Technology* 43: 5383–5390.

Allan, S.E., B.W. Smith, R.L. Tanguay, and K.A. Anderson. 2012. Bridging environmental mixtures and toxic effects. *Environmental Toxicology and Chemistry* 31: 2877–2887.

Allan, S.E., G.J. Sower, and K.A. Anderson. 2011. Estimating risk at a superfund site using passive sampling devices as biological surrogates in human health risk models. *Chemosphere* 85: 920–927.

Allan, I.J., B. Vrana, R. Greenwood, G.A. Mills, B. Roig, and C. Gonzalez. 2006. A "toolbox" for biological and chemical monitoring requirements for the European Union's Water Framework Directive. *Talanta* 69: 302–322.

Anderson, K.A., and E.R. Johnson. 2001. Bioavailable organochlorine pesticides in a semi-arid region of eastern Oregon, USA, as determined by gas chromatography with electron capture detector. *Journal of AOAC International* 84: 1371–1382.

Anderson, K.A., D. Sethajintanin, G. Sower, and L. Quarles. 2008. Field trial and modeling of uptake rates of in situ lipid-free polyethylene membrane passive sampler. *Environmental Science & Technology* 42: 4486–4493.

Baird, W. M. 2015. Interview covering benzo[a]pyrene and other PAHs. In: Paulik, L B. ed.

Baird, W.M., L.A. Hooven, and B. Mahadevan. 2005. Carcinogenic polycyclic aromatic hydrocarbon-DNA adducts and mechanism of action. *Environmental and Molecular Mutagenesis* 45: 106–114.

Bamberger, M., and R.E. Oswald. 2014. Unconventional oil and gas extraction and animal health. *Environmental Science: Processes & Impacts* 16: 1860–1865.

Bamforth, S.M., and I. Singleton. 2005. Bioremediation of polycyclic aromatic hydrocarbons: current knowledge and future directions. *Journal of Chemical Technology & Biotechnology* 80: 723–736.

Bartkow, M.E., D.W. Hawker, K.E. Kennedy, and J.F. Muller. 2004. Characterizing uptake kinetics of PAHs from the air using polyethylene-based passive air samplers of multiple surface area-to-volume ratios. *Environmental Science & Technology* 38: 2701–2706.

Baussant, T., S. Sanni, G. Jonsson, A. Skadsheim, and J.F. Borseth. 2001. Bioaccumulation of polycyclic aromatic compounds: 1. Bioconcentration in two marine species and in semipermeable membrane devices during chronic exposure to dispersed crude oil. *Environmental Toxicology and Chemistry* 20: 1175–1184.

Birnbaum, L.S. 2012. NIEHS's new strategic plan. *Environmental Health Perspectives* 120: A298.

Booij, K., F. Smedes, E.M. Van Weerlee, and P.J.C. Honkoop. 2006. Environmental monitoring of hydrophobic organic contaminants: The case of mussels versus semipermeable membrane devices. *Environmental Science & Technology* 40: 3893–3900.

Booij, K., B.N. Zegers, and J.P. Boon. 2002. Levels of some polybrominated diphenyl ether (PBDE) flame retardants along the Dutch coast as derived from their accumulation in SPMDs and blue mussels (Mytilus Edulis). *Chemosphere* 46: 683–688.

Brack, W. 2003. Effect-directed analysis: A promising tool for the identification of organic toxicants in complex mixtures? *Analytical and Bioanalytical Chemistry* 377: 397–407.

Brasier, K.J., M.R. Filteau, D.K. Mclaughlin, J. Jacquet, R.C. Stedman, T.W. Kelsey, and S.J. Goetz. 2011. Residents' perceptions of community and environmental impacts from development of natural gas in the Marcellus shale: A comparison of Pennsylvania and New York cases. *Journal of Rural Social Sciences* 26: 32–61.

Brun, G.L., O.C. Vaidya, and M.G. Léger. 2004. Atmospheric deposition of polycyclic aromatic hydrocarbons to Atlantic Canada: Geographic and temporal distributions and trends 1980-2001. *Environmental Science & Technology* 38: 1941–1948.

Bunch, A., C. Perry, L. Abraham, D. Wikoff, J. Tachovsky, J. Hixon, J. URBAN, M. Harris, and L. Haws. 2014. Evaluation of impact of shale gas operations in the Barnett shale region on volatile organic compounds in air and potential human health risks. *Science of the Total Environment* 468: 832–842.

Burgess, R.M., K.T. Ho, W. Brack, and M. Lamoree. 2013. Effects-directed analysis (EDA) and toxicity identification evaluation (TIE): Complementary but different approaches for diagnosing causes of environmental toxicity. *Environmental Toxicology and Chemistry* 32: 1935–1945.

Burgess, R.M., S.B. Kane Driscoll, A. Burton, P.M. Gschwend, U. Ghosh, D. Reible, S. Ahn, and T. Thompson. 2017. *Laboratory, field, and analytical procedures for using passive sampling in the evaluation of contaminated sediments: User's manual*. Washington, DC: U.S. Environmental Protection Agency. EPA/600/R-16/357.

Burstyn, I., H. Kromhout, T. Partanen, O. Svane, S. Langard, W. Ahrens, T. Kauppinen, I. Stucker, J. Shaham, D. Heederik, G. Ferro, P. Heikkila, M. Hooiveld, C. Johansen, B.G. Randem, and P. Boffetta. 2005. Polycyclic aromatic hydrocarbons and fatal ischemic heart disease. *Epidemiology* 16: 744–750.

Cedergreen, N., A.M. Christensen, A. Kamper, P. Kudsk, S.K. Mathiassen, J.C. Streibig, and H. Sørensen. 2008. A review of independent action compared to concentration addition as reference models for mixtures of compounds with different molecular target sites. *Environmental Toxicology and Chemistry* 27: 1621–1632.

Chung, A., D.P. Chang, M.J. Kleeman, K.D. Perry, T.A. Cahill, D. Dutcher, E.M. Mcdougall, and K. Stroud. 2001. Comparison of real-time instruments used to monitor airborne particulate matter. *Journal of the Air & Waste Management Association* 51: 109–120.

Colborn, T., K. Schultz, L. Herrick, and C. Kwiatkowski. 2014. An exploratory study of air quality near natural gas operations. *Human and Ecological Risk Assessment* 20: 86–105.

Dejmek, J., S.G. Selevan, I. Benes, I. Solansky, and R.J. Sram. 1999. Fetal growth and maternal exposure to particulate matter during pregnancy. *Environmental Health Perspectives* 107: 475–480.

Dejmek, J., I. Solansky, I. Benes, J. Lenicek, and R.J. Sram. 2000. The impact of polycyclic aromatic hydrocarbons and fine particles on pregnancy outcome. *Environmental Health Perspectives* 108: 1159–1164.

U.S. EIA. 2011. Review of emerging resources: U.S. Shale gas and shale oil plays. Energy Information Administration, U.S. Department of Energy.

U.S. EPA. 2011. Exposure factors handbook. Available: https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=236252.

———. 2016. *Rapid Chemical Exposure and Dose Research* [online]. Available: http://www.epa.gov/chemical-research/rapid-chemical-exposure-and-dose-research.

———. 1993. Provisional guidance for quantitative risk assessment of polycyclic aromatic hydrobarcons.

———. 2000. Supplementary guidance for conducting health risk assessment of chemical mixtures. Washington, D.C.

———. 2010. *Development of a Relative Potency Factor (RPF) approach for Polycyclic Aromatic Hydrocarbon (PAH) mixtures*. Washington DC.: Intregrated Risk Information Systems (IRIS).

———. 2015. *Toxicity forecasting: Advancing the next generation of chemical evaluation* [Online]. Available: http://www.epa.gov/chemical-research/toxicity-forecasting.

Escher, B.I., and J.L. Hermens. 2004. Internal exposure: Linking bioavailability to effects. *Environmental Science & Technology* 38: 455A–462A.

Fernandez, L.A., and P.M. Gschwend. 2015. Predicting bioaccumulation of polycyclic aromatic hydrocarbons in soft-shelled clams (Mya Arenaria) using field deployments of polyethylene passive samplers. *Environmental Toxicology and Chemistry* 34: 993–1000.

Forsberg, N.D., R. Rodriguez-Proteau, L. Ma, J. Morré, J.M. Christensen, C.S. Maier, J.J. Jenkins, and K.A. Anderson. 2011. Organophosphorus pesticide degradation product in vitro metabolic stability and time-course uptake and elimination in rats following oral and intravenous dosing. *Xenobiotica* 41: 422–429.

Forsberg, N.D., B.W. Smith, G.J. Sower, and K.A. Anderson. 2014. Predicting polycyclic aromatic hydrocarbon concentrations in resident aquatic organisms using passive samplers and partial least-squares calibration. *Environmental Science & Technology* 48: 6291–6299.

Forsberg, N.D., D. Stone, A. Harding, B. Harper, S. Harris, M.M. Matzke, A. Cardenas, K.M. Waters, and K.A. Anderson. 2012. Effect of native American fish smoking methods on dietary exposure to polycyclic aromatic hydrocarbons and possible risks to human health. *Journal of Agricultural and Food Chemistry* 60: 6899–6906.

Gentner, N.J., and L.P. Weber. 2011. Intranasal benzo[a]pyrene alters circadian blood pressure patterns and causes lung inflammation in rats. *Archives of Toxicology* 85: 337–346.

Gimpel, J., H. Zhang, W. Davidson, and A. Edwards. 2003. In situ trace metal speciation in lake surface waters using DGT, dialysis, and filtration. *Environmental Science and Technology* 37: 138–146.

Goldstein, B.D., B.W. Brooks, S.D. Cohen, a.E. Gates, M.E. Honeycutt, J.B. Morris, T.M. Penning, J. Orme-Zavaleta, and J. Snawder. 2014. The role of toxicological science in meeting the challenges and opportunities of hydraulic fracturing. *Toxicological Sciences* 139: 271–283.

Greenberg, M.S., P.M. Chapman, I.J. Allan, K.A. Anderson, S.E. Apitz, C. Beegan, T.S. Bridges, S.S. Brown, J.G. Cargill, and M.C. Mcculloch. 2014. Passive sampling methods for contaminated sediments: Risk assessment and management. *Integrated Environmental Assessment and Management* 10: 224–236.

Harner, T., N.J. Farrar, M. Shoeib, K.C. Jones, and F.A.P.C. Gobas. 2003. Chracterization of polymer-coated glass as a passive air sampler for persistent organic pollutants. *Environmental Science and Technology* 37: 2486–2493.

Hassan, S., and M.I. Khoder. 2012. Gas–particle concentration, distribution, and health risk assessment of polycyclic aromatic hydrocarbons at a traffic area of Giza, Egypt. *Environmental Monitoring and Assessment* 184: 3593–3612.

Herbstman, J.B., D. Tang, D. Zhu, L. Qu, A. Sjodin, Z. Li, D. Camann, and F.P. Perera. 2012. Prenatal exposure to polycyclic aromatic hydrocarbons, benzo [a] pyrene-DNA adducts, and genomic DNA methylation in cord blood. *Environmental Health Perspectives* 120: 733–738.

Hillwalker, W.E., S.E. Allan, R.L. Tanguay, and K.A. Anderson. 2010. Exploiting lipid-free tubing passive samplers and embryonic zebrafish to link site specific contaminant mixtures to biological responses. *Chemosphere* 79: 1–7.

Huckins, J.N., J.D. Petty, and K. Booij. 2006. *Monitors of organic Chemicals in the Environment*. New York: Springer.

Huckins, J.N., H.F. Prest, J.D. Petty, J.A. Lebo, M.M. Hodgins, R.C. Clark, D.A. Alvarez, W.R. Gala, A. Steen, R. Gale, and C.G. Ingersoll. 2004. Overview and comparison of lipid-containing semipermeable membrane devices and oysters (Crassostrea Gigas) for assessing organic chemical exposure. *Environmental Toxicology and Chemistry* 23: 1617–1628.

Huckins, J.N., M.W. Tubergen, J.A. Lebo, R.W. Gale, and T.R. Schwartz. 1990. Polymeric film dialysis in organic solvent media for cleanup of organic contaminants. *Journal of AOAC International* 73: 290–293.

Iarc & Monographs. 2005. *Some non-heterocyclic polycyclic aromatic hydrocarbons and some related exposures IARC working group on the evaluation of carcinogenic risks to humans*. Lyon: France. Worl Health Organization.

Iarc & Monographs. Chemical agents and related occupations. 2009. *A review of human carcinogens. IARC Working group on the evaluation of carcinogenic risks to humans*. Lyon: France. World Health Organization.

Johnson-Logan, L.R., R.E. Broshears, and S.J. Klaine. 1992. Partitioning behavior and the mobility of chlordane in groundwater. *Environmental Science & Technology* 26: 2234–2239.

Jones, K., G. Grimmer, J. Jacob, and A. Johnston. 1989. Changes in the polynuclear aromatic hydrocarbon content of wheat grain and pasture grassland over the last century from one site in the UK. *Science of the Total Environment* 78: 117–130.

Jung, K.H., B. Yan, K. Moors, S.N. Chillrud, M.S. Perzanowski, R.M. Whyatt, L. Hoepner, I. Goldstein, B. Zhang, and D. Camann. 2012. Repeated exposure to polycyclic aromatic hydrocarbons and asthma: Effect of seroatopy. *Annals of Allergy, Asthma & Immunology* 109: 249–254.

Kameda, Y., J. Shirai, T. Komai, J. Nakanishi, and S. Masunaga. 2005. Atmospheric polycyclic aromatic hydrocarbons: Size distribution, estimation of their risk and their depositions to the human respiratory tract. *Science of the Total Environment* 340: 71–80.

Karimi, P., K. Peters, K. Bidad, and P. Strickland. 2015. Polycyclic aromatic hydrocarbons and childhood asthma. *European Journal of Epidemiology* 30: 1–11.

Ke, R., J. Li, M. Qiao, Y. Xu, and Z. Wang. 2007. Using semipermeable membrane devices, bioassays, and chemical analysis for evaluation of bioavailable polycyclic aromatic hydrocarbons in water. *Archives of Environmental Contamination and Toxicology* 53: 313–320.

Keith, L.H. 2014. The source of US EPA's sixteen PAH priority pollutants. *Polycyclic Aromatic Compounds* 35 (1–14): 147–160.

Khairy, M.A., and R. Lohmann. 2012. Field validation of polyethylene passive air samplers for parent and alkylated PAHs in Alexandria, Egypt. *Environmental Science & Technology* 46: 3990–3998.

Kingston, J.K., R. Greenwood, G.A. Mills, G.M. Morrison, and L.B. Persson. 2000. Development of a novel passive sampling system for the time-averaged measurement of a range of organic pollutants in aquatic environments. *Journal of Environmental Monitoring* 2: 487–495.

Knaapen, A.M., D.M. Curfs, D.M. Pachen, R.W. Gottschalk, M.P. De Winther, M.J. Daemen, and F.J. Van Schooten. 2007. The environmental carcinogen benzo[a]pyrene induces expression of monocyte-chemoattractant protein-1 in vascular tissue: A possible role in atherogenesis. *Mutation Research* 621: 31–41.

Knecht, A.L., B.C. Goodale, L. Truong, M.T. Simonich, A.J. Swanson, M.M. Matzke, K.A. Anderson, K.M. Waters, and R.L. Tanguay. 2013. Comparative developmental toxicity of environmentally relevant oxygenated PAHs. *Toxicology and Applied Pharmacology* 271: 266–275.

Kobayashi, R., R.A. Okamoto, R.L. Maddalena, and N.Y. Kado. 2008. Polycyclic aromatic hydrocarbons in edible grain: A pilot study of agricultural crops as a human exposure pathway for environmental contaminants using wheat as a model crop. *Environmental Research* 107: 145–151.

Kuo, C.-Y., Y.-W. Cheng, C.-Y. Chen, and H. Lee. 1998. Correlation between the amounts of polycyclic aromatic hydrocarbons and mutagenicity of airborne particulate samples from Taichung City, Taiwan. *Environmental Research* 78: 43–49.

Lampi, M.A., J. Gurska, K.I.C. Mcdonald, F. Xie, X.-D. Huang, D.G. Dixon, and B.M. Greenberg. 2006. Photoinduced toxicity of polycyclic aromatic hydrocarbons to Daphnia Magna: Ultraviolet-mediated effects and the toxicity of polycyclic aromatic hydrocarbon photoproducts. *Environmental Toxicology and Chemistry* 25: 1079–1087.

Lee, M.-S., S. Magari, and D.C. Christiani. 2011. Cardiac autonomic dysfunction from occupational exposure to polycyclic aromatic hydrocarbons. *Occupational and Environmental Medicine* 68: 474–478.

Lee, M.G., A. Phimister, D. Morin, A. Buckpitt, and C. Plopper. 2005. In situ naphthalene bioactivation and nasal airflow cause region-specific injury patterns in the nasal mucosa of rats exposed to naphthalene by inhalation. *The Journal of Pharmacology and Experimental Therapeutics* 314: 103–110.

Lichtenstein, P., N.V. Holm, P.K. Verkasalo, A. Iliadou, J. Kaprio, M. Koskenvuo, E. Pukkala, A. Skytthe, and K. Hemminki. 2000. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *The New England Journal of Medicine* 343: 78–85.

Lidén, C., L. Skare, G. Nise, and M. Vahter. 2008. Deposition of nickel, chromium, and cobalt on the skin in some occupations – Assessment by acid wipe sampling. *Contact Dermatitis* 58: 347–354.

Liu, Y., S. Tao, H. Dou, T. Zhang, X. Zhang, and R. Dawson. 2007. Exposure of traffic police to polycyclic aromatic hydrocarbons in Beijing, China. *Chemosphere* 66: 1922–1928.

Lohmann, R., B.P. Corrigan, M. Howsam, K.C. Jones, and W. Ockenden. 2001. Further developments in the use of semipermeable membrane devices (SPMDs) as passive air samplers for persistent organic pollutants: Field application in a spatial survey of PCDD/fs and PAHs. *Environmental Science and Technology* 35: 2576–2582.

Lundstedt, S., P.A. White, C.L. Lemieux, K.D. Lynes, I.B. Lambert, L. Berg, P. Haglund, and M. Tysklind. 2007. Sources, fate, and toxic hazards of oxygenated polycyclic aromatic

hydrocarbons (PAHs) at PAH- contaminated sites. *Ambio: A Journal of the Human Environment* 36: 475–485.

Martin, H., B.M. Patterson, and G.B. Davis. 2003. Feild trial of contaminant groundwater monitoring: Comparing time-intergrated ceramic dosimeters and conventional water sampling. *Environmental Science and Technology* 37: 1360–1364.

Mayer, P., J. Tolls, J.L.M. Hermens, and D. Mackay. 2003. Equilibrium sampling devices. *Environmental Science and Technology* 37: 185A–191A.

Mayer, P., T.F. Parkerton, R.G. Adams, J.G. Cargill, J. Gan, T. Gouin, P.M. Gschwend, S.B. Hawthorne, P. Helm, G. Witt, J. You, and B.I. Escher. 2014. Passive sampling methods for contaminated sediments: Scientific rationale supporting use of freely dissolved concentrations. *Integrated Environmental Assessment and Management* 10 (2): 197–209.

Mckenzie, L.M., R. Guo, R.Z. Witter, D.A. Savitz, L.S. NEWMAN, and J.L. ADGATE. 2014. Birth outcomes and maternal residential proximity to natural gas development in rural Colorado. *Environmental Health Perspectives* 122: 412–417.

Mckenzie, L.M., R.Z. Witter, L.S. Newman, and J.L. Adgate. 2012. Human health risk assessment of air emissions from development of unconventional natural gas resources. *Science of the Total Environment* 424: 79–87.

Menzie, C.A., B.B. Potocki, and J. Santodonato. 1992. Exposure to carcinogenic PAHs in the environment. *Environmental Science & Technology* 26: 1278–1284.

Miller, R.L., R. Garfinkel, M. Horton, D. Camann, F.P. Perera, R.M. Whyatt, and P.L. Kinney. 2004. Polycyclic aromatic hydrocarbons, environmental tobacco smoke, and respiratory symptoms in an inner-city birth cohort. *Chest* 126: 1071–1078.

Miller, R.L., R. Garfinkel, C. Lendor, L. Hoepner, Z. Li, L. Romanoff, A. Sjodin, L. Needham, F.P. Perera, and R.M. Whyatt. 2010. Polycyclic aromatic hydrocarbon metabolite levels and pediatric allergy and asthma in an inner-city cohort. *Pediatric Allergy and Immunology* 21: 260–267.

Nawg. 2015. *Wheat Info* [Online]. National Association of wheat growers. Available: http://www.wheatworld.org/wheat-info/fast-facts/.

Nieuwenhuijsen, M. 2005. Design of exposure questionnaires for epidemiological studies. *Occupational and Environmental Medicine* 62: 272–280.

O'Connell, S.G., L.D. Kincl, and K.A. Anderson. 2014a. Silicone wristbands as personal passive samplers. *Environmental Science & Technology* 48: 3327–3335.

O'Connell, S.G., M.A. Mccartney, L.B. Paulik, S.E. Allan, L.G. Tidwell, G. Wilson, and K.A. Anderson. 2014b. Improvements in pollutant monitoring: Optimizing silicone for co-deployment with polyethylene passive sampling devices. *Environmental Pollution* 193: 71–78.

Paulik, L.B., B.W. Smith, A.J. Bergmann, G.J. Sower, N.D. Forsberg, J.G. Teeguarden, and K.A. Anderson. 2016a. Passive samplers accurately predict PAH levels in resident crayfish. *Science of the Total Environment* 544: 782–791.

Paulik, L.B., C.E. Donald, B.W. Smith, L.G. Tidwell, K.A. Hobbie, L. Kincl, E.N. Haynes, and K. A. Anderson. 2016b. Emissions of polycyclic aromatic hydrocarbons from natural gas extraction into air. *Environmental Science & Technology* 50 (14): 7921–7929.

Penn, A., and C. Snyder. 1988. Arteriosclerotic plaque development is 'promoted' by polynuclear aromatic hydrocarbons. *Carcinogenesis* 9: 2185–2189.

Penning, T.M., P.N. Breysse, K. Gray, M. Howarth, and B. Yan. 2014. Environmental Health Research recommendations from the inter-environmental health sciences core center working group on unconventional natural gas drilling operations. *Environmental Health Perspectives* 122: 1155.

Perera, F.P., W. Jedrychowski, V. Rauh, and R.M. Whyatt. 1999. Molecular epidemiologic research on the effects of environmental pollutants on the fetus. *Environmental Health Perspectives* 107 (Suppl 3): 451–460.

Perera, F.P., Z. Li, R. Whyatt, L. Hoepner, S. Wang, D. Camann, and V. Rauh. 2009. Prenatal airborne polycyclic aromatic hydrocarbon exposure and child IQ at age 5 years. *Pediatrics* 124: e195–e202.

Perera, F.P., V. Rauh, W.-Y. Tsai, P. Kinney, D. Camann, D. Barr, T. Bernert, R. Garfinkel, Y.-H. Tu, and D. Diaz. 2003. Effects of transplacental exposure to environmental pollutants on birth outcomes in a multiethnic population. *Environmental Health Perspectives* 111: 201.

Perera, F.P., V. Rauh, R.M. Whyatt, W.-Y. Tsai, D. Tang, D. Diaz, L. Hoepner, D. Barr, Y.-H. Tu, D. Camann, and P. Kinney. 2006. Effect of prenatal exposure to airborne polycyclic aromatic hydrocarbons on neurodevelopment in the first 3 years of life among Inner-City children. *Environmental Health Perspectives* 114: 1287–1292.

Pérez, A.L., and K.A. Anderson. 2009a. DGT estimates cadmium accumulation in wheat and potato from phosphate fertilizer applications. *Science of the Total Environment* 407: 5096–5103.

Pérez, A.L., and K.A. Anderson. 2009. Soil-diffusive gradient in thin films partition coefficients estimate metal bioavailability to crops at fertilized field sites. *Environmental Toxicology and Chemistry* 28: 2030–2037.

Petty, J.D., J.N. Huckins, D.A. Alvarez, W.G. Brumbaugh, W.L. Cranor, R.W. Gale, A. Rastall, T.L. Jones-Lepp, T.J. Leiker, C.E. Rostad, and E.T. Furlong. 2004. A holistic passive integrative sampling approach for assessing the presence and potential impacts of waterborne environmental contaminants. *Chemosphere* 54: 695–705.

Petty, J.D., J.N. Huckins, and J.L. Zajicek. 1993. Application of semipermeable membrane devices (SPMDs) as passive air samplers. *Chemosphere* 27: 1609–1624.

Petty, J.D., C.E. Orazio, J.N. Huckins, R.W. Gale, J.A. Lebo, J.C. Meadows, K.R. Echols, and W.L. Cranor. 2000. Considerations involved with the use of semipermeable membrane devices for monitoring environmental contaminants. *Journal of Chromatography A* 879: 83–95.

Phillips, D.H. 1999. Polycyclic aromatic hydrocarbons in the diet. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 443: 139–147.

Rabinowitz, P.M., I.B. Slizovskiy, V. Lamers, S.J. Trufan, T.R. Holford, J.D. Dziura, P.N. Peduzzi, M.J. Kane, J.S. Reif, and T.R. Weiss. 2015. Proximity to natural gas wells and reported health status: Results of a household survey in Washington County, Pennsylvania. *Environmental Health Perspectives* 123: 21–26.

Ramírez, N., A. Cuadras, E. Rovira, R. Maria Marcé, and F. Borrull. 2011. Risk assessment related to atmospheric polycyclic aromatic hydrocarbons in gas and particle phases near industrial sites. *Environmental Health Perspectives* 119: 1110–1116.

Rappaport, S.M. 2011. Implications of the exposome for exposure science. *Journal of Exposure Science and Environmental Epidemiology* 21: 5–9.

Rappaport, S.M., and M.T. Smith. 2010. Environment and disease risks. *Science* 330: 460–461.

Ravindra, K., R. Sokhi, and R. Van Grieken. 2008. Atmospheric polycyclic aromatic hydrocarbons: Source attribution, emission factors and regulation. *Atmospheric Environment* 42: 2895–2921.

Rohlman, D., L. Syron, K. Hobbie, K.A. Anderson, C. Scaffidi, D. Sudakin, E.S. Peterson, K.M. Waters, E. Haynes, and L. Arkin. 2015. A community-based approach to developing a mobile device for measuring ambient air exposure, location, and respiratory health. *Environmental Justice* 8: 126–134.

Rosa, M.J., K.H. Jung, M.S. Perzanowski, E.A. Kelvin, K.W. Darling, D.E. Camann, S.N. Chillrud, R.M. Whyatt, P.L. Kinney, F.P. Perera, and R.L. Miller. 2011. Prenatal exposure to polycyclic aromatic hydrocarbons, environmental tobacco smoke and asthma. *Respiratory Medicine* 105: 869–876.

Schwartz, D.A., B. Weis, and S.H. Wilson. 2005. The Need for Exposure Health Sciences. *Environmental Health Perspectives* 113: A650.

Sethajintanin, D., and K. Anderson. 2006. Temporal bioavailability of organochlorine pesticides and PCBs. *Environmental Science & Technology* 40: 3689–3695.

Siddens, L.K., A. Larkin, S.K. Krueger, C.A. Bradfield, K.M. Waters, S.C. Tilton, C.B. Pereira, C.V. Löhr, V.M. Arlt, and D.H. Phillips. 2012. Polycyclic aromatic hydrocarbons as skin carcinogens: Comparison of benzo [a] pyrene, dibenzo [def, p] chrysene and three environmental mixtures in the FVB/N mouse. *Toxicology and Applied Pharmacology* 264: 377–386.

Simonich, S.L., and R.A. Hites. 1994a. Importance of vegetation in removing polycyclic aromatic hydrocarbons from the atmosphere. *Nature* 370: 49–51.

Simonich, S.L., and R.A. Hites. 1994. Vegetation-atmosphere partitioning of polycyclic aromatic hydrocarbons. *Environmental Science & Technology* 28: 939–943.

Sower, G.J., and K.A. Anderson. 2008. Spatial and temporal variation of freely dissolved polycyclic aromatic hydrocarbons in an urban river undergoing superfund remediation. *Environmental Science & Technology* 42: 9065–9071.

Spacie, A., L.S. Mccarty, and G.M. Rand. 1995. Bioaccumulation and bioavailability in multiphase systems. In *Fundamentals of aquatic toxicology: effects, environmental fate, and risk assessment*, ed. G.M. Rand, 2nd ed. Washington, D.C.: Taylor & Francis.

Stout, S.A., S.D. Emsbo-Mattingly, G.S. Douglas, A.D. Uhler, and K.J. Mccarthy. 2015. Beyond 16 priority pollutant PAHs: A review of PACs used in environmental forensic chemistry. *Polycyclic Aromatic Compounds* 35: 1–31.

Suffet, I.H., C.T. Jafvert, J. Kukkonen, M.R. Servos, A. Spacie, L.L. Williams, and J.A. Noblet. 1994. Synopsis of discussion session:Influences of particulate and dissolved material on the bioavailability of organic compounds. In *Bioavailability:Physical, chemical, and biological interactions*, ed. J.L. Hamelink, P.F. Landrum, H.L. Bergman, and W.H. Benson. Boca Raton: Lewis Publishers.

Sun, H., C. Wang, C. Huo, and Z. Zhou. 2008. Semipermeable membrane device-assisted desorption of pyrene from soils and its relationship to bioavailability. *Environmental Toxicology and Chemistry* 27: 103–111.

Tao, Y., S. Zhang, Z. Wang, R. Ke, Z. Shan, and P. Christie. 2008. Biomimetic accumulation of PAHs from soils by triolein-embedded cellulose acetate membranes (TECAMs) to estimate their bioavailability. *Water Research* 42: 754–762.

Tidwell, L.G., S.E. Allan, S. O'Connell, K.A. Hobbie, B. Smith, and K.A. Anderson. 2016. PAH and OPAH flux during the Deepwater Horizon Incident. *Environmental Science & Technology* 50 (14): 7489–7497.

Tobiszewski, M., and J. Namieśnik. 2012. PAH diagnostic ratios for the identification of pollution emission sources. *Environmental Pollution* 162: 110–119.

Tomasetti, C., and B. Vogelstein. 2015. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347: 78–81.

Tornero-Velez, R., P.P. Egeghy, and E.A. Cohen Hubal. 2012. Biogeographical analysis of chemical co-occurrence data to identify priorities for mixtures research. *Risk Analysis* 32: 224–236.

Tsai, P.-J., H.-Y. Shieh, W.-J. Lee, and S.-O. Lai. 2002. Characterization of PAHs in the atmosphere of carbon black manufacturing workplaces. *Journal of Hazardous Materials* 91: 25–42.

Tsai, P.-J., and J.H. Vincente. 2001. A study of workers' exposures to the inhalable and 'total'aerosol fractions in the primary nickel production industry using mannequins to simulate personal sampling. *Annals of Occupational Hygiene* 45: 385–394.

Tuominen, J.P., H.S. Pyysalo, and M. Sauri. 1988. Cereal products as a source of polycyclic aromatic hydrocarbons. *Journal of Agricultural and Food Chemistry* 36: 118–120.

USDA. 2014. *Wheat – Overview* [Online]. Available: http://www.ers.usda.gov/topics/crops/wheat/.aspx.

Utvik, T.I.R., G.S. Durell, and S. Johnsen. 1999. Bioavailability of polycyclic aromatic hydrocarbons in the North Sea waters: Comparison of sampling techniques. *Marine Pollution Bulletin* 38: 977–989.

Vrana, B., I.J. Allan, R. Greenwood, G.A. Mills, E. Dominiak, K. Svensson, J. Knutsson, and G. Morrison. 2005. Passive sampling techniques for monitoring pollutants in water. *TrAC Trends in Analytical Chemistry* 24: 845–868.

Wang, W., S. Tao, W. Wang, G. Shen, J. Zhao, and K.C. Lam. 2009. Airborne particulates and polycyclic aromatic hydrocarbons (PAHs) in ambient air in Donghe, northern China. *Journal*

of Environmental Science and Health. Part A, Toxic/Hazardous Substances & Environmental Engineering 44: 854–860.

Wania, F., L. Shen, Y.D. Lei, C. Teixeira, and D.C.G. Muir. 2003. Development and calibration of resin-based passive sampling system for monitoring persistent organic pollutants in the atmosphere. *Environmental Science and Technology* 37: 1352–1359.

Wells, J.B. and Lanno, R.P. 2000. Passive sampling devices (PSDs) as biological surrogates for estimating the bioavailability of organic chemicals in soil. *Environmental toxicology and risk assessment: Science, Policy, and Standardization—Implications for Environmental Decisions: Tenth Volume*. ASTM International.

Wennrich, L., P. Popp, and C. Hafner. 2002. Novel integrative passive samplers for the long-term monitoring of semivolatile organic air pollutants. *Journal of Environmental Monitoring* 4: 371–376.

Wild, C.P. 2005. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology Biomarkers & Prevention* 14: 1847–1850.

Wild, C.P. 2012. The exposome: From concept to utility. *International Journal of Epidemiology* 41: 24–32.

Wild, C.P., A. Scalbert, and Z. Herceg. 2013. Measuring the exposome: A powerful basis for evaluating environmental exposures and cancer risk. *Environmental and Molecular Mutagenesis* 54: 480–499.

Willett, W.C. 2002. Balancing life-style and genomics research for disease prevention. *Science* 296: 695–698.

Wischmann, H., and H. Steinhart. 1997. The formation of PAH oxidation products in soils and soil/compost mixtures. *Chemosphere* 35: 1681–1698.

Wright, D.A., and P. Welbourne. 2001. *Factors affecting toxicity. Environmental toxicology*. Cambridge: Cambridge University Press.

Wu, J., H. Hou, B. Ritz, and Y. Chen. 2010. Exposure to polycyclic aromatic hydrocarbons and missed abortion in early pregnancy in a Chinese population. *The Science of Total Environment* 408: 2312–2318.

Wu, S., S. Powers, W. Zhu, and Y.A. Hannun. 2016. Substantial contribution of extrinsic risk factors to cancer development. *Nature* 529: 43–47.

Yu, H. 2002. Environmental carcinogenic polycyclic aromatic hydrocarbons: Photochemistry and phototoxicity. *Journal of Environmental Science and Health, Part C* 20: 149–183.

Yunker, M.B., R.W. Macdonald, R. Vingarzan, R.H. Mitchell, D. Goyette, and S. Sylvestre. 2002. PAHs in the Fraser River basin: A critical appraisal of PAH ratios as indicators of PAH source and composition. *Organic Geochemistry* 33: 489–515.

Zhang, H., and W. Davison. 1995. Performance characteristics of diffusion gradients in thin films for the in situ measurement of trace metals in aqueous solution. *Analytical Chemistry* 67: 3391–3400.

Zhang, H., W. Davison, S. Miller, and W. Tych. 1995. In situ high resolution measurements of fluxes of Ni, Cu, Fe, and Mn and concentrations of Zn and Cd in porewaters by DGT. *Geochimica et Cosmochimica Acta* 59: 4181–4192.

Chapter 4
Modeling Complex Exposures

Dimosthenis A. Sarigiannis and Spyros P. Karakitsios

Abstract This chapter deals with the aspects of modeling complex exposures, highlighting the integration of various external and internal exposure models, following the exposome paradigm. Several approaches are investigated, relating either to the assessment of the overall chemical mixture as a single compound, or applying the compound-by-compound approach. Identifying the contribution of the various pathways leading to complex exposure requires the precise estimation of the various exposure mechanisms that integrate through the three main exposure routes (inhalation, oral and skin); hence, modeling environmental fate at different scales (such as regional, local or micro-environmental scale) for capturing both far field and near field exposure is essential. Integration of exposure through various pathways and routes occurs at the level of internal dosimetry. This is also reflected in the observed biomonitoring data, highlighting the need for integrated modeling tools that allow the functional link among exposure, internal dose and biomonitoring data. Extrapolation of exposure estimates from individual data to population exposure through advanced probabilistic techniques and agent-based models, as well as the latest advances in personal sensors for tracking activity and location are also presented. The importance of these aspects is highlighted in characteristic case studies regarding indoor air mixtures and multiple pesticide exposure.

Keywords Exposure modeling · Exposome · Complex mixtures · Internal dosimetry · Probabilistic exposure modeling

D. A. Sarigiannis (✉) · S. P. Karakitsios
Aristotle University of Thessaloniki, Chemical Engineering Department, Environmental Engineering Laboratory, Thessaloniki, Greece
e-mail: denis@eng.auth.gr

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_4

4.1 Exposome and the Need for Modeling Complex Exposures

One of the most recent developments in the field of exposure science is the introduction of the exposome concept (Wild 2005). The exposome represents the totality of exposures from conception onwards, simultaneously identifying, characterizing and quantifying the exogenous and endogenous exposures and modifiable risk factors that predispose to and predict diseases throughout a person's life span. The exposome came as a complement to the human genome; although decoding of the human genome (Schmutz et al. 2004) increased our understanding of the underlying causes of disease, the genome explains only a percentage of the population burden of disease. Indeed, according to Rappaport et al. (2014), two thirds of all people worldwide die of chronic disease (mostly heart disease and cancer) which are caused by combinations of the genome and exposome (representing all exposures – internal and external). However, disease risks attributed to the genome alone are modest, representing less than 15%, suggesting that more than 85% of risks results from exposome and interaction of the genome and the exposome. Thus, it is evident that environmental factors are equally or potentially more important than genetic traits for characterizing health risk. What is truly critical is the interaction between environmental determinants of disease with biological systems. Characterizing the exposome will carry us along towards a better understanding of the causal links between the genome, environment, and disease. To do this, both environmental exposures and genetic variation should be assessed simultaneously; properly determining and quantifying complex exposures is one of the main pillars of the current efforts towards unraveling the exposome. It is very important to keep in mind that the several responses observed at the molecular level, are the result of the combined exposures to several compounds in a dynamic, time-dependent manner. Thus, to properly identify the differences in the responses at different levels of biological organization it is critical:

- To properly account for different sources of exposure at different time resolutions and scales
- To translate these exposures into internal and biologically effective doses.

Exposome studies require novel tools to address the complexity of emerging environmental health issues. Critical for success will be the ability to bring together existing geospatial, environmental, health and socioeconomic data, and to collect new high resolution data. Innovative environmental micro-sensors, remote sensing or other community- and -omics/systems biology-based approaches can be used to describe the exposome and how it relates to the advent of multiparametric and multicausal human disease, such as endocrine disruption-related syndromes and sex-related changes (menopause), neurodegenerative or respiratory diseases. It is important to focus on: susceptibility windows during growth (including pregnancy) and development; the unequal distribution of the burden of food and environment-related disease to vulnerable populations (e.g., the young, the elderly, socio-

economic disadvantaged, and gender and ethnic minorities); and epigenetic influences.

The individual exposome is dynamic and continually changing. Indeed, all exposures and their determinants and modifiers can vary over the course of a day, not to mention over the weeks, months, and years that make up a lifetime, as our bodies, diets, risk factors and lifestyles change. Sources and levels of exposure change over time and capturing all these changes verges on the impossible in the impracticality of "high-resolution real-time" monitoring of all the exposures for the entire lifetime. Thus, the exposome has to be constructed by assessing the exposures at critical life periods through representative snapshots that act as demonstrative measures of these critical periods. Indeed, mapping the entire lifecycle of an individual may not be necessary if critical lifetime events where an individual's geospatial lifeline crosses a noteworthy environmental event (Sabel et al. 2009) are recognized and understood. Thus, one major challenge consists of identifying critical life stages that are most informative, as well as forming a picture of a person's overall exposure using sets of short-term measurements during these critical life stages, and relating these to downstream consequences. The latter may include both observable health outcomes and subtler changes in biomarkers. The most relevant exposure episodes in an individual's life could be reconstructed and linked to socio-economic conditions at critical life stages such as prenatal exposure, puberty, or the reproductively active period. Whereas exposure during all life stages may entail adverse effects, *fetuses*, *children, pregnant women* and the *elderly* are particularly susceptible. Modeling the mobility patterns of the population at risk at the individual level is challenging. There are considerable conceptual and computational difficulties involved in intersecting data on the distributions of pollutants, and/or the patterns of movements of recipient individuals or groups, reflecting the limitations of available data on environmental conditions and human distributions. Complex exposure patterns can be disentangled by fusing mobility and behavior data with the corresponding environmental data. The most appropriate such data are themselves the result of fusion of environmental monitoring data derived from the use of personal and remote sensors, including air- and satellite-borne ones; conventional monitoring systems used for regulatory compliance across several jurisdictions; and environmental modeling used to fill the observed data gaps. With the advent of geographic information systems (GIS), global positioning systems (GPS) to track individuals, and personal environmental monitoring, undertaking such analyses throughout an individual's lifetime is now possible.

Complex exposures are rarely measurable at sufficient levels of resolution and precision to allow proper exposure assessment for risk assessment purposes. For this reason, exposure modeling is necessary to support accurate exposure assessment and thus provide the currently missing link in complex mixtures risk assessment. In this context external exposure modeling needs to be coupled with internal exposure modeling to properly account for potential interactions influencing both:

- Uptake and intake rate of mixture components (dealt with individually or in chemical groups) and
- Exposure route and the relative importance of multiple routes.

The latter, in turn, influences the bioavailability of the mixture components and may determine the dominant phase of metabolism and consequently the internal and biologically effective dose of mixture components (assessed individually or in chemical classes on the basis of clustering criteria coupling physical/chemical properties, biokinetic features and toxicity profiles).

Internal and external exposure modeling brings out the hidden added value of human biomonitoring data. Human biomonitoring alone has a limited ability to identify specific exposure sources, and may not be applicable to all relevant environmental stressors (e.g., particulate matter (PM_x) or noise). Coupling it with internal exposure modeling may help to reconstruct external exposure profiles even in cases with high complexity either due to the nature of co-exposure patterns to multiple stressors or due to the complexity of the exposure patterns in time or space. Integrating human biomonitoring data, exposure models and environmental monitoring and modeling data will lead to a more comprehensive view of the exposome and of related health outcomes and will be of use in future large-scale population studies.

Information on *lifestyle/behavior patterns* (such as time-activity-location information, food consumption, use of consumer products, etc.) is needed to understand individual and population-based geospatial lifelines and the corresponding exposure profiles.

- Spatial information and initiatives to harmonize their collection (e.g., Infrastructure for Spatial Information in Europe (INSPIRE), Copernicus) have the ability to transform the way scientists and policy makers think about exposure to environmental stressors.
- At the same time, behavioral information functions as the most accessible and direct way for policy makers and risk assessors to understand and manage an individual's exposure patterns.

4.2 Complex Exposure Modeling

Complex exposures include exposures to complex mixtures of chemical stressors (e.g., mixtures of more than fifty chemicals) as well as to combinations of stressors (chemical, physical, biological) on the basis of complex exposure patterns in space and time. Modeling complex exposures requires describing and mathematically capturing both the fate and exposure characteristics of key individual mixture components and taking into account potential interactions of the latter. Thus the related models need to pertain to the combination of aggregate (all pathways and routes) and cumulative (all stressors) exposure. In practice, people live in a continuously dynamic environment, encountering different locations and performing different activities within the day, ingesting several food items and using several consumer products. Thus, individual exposure dynamics are greatly affected by personal behavior and practices such as transportation mode and nutritional habits.

Although exposure assessment historically used to pay special attention to specific exposure scenarios (e.g., a particular occupational exposure) and well defined mixtures (e.g., circumscribed chemical classes such as polycyclic aromatic hydrocarbons (PAHs) or polychlorinated biphenyls (PCBs)), now attention has been paid to the cumulative exposure to compounds or groups of compounds that pose additive or greater than additive effects. That means that when exploring the causality of an adverse outcome, e.g., endocrine disruption related outcomes, a long list of compounds needs to be addressed such as plasticizers (bisphenol-A, phthalates), pesticides, dioxins and PCBs. This implies that an exposure-driven approach should be established, seeking identification of the complex and cumulative exposures of stressors that result in combined toxicity. Thus, it is imperative to properly account for exposure of each individual compound (or sub-groups of compounds) before estimating risk based on established methods (e.g., concentration/dose addition, independent action, effect summation) or seeking for associations between exposure and disease.

One of the main dimensions of complex exposure modeling is accounting for the fate of and exposure to complex chemical mixtures (Kinerson 1987). Kinerson (1987) has identified the following possibilities to model complex chemical mixtures:

(a) Based on bulk properties of the overall mixture
(b) Based on chemical classes that are representative of the mixture
(c) Based on chemical fractions clustered by physical/chemical and biokinetic properties (the latter is of particular importance when considering the link between external exposure levels into internal and, eventually, biologically effective dose.
(d) As individual components (one compound at a time).

(a) *Model bulk properties of the mixture:* In this approach, the overall mixture of components is treated as a single component. This approach has to be carefully applied and it is appropriate only for mixtures where the properties governing environmental and biokinetic fate are very similar for the individual components of the mixture. The advantage of the method is that model estimates are obtained through a single run.
(b) *Model by (representative) chemical classes:* This approach represents an intermediate solution between the one compound and the bulk properties approach. In practice, the overall mixture is broken down into representative chemical classes, and each class is represented by one characteristic compound. Although ideally the representative compounds should be selected as being of biological significance, frequently, they are selected because their properties have been determined.

A good example of such an approach is given by the work of Pistocchi and Bidoglio (2009) who attempted to model the spatial extent of exposure to pesticides in Europe on the basis of a compound that has been considered as representative of each chemical class of pesticides based on the toxic potency of the class. Similarly, Sarigiannis et al. (2013) have derived a highly granular inventory of pesticide

emissions into the air and the corresponding bystander and farmer/applicator exposure maps across Europe. An example of this application is given later in this chapter to exemplify the advantages and disadvantages of the approach as well as possibilities to optimize the outcome by integrating models of different complexity over space and time.

(c) *Model by chemical fractions:* a similar approach to the above is to chemically fractionate the mixture, by dividing it physically into several fractions, each containing chemicals of greater similarity, based on determinants of environmental fate (water solubility, volatility, degradability, etc.). The advantage of this approach compared to modeling by chemical class is that chemicals are clustered based on their properties rather than the designated class; thus they provide more realistic estimates about the compounds included within the same fraction. Highly toxic fractions could be further fractionated and tested to determine their chemical properties in an iterative process to optimize the modeling outcome without excessive demand in computational resources. The chemical fractionation approach is discussed further in Chap. 3.

(d) *Model one compound at a time:* The most common approach for addressing complex exposure modeling is to model the fate and exposure of one single compound at a time. This includes execution of the model for each compound identified in the mixture and collation of the results for assessing the overall mixture fate. Although this approach seems ideal, sufficient data on the properties (volatilization, hydrolysis, photolysis, and biodegradation) governing environmental behavior and exposure are lacking, and there are very high requirements of modeling and computational resources for such an endeavor to succeed.

The choice of complex exposure modeling tier could be based on the availability of data to characterize the chemicals constituting the complex mixture of interest or the intended use of the modeling outcome. This follows the two-tiered approach to cumulative risk assessment of chemicals proposed by Sarigiannis and Hansen (2012). The authors suggest using a dose addition assumption to calculate a hazard index taking into account interactions as a default option for hazard quantification and risk assessment. The hazard index formulation takes into account potential non-linear effects from the interaction of mixture components if the necessary information is available, while simplifying down to dose addition if interaction data do not exist. Overall, it would give a reasonable approximation of the toxic potency of a mixture if the necessary data were available; and it would allow conservative assumptions about effects of combined exposure to multiple chemicals if no such data exist. For further detail, please read the chapter on dose addition (Chap. 9) or the chapter on component-based risk assessment (Chap. 14) in this volume. As a second tier assessment (i.e., when dealing with data-rich situations) more sophisticated tools can be used, including mechanistic, biology-based modeling that accounts for the biologically effective dose of mixture components at the target tissue and incorporates system-wide data coming from –omics technologies. The authors call this the *connectivity* approach. By the same token, exposure

modeling of complex mixtures and scenarios could be designed to follow these tiers by moving gradually from whole mixture, to chemical class, fraction and finally individual compounds considering the potential interactions amongst them.

4.2.1 Modeling Exposure Using the Intake Fraction

Intake fraction (iF) is a metric of the emission-to-intake relationship, facilitating comparisons among sources in terms of their exposure potential (Bennett et al. 2002b). For a given emission source and pollutant, iF is the cumulative mass taken in by the exposed population divided by the cumulative emissions. Considering that iF depends on several parameters affecting the emission-to-intake process, (e.g. prevalent wind, emissions strength, population density), it is expected that it would vary with location and time.

For a primary pollutant, iF can be expressed as (Marshall et al. 2003):

$$\mathrm{iF} = \frac{\text{Population Intake}}{\text{Total Emissions}} = \frac{\int_{T_1}^{\infty} \left(\sum_{i=1}^{P} C_i(t) Q_i(t)\right) dt}{\int_{T_1}^{T_2} E(t)\, dt} \tag{4.1}$$

Where, $T1$ and $T2$ are the starting and ending times of the emission; P is the number of people in the exposed population; $Q_i(t)$ is the intake rate for individual i at time t; $C_i(t)$ is the incremental concentration, attributable to a specific source at time t and $E(t)$ is that source's emissions at time t.

The mathematical expression for exposure to a release through all exposure pathways is given by the following equation (Bennett et al. 2002a):

$$\mathrm{iF(total)} = \mathrm{iF(inhalation)} + \mathrm{iF(ingestion)} + \mathrm{iF(dermal)} \tag{4.2}$$

where the term "total" indicates that intake is summed across all exposure routes.

In a multimedia, multipathway model, a source to-intake relationship is typically expressed as an intake rate (mg/kg-BW/d) per unit emission rate (mg/d). To convert from a source-to-intake relationship to an iF, the following conversion must be made (Bennett et al. 2002a):

$$\mathrm{iF} = \text{Source to Dose}\left(\frac{\mathrm{mg/kg/day}}{\mathrm{mg/day}}\right) \times \mathrm{BW(kg)} \times \text{Population} \tag{4.3}$$

Where BW is the population average body weight (kg) and Population is the size of the exposed population.

Typical values for the iF vary greatly based upon the environmental fate of the mixture of interest and the population density of the area of release; this is the reason why iFs of mixtures released into the indoor environment are usually two to three orders of magnitude higher than the respective values of iF for mixtures emitted into the external environment in urban areas. iF is especially useful in obtaining a quick

overview of the emission-to-intake pathway without returning to detailed environmental fate and exposure modeling. This is of particular interest when exposure modeling aims at identifying the extremes of the exposure probability distribution especially when the latter affects the most socioeconomically disadvantaged population (Marshall and Nazaroff 2007).

Parameters affecting iF include (a) the location where the release occurs (i.e., indoor or outdoor) and the area of interest, (b) the population density and size of the exposed population close to the area associated with the quantity released, (c) dispersion parameters related to the natural or the built environment, (d) the compound-specific environmental fate parameters and (e) based on the different environmental media distribution, the exposure pathway(s) of relevance.

The iF may be an important tool for complex exposure modeling, since it can be applied to groups of pollutants. This pertains to compounds emitted from the same source characterized by similar properties with regard to environmental fate and transport. Thus, by breaking the mixture into several fractions, each of which comprises chemicals with similar environmental fate features, one can be certain that the compounds within the same group will have similar iFs, irrespective of the overall mixture chemical composition and mass emission rates.

4.2.2 Environmental Fate Modeling of Real Life Chemical Mixtures

To properly address complex chemical exposures, it is essential to develop modeling tools that cover a wide chemical space including a large number of industrial chemicals and metals characterized by significantly diverse physicochemical properties. These affect the distribution of mixture components in different environmental media (air, soil, water, sediment), their persistence (regulating processes such as biodegradation and photo degradation) and the respective bioaccumulation and biomagnification potentials.

4.2.2.1 Environmental Fate and Exposure Models for Complex Mixtures

Environmental fate models describe the interactions between different environmental scales and media (air, soil, water, sediment) using physicochemical properties, such as the octanol water partition coefficient (K_{ow}) and octanol-air partition coefficient (K_{oa}), to describe transfer, partitioning, and degradation (Mackay et al. 1992, 2001). The main inputs of multimedia models relate to environmental releases and mode of entry in the environment, properties of the environment or landscape receiving the contaminants (e.g., organic content of soil, distribution of land cover) and compound specific physicochemical properties.

The European Union System for the Evaluation of Substances (EUSES) (Lijzen and Rikken 2004) provides a comprehensive framework for evaluating human and ecosystem exposures and health risks from new and existing chemicals in the European Union. EUSES directly links the overall uptake to probable health endpoints through exposure/response relations without taking into account the toxicokinetics/toxicodynamics and the related internal dose (Fryer et al. 2004). The Calendex™ model system is currently used by the U.S. EPA to evaluate aggregate and cumulative human exposures to pesticides. Calendex™ is similar in both scope and approach to the LifeLine™ (Hampshire 2002) and CARES: Cumulative and Aggregate Risk Evaluation System (CropLife 2002) models. Out of these three models, Calendex™ is generally the least complex in terms of the methodology and techniques it adopts to conduct exposure assessments. However, the proprietary nature of the model and its dependence on expert judgement would appear to limit its potential for widespread adoption (Fryer et al. 2004). The CARES model is fundamentally similar in scope and approach to both the LifeLine™ and Calendex™ models. All three models focus on predicting risks to the U.S. population from dietary, drinking water and residential pesticide exposures. However, the exact methodology adopted in CARES is different from that used by the other models, particularly with regard to the use by CARES of a reference population. The source code of the model has been published and is freely available (Farrier and Pandian 2002). This means that although the model has been developed for use in the USA, it could be updated and adapted to be representative of situations in the EU and other world regions. The LifeLine™ model provides a comprehensive, in-depth tool for assessing human exposures to pesticides and subsequent health risks. LifeLine™ focuses on intra-individual variability in exposure levels in more detail than both Calendex and CARES. The ConsExpo model (Vermeire et al. 1993) provides a framework for evaluating exposures to chemicals in consumer products. The inclusion of models of varying degrees of complexity means that ConsExpo provides a useful tool for assessing consumer product exposures at all tiers of the risk assessment framework, from screening level to specific exposure situations. Validation studies have assessed the performance of some of the individual ConsExpo models against measured datasets (Van Veen et al. 1999; Wilschut et al. 1995) and calculated exposure estimates were generally found to be within an order of magnitude of the measured values. A multimedia modeling approach focusing on spatially explicit modeling of chemical fate and transport processes has been proposed by Pistocchi et al. (2010). The basic idea of this approach is to replace the numerical solution to the advection–dispersion equation with a series of local analytical solutions. Such simplified models comprise the box model, or "continuous stirred tank reactor" (CSTR), the plug flow (PF) and Gaussian plume (GP) models. This set of models judiciously combined may represent most of the typical environmental distributions. E-FAST (Exposure and Fate Assessment Screening Tool) is a model developed by U.S. EPA aimed at providing screening-level estimates of the concentrations of chemicals released to air, surface water, landfills, and from consumer products (Egeghy et al. 2011). E-FAST intentionally provides reasonable overestimations

(90% confidence limit of the upper bound of the estimate) of exposures (ERG 2001), for use in screening level assessment.

For more elaborate calculations, the Stochastic Human Exposure Dose Simulation for multimedia, multipathway chemicals (SHEDS-Multimedia) system is available for download from the U.S. EPA website (https://www.epa.gov/chemical-research/stochastic-human-exposure-and-dose-simulation-sheds-estimate-human-exposure). The residential scenarios included in the SHEDS-Pesticides model focus on organophosphate pesticides (Hore et al. 2006). The SHEDS models do however allow detailed assessments of specific exposure scenarios to be made and are non-proprietary in nature, also providing some links to biomonitoring data (Zartarian et al. 2002). To further improve the exposure assessment approach, the methodology first developed for the SHEDS model was enhanced and incorporated through new, generalized code into the Modeling ENvironment for TOtal Risk studies (MENTOR) (Georgopoulos et al. 2005; Georgopoulos and Lioy 2006; Georgopoulos et al. 2006; Lioy et al. 2007; Georgopoulos et al. 2008c), which was designed to analyze not only exposures to individual contaminants but to assess physiologically based target tissue doses of Multiple co-occurring contaminants via Multimedia, Multipathway, Multiroute exposures (4 M) for specific individuals or for study-specific populations. The GIS extension module of MENTOR, is the Prioritization/Ranking of Toxic Exposures (Royce et al. 2014; Georgopoulos et al. 2014), that utilizes simplified versions of MENTOR components to provide screening level analyses. Similarly to MENTOR, INTEGRA (Sarigiannis et al. 2014a) provides a multimedia environmental model similar to EUSES (Vermeire et al. 1997), following the European Chemicals Agency (ECHA) recommendations, a detailed micro environmental multi-zone model (Sarigiannis et al. 2012a, b), and addresses in detail multi-route exposure and internal dosimetry. Among all the above models, MENTOR-4 M and INTEGRA provide the most complete methodological framework for assessing aggregate exposure from environmental and consumer sources. In addition, INTEGRA integrates a large database for industrial chemicals and additional QSAR models, enabling environmental modeling for multiple chemicals.

4.2.2.2 Specific Considerations for Addressing Complex Exposures

Despite the limitations of the above models, many of them are able to address the issue of complex exposures as a substance by substance problem. One significant issue that relates to complex exposure is the effect of biotransformation in the environment and consequently the fate of transformation products. This results in exposure to additional compounds than the one that was initially released in the environment. The next step is the incorporation of interactions between compounds released in the environment. At present, these types of interactions are limited to consideration in atmospheric (Morris et al. 2004) and indoor air chemistry, where these types of transformation play a significant role.

4.2.3 *Addressing Multi-pathway and Multi-route Complex Exposures with the Compound-by-Compound Approach*

To properly account for the contribution of different pathways and routes of exposure, the key parameters affecting the major exposure mechanisms have to be briefly described. It is important to understand that differences in behavioral patterns are also relevant to exposure, e.g., infants and children are more likely to be exposed to compounds found in settled dust than adults due to significantly more frequent hand to mouth behavior, or people performing intensive exercise close to busy roads are more likely to be exposed to ambient air mixtures.

4.2.3.1 Inhalation

Inhalation is a major route for numerous outdoor (e.g., CO, NOx, SO_2, PM, PAHs) and indoor (e.g., aldehydes, phthalates, and benzene, toluene, ethylbenzene, and xylenes (BTEX)) air pollutants. Personal exposure is equal to the average concentration of a pollutant that a person is exposed to over a given period of time. If over the given period of time, T, the person passes through n locations, spending a fraction f_n of the period T in location n where the concentration of the pollutant under consideration is C_n, then the personal exposure for this period T, represented by the concentration C_T, is given by Eq. 4.4:

$$C_T = \sum_n f_n \cdot C_n \tag{4.4}$$

Inhalation uptake is estimated by the area under the curve of exposure E multiplied by the inhalation rate inh, divided by the bodyweight BW and for the desired simulation time.

$$\text{Uptake}_{\text{inh}} = \frac{\sum_n E_n \cdot \text{inh}_n}{\text{BW}} \tag{4.5}$$

where inh_n is the inhalation rate which is age and activity dependent (ICRP 2002) for each type of microenvironment n encountered.

To properly estimate inhalation exposure, age, gender and activity intensity differences have to be taken into account. There are databases that categorize the majority of daily activities based on their intensity; intensity of activity is associated with age- and gender-dependent inhalation rates (Sarigiannis et al. 2012a, b). In the absence of data, default daily activity patterns can be used. Intensity of activity can also be measured using personal wearable sensors such as Fitbit or Actigraph. This significantly alters the outcome of actual exposure and intake, either between

different individuals encountering the same locations (Sarigiannis et al. 2012a, b), or the intra-day variability for a given individual (Sarigiannis et al. 2014b).

4.2.3.2 Dietary Ingestion

Dietary exposure sources include water and food, and may occur through environmental contamination and bioaccumulation (e.g. pesticides, mercury) or leaching from food contact materials (e.g. bisphenol A from can lining).

To estimate human exposure through diet, contaminant concentrations in foods are multiplied by the corresponding intake rates. The sum of these individual food contaminant intake values is corrected for bodyweight to obtain the daily contaminant exposure via the diet (Lambe 2002):

$$E_{\text{diet}} = \sum_{x=1}^{n} \frac{\left(C_{\text{food}_x} \cdot q_{\text{food}_x}\right)}{\text{BW}} \tag{4.6}$$

E_{diet}: daily contaminant exposure through diet (mg/kg/day)
$C_{\text{food}x}$: contaminant concentration in food item x (mg/kg)
$q_{\text{food}x}$: food item x consumption (kg/day)
BW: body weight (kg)

To properly account for dietary exposure, detailed food consumption databases have to be used that take into account ethnicity, age, gender and socioeconomic status differences (EFSA 2011). Food residues are estimated as the sum of the contribution of the contamination transferred through the food web and migration from food contact materials.

4.2.3.3 Non-dietary Ingestion

4.2.3.3.1 Dust and Soil Ingestion

Scenarios simulating the ingestion of dust and soil combine amounts of dust and soil ingested daily with concentrations of chemicals in these media. The amount of soil and dust ingested daily might be estimated either from daily determinations of trace elements in food intake and fecal output (Stanek and Calabrese 1995, 2000), or by predicting (modeling) soil and dust ingestion by pathway, source type, population group, geographic location, and other factors (Ozkaynak et al. 2011). These exposure pathways are particularly relevant for infants and toddlers who are known to incidentally ingest small amounts of dust and soil daily. Such quantities are higher than the ones for adults by one to two orders of magnitude.

Average daily dose from non-dietary ingestion of chemicals from dust is estimated by the following formula (Wormuth et al. 2006):

$$E_{\text{dust\_ing}} = \frac{C_{\text{dust}} \cdot q_{\text{dust\_ing}}}{\text{BW}} \cdot r_{\text{uptake}} \tag{4.7}$$

where,

$E_{\text{dust\_ing}}$: the internal exposure to chemical (μg/kg/day);
C_{dust}: Concentration of the chemical in dust (μg/mg)
$q_{\text{dust\_ing}}$: Amount of dust ingested (mg/day)
r_{uptake}: absorbed fraction from the ingested quantity
BW: body weight (kg)

Similarly, for soil ingestion, the following formula is used (Wormuth et al. 2006)

$$E_{\text{soil\_ing}} = \frac{C_{\text{soil}} \cdot q_{\text{soil\_ing}}}{\text{BW}} \cdot r_{\text{uptake}} \tag{4.8}$$

where,

$E_{\text{soil\_ing}}$: the internal exposure to chemical (μg/kg/day);
C_{soil}: Concentration of the chemical in soil (μg/mg)
$q_{\text{soil\_ing}}$: Amount of soil ingested (mg/day)
r_{uptake}: absorbed fraction from the ingested quantity
BW: body weight (kg)

4.2.3.3.2 Object-to-Mouth

Several literature sources that describe non-dietary ingestion exposure to chemical residues in objects contacted via object-to-mouth activity can be found. One of them is the U.S. EPA's EXPOsure toolBOX (EPA-Expo-Box[1]), a toolbox created to assist individuals from government, industry, academia, and the general public with assessing exposure (USEPA 2013). To estimate the average daily potential dose from ingestion of surface residues from object-to-mouth contact, the U.S. EPA proposes the following algorithm:

$$\text{ADD} = C_{\text{surface residue}} \cdot \text{CR} \cdot \text{EV} \cdot \text{ET} \cdot \text{EF} \cdot \frac{\text{ED}}{\text{BW} \cdot \text{AT}} \tag{4.9}$$

where,

ADD: Average daily potential dose (mg/kg/day)
$C_{\text{surface residue}}$: Concentration of contaminant on the surface of the hands or objects that are mouthed (mg/cm^2)
CR: Contact rate with contaminated surface (cm^2/event)
EV: Event frequency (events/h)

[1] http://www.epa.gov/risk/expobox/index.htm.

ET: Exposure time (h/day)
EF: Exposure frequency (days/year)
ED: Exposure duration (years)
BW: body weight (kg)
AT: Averaging time (days)

This could be further refined if more detailed information is available for the microactivity (i.e., a specific mouthing event in a specific microenvironment) resulting in indirect ingestion over a given exposure period. In this case, total indirect ingestion exposure is estimated in two steps (Tulve et al. 2002): (a) individually for each microactivity, and/or (b) summed for all activities for an exposure duration of interest (i.e., 24-h).

For each microactivity resulting in indirect ingestion, exposure over a 24-h period can be defined as:

$$E_{\mathrm{nd}} = C_X \cdot \mathrm{TE}_X \cdot \mathrm{SA}_X \cdot \mathrm{EF} \tag{4.10}$$

where,

X: body, hand, surface, toy, or any other object that is mouthed
E_{nd}: indirect ingestion exposure from a specific mouthing event over a 24-h period (μa_x) (μg/day)
C_x: total contaminant loading on object x (μg/cm^2)
TEx: transfer efficiency, fraction transferred from object x to mouth
SA$_x$: surface area of object x that is mouthed (cm^2/event)
EF: frequency of mouthing events over a 24-h period (event/day)

The total indirect ingestion exposure over a 24-h period can be estimated by summing exposures for all microactivities. For any particular microenvironment being modeled, the potential exposure is the sum of all exposures for all microactivities conducted in that microenvironment (e.g., indoors, at home, on carpet).

4.2.3.3.3 Unintentional Swallowing of a Substance in a Product During Normal Use

Here, it is assumed that consumers incidentally ingest small amounts of a chemical substance in a consumer product. The best known application of this exposure scenario is the unintentional ingestion of personal care products (PCPs). Scenarios for ingestion of PCPs use information on amounts of products ingested daily and on chemical concentrations in such products. Usually detailed information on how much PCPs are ingested daily is not available; thus, a worst-case assumption could be used here: infants, toddlers, children, and female teenagers and adults ingest 50 mg product per day; male teenagers and adults ingest 25 mg product per day (Wormuth et al. 2006). The higher amounts ingested should reflect the more

careless use of PCPs by infants, toddlers and children and the more frequent use of PCPs by female consumers.

The mathematical formulation is

$$E_{\text{prod\_ing}} = \frac{C_{\text{prod}} \cdot q_{\text{prod\_ing}}}{\text{BW}} \cdot r_{\text{uptake}} \tag{4.11}$$

where,

$E_{\text{prod\_ing}}$: the internal exposure to chemical (μg/kg/day);
C_{prod}: Concentration of the chemical in the product (μg/mg)
$q_{\text{prod\_ing}}$: Amount of product ingested (mg/day)
r_{uptake}: absorbed fraction from the ingested quantity
BW: body weight (kg)

4.2.3.4 Dermal Exposure

Dermal exposure is determined by the processes involved in contact between the skin and the product or article. Since processes and exposure determinants differ largely between articles and products, different approaches are needed. Also within the category of products, a number of processes play an important role in dermal exposure, depending on the type of product and its use. For example, dermal exposure to substances in personal care products can be approximated by the applied dose and the application surface, whereas dermal exposure to substances in household products is not only affected by the amount of product used, but also depends on duration and type of contact between the product and skin during the application phase, and by the contact between skin and the surface on which the product is applied (post-application phase). Therefore, mathematical description of dermal exposure is split up for different types of products, and articles as a separate category.

Exposure through skin includes several mechanisms that relate to the different uses of industrial chemicals. Major mechanisms of dermal exposure include (Delmaar et al. 2005):

- *Instant application:* The instant application mode assumes that all compounds in the product are directly applied to the skin. This is the situation for personal care products, but can also be used as a first tier worst-case approach or if details on how the skin is exposed to the compound are not known.
- *Constant rate of application to the skin:* This mode of dermal loading describes a situation in which a compound is loaded onto the skin during a certain time, with a constant rate (e.g., when skin comes into contact with a clothing).
- *Rubbing off mechanism:* Contrary to the previous dermal exposure modes, the rubbing off mode describes a secondary exposure situation. Instead of direct application of a product to the skin, the rubbing off mode describes a situation

in which a surface (e.g., table top, floor) is treated with a product and dermal exposure arises from contact with the treated surface.
- *Exposure during showering or swimming pool*
- *Deposition of particles onto the skin*

4.3 Individual and Population Exposure Modeling

Exposure to chemicals is rarely characterized by regular, uniform events; thus exposure assessment needs to account for the frequency, duration and level (magnitude) of exposure (Nieuwenhuijsen 2003). Since the degree of exposure often varies with time, the period during which an exposure estimate is based can have a large influence on the result (Benford and Tennant 1997). Thus, exposure assessment may target either an individual, or the population at large; the latter is usually the target group of regulatory bodies.

4.3.1 Deterministic Exposure Modeling

Theoretically, there is no single risk for a particular exposure circumstance; rather, there are as many different risk values as there are individuals (Harper 2004). To overcome the problem of addressing variability in exposure and risk assessment, regulatory authorities have traditionally characterized the risks to individuals in a population who are likely to encounter the greatest exposure. The approach they have used, frequently referred to as a 'point-estimate' or 'deterministic' approach, uses single values to represent each exposure variable and produces a single risk estimate. In chemical risk assessments, initial screening of potential human health risks from chemicals of concern is often carried out by calculating 'worst-case', (a.k.a. 'high-end' or 'upper bound') point estimates of exposure using maximum or upper percentile values for exposure variables. In risk characterization, these point estimates of exposure are then combined with an appropriate toxicological end-point to determine whether a hypothetical 'worst' case individual exceeds the regulatory threshold of concern (or other calculated margins of safety). Where worst-case exposure estimates exceed regulatory thresholds, refined point-estimate exposure estimates (or 'best-case' estimates), are sometimes derived using average, mean or median values for exposure variables to provide a more realistic estimate of exposure.

The main advantages of using deterministic approaches for modeling exposure are that these are generally simple, quick and inexpensive and can be used as a screening tool for assessing chemical health risks. These approaches have, however, a number of disadvantages, which can undermine their use in regulatory decision-making. Deterministic approaches provide little information on the extent to which exposure or risk varies within a population or subgroup under investigation; certain

models are inflexible and do not allow different assumptions or scenarios to be considered and they can provide conservative or unrealistic exposure estimates. For example, it is not possible to determine from a 'worst-case' point estimate whether this represents an exposure likely to be encountered by the 95th, 99th or 99.999th percentile individual in a given population or is so extreme that it is unlikely ever to take place. If a high-end point estimate significantly exceeds the maximum (100th percentile) exposure likely to be encountered by a real population, it is likely to be highly unrealistic and provide an extremely conservative basis upon which to regulate safety to chemicals.

4.3.2 Probabilistic Exposure Modeling

To expand exposure assessment from the single individual to the wider population groups, probabilistic modeling techniques can be implemented (Bogen et al. 2009; Mutshinda et al. 2008; Zidek et al. 2005). Probabilistic analysis is an alternative approach used in exposure modeling which addresses the shortcomings of deterministic, point-estimate methods in terms of variability and uncertainty and produces more accurate and realistic estimates of exposure across the populations under investigation (Harper 2004). Depending on the availability and quality of data, distributions for any exposure variable relevant to a given exposure assessment scenario can be used in a probabilistic exposure model. In probabilistic modeling, distributions of exposure variables are combined in such a way as to give an exposure distribution. Exposure variables are also sometimes combined with toxicological endpoint levels to give risk distributions. Although there are several ways to combine exposure input distributions, the most common approach involves the use of a mathematical sampling technique called Monte Carlo simulation. The Monte Carlo technique, as applied to exposure assessment, involves combining the results of hundreds or thousands of random samplings of values from input distributions to produce an output distribution, which reflects the expected range and frequency of exposures.

Monte Carlo analysis is used to determine the probability of occurrence for the point estimates of a deterministic risk assessment and, in this way, deal with the uncertainty associated with these assessments (Hayes 2000). Whilst deterministic risk assessment applies a single value for each of the model's input parameters and calculates a single output value, probability risk assessment assigns a probability distribution to these input parameters, either as a probability density function, which is an analytical continuous function, or as a probability mass function, which is a discretized distribution. For a continuous random variable (*i*.e., a variable that can assume any value within some defined range), the probability density function expresses the likelihood that the value for a random sample will fall within a particular very small interval. Well known probability density functions are: normal, triangular, uniform and lognormal (Wilson et al. 2013).

A key part of developing a comprehensive probabilistic exposure model is to conduct a global sensitivity analysis of the exposure determinants. Sensitivity analysis is a technique that allows determination of the effect on the overall outcome of altering the value of one variable. The relative importance of each variable in determining the values of the output distribution can then be independently assessed.

4.3.2.1 Hierarchical Population Modeling Based on Bayesian Statistics

The Bayesian approach provides a formal way to incorporate prior knowledge on model parameters together with observed data in the modeling process. The analysis starts with the construction of prior probability distributions of the model parameters of interest, usually based on studies available in the literature. These distributions are then evaluated on the basis of their likelihood given observed data to compute posterior distributions of the model parameters. Hierarchical modeling with Bayesian Markov Chain Monte Carlo simulation is suitable for population exposure and internal dose (PBBK) models because the development of these models often involves non-linear processes, small datasets, high uncertainty, and biological variability (Bernillon and Bois 2000).

Markov Chain Monte Carlo methods are a class of algorithms for sampling from generic probability distributions (for more detailed information, see Robert and Casella 2004). A basic concept of the method is that of a Markov chain, i.e., a sequence of random variables:

$$Y_0, Y_1, Y_2, \ldots,$$

for which the distribution of the future state of the process, given the current and the past values, depends only on the immediately preceding state:

$$p(Y_{t+1} \mid Y_0, Y_1, \ldots, Y_t) = p(Y_{t+1} \mid Y_t)$$

Markov Chain Monte Carlo methods are based on the construction of a Markov chain that converges to the desired target distribution p (i.e., the one from which one wants to simulate from, for instance the unknown distribution of a parameter of interest). More formally, we say that p is the stationary distribution of the Markov chain. In most practical cases, after a sufficiently large number of iterations, referred to as "burn-in," the chain will forget the initial state and will converge to a unique stationary distribution, which does not depend on state t or Y_0. Once convergence is reached, it is possible to calculate any required statistic using Monte Carlo integration.

Figure 4.1 shows an intuitive representation of the process of convergence for a Markov Chain. Initially, the values sampled for two chains are dependent on the two different starting points. However, after the burn-in period, they tend to converge to the same distribution (this process is also known as mixing up). The first set of simulated values can then be discarded and the ones after convergence used as a

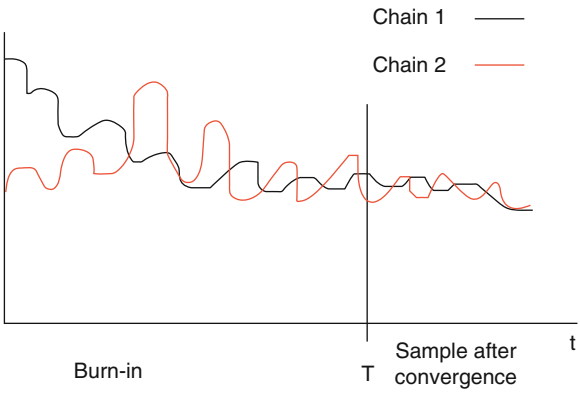

Fig. 4.1 A graphical representation of the process of convergence of Markov Chains: the two chains start from very different points but after the burn-in they converge to the stationary distribution

sample from the target distribution. One of the most popular Markov Chain Monte Carlo methods is Gibbs sampling (Geman and Geman 1984). The steps needed to perform the simulation via Gibbs sampling are schematically described in the following.

1. Define an initial value to be arbitrarily assigned to the parameter of interest. The sampling procedure starts from that value.
2. Perform a set of simulations during which the Markov chain converges to the stationary distribution, *i.e.* the required posterior. It is usually convenient to define more than one chain (two are generally sufficient), starting from distant initial values, to assess the convergence more efficiently (see Fig. 4.1).

Once convergence is reached (this process can be monitored by suitable statistics, such as that proposed by Gelman and Rubin (1996), a sample of values is drawn from the estimated target distribution. Using this sample all the inferences of interest can be performed; for instance, the whole distribution might be analyzed (i.e., by means of graphical methods, such as histograms or kernel density estimations), or point estimations such as the posterior mean or median can be computed.

4.3.2.2 Maximum Likelihood Estimates

Maximum likelihood estimation is a statistical method used for fitting a statistical model to data, and providing estimates for the model parameters. The method of maximum likelihood corresponds to many well-known estimation methods in statistics. Given a sample of some number of exposure attributes, but not the entire population, with knowledge that their values are normally distributed with some unknown mean and variance, the sample mean is the maximum likelihood estimator of the population mean, and the sample variance is a close approximation to the maximum likelihood estimator of the population variance.

For a fixed set of data and underlying probability model, maximum likelihood picks the values of the model parameters that make the data "more likely" than any

other values of the parameters would make them. Maximum likelihood estimation gives a unique and easy way to find a solution in the case of the normal distribution and many other problems, although in very complex problems this may not be possible. If a uniform prior distribution is assumed over the parameters, the maximum likelihood estimate coincides with the most probable values thereof.

4.3.3 Agent-Based Modeling

Using data fusion techniques, health and exposure data derived from fixed monitoring networks may be supplemented by a range of emerging novel techniques and technologies such as agent-based modeling, mobile phone apps, environmental sensor-webs, micro-sensors and satellite remote sensing. The information from the coupled use of agent-based models and sensor webs improves exposure modeling using deterministic and/or probabilistic approaches, and supports the application of new epidemiological and biostatistical methods to relate modeled exposure to health outcomes. The input to agent-based models consists of data relating to an individual's behavior within his/her environment (such as movement data within specific micro-environments) and between individuals exploring interactions around health-related behaviors and key risk determinants such as low socio-economic status. Using these parameters and the evolution of the virtual agents, simulations produce detailed information relating to the overall societal systems and populations considered. The estimated values produced can be used to fill the gaps of traditional datasets. This holistic approach is highly novel, taking the best from existing monitoring and sensor technology, but supplementing it with computational modeling. It is of particular relevance where real-world data are unavailable at the spatial and temporal scales that modeling complex exposures at the individual or population subgroup level requires. Although commonly used elsewhere, agent-based models and fusion methods have not been regularly applied in exposure assessment yet. This array of novel technologies, coupled with state-of-the-art fate modeling of chemicals will provide a complete and dynamic picture of external exposure to environmental chemicals in the near future supporting comprehensive, yet refined exposure and health risk assessment.

4.3.3.1 Improving Assessment of Activity Patterns: Use of Personal Sensors

Technological advances in recent years have produced sophisticated monitoring devices which can be carried or worn by a person during his/her regular daily routine, allowing for personal exposure to be monitored explicitly. Smartphone apps, wireless devices and the downsizing of monitoring technologies and costs make it possible for various environmental stressors and exposure factors to be measured more easily and frequently, thus providing a more reliable "time–

geography of exposure" shifting the current paradigm from population to individual level exposure.

Direct reading monitors help us to identify whether peak exposures are more important than average exposure values, identify specific exposure pathways that dominate in critical time windows over an individual's lifetime, and finally build individual exposure profiles. The advent of multiple sensor classes makes the use of sophisticated data and model fusion schemes necessary if the full potential of remote and personal sensing is to be harvested for improved cumulative exposure assessment. Such algorithmic schemes include the use of advanced statistical models such as random forest optimization, artificial intelligence techniques such as back-propagation artificial neural networks or data clustering techniques such as fuzzy set modeling.

Combining information on individual position with spatially resolved pollution levels allows assignment of pollutant concentrations to persons as they move through different microenvironments. Moreover, information on individual physical activity as tracked by personal sensors supports the estimation of breathing rates during different activities, which, in turn, translate into inhaled dose. The possibility to use personal sensors able to provide real-time data on air pollution exposure (CO_2, CO, NO_2, O_3, PM_x of different size fractions) has been explored by several investigators (Snyder et al. 2013; De Nazelle et al. 2013). If proven to be reliable, these sensors will constitute an added value to the array of remote sensing instrumentation building the sensor web of exposome related studies (Sarigiannis and Gotti 2014; Nieuwenhuijsen et al. 2014).

4.3.3.2 The Development of Space-Time-Exposure Trajectories

Time-Geography provides a coherent ontological framework within which to explore spatio-temporal behavior of individuals and their interaction with the environment. By analyzing and modeling these trajectories an individual's behavior can be determined in terms of time-geography, thus beginning to estimate individual level exposure. In Fig. 4.2 one sees conceptually how an individual can coincide spatially in X,Y and time, either with an environmental hazard prism (left) or vector (right).

In this ontological modeling framework an individual who resides in one place may be represented by a vertical line (a process in the time dimension alone) while horizontal lines show changes of place (processes in the spatial dimension as well). Time periods usually contain innumerable moves in space, which in turn create trajectories. By analyzing and modeling these trajectories, one could determine an individual's behavior in terms of time geography, and thus begin to estimate individual level exposure. With increasing access to individual residential history data, and computational power (e.g., exploring the possibilities offered by cloud-based and distributed computing), the time-geography approach has recently regained popularity in environmental health sciences. An example of this multi-layered data fusion coupled with agent-based modeling for the estimation of exposure to particulate matter through the ambient air is given graphically in Fig. 4.3.

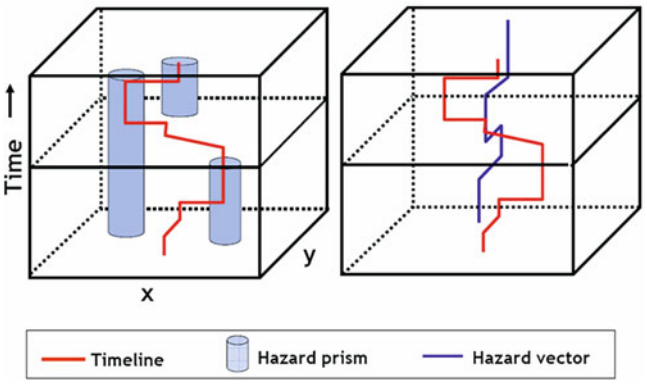

Fig. 4.2 Space-time trajectories through environmental hazard prisms (left) and vectors (right)

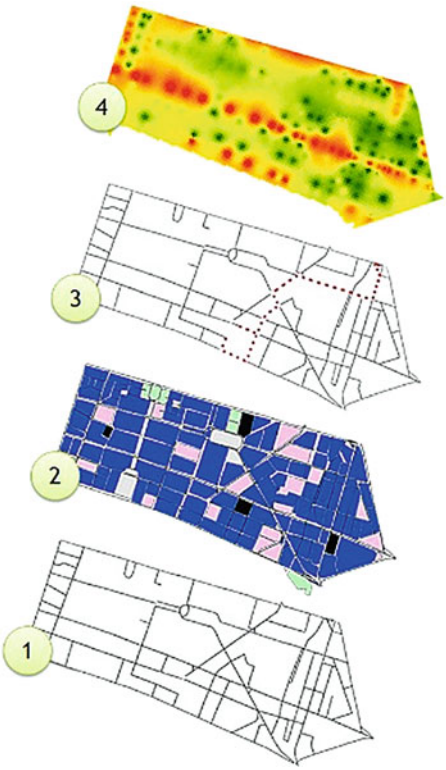

Fig. 4.3 Layer 1: road network, layer 2: buildings network, layer 3: agent's trajectory, layer 4: daily average PM concentration map

Administrative and spatially resolved infrastructure information such as the road and building networks in the area of interest (e.g., city, district) are used as knowledge substrates upon which the agent-based model estimates space-time trajectories of individual agents within the exposure time frame of reference. The emission and

environmental concentration model or the data fusion model that integrates multi-platform environmental monitoring data across the area of interest is then coupled with the space-time trajectories to reckon personal exposure estimates. Allowing the simulation to unfold and running it using a Monte Carlo algorithm to perturb the initial conditions stochastically within specified limits (determined to ensure plausibility of exposure scenarios) results in emerging patterns of behavior and corresponding combined exposure burden to the pollutants of interest. Exposure estimates are differentiated by type of population subgroup modeled, and time-dependent exposure profiles for characteristic individuals can be drawn; the method gives an explicit account of the residual uncertainty and variability in exposure profiles.

4.3.4 Critical Time Windows of Exposure

Vulnerability (defined here as variations in exposure between individuals or groups) and susceptibility (the degree to which individuals or groups may respond to a given exposure) related to complex exposure vary significantly during an individual's lifespan. Thus, it is of great importance to identify the critical periods and the types of complex exposures that require special attention during certain life stages. Within the frame of the HEALS[2] (Health and Environment-wide Associations based on Large population Surveys) project, ten critical periods of exposure were identified (Table 4.1). Critical periods include early developmental stages such as preconception, the three trimesters of pregnancy and the age before and after 3 years of age. Puberty is a period with significant hormonal alterations, and as such, it has been proven to be crucial for asthma, weight and behavioral variations. Middle age lifestyle parameters (e.g., nutrition, exercise, smoking), health status (hypertension, diabetes) and use of drugs are determinant for the onset and the progress of

Table 4.1 Critical periods of exposure

Preconception
1st semester of pregnancy
2nd semester of pregnancy
3rd semester of pregnancy
3rd year of age
Puberty
Middle age
Menopause
Age of 50
Age of 65
Age of 80

[2] www.heals-eu.eu.

neurodegenerative diseases. Change in lifestyle choices after the age of 30 introduces new conditions that increase the risk of metabolic disorders that may eventually lead to obesity and type 2 diabetes. Menopause in women (between 45 and 55 years of age) is a period of significant change in the hormonal system, related to a cascade of effects, asthma, and increased susceptibility to metabolic disorders, as well as to neuroinflammation. At the age of 50 significant changes in gene expression involved in brain-related function seem to be determinant for the onset of neurodegenerative disorders. After 65 years both males and females are more susceptible to environmental insults, due to reduced detoxification capacity, as well as reduced capacity of maintaining homeostasis. After 80 and 85 years, normal ageing is accompanied by pathological ageing.

As a general rule, for stages related to development or significant hormonal changes, assessment of complex exposures should be more focused on endocrine disrupting chemicals (EDCs), including several chemical classes, (e.g. PCBs and dioxins, phthalates, BFRs), each of which includes multiple individual mixture components. To properly account for these compounds, multiple pathways and exposure routes have to be addressed; their relative importance is also age dependent. At later stages, complex air quality mixtures (PAHs, BTEX, CO, NO_x ozone and PM_x) that relate to oxidative stress (and the related cascade of effects) are more important than exposure to EDCs. Thus, modeling efforts of complex exposures should account for the specific needs of the critical windows of exposure that pertain to the individuals or the population at risk.

4.4 Internal Exposure Modeling of Real Life Chemical Mixtures

4.4.1 Overview of Physiology Based BioKinetic (PBBK) Models

PBBK models are continuously gaining ground in regulatory toxicology, describing in quantitative terms the absorption, metabolism, distribution and elimination processes in the human body, with a focus on the effective dose at the expected target site (Bois et al. 2010). This trend is further amplified by the continuously increasing scientific and regulatory interest about aggregate and cumulative exposure; PBBK models translate external exposures from multiple routes (Yang et al. 2010) into internal exposure metrics, addressing the effects of exposure route in the overall bioavailability (Sarigiannis and Karakitsios 2011; Valcke and Krishnan 2011) or the dependence on critical developmental windows of susceptibility, such as pregnancy (Beaudouin et al. 2010), lactation (Verner et al. 2008) and infancy (Edginton and Ritter 2009). With regard to cumulative exposure, PBBK models offer the advantage of calculating the effect of the interactions among the mixture compounds at the level of metabolism, however due to the inherent difficulties arising, the existing

applications are currently limited mainly to VOCs (Haddad et al. 2000; Sarigiannis and Gotti 2008) and metals (Sasso et al. 2010). Recently, efforts have shifted towards the integration of whole-body physiology, disease biology, and molecular reaction networks (Eissing et al. 2011), as well as integration of cellular metabolism into multi-scale whole-body models (Krauss et al. 2012).

The use of internal dose modeling aims at integrating exposure data and modeling output with human biomonitoring data. Its goals are to (a) provide the time history of the exposure profile, focusing on susceptible developmental stages; (b) assimilate the biomonitoring data related to the cohorts to estimate the individual exposome in quantitative terms; and (c) derive reliable biologically effective dose values for the compounds of interest so that they can be associated to observed health outcomes. The key component of the above is the development of a lifetime (including gestation and breastfeeding) generic PBBK model (Sarigiannis and Karakitsios 2012) incorporating mixtures interaction (Sarigiannis and Gotti 2008) and a framework for biomonitoring data assimilation (Georgopoulos et al. 2008b). Aiming to expand the applicability of the generic PBBK model to cover the chemical space as much as possible, parameterization of the model for known and new chemicals with limited information is done through the development of QSAR models. The generic PBBK model will also be used to reconstruct exposure from human biomonitoring data (Andra et al. 2015). A tiered approach will be followed as a function of data availability (periodicity and size of sampling, specimen type) and requirements of the exposure reconstruction analysis (temporal analysis of exposure, contribution from different routes), ranging from Exposure Conversion Factors (Tan et al. 2006), up to Markov Chain Monte Carlo analysis. Inputs involve spatial and temporal information on micro-environmental media concentrations of xenobiotics and corresponding information on human activities, food intake patterns or consumer product use that results in intakes; outputs are the observed biomarkers; and the error metric can be defined in terms of population variation (the latter has to be lower than the intra-individual variation, which may be associated with measurement or other random error source). On the individual level, PBBK will be combined with multimedia models and survey questionnaires to identify exposure sources. PBBK modeling will also be used to estimate the internal doses of xenobiotics that exceed levels associated with biological pathway alterations (Judson et al. 2011) and, eventually, health risk. The latter can involve the use of specific omics results (e.g., metabolomics analysis) and associations of biologically effective doses to early biological responses. In addition, biologically effective doses would be used to quantify the effect of compound-induced extracellular perturbations on metabolic states, so as to directly couple the PBBK model with metabolic regulatory networks. Direct coupling defines a feedback loop that connects clearance and metabolite production rates to metabolism regulation (Eissing et al. 2011) via dynamic flux balance analysis (Krauss et al. 2012).

Considering the opportunities offered by the use of PBBK models in exposure/risk characterization, several research groups are developing generic PBBK models, either as stand-alone models such as PK-Sim (Willmann et al. 2003) and IndusChem (Jongeneelen and Berge 2011), or incorporated within integrated computational platforms for exposure assessment such as INTERA (Sarigiannis et al. 2011)

and MENTOR (Georgopoulos et al. 2008c). The development of generic PBBK models is substantiated by the recent advances in quantitative structure–activity relationships (QSARs) and quantitative structure–property relationships (QSPRs) (Price and Krishnan 2011; Peyret and Krishnan 2011), providing the basis for development of relevant PBBK models for data-poor or new chemicals.

The INTEGRA methodology is advancing the existing state of the art by integrating all of the above elements, with a plan to develop a generic lifetime (including pregnancy) (Sarigiannis and Karakitsios 2012) multi-route PBBK model. The integration of this generic PBBK model into a wider modeling framework will allow forward (internal exposure) or reverse calculations (exposure reconstruction) so as to provide the link among exposure components and biomonitoring data. Additional elements of using physiologically based modeling to understand the kinetics and effects of chemical mixtures are covered in Chap. 12.

4.4.2 Internal Dosimetry Models

PBBK models are tools that describe the mechanisms of absorption, distribution, metabolism and elimination of chemicals in the body resulting from acute and/or chronic exposure regimes. They are independent structural models, comprising the tissues and organs of the body with each perfused by, and connected via, the blood circulatory system. In PBBK models the organism is frequently represented as a network of tissue compartments (e.g., liver, fat, slowly perfused tissues, and richly perfused tissues) interconnected by systemic circulation. A generic PBBK model, reflects the incorporation of basic physiology and anatomy. The compartments actually correspond to anatomic entities such as liver, lung, etc., and the blood circulation conforms to the basic mammalian physiology. The primary means of transport for xenobiotic chemicals that enter the body through one or more of these routes is via blood, the main vehicle for nutrient supply and waste removal from tissues. In the basic PBPK model, transport of chemicals between blood and tissues is assumed to be flow-limited, which implies that the transport barriers between the free molecules of chemical in blood and tissue are negligible, and equilibration between free and bound fractions in blood and tissue is rapid. Concentrations of chemical in venous blood exiting a tissue, and tissue concentrations are assumed to be at equilibrium, and the tissue is assumed to be homogeneous with respect to the concentration of the chemical. The flow-limited assumption is usually appropriate for lipophilic or low molecular weight compounds, which easily partition or diffuse through cell membranes. Every PBBK model requires several parameters that are critical determinants of chemical uptake and disposition. These determinants can be classified into three main categories, namely, anatomical/physiological, physicochemical, and biochemical. A partial list of anatomical/physiological parameters includes cardiac output, tissue blood flow rate, organ and tissue weight and volumes. In addition to physiological/anatomical data, PBBK models require information on the ability of the body to metabolize chemicals – these are known as biochemical

parameters. Typical biochemical parameters include the maximal velocity for metabolism (V_{max}), binding association constant (Kb) and Michaelis affinity constant (K_m). The third type of data required by these models is the solubility of pollutants in the organs and tissues of the body. These are physicochemical data known as partition coefficients (P). Partition coefficients are experimentally determined parameters that give an indication of the distribution of a chemical between two different phases, e.g. air and blood, blood and liver, blood and muscle, blood and fat, etc. The fundamentals of PBBK modeling are to identify the principal organs or tissues involved in the disposition of the chemical of interest and to correlate the chemical absorption, distribution, metabolism, and excretion within and among these organs and tissues in an integrated and biologically plausible manner.

A scheme is usually formed where the normal physiology is followed in a graphical manner. Within the boundary of the identified compartment (e.g., an organ or tissue or a group of organs or tissues), whatever inflows must be accounted for via whatever outflows or whatever is transformed into something else. This mass balance is expressed as a mathematical equation with appropriate parameters carrying biological significance. A generic equation, for any tissue or organ, is:

$$V_i \frac{dC_{ij}}{dt} = Q_i\left(\mathrm{CA}_j - \mathrm{CV}_{ij}\right) - \mathrm{Metab}_{ij} - E\mathrm{lim}_{ij} + \mathrm{Absorp}_{ij} - \mathrm{PrBinding}_{ij} \quad (4.12)$$

where V_i represents the volume of tissue group i, Q_i is the blood flow rate to tissue group i, CA_j is the concentration of chemical j in arterial blood, and C_{ij} and CV_{ij} are the concentrations of chemical j in tissue group i and in the effluent venous blood from tissue i, respectively. Metab_{ij} is the rate of metabolism for chemical j in tissue group i; liver, is the principal organ for metabolism and, with some exceptions, Metab_{ij} is usually equal to zero in other tissue groups. Elim_{ij} represents the rate of elimination from tissue group i (e.g., biliary excretion from the liver), Absorp_{ij} represents uptake of the chemical from dosing (e.g., oral dosing), and $\mathrm{PrBinding}_{ij}$ represents protein binding of the chemical in the tissue. All these terms are zero unless there is definitive knowledge that the particular organ and tissue of interest has such processes.

A series of similar mass balance differential equations representing all of the interlinked compartments are formulated to express a mathematical representation, or model, of the biological system. This model can then be used for computer simulation to predict the time course behavior of any given parameter in the model. See Chap. 12 for more information on PBBK development.

The generic model developed in INTEGRA is designed to describe as closely as possible the actual absorption, distribution, metabolism and elimination processes occurring in the human body, so that it can be easily applicable for a broad variety of chemicals assuming proper parameterization. The model includes the parent compounds and at least three potential metabolites for each of the compounds in the mixture. For each compound/metabolite all major organs are included and the link among the compounds and the metabolites is through the metabolizing tissues. This is mainly the liver, but also other sites of metabolism (e.g., gut, skin) might be

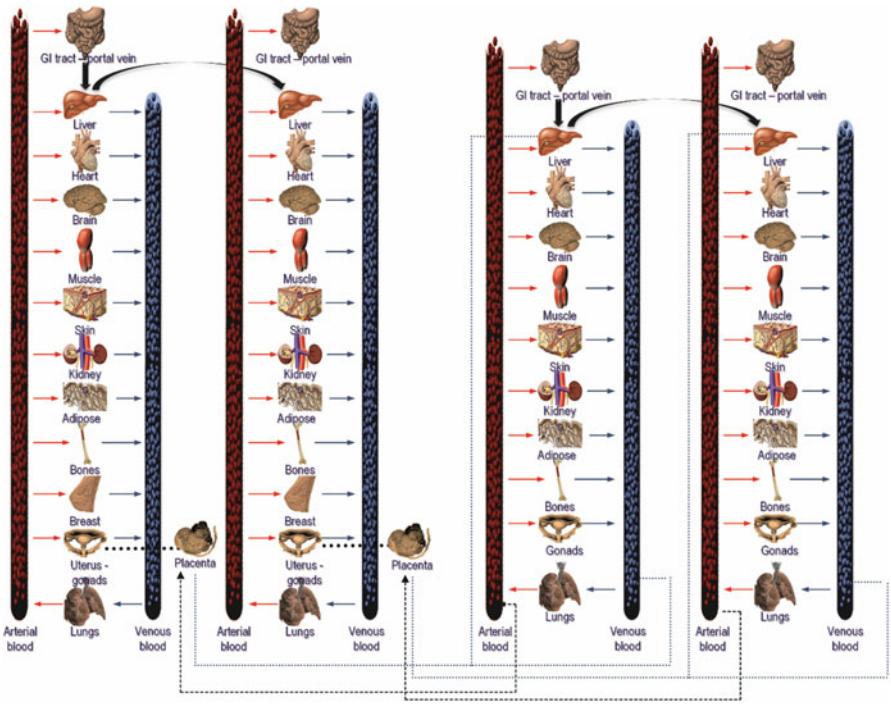

Fig. 4.4 Conceptual representation of the Mother-Fetus PBBK model, including both the parent compound and one metabolite

considered based on the presence of the enzymes involved in the metabolism of the compound of interest. To capture in utero exposure, the model is replicated to describe the functional interaction of the mother and the developing fetus through the placenta (Fig. 4.4). The anthropometric parameters of both the mother and the fetus models are age-dependent, so as to provide a life stage-dependent internal dose assessment.

4.4.3 Expanding the Chemical Space to Assess Internal Dose for Multiple Chemicals

A critical limiting factor in describing ADME processes accurately for a large chemical space is the proper parameterization of PBBK models for "data poor" compounds. Advanced Quantitative Structure-Activity Relationships (QSARs) can be used to predict input parameters for these models allowing PBBK models to cover a large number, and several classes, of chemicals. In silico approaches, including QSARs, are widely used for the estimation of physicochemical and biochemical properties and predicting how they might lead to biological responses (Puzyn et al.

2010). QSARs are described as regression or classification models, which form a relationship between the biological effects and chemistry of each chemical compound (Puzyn et al. 2010). Significant progress in expanding the chemical space for industrial chemicals has been made by the INTEGRA project, where parameterization of essential parameters such as blood:tissue partition coefficients for several tissues, maximum initial velocity of the enzyme catalyzed reaction (V_{max}) and the substrate concentration that gives half maximal velocity of an enzymatic reaction (K_m or Michaelis-Menden constant) has been carried out for a large number of chemicals. The mathematical formulation coupled Abraham's solvation equation with Artificial Neural Networks of variable geometry in order to optimize the performance of the model. Abraham's solvation equation (Linear Free Energy Relationship) describes the process of the transfer of chemicals from the liquid phase to a large number of solvents or other condensed phases, including biophases. The descriptors, which characterize these physicochemical and biochemical phenomena, are combined into Eq. 4.13,

$$\log SP = c + e \cdot E + s \cdot S + a \cdot A + b \cdot B + v \cdot V \tag{4.13}$$

Where SP is a biological property for a set of chemicals in a given system. The independent descriptors are the properties of the examined chemicals, E is the excess molar refractivity of the chemical, S is the chemical's dipolarity/polarizability, A and B are the chemical's effective or summation hydrogen bond acidity and basicity, respectively, and V is the McGowan characteristic volume of the chemical (Abraham 1993; Payne and Kenny 2002). The coefficients c, e, s, a, b and v reflect the properties of chemicals, so e corresponds to the tendency of the chemical to interact with solute π- and n- electrons, s corresponds to the chemical's dipolarity/polarizability, a and b correspond to the chemical's hydrogen bond basicity and acidity, respectively, and v is a measure of the chemical's lipophilicity. Artificial Neural Networks were used to develop a non-linear model based on Abraham's solvation equation.

The calculated values of metabolic constants using the statistical method described above (Abraham's solvation equation coupled with Artificial Neural Networks) were compared to experimental values and the results obtained by Price and Krishnan (2011) in Fig. 4.5. The methodology followed by Price and Krishnan (2011) was based on the group contribution method, implying that each fragment in the molecular structure contributes to the metabolic parameters, depending on its frequency of occurrence in the given molecule (Gao et al. 1992). In previous studies, the parameters used to describe the interactions between chemicals and tissues were mainly related to chemical structure or tissue composition in water, proteins and lipids Price and Krishnan 2011; Zhang 2004). In the present example, Abraham's equation descriptors are not linked directly with tissue composition. They encode specific chemical information regarding the size, polarizability and hydrogen bonding of the examined chemicals and each term can reveal the factors that influence a particular interaction. The modeling results indicate that the molecular descriptors of the equation can be suitable for the estimation of the parameters that characterize

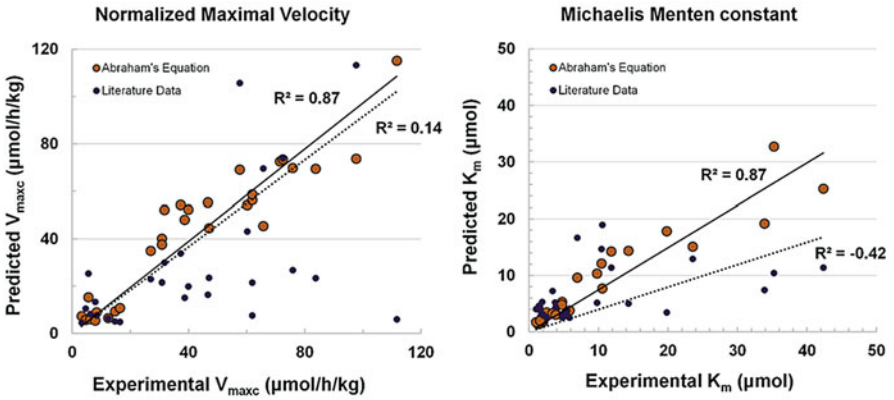

Fig. 4.5 Predicted vs experimental values of normalized maximal velocity and Michaelis – Menten constant under Abraham's equation (orange dots) and a group contribution method (Literature data; blue dots)

relevant physicochemical and biochemical phenomena. The improved performance of Abraham's equation compared to the group contribution method can be attributed to its capacity to represent the complex interactions of the micro-processes of chemicals' distribution and metabolism into several tissues.

4.5 Complex Exposure Modeling Using Human Biomonitoring Data

4.5.1 Overview of Biomonitoring

The main achievement of human biomonitoring is that it provides an integrated overview of the pollutant load to which an individual is exposed, and hence serves as an excellent approximation of aggregate exposure including all pathways, mechanisms and routes of exposure. For additional information on biomonitoring and its utility in measuring exposure to mixtures, see Chap. 2. The internal dose of a chemical, following aggregate exposure has a much greater value for environmental health impact assessment as the internal body concentration is much more relevant to the impact on human health than mere exposure data. However, it needs to be stressed that HBM in itself cannot replace environmental monitoring and modeling data. At the same time, mathematical approaches to describe the pharmacokinetic and toxicokinetic behavior of environmental agents (i.e. PBBK models) offer a more mechanistic insight into the behavior and fate of environmental agents following exposure. As biomarker data also reflect individual ADME characteristics of chemicals, HBM data offer an excellent opportunity to validate PBBK models. Ultimately, coupling both lines of evidence to assess exposure proves to be the

optimal solution towards relating complex exposure to environmental stressors to potential adverse health effects assessment.

There are three approaches for linking biomonitoring data to health outcomes: direct comparison to toxicity values, forward dosimetry, and reverse dosimetry. Biomonitoring data can be directly compared to toxicity values when the relationship of the biomarker to the health effect of concern has been characterized in the human. In forward dosimetry, pharmacokinetic data in the experimental animal can be used to support a direct comparison of internal exposure in humans derived through the application of PBBK models, providing an estimate of the Margin of Safety in humans. It is possible to determine the relationship between biomarker concentration and effects observed in animal studies. An evolution of this concept is the biomonitoring equivalents. Alternatively, reverse dosimetry can be performed to estimate the external exposure that is consistent with the measured biomonitoring data through the backward application of PBBK models. In a more elaborate scheme, the reconstructed exposure could be used to run the PBBK model in forward mode, so as to estimate the biologically effective dose at the target tissue.

4.5.2 Exposure Reconstruction in Practice

Human biomonitoring typically is an integrative measure of different exposure episodes along various routes and over different time scales; thus, it is often difficult to reconstruct the primary exposure routes from human biomonitoring data alone. This uncertainty limits the interpretative value of biomarker data. However, several mathematical approaches have been developed to reconstruct exposures related to population biomonitoring studies, and can be subdivided into a number of different approaches. Exposure reconstruction techniques combined with PBBK models can be divided into Bayesian and non-Bayesian approaches (Georgopoulos et al. 2008a). Moreover, computational inversion techniques (and exposure reconstruction techniques as well), can be classified as deterministic or stochastic (Moles et al. 2003) based on the identification of a global minimum of the error metric, the input parameters and the model setup.

The deterministic methods aim to achieve convergence on a global minimum. The problem is solved using an "objective function" based on biomarkers. Additionally, constraints in the form of bounds, equalities and inequalities are incorporated. Deterministic models have been used in several biological applications using different methods. Muzic Jr and Christian (2006) have applied a regression technique to estimate pharmacokinetic parameters. A gradient method has been used by Isukapalli et al. (2000) to calculate the uncertainty in PBBK models. A maximum likelihood method has been carried out for short- and long-term exposure reconstruction using a PBBK model for chloroform (Roy et al. 1996).

In contrast, stochastic methods aim to provide a reasonable solution, not a mathematically optimal one. A probabilistic framework for the inverse computation problem is the Bayesian approach, which is based on Bayes' theorem. According to

the methodology developed in the frame of the INTEGRA project, the analysis of exposure reconstruction problems based on the Markov Chain Monte Carlo and Differential Evolultion Markove Chain technique is realized according to the following steps:

1. The process starts from exposure related data which are fed into the INTEGRA exposure model;
2. This in turn provides input to the PBBK model, taking into account the duration and the magnitude of exposure from all exposure routes (inhalation, skin and oral route);
3. The result of the PBBK model simulation (also taking into account the distribution of PBBK parameters, e.g., inter-individual variability in clearance), is then evaluated against the human biomonitoring data distributions. Based on the outcome of the comparison, the optimization algorithm changes the exposure model input parameters after each iteration, so as to achieve convergence to biomonitoring data;
4. More detailed information on exposure parameters reduces uncertainty in back-calculating doses from biomarker information, resulting in faster and more efficient convergence;
5. Several iterations are repeated, until the error between the predicted and the actual biomonitored data is minimized.

The Bayesian Markov Chain Monte Carlo technique described above simulates and calculates the investigated exposure conditions. The sampling scheme is set appropriately according to the problem and to the available data for the proposed function. The flowchart of the overall process is shown in Fig. 4.6.

4.6 Case Studies of Complex Mixtures Modeling

4.6.1 Exposure Assessment of Indoor Air Complex Mixtures

Indoor air is one of the most typical examples related to complex exposures. The combination of building materials (e.g., paint, floors, doors and windows), consumer products (e.g., electronic devises, furniture, carpets) and activities (e.g., biomass combustion, smoking, cooking) creates a variable and complex mixture of chemical and biological health stressors (e.g., mold, pollen). The multitude of compounds found in indoor air (Sarigiannis 2014), as well as the respective health risks are graphically illustrated in Fig. 4.7.

Modeling complex exposures in the indoor environment requires a virtual reconstruction of the actual environmental setting of interest. This implies the virtual reconstruction of the indoor environment, including all potential emission sources. After calculating emissions, the next step is to calculate indoor concentrations in the three media of exposure relevance, meaning gaseous phase, particles and settled dust. The latter is of particular interest, since based on the physicochemical

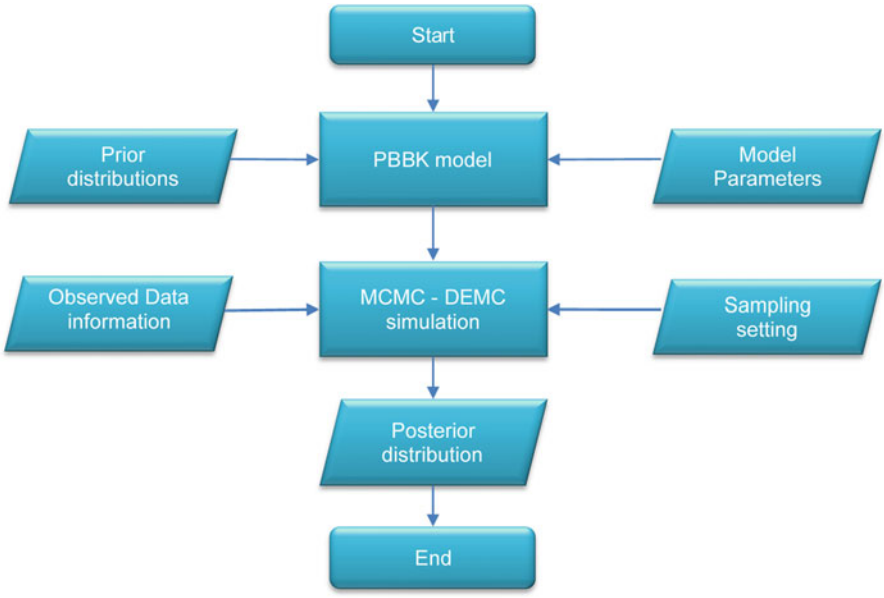

Fig. 4.6 Exposure reconstruction flowchart

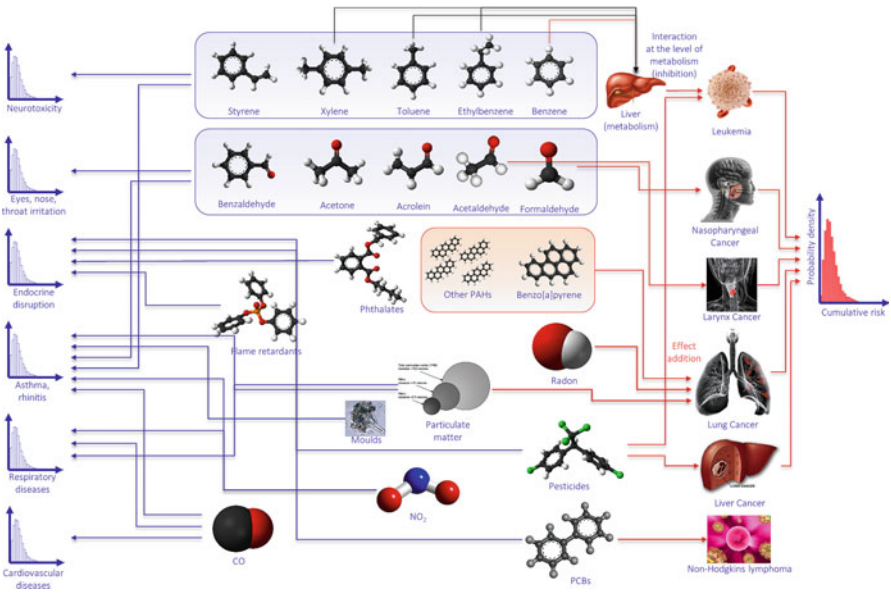

Fig. 4.7 Multiple stressors found in indoor environment and related health endpoints

properties of the compound (e.g., K_{ow} and Henry constant), significant differences in the respective phase distribution are expected; more volatile compounds such as aromatics and aldehydes are found only in the gaseous phase, semi-volatile compounds (e.g., phthalates) tend to distribute in all phases, while heavier and more lipophilic compounds (e.g. PBDEs) are found mostly in dust (Weschler and Nazaroff 2010; Weschler and Nazaroff 2008). In turn, the way compounds are distributed in different phases determines the pathways and routes of exposure involved, e.g., non-dietary ingestion and inhalation respectively. This allows the proper estimation of external exposure, which in turn provides input to the internal exposure model. The full series of calculations, starting from emissions, calculating indoor environmental levels, exposure and internal dose for many chemicals can be performed with the INTERA computational platform[3] (Sarigiannis et al. 2012a), an open access online computational platform running via the world wide web at the Centre for Research and Technology Hellas (CERTH). A special case of complex exposure in the indoor environment is tobacco smoke. During smoking, several compounds are emitted, including particles and organic compounds such as alkenes, nitrosamines, aromatic and heterocyclic hydrocarbons and amines. Some of these compounds are emitted from several other sources as well; thus, it is not always easy to attribute poor indoor air quality to cigarette side stream smoke. However, nicotine, serves as a unique marker of exposure to environmental tobacco smoke. Nicotine is rapidly metabolized to cotinine upon entering the human body, which is excreted through urine. Urinary cotinine serves as an exposure biomarker to environmental tobacco smoke. Using complex exposure modeling the amount of cotinine found in urine could be used as a starting point for reconstructing exposure to nicotine. This would allow identification of exposure levels and in turn the indoor concentration of nicotine that resulted in the observed biomonitored cotinine levels (Sarigiannis et al. 2009). By continuing the reverse calculation, smoking intensity is estimated. At this point, estimated smoking intensity can be used to estimate the emissions and concentration levels for the hundreds of compounds present in environmental tobacco smoke...

Exposure reconstruction of urinary cotinine levels allows us to further identify the exposure and effects of individual carcinogenic compounds e.g., benzene, formaldehyde, Nicotine-derived nitrosamine ketone (NNK), B[a]P and their interactions at different levels among the individual compounds (Sarigiannis et al. 2009):

- Interaction at the level of metabolism (using PBBK modeling) among benzene, toluene, ethylbenzene and xylene
- Effect summation of lung cancer related to NNK and B[a]P
- Independent action in terms of cumulative cancer risk (at different sites) among benzene (leukemia), formaldehyde (nasopharyngeal cancer), NNK and B[a]P (lung cancer)

[3] http://www.intera.cperi.certh.gr/auth/login.

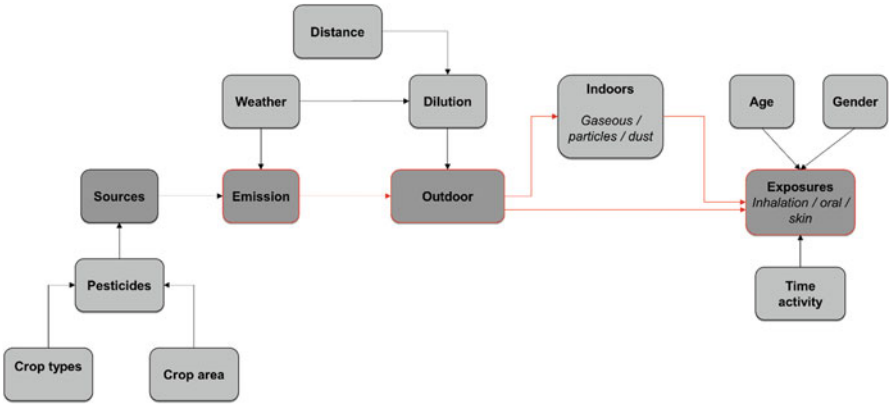

Fig. 4.8 Full chain complex exposure assessment from pesticides

4.6.2 Pesticides: Multi-pathway and Multi-route Exposure by Chemical Class

Complex exposure to pesticides for bystanders has been computed on the European scale at a very high spatial resolution using a multi-compartment model, the information flow of which is depicted in Fig. 4.8. The modeling methodology has three main components: (1) modeling the emission of active substances (AS) (i.e., AS emissions to air per km^2 extracted from the emission inventory), (2) modeling the fate and transport of the AS in the environment to estimate concentrations (expressed in computed AS concentration per hour in 1 year), and (3) modeling population exposure (expressed as intake computed from daily average AS concentrations) differentiated by age and gender for all AS. The overall model is spatially resolved and all estimates are given at a pan-European 0.1 × 0.1 km grid.

The multi-step methodology used included the following steps:

1. Starting from the emission inventory, annual emission data per AS were extracted for 25 EU Member States.
2. A typical emission profile was used in accordance to the local agriculture practices for a time window that coincided with the growing season in each country/region.
3. A pesticide dispersion model was developed to compute concentration into the ambient air at a 1 × 1 km grid using as input the variable emission profile, local meteorological data and AS physicochemical characteristics.
4. Outdoor to indoor penetration modeling was used to estimate the indoor concentration of AS and its partitioning among the different phases (gaseous, particles and settled dust).
5. An exposure model was developed, based on which intake rates per population group differentiated by age and gender (i.e., adult male-female, children 0–4 yr, 5–9 yr and 10–14 yr) was computed comprising all exposure pathways

(inhalation of gaseous and particles, dust ingestion, particles deposition on skin, dust rubbing off) and routes. The effect of changing the daily duration of pesticide application, the total application window as well as uncertainty in the meteorological conditions and variability in the physiological parameters were incorporated in the assessment.

The pesticide release inventory model (Sarigiannis et al. 2013) comprised five crop types (three seasonal and two permanent), the list of pesticides used per crop and their usage quantities at the country level (source: Eurostat 2011), a pesticide disaggregation algorithm to distribute quantities at the grid and the computed annual emission data based on wind drift, volatilization during application and from the crop canopy. This inventory model was based on crop data extracted from the Common Agricultural Policy Regionalised Impact (CAPRI) Modeling System (CAPRI 2012).

According to the methodology followed to create this pesticide inventory (Sarigiannis et al. 2013), usage quantities for each AS per cell and crop were disaggregated from the country level to the 1 × 1 km grid, using an area weighing algorithm and assuming a constant annual 'area reduced dosage' per AS for each cell in the same country. The 'area reduced dosage' is a measure of pesticide use per crop area based on country data, and incorporates estimates of uncertainty in the actual crop area to which a specific AS is applied within the spatial grid. Average annual emissions, ER_{ijk},(in kg/yr), of each AS applied in a field to the air of an AS i applied on crop j for a country k, at short range from the site of application were computed from the sum of spray wind drift D_{ijk}, volatilization during application $E_{\text{app,air},ijk}$ and volatilization from the crop canopy $E_{\text{crop,air},ijk}$ as shown by Eq. 4.14.

$$\text{ER}_{ijk} = D_{ijk} + E_{\text{app, air}, ijk} + E_{\text{crop, air}, ijk} \tag{4.14}$$

The annual emission data generated from Eq. 4.14 per grid cell were fed to the concentration model described in the following sections, assuming typical emission profiles.

The pesticides were prioritized on the basis of a hazard factor that accounts for both toxicity and persistence in the environment based on the methodology of Gunier et al. (2001). According to this methodology, the hazard factor (HF) is multiplied by AS quantities and then the AS with the highest score (HF × quantity) from each group (i.e., herbicides, fungicides, insecticides, other) was selected for more detailed presentation. The top chemicals in each group were glyphosate (herbicide), chlorpyrifos (insecticide), mancozeb (fungicide) and 1,3-dichloropropene (other). 1,3-dichloropropene has one of the largest hazard factors due to high toxicity and high volatilization flux.

In practice, application periods are limited to 1–3 months during the year and correspond to specific crop types, climatic conditions and agricultural practices that differ among countries, even among regions. It is assumed that the applicators of pesticides and farmers use the total quantity of pesticides in a specific time period, regardless of weather conditions to render the assessment conservative. Therefore, the annual quantity of AS i for crop j for a country k (ER_{ijk}) is applied, in the form of

a 'pulse' with a 10-h period for a total window of several months, in accord with the estimated country annual emission estimates and the typical AS uses. Moreover, since application practices in Europe vary, pesticide drift does occur in many cases, differentiated between primary drift, off-site movement of spray at the time of application, and secondary drift associated with pesticide vapor. The effects of pesticide drift were included in this assessment via the AgDrift (Teske et al. 2002) and AgDisp (Bird et al. 2002) models. They were used to evaluate the average deposition fraction (i.e., implicitly drift), under different operational and environmental conditions, focusing in particular, on the droplet size in accordance, to the ASAE S572 standard (very fine <150 μm, fine 150–250 μm, medium 250–350 μm and coarse 350–425 μm), the wind speed, the temperature and the relative humidity. Hence, changes in emissions over time were deduced and used as input to the concentration model.

A critical step for calculating exposure was the estimation of outdoor concentrations. The concentration model employed at each cell was of a box-volume form, described by the differential equation:

$$V \cdot \left(\frac{dC_{ijk}}{dt}\right) = \mathrm{ER}_{ijk} - C_{ijk} \cdot I \cdot V - K_i \cdot C_{ijk} \cdot V \tag{4.15}$$

where C_{ijk} is the concentration of an AS i applied on crop j for a country k, in g/m^3, ER_{ijk} is the average emission rate of an AS i, in g/h during application, I is the air changes per hour in the volume (i.e., $I = u/L$ with u the average wind speed in m/s, L the lateral distance covered in m), V is mixing volume, in m^3 (i.e., $V = L^2 \cdot H$, with L the lateral distance, in m and H the mixing height, in m), t is the time, in h, Ki is the decay rate of an AS i, in h^{-1} (i.e., $Ki = ln2/(\mathrm{HL}_i)$, with HL_i the half life in air of an AS i, in h). The following solution of Eq. 4.15 is obtained for discrete time steps Δt:

$$C_{ijk}(t) = \frac{1}{u/L + \ln 2/\mathrm{HL}_i} \cdot \left(\frac{\mathrm{ER}_{ijk}(t)}{L^2 \cdot H}\right) \cdot \left(1 - \exp\left(-\left(\frac{u}{L} + \frac{\ln 2}{\mathrm{HL}_i}\right) \cdot \Delta t\right)\right) + C_{ijk}(t-1) \cdot \exp\left(-\left(\frac{u}{L} + \frac{\ln 2}{\mathrm{HL}_i}\right) \cdot \Delta t\right) \tag{4.16}$$

In addition, when the application rate is zero, Eq. 4.16 becomes,

$$C_{ijk}(t) = C_{ijk}(t-1) \cdot \exp\left(-\left(\frac{u}{L} + \frac{\ln 2}{\mathrm{HL}_i}\right) \cdot \Delta t\right) \tag{4.17}$$

Pesticides in the particle phase were also estimated. This calculation was based on the partition coefficient K_p between gaseous and particles phase based on the Pankow model (Pankow 1994):

$$K_p = \frac{N_s \cdot \alpha_{\mathrm{TSP}} \cdot T \cdot e^{\frac{(Q_1 - Q_v)}{R \cdot T}}}{1600 \cdot p_L^{\circ}} \tag{4.18}$$

where N_s (cm^{-2}) is the available surface for adsorption, a_{tsp} ($m^2\ g^{-1}$) is the special surface of aerosols, Q_I (kJ mol^{-1}) is the enthalpy of adsorption from the surface, Q_V is the enthalpy of vaporization of the subcooled liquid, R is the ideal gas constant, T *is the temperature* (°K), and p_L° is the vapor pressure at 25 °C. Pesticide concentrations in the particle phase were then estimated by the following relationship:

$$K_p = \frac{F/\mathrm{TSP}}{A} \tag{4.19}$$

where F is the concentration of pesticides in the particles phase, TSP is the total suspended particles (in practice all the amount of pesticides is adsorbed in PM up to 10 μm aerodynamic diameter) and A is the concentration of pesticides in the gaseous phase. This calculation was done for each of the AS using EPISuite v4.11 (EPA 2012)

Concentration estimates were obtained with a time step of 1 h. This ambient air concentration was used as input to the microenvironmental model, allowing the estimation of the concentration in the different exposure relevant indoor environmental media (gaseous, particles and dust phase).

The inhalation exposure model is described by Eq. 4.20, where the daily average intake rate IR_{ijkg} (in mg/kg_bw/day) for each AS at each cell, was computed from the pesticide concentration, both outdoor and indoor (integrated over a year), the exposed group's inhalation rate and body weight and from the total time of exposure. At each time step, the respective outdoor or indoor concentration was estimated based on the activity pattern of the exposed individuals. The exposed population groups considered, included infants, children aged 4–9 years, 10–14 years, adult females and males. For each age, gender and ethnicity group, different inhalation rates (ICRP 2002), amount of dust ingested (Wormuth et al. 2006) and body weights (Sarigiannis et al. 2012b) were used.

$$IR_{ijkg} = \left(\frac{Q_{\mathrm{inh},g} \cdot t_{\mathrm{exp}}}{\mathrm{BW}_g}/365\right) \cdot \left(\int_{t_1}^{t_2} C_{ijk}(t)dt + \int_{t_2}^{t_3} C_{ijk}(t)dt + \ldots + \int_{t_{n-1}}^{t_n} C_{ijk}(t)dt\right) \tag{4.20}$$

where C_{ijk} is the average pesticide concentration in the exposure medium (in mg AS/m^3) over the exposure period (t_{exp}), $Q_{inh,g}$ is the daily inhalation rate per gender category g (in m^3 air/d), BW_g is the body weight (in kg) per gender category g, t_{exp} is 1 day and t_n is the total simulation time (in hours). The same approach was used for the inhaled pesticides adsorbed in particles. In this case, the actual intake (taking into account the deposition fractions based on PM size distribution) was estimated. Similar considerations (in terms of exposure duration and age and gender dependence) were made for the other pathways and routes, which in practice included (a) particles deposition on the skin, (b) dust exposure to skin through rubbing off and (c) dust ingestion due hand to mouth behavior. Overall intake on a daily basis was the sum of the intake rate from all exposure pathways.

4.7 Conclusions

Exposure assessment is the weak link in the chain of calculations required for assessing the risk of chemical mixtures. The current understanding of the need to capture exposures that take place during different key periods of one's life (i.e. the exposome) to properly investigate the link between chemical mixtures and human health warrants the use of complex models. Intake fraction modeling is a good start for screening purpose modeling. However, more detailed insights on exposure drivers and patterns, dynamics in space and time and variation by gender, age, socio-economic status, location and other determinants are needed to properly account for co-exposure to multiple chemicals in real life. Our work has shown that the real integrator is the human body, i.e. that internal exposure should be considered to properly capture the health effects of complex chemical exposure. Indeed, age, physiology, metabolic capacity, pre-existing health condition and exposure history (especially to persistent and biocumulative compounds) affect significantly how uptake dose of chemicals is transformed into biologically effective dose at the relevant target tissue. Integrating external with internal exposure is key to improving health risk assessment of chemical mixtures. Integrated complex exposure modeling facilitates the assimilation of human biomonitoring data in the actual exposure estimation.

Complex exposure modeling helps the assimilation of human biomonitoring data for exposure estimation. It also helps capture and quantify potential interactions between mixture components at realistic / actual exposure doses. Biokinetics and biodynamics of active xenobiotics may be perturbed from co-exposure to chemicals, which compete for the same metabolic receptor sites or induce allosteric effects perturbing metabolic pathways that may be linked to adverse outcome pathways. Being able to mathematically describe such perturbatory mechanisms avoiding the complexity and cost of extensive experimentation helps to tackle mechanistically the effects of co-exposure to multiple compounds and/or elements. Modeling platforms such as INTEGRA and MENTOR provide the necessary computational infrastructure to perform high performance computing so as to reckon the biologically effective dose of xenobiotics in a mixture and their toxic metabolites.

References

Abraham, M.H. 1993. Application of solvation equations to chemical and biochemical processes. *Pure and Applied Chemistry* 65: 2503–2512.

Andra, S.S., P. Charisiadis, S. Karakitsios, D.A. Sarigiannis, and K.C. Makris. 2015. Passive exposures of children to volatile trihalomethanes during domestic cleaning activities of their parents. *Environmental Research* 136: 187–195.

Beaudouin, R., S. Micallef, and C. Brochot. 2010. A stochastic whole-body physiologically based pharmacokinetic model to assess the impact of inter-individual variability on tissue dosimetry over the human lifespan. *Regulatory Toxicology and Pharmacology* 57: 103–116.

Benford, D.J., and D.R. Tennant. 1997. Food chemical risk assessment. In *Food chemical risk analysis*. London: Kluwer Academic Publishers.

Bennett, D.H., M.D. Margni, T.E. Mckone, and O. Jolliet. 2002a. Intake fraction for multimedia pollutants: A tool for life cycle analysis and comparative risk assessment. *Risk Analysis* 22: 905–918.

Bennett, D.H., T.E. Mckone, J.S. Evans, W.W. Nazaroff, M.D. Margni, O. Jolliet, and K.R. Smith. 2002b. Defining intake fraction. *Environmental Science and Technology* 36: 207a–211a.

Bernillon, P., and F.Y. Bois. 2000. Statistical issues in toxicokinetic modeling: A Bayesian perspective. *Environmental Health Perspectives* 108: 883–893.

Bird, S.L., S.G. Perry, S.L. Ray, and M.E. Teske. 2002. Evaluation of the AgDISP aerial spray algorithms in the AgDRIFT model. *Environmental Toxicology and Chemistry* 21: 672–681.

Bogen, K.T., A.C. Cullen, H.C. Frey, and P.S. Price. 2009. Probabilistic exposure analysis for chemical risk characterization. *Toxicological Sciences* 109: 4–17.

Bois, F.Y., M. Jamei, and H.J. Clewell. 2010. PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. *Toxicology* 278: 256–267.

CAPRI. 2012. *Common agricultural policy regionalised impact modelling system.* Available at http://www.capri-model.org/dokuwiki/doku.php [Online]. Available: http://www.capri-model.org/dokuwiki/doku.php 2012.

CROPLIFE. 2002. *Croplife America CARES* [Online]. Available: http://www.croplineamerica.org 2010.

De Nazelle, A., E. Seto, D. Donaire-Gonzalez, M. Mendez, J. Matamala, M.J. Nieuwenhuijsen, and M. Jerrett. 2013. Improving estimates of air pollution exposure through ubiquitous sensing technologies. *Environmental Pollution* 176: 92–99.

Delmaar, J.E., M.V.D.Z. Park, and J.G.M. van Engelen. 2005. *Consexpo 4.0: Consumer exposure and uptake models: Program manual*. Bilthoven: National Institute for Public Health and the Environment (RIVM).

Edginton, A.N., and L. Ritter. 2009. Predicting plasma concentrations of bisphenol a in children younger than 2 years of age after typical feeding schedules, using a physiologically based toxicokinetic model. *Environmental Health Perspectives* 117: 645–652.

EFSA. 2011. Guidance of EFSA: Use of the EFSA comprehensive European food consumption database in exposure assessment. *EFSA Journal* 9: 2097–2131.

Egeghy, P.P., D.A. Vallero, and E.A. Cohen Hubal. 2011. Exposure-based prioritization of chemicals for risk assessment. *Environmental Science & Policy* 14: 950–964.

Eissing, T., L. Kuepfer, C. Becker, M. Block, K. Coboeken, T. Gaub, L. Goerlitz, J. Jaeger, R. Loosen, B. Ludewig, M. Meyer, C. Niederalt, M. Sevestre, H.U. Siegmund, J. Solodenko, K. Thelen, U. Telle, W. Weiss, T. Wendl, S. Willmann, and J. Lippert. 2011. A computational systems biology software platform for multiscale modeling and simulation: Integrating whole-body physiology, disease biology, and molecular reaction networks. *Frontiers in Physiology* 2: 1–10.

EPA. 2012. EPI Suite™ V4.11. http://www.epa.gov/oppt/exposure/pubs/episuitedl.htm

ERG. 2001. *Final exposure and fate assessment screening tool peer review summary, EPA contract no. 68-W6-0022*. Chantilly: Eastern Research Group, Inc.

EUROSTAT. 2011. Population data [Online]. Available: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home

Farrier, D. S., and M. D. Pandian. 2002. In: America, C. (ed.) *CARES cumulative aggregate risk evaluation system: User guide*. Washington, DC: CropLife America.

Fryer, M., C. Collins, R. Colvile, H. Ferrier, and M. Nieuwenhuijsen. 2004. *Evaluation of currently used exposure models to define a human exposure model for use in chemical risk assessment in the UK*. London: Department of Environmental Science and Technology, Imperial College.

Gao, C., R. Goind, and H.H. Tabak. 1992. Application of the group contribution method for predicting the toxicity of organic chemicals. *Environmental Toxicology and Chemistry* 11: 631–636.

Gelman, A., and D.B. Rubin. 1996. Markov chain Monte Carlo methods in biostatistics. *Statistical Methods in Medical Research* 5: 339–355.

Geman, S., and D. Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6: 721–741.

Georgopoulos, P.G., and P.J. Lioy. 2006. From a theoretical framework of human exposure and dose assessment to computational system implementation: The Modeling ENvironment for TOtal Risk Studies (MENTOR). *Journal of Toxicology and Environmental Health – Part B: Critical Reviews* 9: 457–483.

Georgopoulos, P.G., S.W. Wang, V.M. Vyas, Q. Sun, J. Burke, R. Vedantham, T. Mccurdy, and H. Ozkaynak. 2005. A source-to-dose assessment of population exposures to fine PM and ozone in Philadelphia, PA, during a summer 1999 episode. *Journal of Exposure Analysis and Environmental Epidemiology* 15: 439–457.

Georgopoulos, P.G., S.W. Wang, I.G. Georgopoulos, M.J. Yonone-Lioy, and P.J. Lioy. 2006. Assessment of human exposure to copper: A case study using the NHEXAS database. *Journal of Exposure Science and Environmental Epidemiology* 16: 397–409.

Georgopoulos, P.G., S. Balakrishnan, A. Roy, S. Isukapalli, A. Sasso, Y.-C. Chien, and C.P. Weisel. 2008a. *A comparison of maximum likelihood estimation methods for inverse problem solutions employing PBPK modeling with biomarker data: Application to tetrachloroethylene*. New Jersey, Rutgers, USA: CCL.

Georgopoulos, P.G., A.F. Sasso, S.S. Isukapalli, P.J. Lioy, D.A. Vallero, M. Okino, and L. Reiter. 2008b. Reconstructing population exposures to environmental chemicals from biomarkers: Challenges and opportunities. *Journal of Exposure Science and Environmental Epidemiology* 19: 149–171.

Georgopoulos, P.G., S.W. Wang, Y.C. Yang, J. Xue, V.G. Zartarian, T. Mccurdy, and H. Ozkaynak. 2008c. Biologically based modeling of multimedia, multipathway, multiroute population exposures to arsenic. *Journal of Exposure Science and Environmental Epidemiology* 18: 462–476.

Georgopoulos, P.G., C.J. Brinkerhoff, S. Isukapalli, M. Dellarco, P.J. Landrigan, and P.J. Lioy. 2014. A tiered framework for risk-relevant characterization and ranking of chemical exposures: Applications to the National Children's Study (NCS). *Risk Analysis* 34: 1299–1316.

Gunier, R.B., M.E. Harnly, P. Reynolds, A. Hertz, and J. Von Behren. 2001. Agricultural pesticide use of California: Pesticide prioritization, use densities, and population distributions for a childhood cancer study. *Environmental Health Perspectives* 109: 1071–1078.

Haddad, S., G. Charest-Tardif, R. Tardif, and K. Krishnan. 2000. Validation of a physiological modeling framework for simulating the toxicokinetics of chemicals in mixtures. *Toxicology and Applied Pharmacology* 167: 199–209.

Hampshire. 2002. *Hampshire Research Institute* [Online]. Available: http://www.hampshire.org/ 2010.

Harper, D. 2004. *Guidelines for good exposure assessment practice for human health effects of chemicals*. Leicester: The Interdepartmental Group on Health Risks from Chemicals.

Hayes, A.W. 2000. *Principles and methods of toxicology*. New York: Informa Healthcare USA, Inc.

Hore, P., V. Zartarian, J. Xue, H. Özkaynak, S.W. Wang, Y.C. Yang, P.L. Chu, L. Sheldon, M. Robson, L. Needham, D. Barr, N. Freeman, P. Georgopoulos, and P.J. Lioy. 2006. Children's residential exposure to chlorpyrifos: Application of CPPAES field measurements of chlorpyrifos and TCPy within MENTOR/SHEDS-pesticides model. *Science of the Total Environment* 366: 525–537.

ICRP. 2002. *Basic anatomical and physiological data for use in radiological protection: Reference values*. In: Valentin, J. (ed.) The International Commission on Radiological Protection. Ottawa, Ontario, Canada.

Isukapalli, S., A. Roy, and P. Georgopoulos. 2000. Efficient sensitivity/uncertainty analysis using the combined stochastic response surface method and automated differentiation: Application to environmental and biological systems. *Risk Analysis* 20: 591–602.

Jongeneelen, F.J., and W.F.T. Berge. 2011. A generic, cross-chemical predictive PBTK model with multiple entry routes running as application in MS excel; design of the model and comparison of predictions with experimental results. *Annals of Occupational Hygiene* 55: 841–864.

Judson, R.S., R.J. Kavlock, R.W. Setzer, E.A. Cohen Hubal, M.T. Martin, T.B. Knudsen, K.A. Houck, R.S. Thomas, B.A. Wetmore, and D.J. Dix. 2011. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. *Chemical Research in Toxicology* 24: 451–462.

Kinerson, R.S. 1987. Modelling the fate and exposure of complex mixtures of chemicals in the aquatic environment. In *Methods for assessning the effects of mixtures of chemicals*, ed. G.C. Butler, V.B. Vouk, A.C. Upton, D.V. Parke, and S.C. Asher. New York: Wiley.

Krauss, M., S. Schaller, S. Borchers, R. Findeisen, J. Lippert, and L. Kuepfer. 2012. Integrating cellular metabolism into a multiscale whole-body model. *PLoS Computational Biology* 8: e1002750.

Lambe, J. 2002. The use of food consumption data in assessments of exposure to food chemicals including the application of probabilistic modelling. *Proceedings of the Nutrition Society* 61: 11–18.

Lijzen, J., and M. Rikken. 2004. *Euses version 2.0*. Bilthoven: RIVM.

Lioy, P.J., D. Vallero, G. Foley, P. Georgopoulos, J. Heiser, T. Watson, M. Reynolds, J. Daloia, S. Tong, and S. Isukapalli. 2007. A personal exposure study employing scripted activities and paths in conjunction with atmospheric releases of perfluorocarbon tracers in Manhattan, New York. *Journal of Exposure Science and Environmental Epidemiology* 17: 409–425.

Mackay, D., S. Paterson, and W.Y. Shiu. 1992. Generic models for evaluating the regional fate of chemicals. *Chemosphere* 24: 695–717.

Mackay, D., E. Webster, I. Cousins, T. Cahill, K. Foster, and T. Gouin. 2001. *An introduction to multimedia models*. Ontario: Canadian Environmental Modelling Centre.

Marshall, J.D., and W.W. Nazaroff. 2007. Intake fraction. In *Exposure analysis*, ed. R.W. Ott, A.C. Steinemann, and L.A. Wallace. Boca Raton: Taylor & Francis Group, LLC.

Marshall, J.D., W.J. Riley, T.E. Mckone, and W.W. Nazaroff. 2003. Intake fraction of primary pollutants: Motor vehicle emissions in the South Coast Air Basin. *Atmospheric Environment* 37: 3455–3468.

Moles, C.G., P. Mendes, and J.R. Banga. 2003. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Research* 13: 2467–2474.

Morris, R.E., G. Yarwood, C.A. Emery, and B. Koo. 2004. Development and application of the CAMx regional one-atmospheric model to treat ozone, particulate matter, visibility, air toxics and mercury. Proceedings of the Air and Waste Management Association's annual meeting and exhibition, 503–516.

Mutshinda, C.M., I. Antai, and R.B. O'hara. 2008. A probabilistic approach to exposure risk assessment. *Stochastic Environmental Research and Risk Assessment* 22: 441–449.

Muzic, R.F., Jr., and B.T. Christian. 2006. Evaluation of objective functions for estimation of kinetic parameters. *Medical Physics* 33: 342–353.

Nieuwenhuijsen, M.J. 2003. Introduction to exposure assessment. In *Exposure assessment in occupational and environmental epidemiology*. Oxford, UK: Oxford University Press.

Nieuwenhuijsen, M.J., D. Donaire-Gonzalez, M. Foraster, D. Martinez, and A. Cisneros. 2014. Using personal sensors to assess the exposome and acute health effects. *International Journal of Environmental Research and Public Health* 11: 7805–7819.

Ozkaynak, H., J. Xue, V.G. Zartarian, G. Glen, and L. Smith. 2011. Modeled estimates of soil and dust ingestion rates for children. *Risk Analysis* 31: 592–608.

Pankow, J.F. 1994. An absorption model of gas/particle partitioning of organic compounds in the atmosphere. *Atmospheric Environment* 28: 185–188.

Payne, M.P., and L.C. Kenny. 2002. Comparison of models for the estimation of biological partition coefficients. *Journal of Toxicology and Environmental Health Part A* 65: 897–931.

Peyret, T., and K. Krishnan. 2011. QSARS for PBPK modelling of environmental contaminants. *SAR and QSAR in Environmental Research* 22: 129–169.

Pistocchi, A., and G. Bidoglio. 2009. Is it presently possible to assess the spatial distribution of agricultural pesticides in Europe? A screening study based on available data. *EFSA PPR stakeholder workshop "Improved realism in soil risk assessment (IRIS) – How will pesticide risk assessment in soil be tackled tomorrow?"*. Ispra, Italy.

Pistocchi, A., D.A. Sarigiannis, and P. Vizcaino. 2010. Spatially explicit multimedia fate models for pollutants in Europe: State of the art and perspectives. *Science of the Total Environment* 408: 3817–3830.

Price, K., and K. Krishnan. 2011. An integrated QSAR-PBPK modelling approach for predicting the inhalation toxicokinetics of mixtures of volatile organic chemicals in the rat. *SAR and QSAR in Environmental Research* 22: 107–128.

Puzyn, T., J. Leszczynski, and M.T.D. Cronin. 2010. *Recent advances in QSAR studies*. New York: Springer Science+Business Media.

Rappaport, S.M., D.K. Barupal, D. Wishart, P. Vineis, and A. Scalbert. 2014. The blood exposome and its role in discovering causes of disease. *Environmental Health Perspectives* 122: 769–774.

Robert, C., and G. Casella. 2004. *Monte Carlo statistical methods*. 2nd ed. New York: Springer.

Roy, A., C.P. Weisel, M. Gallo, and P. Georgopoulos. 1996. Studies of multiroute exposure/dose reconstruction using physiologically based pharmacokinetic models. *Journal of Clean Technology, Environmental Toxicology and Occupational Medicine* 5: 285–295.

Royce, S.G., D. Mukherjee, T. Cai, S.S. Xu, J.A. Alexander, Z. Mi, L. Calderon, G. Mainelis, K. Lee, P.J. Lioy, T.D. Tetley, K.F. Chung, J. Zhang, and P.G. Georgopoulos. 2014. Modeling population exposures to silver nanoparticles present in consumer products. *Journal of Nanoparticle Research* 16: 1–38.

Sabel, C.E., P. Boyle, G. Raab, M. Löytönen, and P. Maasilta. 2009. Modelling individual space-time exposure opportunities: A novel approach to unravelling the genetic or environment disease causation debate. *Spatial and Spatio-Temporal Epidemiology* 1: 85–94.

Sarigiannis, D.A. 2014. *Combined or multiple exposure to health stressors in indoor built environments*. Copenhagen: WHO Regional Office for Europe.

Sarigiannis, D.A., and A. Gotti. 2008. Biology-based dose-response models for health risk assessment of chemical mixtures. *Fresenius Environmental Bulletin* 17: 1439–1451.

———. 2014. New methods for personal monitoring of air pollution through the use of passive sensors during childhood. *Pneumologia Pediatrica* 54: 37–43.

Sarigiannis, D.A., and U. Hansen. 2012. Considering the cumulative risk of mixtures of chemicals – A challenge for policy makers. *Environmental Health: A Global Access Science Source* 11: S18.

Sarigiannis, D., and S. Karakitsios. 2011. Perinatal exposure to bisphenol A: The route of administration makes the dose. *Epidemiology* 22: S172.

Sarigiannis, D.A., and S.P. Karakitsios. 2012. A dynamic physiology based pharmacokinetic model for assessing lifelong internal dose. AIChe 2012, October 28 – November 2 2012. Pittsburgh, PA.

Sarigiannis, D., S. Karakitsios, and A. Gotti. 2009. Mechanistic full chain approach for ETS carcinogenicity impact assessment in the EU. *Epidemiology* 20: S88.

Sarigiannis, D., A. Gotti, and S. Karakitsios. 2011. A computational framework for aggregate and cumulative exposure assessment. *Epidemiology* 22: S96–S97.

Sarigiannis, D., A. Gotti, S. Karakitsios, P. Kontoroupis, and S. Nikolaki. 2012a. Intera platform: A tool for mechanistic risk assessment of indoor air pollutants. *Epidemiology* 23: 639.

Sarigiannis, D.A., S.P. Karakitsios, M.P. Antonakopoulou, and A. Gotti. 2012b. Exposure analysis of accidental release of mercury from compact fluorescent lamps (CFLs). *Science of the Total Environment* 435–436: 306–315.

Sarigiannis, D., P. Kontoroupis, E. Solomou, S. Nikolaki, and A. Karabelas. 2013. Inventory of pesticide emissions into the air in Europe. *Atmospheric Environment* 75: 6–14.

Sarigiannis, D., S. Karakitsios, A. Gotti, G. Loizou, J. Cherrie, R. Smolders, K. De Brouwere, K. Galea, K. Jones, E. Handakas, K. Papadaki, and A. Sleeuwenhoek. 2014a. Integra: From global scale contamination to tissue dose. Proceedings – 7th international congress on

environmental modelling and software: Bold visions for environmental modeling, iEMSs 2014, 1001–1008.

Sarigiannis, D.A., S.P. Karakitsios, M. Kermenidou, S. Nikolaki, D. Zikopoulos, S. Semelidis, A. Papagiannakis, and R. Tzimou. 2014b. Total exposure to airborne particulate matter in cities: The effect of biomass combustion. *Science of the Total Environment* 493: 795–805.

Sasso, A.F., S.S. Isukapalli, and P.G. Georgopoulos. 2010. A generalized physiologically-based toxicokinetic modeling system for chemical mixtures containing metals. *Theoretical Biology and Medical Modelling* 7: 17.

Schmutz, J., J. Wheeler, J. Grimwood, M. Dickson, J. Yang, C. Caoile, E. Bajorek, S. Black, Y.M. Chan, M. Denys, J. Escobar, D. Flowers, D. Fotopulos, C. Garcia, M. Gomez, E. Gonzales, L. Haydu, F. Lopez, L. Ramirez, J. Retterer, A. Rodriguez, S. Rogers, A. Salazar, M. Tsai, and R.M. Myers. 2004. Quality assessment of the human genome sequence. *Nature* 429: 365–368.

Snyder, E.G., T.H. Watkins, P.A. Solomon, E.D. Thoma, R.W. Williams, G.S. Hagler, D. Shelow, D.A. Hindin, V.J. Kilaru, and P.W. Preuss. 2013. The changing paradigm of air pollution monitoring. *Environmental Science & Technology* 47: 11369–11377.

Stanek, E.J., 3rd, and E.J. Calabrese. 2000. Daily soil ingestion estimates for children at a superfund site. *Risk Analysis* 20: 627–635.

Stanek, E.J., and E.J. Calabrese. 1995. Daily estimates of soil ingestion in children. *Environmental Health Perspectives* 103: 276–285.

Tan, Y.M., K. Liao, R. Conolly, B. Blount, A. Mason, and H. Clewell. 2006. Use of a physiologically based pharmacokinetic model to identify exposures consistent with human biomonitoring data for chloroform. *Journal of Toxicology and Environmental Health – Part A: Current Issues* 69: 1727–1756.

Teske, M.E., S.L. Bird, D.M. Esterly, T.B. Curbishley, S.L. Ray, and S.G. Perry. 2002. AgDRIFT: A model for estimating near-field spray drift from aerial applications. *Environmental Toxicology and Chemistry* 21: 659–671.

Tulve, N.S., J.C. Suggs, T. Mccurdy, E.A. Cohen Hubal, and J. Moya. 2002. Frequency of mouthing behavior in young children. *Journal of Exposure Analysis and Environmental Epidemiology* 12: 259–264.

USEPA. 2013. *EPA – EXPO – BOX: A toolbox for exposure and risk assessors*. Washington: U.S. Environmental Protection Agency.

Valcke, M., and K. Krishnan. 2011. Evaluation of the impact of the exposure route on the human kinetic adjustment factor. *Regulatory Toxicology and Pharmacology* 59: 258–269.

Van Veen, M.P., F. Fortezza, H.J.T. Bloemen, and J.J. Kliest. 1999. Indoor air exposure to volatile compounds emitted by paints: Experiment and model. *Journal of Exposure Analysis and Environmental Epidemiology* 9: 569–574.

Vermeire, T.G., P. Van Der Poel, R.T.H. Van De Laar, and H. Roelfzema. 1993. Estimation of consumer exposure to chemicals: Application of simple models. *Science of the Total Environment* 136: 155–176.

Vermeire, T.G., D.T. Jager, B. Bussian, J. Devillers, K. Den Haan, B. Hansen, I. Lundberg, H. Niessen, S. Robertson, H. Tyle, and P.T.J. Van Der Zandt. 1997. European Union System for the Evaluation of Substances (EUSES). Principles and structure. *Chemosphere* 34: 1823–1836.

Verner, M.A., M. Charbonneau, L. Lopez-Carrillo, and S. Haddad. 2008. Physiologically based pharmacokinetic modeling of persistent organic pollutants for lifetime exposure assessment: A new tool in breast cancer epidemiologic studies. *Environmental Health Perspectives* 116: 886–892.

Weschler, C.J., and W.W. Nazaroff. 2008. Semivolatile organic compounds in indoor environments. *Atmospheric Environment* 42: 9018–9040.

———. 2010. SVOC partitioning between the gas phase and settled dust indoors. *Atmospheric Environment* 44: 3609–3620.

Wild, C.P. 2005. Complementing the genome with an "Exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology Biomarkers and Prevention* 14: 1847–1850.

Willmann, S., J. Lippert, M. Sevestre, J. Solodenko, F. Fois, and W. Schmitt. 2003. Pk-Sim®: A physiologically based pharmacokinetic 'whole-body' model. *Drug Discovery Today: BIOSILICO* 1: 121–124.

Wilschut, A., W.F. Ten Berge, P.J. Robinson, and T.E. Mckone. 1995. Estimating skin permeation. The validation of five mathematical skin permeation models. *Chemosphere* 30: 1275–1296.

Wilson, R., A.V. Hill, and H. Glazer. 2013. *Tools and tactics for operations managers (collection)*. Upper Saddle River, New Jersey, Pearson Education.

Wormuth, M., M. Scheringer, M. Vollenweider, and K. Hungerbühler. 2006. What are the sources of exposure to eight frequently used phthalic acid esters in Europeans? *Risk Analysis* 26: 803–824.

Yang, Y., X. Xu, and P.G. Georgopoulos. 2010. A Bayesian population PBPK model for multiroute chloroform exposure. *Journal of Exposure Science and Environmental Epidemiology* 20: 326–341.

Zartarian, V.G., J. Xue, H. Özkaynak, G. Glen, C. Stallings, L. Smith, W. Dang, N. Cook, D. Aviado, S. Mostaghimi, and J. Chen. 2002. SHEDS-wood stochastic human exposure and dose simulation model for a wood preservative exposure scenario, technical manual: Using SHEDS-wood for the assessment of children's exposure and dose from treated wood preservatives on playsets and residential decks. Washington: U.S. EPA.

Zhang, H. 2004. A new nonlinear equation for the tissue/blood partition coefficients of neutral compounds. *Journal of Pharmaceutical Sciences* 93: 1595–1604.

Zidek, J.V., G. Shaddick, R. White, J. Meloche, and C. Chatfield. 2005. Using a probabilistic model (pCNEM) to estimate personal exposure to air pollution. *Environmetrics* 16: 481–493.

Part II
Prioritizing Mixtures for Study

Chapter 5
Introduction to Environment and Exposome-Wide Association Studies: A Data-Driven Method to Identify Multiple Environmental Factors Associated with Phenotypes in Human Populations

Chirag J. Patel

Abstract It is a priority to identify multiple environmental factors, or mixtures, associated with disease phenotypes in human populations. However, high-throughput computational methods to identify mixtures that are important in human disease are lacking. This chapter describes the "environment-wide association study" (EWAS) analytic approach to identify a number of environmental exposures in human disease. With the advent of high-throughput environmental exposure information (e.g., exposome), methods such as EWAS will be instrumental to accelerate discovery in disease.

Keywords Environment-wide association study · Genome-wide association study · High-throughput · Biostatistics · Bioinformatics

5.1 Introduction

The phenomena of environmental exposure are complex, and humans are exposed to not a handful but many heterogeneous exposures simultaneously (Patel and Ioannidis 2014a, b). As discussed throughout this book, it is a priority to identify combinations of factors, or *mixtures*, associated with disease in human populations. However, most epidemiological investigation studies to date consider one or a few exposures at a time, and we currently lack data-driven methods to associate, and discover, numerous environmental exposures with phenotype and disease.

This is of concern as complex diseases are multifactorial, and it is hypothesized that diseases arise due to the contribution of multiple interacting genetic and environmental factors (Schwartz and Collins 2007). For example, the genome-

C. J. Patel (✉)
Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
e-mail: chirag_patel@hms.harvard.edu

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_5

wide association study (GWAS) is a commoditized, standardized, and popular framework used by researchers to evaluate genetic factors in disease or phenotype along the entirety of the genome (Burdett et al. 2015; Salonen et al. 2007; Saxena et al. 2007; Sladek et al. 2007). Due to wide accessibility of genome-scale GWAS assays, which can ascertain millions of genetic variants simultaneously, human geneticists have now moved from studying a handful of genetic variants at a time to a more data-driven, comprehensive, systematic, and agnostic reporting of robust genetic associations and their replication in independent populations. As a result, over 2000 GWAS have been published, often over 20 for specific diseases such as type 2 diabetes (T2D) (Burdett et al. 2015) in sample sizes now reaching in the hundreds of thousands. GWAS has strengthened the epidemiological process and methodology of screening and validating genetic variants. In this chapter, analogous methods are described for environmental exposure to enable data-driven discovery of multiple environmental exposures that reflect a mixture of individual factors associated with disease.

Specifically, in this chapter, an analogous framework to GWAS is proposed, called "environment-wide association study" (EWAS), to search for and analytically validate environmental factors associated with continuous phenotypes or discrete ones such as disease. This type of question is different from a hypothesis-driven approach in which a single candidate or a handful of environmental factors are chosen a priori and tested individually for their association to a phenotype and analogous to questions facilitated by GWAS.

But first, what accelerated genomic discovery through GWAS? The primary driver for GWAS was accessibility of decreased-cost, high-throughput, and unified genome-wide assays, giving human geneticists and genetic epidemiologists (Wild 2005, 2012; Wild et al. 2013) a more comprehensive way to search for genetic factors in disease. Currently, this is elusive in environmental epidemiological studies. To meet this challenge, scientists have called for efforts to elucidate and measure the *exposome* (Chap. 3), the environmental exposure analog to the genome, whereby a higher-throughput battery of environmental exposures is ascertained in humans including infection, pollutants, and nutrients simultaneously throughout the life course (Rappaport 2012; Rappaport et al. 2014; Rappaport and Smith 2010; Buck Louis and Sundaram 2012; Miller and Jones 2014; Patel and Ioannidis 2014b).

As such, there are few epidemiological cohorts that ascertain the *exposome* to facilitate the identification of mixtures associated with disease. However, at least two studies, the National Health and Nutrition Examination Survey (NHANES) (Centers for Disease Control and Prevention (CDC) 2013, http://www.cdc.gov/nchs/nhanes.htm and http://nhanes.hms.harvard.edu) and DEMOCOPHES (http://www.eu-hbm.info/democophes), provide opportunities for exposome and mixtures research. The NHANES is a cross-sectional survey representative of the United States. It is comprised of both health questionnaire and laboratory and clinical data using a multistage probability sampling design, and the Centers for Disease Control and Prevention – the administrators of the survey – collected information through in-person interviews, physical measurement at mobile examination centers, and human samples (Chap. 2). To date, the NHANES is a gold standard for

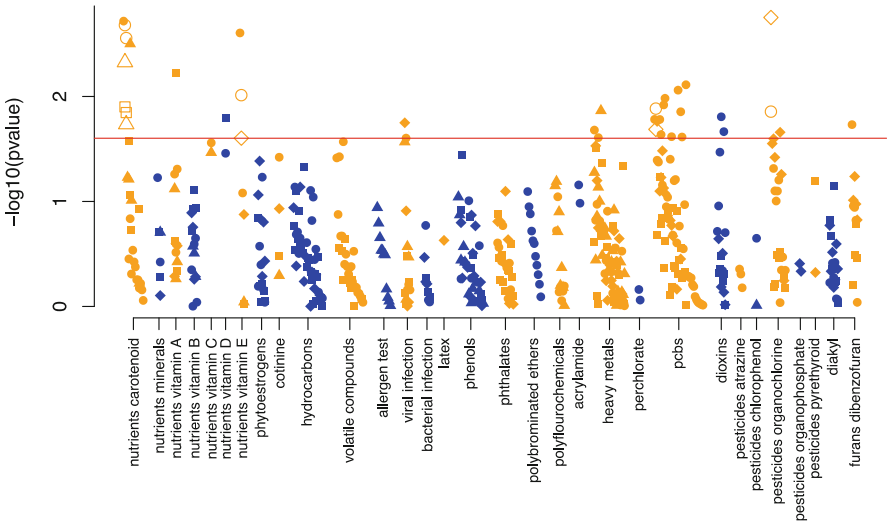

Fig. 5.1 A "Manhattan plot" visualization of an EWAS for type 2 Diabetes (T2D). Briefly, NHANES was utilized to search for environmental exposure factors associated with T2D. Individuals classified as T2D had a fasting blood glucose greater or equal to 126 mg/dL, and controls were individuals lower than this threshold. Using methods described in this chapter, each of 266 exposures in four independent surveys of NHANES (depicted as diamonds [1999–2000 survey], squares [2001–2002], circles [2003–2004], and triangles [2005–2006]) was associated. The y-axis depicts the *p*-value of significance of the correlation; the x-axis arranges each of the exposures tested in categories of exposure, such as nutrients, PCBs, and heavy metals. The red line depicts a false discovery rate (FDR) threshold of 10%, and open symbols are ones that are found in more than one survey and are replicated (Reproduced from Patel et al. 2010)

ascertainment of quantitative exposome measurements in human tissue, consisting of over 300 individual biomarkers of environmental exposure (see http://www.cdc.gov/nchs/data/nhanes/survey_content_99_14.pdf). For a subset of the population, longer-term follow-up information such as cause of death and utilization of Medicare is also available to the public.

In fact, NHANES has been utilized to conduct EWAS for multiple phenotypes, including to identify in a data-driven manner multiple exposures associated with type 2 diabetes (Patel et al. 2010), blood pressure (Tzoulaki et al. 2012), serum lipid levels (Patel et al. 2012), all-cause mortality (Patel et al. 2013a), telomere length (Patel et al. 2016) mothers with preterm birth (Patel et al. 2013b), and even correlates of income (Patel et al. 2014). See Fig. 5.1 for examples. These analyses provide illustrations of the potential for identifying mixtures in populations, and readers are encouraged to read them for broader details. By the time that this chapter is published, new technologies and cohorts will likely have emerged that consider the exposome and its association to disease as it is an obvious venue to ascertain mixtures in humans. One promising avenue includes the *Children's Health Analysis Resource* (or CHEAR: http://www.niehs.nih.gov/research/supported/dert/programs/chear/), a NIH/NIEHS funded program to develop new technologies to measure

children's exposomes and computational methods and standards to associate the exposome to critical health outcomes, such as development and growth.

The description of EWAS begins by briefly introducing the genome-wide analog, GWAS. Second, the EWAS framework and the current EWAS methodology are described. Last, we discuss our results and posit ways to extend the EWAS methodology.

5.2 Methods Background

5.2.1 Genome-Wide Association to Disease

With the sequencing of the genome and projects that characterized common genetic variation such as the HapMap, investigators are now able to interrogate how genome-wide genetic differences are associated with disease and disease-related phenotypes on an epidemiological scale (Hardy and Singleton 2009; International HapMap Consortium 2005). These revolutionary studies, known as GWAS, have enabled investigators to ask what common genetic loci are associated with a particular phenotype in an agnostic, systematic, and comprehensive way with explicit control of multiple test correction to mitigate possibilities of false positive reporting.

Specifically, during the HapMap project, common single nucleotide (SNP) variants were catalogued on the basis of their population frequency ($\geq$ 10% population frequency) and major and minor allele versions (Manolio et al. 2008). The location of each SNP along the genome is referred to as a "locus," and the presence of variation at a particular locus denotes a "polymorphism" or a "polymorphic" locus. "Common" polymorphisms are those that occur in approximately greater than 5–10% in the population. Thus, by definition, a "common" SNP must reside at a polymorphic locus. There are greater than 1 million common SNPs in the genome (International HapMap Consortium 2005). While SNPs are the most common type of polymorphism in the genome accounting for 90% of genetic variation, many other types of genetic variation exist, such as copy number variants, insertions, and deletions.

GWAS relates traits to variation at each – or a large subset of – common polymorphic locus in the genome and is enabled by genomic technologies, known as "SNP microarrays," which can assay greater than 1 million loci simultaneously for an individual. These microarrays are now mere commodity items, like computers, making accessible genome-wide measurements on a large number of individuals (Wetterstrand 2011). Further, these technology platforms are known to have very low measurement error (Ioannidis et al. 2009).

GWAS is constructed by recruiting thousands of individuals with ("cases") and without ("controls") a trait or disease. Genotype frequencies at each locus across the genome are then compared between cases and controls using common statistical tests such as a chi-squared test (Pearson and Manolio 2008), assuming independence

between each locus. Loci can be correlated with each other, a phenomenon known as "linkage disequilibrium." Loci that are in "linkage disequilibrium" occur more frequently together than would be expected if they were independent (discussed in detail in Sect. 2.2). When multiple loci are correlated, their associations with the disease or trait may be shared, and association tests are updated by "adjusting" the statistical models with another locus in linkage. Genetic loci can also be associated with continuous traits, such as levels of a biomarker (e.g., blood pressure or serum glucose), by modeling the continuous trait in a linear regression model (Frayling et al. 2007). This linkage is expanded upon below.

Multiple comparisons are accounted for through conservative Bonferroni adjustment, and significant loci are validated in independent populations. As will be described in more detail below, corrections to significance thresholds, such as the Bonferroni correction, are required to guard against spurious findings when conducting multiple tests of association. For example, imagine conducting an association study between 100 pollutants and blood pressure. Also imagine that an oracle has told us the true scenario that none of the pollutants are truly associated with blood pressure (and therefore no correlations should emerge). At a *p*-value threshold of 0.05 under the distribution of no correlation between the pollutant factors and blood pressure, five pollutants will emerge as significant! These are known as false positives and may emerge when conducting more than one test of association. Go ahead and test this out for yourself in your favorite statistics software by generating 100 random "pollutant" factors, and correlate those each with a random "phenotype," and count how many times a significant correlation occurs at *p*-values 0.1, 0.05, and 0.01. How many associations are significant due to chance?

Preceding GWAS were "candidate gene studies," a hypothesis-driven study to correlate a handful of genetic variants to a trait of interest using a "smaller" sample size. As a consequence of lack of power and prohibitive genotyping cost, the agnostic, comprehensive, and systematic analytical and validation procedure of GWAS eluded traditional genetic association studies (Goldstein 2009; Manolio et al. 2009; Mccarthy et al. 2008; NCI-NHGRI Working Group on Replication in Association Studies 2007; Ioannidis et al. 2001). To facilitate discussion regarding "environment-wide association," we describe these "agnostic," "systematic," or "comprehensive" characteristics of GWAS. The agnostic or comprehensive characteristic means all possible hypotheses (all environmental factors) are tested and the investigators do not "cherry-pick" what factors to correlate. Systematic refers to the similar treatment of each association test: each association is modeled in the same way, using the same model, population samples, and/or adjustment variables. How this is achieved in EWAS is described below.

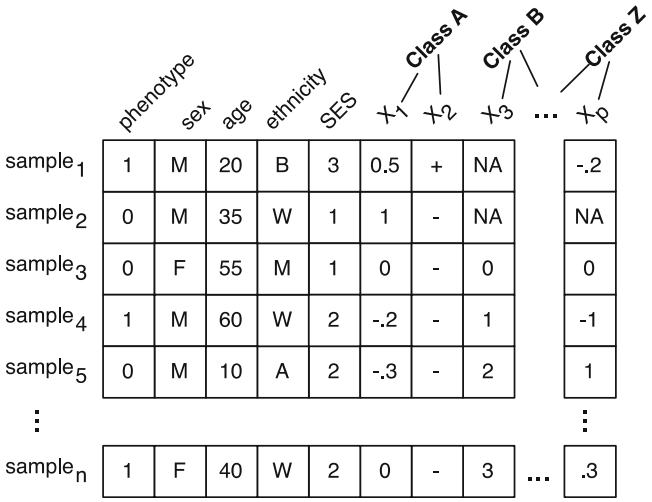

Fig. 5.2 Sample data structure for EWAS. "Phenotype" is the dependent variable. "Sex," "age," "ethnicity," and "SES" (socioeconomic status) are examples of adjustment variables. X_1 through X_p are environmental factors; sample$_1$. . .sample$_n$ are the individuals that make up the population. Values inside each cell denote an example of the data type for the variable. For example, "phenotype" here is a binary variable taking on 1 if the phenotype is present, 0 if absent; "sex" is a categorical variable for males and females. *X* variables representing environmental factors may be continuous (e.g., X_1, X_p), positive/negative (e.g., X_2), or ordinal (X_3). Data might be missing (e.g., NA cells). The vertical axis denotes individuals in the sample. Each environmental factor belongs to a "class," or grouping, that represents a common characteristic of those factors, represented in the figure as "Class A," "Class B," and "Class Z"

5.2.2 *Environment-Wide Association to Disease*

In the following, a study design analogous to GWAS is proposed called "environment-wide association study" (EWAS) to search for and analytically validate environmental factors associated with complex diseases and phenotypes.

EWAS assumes a similar "data structure" to that of GWAS. Recall that in GWAS, multiple genetic factors are assayed along with phenotypic information on each individual (Fig. 5.2). In other words, the genetic factors are the independent variables, and the phenotype is the dependent variable. In EWAS, the genome domain is substituted with the environmental domain or *exposome* domain (Rappaport et al. 2014; Rappaport and Smith 2010).

Specifically, the quantity or presence of environmental factors is directly measured on each individual, such as the amount of a chemical in bodily tissue, or a proxy measure, such as self-report historical exposure. It is important to note that the environment is a dynamic entity, unlike the data structure of GWAS. Thus, the dimension of time may also be added to the structure of EWAS data, framing it in a longitudinal context.

GWAS variables are "binned" by their chromosomal location, facilitating the description of their correlation structure – known as linkage disequilibrium (LD) – when visualizing associations. Specifically, LD is the correlation of two or more loci in the genome. Further, LD is a function of relative location of the two loci; that is, the closer together two loci are on a chromosome in general, the higher their LD. Suppose one locus is being considered: in this scenario, individuals inherit alleles from their parents, one from the mother and one from the father. The genotype at one locus is a random event and is dependent on the frequency of alleles present at that one locus in the mother and father. Now suppose two loci (two sets of genotypes) are in "LD." This means that their pattern of inheritance is correlated; that is the occurrence of a particular allele "A" at a locus *A* and "B" at a locus *B* is *nonrandom* or dependent with respect to one another. In other words, the presence of one allele can predict the presence of another. LD among different populations has been characterized by the HapMap project and is ongoing with the 1000 Genomes Project (International HapMap Consortium 2005). In GWAS, LD structure is important. First, since only a prevalent subset of polymorphic loci is being assayed, LD allows narrowing down of what variants might be causal. For example, given an association signal for a variant, the *causal* variant might be one in strong LD with it. LD also provides an internal gauge of validity; for example, given a strong association signal of a variant at loci *X*, one would expect measured common variants that are also in LD with *X* to harbor some signal.

At present, LD in EWAS is qualitative not quantitative as in GWAS. In application of EWAS, factors are binned according to categories that described the compound "class," had shared environmental health "relevance," or described some other arbitrary shared characteristic as a group of factors. A nonarbitrary shared characteristic would be co-occurrence or correlation (environmental exposure factors that are correlated may co-occur with one another). A research effort will be to fully characterize the LD of the exposome including their correlation/covariance structure and population-wide prevalence as has been done with the HapMap.

EWAS achieves the agnostic, systematic, and comprehensive qualities that characterize GWAS. First, instead of testing a few environmental associations at a time, EWAS evaluates multiple environmental factors agnostically. EWAS is comprehensive in that each factor measured is assessed for possible association with the target phenotype. Next, associations are systematically adjusted for multiplicity of comparisons. Further, EWAS calls for validation of significant associations in an independent population.

The EWAS framework calls for systematic and comprehensive sensitivity analyses of highly significant or validated factors. Specifically, all possible measured confounders are included in final models, and their effect on the estimate of the environmental factor is assessed. Last, given the dense web of correlation for nongenetic measures, such as between environmental factors and clinical measures, the correlation structure between validated environmental factors and risk factors is systematically computed and visualized to understand the degree of their interdependence called "exposome globes" (Patel and Ioannidis 2014b; Patel and Manrai 2015). By visualizing relationships in this way, mixtures of nonindependent

exposures associated with phenotype can be inferred, similar to "relevance network" or clustering analyses (Butte and Kohane 2000; Butte et al. 2000).

5.2.3 *Conceptual Challenges in Data-Driven Studies of Environmental Exposures in Disease*

Challenges and biases related to observational studies influence all association studies, be it from hypothesis-driven candidate factor study, GWAS, or EWAS. In contrast to "gold-standard" randomized trial study data, environmental epidemiological studies, including EWAS, rely on observational study data, such as longitudinal cohort, case-control, or cross-sectional data. These types of epidemiological studies are subject to confounding biases that hinder causal inference and are avoided, to some degree, in randomized studies (Greenland 1990); however, the gold standard scenario of a clinical trial is not suited for agnostic study of the exposome as it is impossible to randomize such a matrix of factors.

"Confounding" is used to describe a scenario in which a variable is correlated with both the factor of interest (the independent variable) and phenotype (dependent variable) (Greenland and Morgenstern 2001); in EWAS analyses, the factor acts as a "proxy" to the confounding variable, resulting in a false association between the dependent and independent variable. A partial solution to this type of bias is including the confounder as a covariate in the statistical model or "controlling" for the confounder. This, of course, is only possible when the confounder is known and measured.

Famous examples include associations derived from observational studies later contradicted by randomized control trials (RCT): (1) β-carotene, thought to have muted the risk for smoking-induced cancer (Peto et al. 1981), only to be refuted by a RCT later (Omenn et al. 1996), (2) vitamin E and decreased risk of coronary heart disease (CHD) (Hooper et al. 2001), and, notably, (3) vitamin C and CHD, where relative risks from observational studies indicated a protective effect of vitamin C, whereas vitamin C was found to increase relative risk of CHD in a large RCT (Davey Smith and Ebrahim 2003)!

Another source of "bias" includes "reverse causality" or reverse association. Reverse causality leads to the failure to infer proper "forward" direction between the independent variable (e.g., environmental factor) and dependent variable (phenotype). Specifically, it occurs when the independent variable comes directly or indirectly as a result of the dependent variable. An example of this is a sample-wide behavioral shift due to the dependent variable, such as increased intake of a vitamin due to an adverse phenotype. If we were to associate the environmental factor, the vitamin, with the phenotype as the dependent variable, the interpretation of the model would suggest that a change in vitamin exposure leads to a change in phenotype when in fact the opposite is true. These biases are especially prevalent in case-control or cross-sectional studies in which individuals are measured at one

point in time. A way to take into account the dynamic nature of nongenetic variables and biases such as reverse causality includes conducting a longitudinal study in which we may observe jointly changes in phenotype and exposure pattern as a function of time (Rothman et al. 2008).

The nature of the environmental factors themselves also biases results. First, the assessment of the quantity of environmental factors in blood and serum is subject to measurement error (Ioannidis et al. 2009), and self-report variables are subject to recall bias. Further, physiological characteristics of factors themselves influence estimates, including the variability of the kinetics of chemical factors, such as how long they are retained in accessible body tissue. For example, chemical compounds that are easily measured include those that are lipophilic and persistent in fatty tissue. As adiposity is related to both the measurement of the factor and often the phenotype of interest (e.g., metabolic syndrome), a positive correlation might indicate confounding. On the other hand, many types of factors are excreted quickly, also affecting their measurement and association to the phenotype of interest; however, "steady-state" or constant exposure might allay a kinetic effect of environmental chemicals (Bartell et al. 2004).

Importantly, current epidemiological investigations may be fraught with issues in multiplicity, and these issues may be exacerbated by the dense correlational web of environmental exposure (Patel and Ioannidis 2014a). Multiplicity signifies the space of hypothesis tests that can occur: for example, given a database of 300 exposures, there are 300 potential hypothesis tests for a disease phenotype. Testing multiple variables for associations with other exposures and outcomes makes possible the prospect of making discoveries, but the cost of multiplicity can lead to type I error or false positives. Multiplicity needs to be taken into consideration when identifying multiple exposures associated with disease, and adjustment for multiple hypotheses is recommended to "pay the cost" for searching for numerous exposures. One way to adjust includes the simple but constraining, Bonferroni correction, and yet another includes the false discovery rate (FDR) approach. These topics are described at length below.

5.3 EWAS Method

The EWAS methodology and analysis framework are analogous to that utilized in GWAS. First, an initial scan is conducted for environmental factors associated with a phenotype of interest through general linear modeling, such as logistic or linear regression (see example below). Since environmental association occurs in the observational (vs. randomized scenario), these models include variables that adjust for known confounders, such as clinical risk factors. Second, multiple hypotheses are accounted for by estimating the false discovery rate (FDR). Third, factors that are deemed significantly associated with the phenotype beyond the region of false discovery are "validated" in independent cohorts. Factors that are validated are

considered true discoveries. R software to conduct EWAS can be found here: https://github.com/chiragjp/xwas.

The EWAS framework also calls for systematic sensitivity analyses, whereby validated factors are modeled under different assumptions or with additional covariates. Further, the pair-wise correlation between each validated factor is computed and examined to determine their dependence, which can be interpreted as co-occurrence of a pair of factors or indicative of confounding. Each step is described further below.

5.3.1 Stage 1: Linear Modeling

Each environmental factor is associated with a phenotype of interest using general linear models; for example, each is associated with disease status (i.e., case or control) using logistic regression. Normally distributed continuous phenotypes are correlated to environmental factors with linear regression. Common risk, demographic, and clinical factors are added as adjusting variables, such as age, sex, ethnicity, and socioeconomic status, as phenotypic states and environmental factors are confounded by these variables. Thus, for an environmental factor X_i in the list of measured factors $X_i \ldots X_{p,}$ the disease state (Y) is modeled as a linear function of environmental factors and adjustment variables (represented by Z):

$$Y = \alpha + \beta_i X_i + \zeta Z \tag{5.1}$$

X_i corresponds to the environmental factor, and β_i corresponds to the effect size of that factor, adjusted by other variables.

The strength of association is computed by the two-sided p-value for β_i, which tests the "null hypothesis" that β_i is equal to zero. When modeling the phenotype as the logit (logistic regression), the exponentiation of β_i serves as the odds ratio or the change in the odds for disease versus un-diseased status for a unit change of the factor. In the linear regression setting, β_i can be interpreted as the change in phenotype per unit change of the factor. p-values are computed through common tests of significance, such as Wald tests.

Continuous factors are z-transformed (centered about the mean and divided by their standard deviation) in order to compare the effect sizes. A z-transformation allows one to compare across exposures as the new units are in terms of standard deviations of exposure factors. Many factors measured in tissue have a right skew and thus are log transformed prior to z-transformation. Binary factors (such as presence or absence of a factor) are standardized such that effect size reflects a unit change between exposed and unexposed status; that is, the referent is consistently the "negative" result of a binary test. Ordinal factors are left untransformed.

5.3.2 *Stage 2: Controlling for Multiple Hypotheses by Estimating the False Discovery Rate*

Given a set of "discoveries," or a list of potentially significant factors, how can those that are false discoveries be determined? In the GWAS setting, Bonferroni correction is utilized to adjust for multiple comparisons. The Bonferroni adjustment is straightforward: it simply divides the significance threshold α for the total number of tests conducted. This adjustment guarantees the "family-wide error rate" – the probability of having one or more false positive(s) in a set of results is equivalent to a setting in which only one hypothesis was tested at level α. However, the threshold is conservative, and therefore power for detection is lost.

To account for multiple comparisons in EWAS, an empirical estimate is computed of the false discovery rate (FDR) derived through permutations of the phenotype multiple times, effectively creating a "null distribution" of test statistics. In contrast to the Bonferroni correction, the FDR provides a quantitative estimate of the number of false positives in a set of "discoveries." The FDR is less conservative and therefore more powerful than the Bonferroni correction (Noble 2009). Further, since the estimate of the FDR utilizes the data itself, it inherently considers the covariance structure of the data, an important quality given the dense correlation of nongenetic factors (Noble 2009).

The FDR is the estimated proportion of false discoveries made versus the number of real discoveries made for a given significance level α, to control for multiple hypothesis testing. To estimate the number of false discoveries, a "null distribution" of regression test statistics is created by shuffling the phenotype a large number of times (100–1000) and refitting the regression models. The FDR is the ratio of the proportion of results that were called significant at a given level α in the null distribution and the proportion of results called significant from the real tests. A significance level is used that corresponds to a FDR of 5–10% to select associations.

The example algorithm to compute the FDR follows:

```
1. Do: Stage 1, Linear Modeling.
2. nullPvalues <- NewList()
3. For i in [1...numberPermutations]:
4 randomPheno <- permutePhenotypeWithoutReplacement(phenotype)
5. For xi in [X1...Xp]:
6. Modi <-GeneralLinearModel(randomPheno,xi,Xses,Xeth,Xsex,
      Xage)
7. ListAppend(nullPvalues, getPvalue(Modeli, xi))
8. fdrRaw <- []
9. for pvalue in Pvalues:
10. numerator <- sum(nullPvalues < pvalue)/numberPermutations
11. denominator <- sum(Pvalues < pvalue)
12. listAppend(fdrRaw, numerator/denominator)
```

```
13. fdrs <- []
14. for I in [1...p]:
15. fdr <- min(rawFdr[i...p])
16. ListAppend(fdrs, fdr)
```

Algorithm 1 Computing the FDR (q-value) for each p-value during Stage 1 of EWAS.

To begin algorithm 1, the p-values for association for each factor need to be established. Then, for a number of permutations, the regression model is refit for the random phenotype for each environmental factor, and all of these "null" p-values (line 3–7) are collected. For each p-value computed, the raw FDR, or the ratio of the raw number of results that exceeds that p-value threshold in the permuted data and the number of results that exceeds that p-value in stage 1 (line 11, 12, and 13), is calculated. As FDR should be a monotonically increasing function of the p-value, the researcher ensures that the FDR for a p-value is the minimum of the FDRs for all p-values equal to or greater than that p-value (line 15). The resulting array of FDR values corresponds to the FDR for each p-value computed in Stage 1.

Of course, the original method for estimating the FDR can be used (Benjamini and Hochberg 1995), eliminating the need for algorithm 1. However, as discussed earlier, estimating the FDR through permutations of the dependent variable is preferred in the scenario in which the variables are correlated. In addition, much has been documented about what variables to permute or bootstrap. For example, it has been suggested that model residuals, the difference between the predicted and true values, should be permuted (or bootstrapped) as opposed to the original outcome variables (replacing line 4 in algorithm 1 (Efron 2010)). In our experience, similar estimates of the FDR were obtained under different documented methods of permuting. The reader is advised to refer to Manly, Efron, and Westfall and Young for more in this area (Manly 2007; Efron 2010; Westfall and Young 1993).

5.3.3 Stage 3: Validation

Findings deemed significant corresponding to some nominal FDR level are validated or replicated in one or more additional independent cohorts with a nominal p-value (e.g., $p = 0.05$ in the independent replication cohort). Importantly, the direction of the effect size in the validation cohort must be equivalent to that in the initial screen.

5.3.4 Stage 4: Sensitivity Analyses

Confounding and reverse causality influences the strength of association, biases the effect size estimate, and in general affects causal inference of environmental factors

to phenotypes. Thus, a method is proposed to begin to approximate these biases. However, it cannot be claimed that these biases will be found nor that confounding will be eliminated; nevertheless, methods are described to assess bias given that they were measured.

In the first, all measured variables are systematically reviewed that were not considered in the list of environmental factors – but could influence the association – and sequentially added to the linear model as an additional covariate. Then the *p*-value of association and the effect size corresponding to the environmental factor calculated from the extended model are compared to the original model computed in Stage 2. The difference between the extended and original factor coefficients quantifies the contribution of the bias due to the new variable.

Types of variables that might bias the associations depend on the phenotype and environmental factors under study but often include knowledge of clinical status (e.g., diagnosis of a disease), recent food, supplement or drug intake, and physical activity. For example, knowledge regarding one's disease state might induce behavioral change, resulting in increased exposure to foods high in vitamins and certain nutrients; association between these vitamin factors and disease might then be attributed to reverse causality. Or, the use of a drug might induce phenotypic change, biasing estimated effects toward the null. This method is dependent on a multitude of measured potential confounders. Large epidemiological datasets arising from the public domain or of large consortia often measure many of these other clinical and behavioral nongenetic variables which can be utilized to test the "sensitivity" of the final validated effects of environmental factors associated with a phenotype.

5.3.5 Stage 5: Correlation Globes

The correlation/covariance structure between nongenetic measures is known to be "dense" (many factors are correlated with one another), and this structure also influences the ability to infer the independent effect of factors on phenotype as discovered in EWAS. Furthermore, the initial screen methodology assumes independence between factors and therefore provides little information about their correlation.

Concretely, given a list of discovered factors, their joint association to the phenotype of interest might be due to their co-occurrence, such as similar routes of exposure. The degree of dependency between validated factors is assessed by computing their raw correlation coefficient (Spearman's ρ) and visualizing this with a correlation "globe" (Patel and Manrai 2015). Briefly, correlation between exposures allows analysts to describe how certain exposures can lead to other exposures. For example, many nutrients are consumed together. A nonoptimal diet (however it may be defined) may lead to a deficiency in a whole group of vitamins and nutrients. As another example, individuals who are exposed to air pollution may also be exposed to or have high body burdens of hydrocarbons, volatile compounds, and heavy metals.

Many methods have been proposed to describe the correlation between multiple variables and have been used successfully in the genomics field (e.g., Horvath 2011; Eisen et al. 1998; Butte and Kohane 2000). These methods have yet to be applied to describe relationships between exposures. The exposome globe is utilized to visualize clusters of measured exposures correlated with exposures identified in EWAS ("EWAS-identified exposure"). Here it is hypothesized that it is possible to attain a broader and more interpretable view of EWAS-identified exposures with an exposome globe.

Briefly, to compute and visualize an exposome globe, a nonparametric correlation coefficient between each pair of environmental factors is estimated. These coefficients are tetrachoric correlations between pairs of binary factors and Spearman correlations for continuous factors. There are many ways to compute correlations between variables. A nonparametric approach was chosen to avoid distributional assumptions regarding the environmental factors. Next, a permutation-based approach – similar to that described in algorithm 1 – is used to estimate the two-sided p-value of significance for each pair of correlations. Specifically, each environmental factor is randomly permuted (sampled without replacement), and the correlations were recomputed to create a set of correlations that reflected the null distribution of no correlation. The p-value for an individual correlation is then estimated by counting the number of permuted correlations (corresponding to the null distribution) that exceed the correlation in question. Next, the FDR is estimated using the Benjamini-Hochberg step-down approach (Benjamini and Hochberg 1995).

This exercise creates an array of pair-wise exposures whose interdependency is captured by a correlation coefficient. Each pair-wise correlation is visualized with the Circos visualization toolkit version 0.67 (Krzywinski et al. 2009). Each individual environmental factor is grouped and arranged in a circle. Lines between factors on the inside of the circle depict replicated correlations between factors, and the thicknesses of the lines depict the absolute values of the correlations. Red and blue lines represent positive and negative correlations, respectively. By visualizing relationships in this way, nonindependent exposures associated with phenotype can be inferred (Butte and Kohane 2000; Butte et al. 2000). For an example, see Fig. 5.3. Code to estimate correlation globes in the National Health and Nutrition Examination Survey (NHANES) can be found here: https://github.com/chiragjp/exposome_correlation. Further, one can peruse correlation globes estimated from the NHANES data here: http://www.chiragjpgroup.org/exposome_correlation/html/. The exposome globe is claimed here as a first step toward identifying putative mixtures of exposures associated with disease. These mixtures may be a result of common routes of exposure or behaviors (e.g., foods are mixtures of nutrients, or smoking behavior can result in a mixture of hydrocarbons and heavy metals). These systematic correlations may also help identify shared characteristics of exposures; for example, chlorinated persistent pollutants were all densely correlated with one another perhaps due to shared routes of exposure but also because they happen to be lipophilic and have similar metabolic fates.

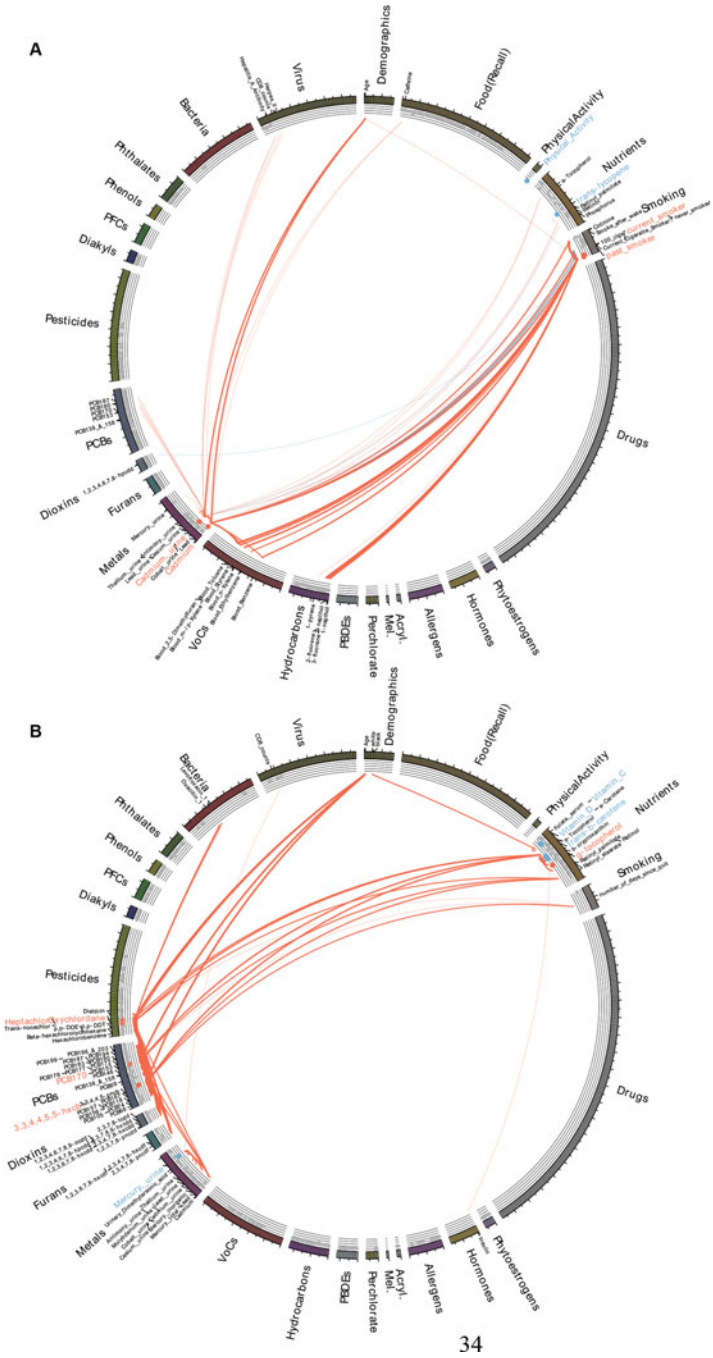

Fig. 5.3 (**a**) Exposome correlation globes for EWAS in all-cause mortality and (**b**) type 2 diabetes (T2D). Association *p*-values from EWAS are shown as a separate track ("EWAS track") above each exposure (red points denote EWAS validated associations with positive effect size [indicating risk]; blue points indicate an EWAS validated negative effect size [indicating protective]). Each line

5.4 Discussion

As described above, EWAS may facilitate many different ways of screening for factors. But, in the end, detecting mixtures will be an analytically complex exercise, requring new machine learning methods, but most importantly, large sample sizes. Previously, we recommended documenting prevalent exposures associated with disease using EWAS and replication before attempting to investigatehow combinations of exposures may influence phenotype or disease risk (Patel 2017). Extensions are described that might be used off-the-shelf to accommodate longitudinal data and statistical learning methods that consider the entire matrix of dependent variables at once.

5.4.1 Longitudinal Data

As discussed, environmental factors are dynamic. One way to capture the dynamic relationship between environmental factors and a phenotype of interest includes repeatedly measuring individuals over time. An example includes a longitudinal cohort study, in which a cohort is followed for a certain amount of time beginning prior to disease onset, such as childhood or adolescence. This type of study design might lessen the bias of reverse causality, but not completely (Rothman et al. 2008).

For a binary dependent variable, the Cox proportional hazard model is a common analytic model that can accommodate both time-independent and time-dependent variables. With this model, line 4 of algorithm 1 is substituted with the Cox model that inputs time-dependent variables. For both continuous and dependent variables, hierarchical modeling techniques such as generalized estimating equations may be utilized. The EWAS as described by algorithm 1 depends on the distribution of the *p*-values and effect sizes for the environmental factors, and statistical tests for these modeling techniques provide this requirement. Calculation of the empirical FDR proceeds also in the same way (Witten and Tibshirani 2010).

Fig. 5.3 (continued) represents *p*-values from high to low, and the significance increases as you move inward. Validated EWAS associations for T2D and all-cause mortality are offset in red or blue text. Only "first-degree" correlations (correlations for validated EWAS findings) are displayed in the globes and displayed in black text. *Acryl* acrylamide, *Mel* Melamine, *VOC* volatile organic compounds, *PCBs* polychlorinated biphenyls, and *PFCs* polyfluorinated compounds (Reproduced with permission from Patel and Manrai 2015)

5.4.2 *Feature Selection: Shrinkage Methods*

The EWAS screening method considers each environmental factor in a separate linear model iteratively (algorithm 1). This makes feasible the screening and interpretation of many variables without over-fitting the linear model (i.e., $p << n$, where p is the number of predictors, n is the number of individuals). However, this falsely assumes independence between environmental factors. Statistical learning methods, such as "shrinkage" methods, enable one to model the dependent variables simultaneously in the "overdetermined" ($p \geq n$) setting.

Two such popular shrinkage methods include the "lasso" (Tibshirani 1996) and "elastic net" (Zou and Hastie 2005). These methods are extensions of multivariate regression and have some relation to tree "boosting" methods (Hastie et al. 2009) and are applicable over the generalized linear model family, including Cox proportional hazards for longitudinal data (Witten and Tibshirani 2010). Both the lasso and elastic net are able to fit an overdetermined model by constraining the size of coefficients ("shrinking"). Because these methods consider the entire set of independent variables simultaneously (i.e., multiple regression), algorithm 1 is supplanted with the shrinkage procedure. Further, k-fold cross-validation is utilized to select features that have the lowest prediction variability on k number of datasets held out of the model building process (Hastie et al. 2009).

Feature selection operates through optimizing prediction accuracy of the dependent variable and not through ordering of test statistics of individual coefficients used in inference. Thus, parts of Stage 1 (FDR estimation) and Stage 2 (Validation) are reconfigured to accommodate this. Reconfiguring Stage 1, one cohort is used as the "discovery" or "prediction" cohort, applying the shrinkage method to find factors associated with the phenotype. Within this cohort, k-fold cross-validation is applied to optimize prediction accuracy with prediction cohort. Thereafter, the top factors found through this method are "validated" individually in additional validation cohorts using common tools for inference (e.g., GLM). Successful validation requires low nominal p and FDR values for the validation analyses.

Of course, "classical" methods for feature selection exist in the linear regression domain, such as "forward-stepwise" and "backward-stepwise." These methods may be used to select environmental factors, but they are not discussed here due to their high variability in subset selection due to the stepwise procedure, ultimately reducing their prediction accuracy (Vittinghoff et al. 2005). The shrinkage methods discussed above avoid this problem.

In this chapter, a straightforward and generalizable way to associate environmental variables toward identification of relevant mixtures in disease is presented. Furthermore, a way of ranking what variables may be worthwhile to pursue for further study through computation of the FDR is provided. Because of its proposed utility, the method has become a center point of discussion and debate (Todd 2010; Fallin and Kao 2011; Mak 2011; Heard et al. 2010; Borrell 2011; Rappaport and Smith 2010).

Internet References

- National Health and Nutrtition and Examination Survey (NHANES): http://www.cdc.gov/nchs/nhanes.htm
- NHANES data can also be downloaded here: http://nhanes.hms.harvard.edu
- International HapMap project: http://hapmap.ncbi.nlm.nih.gov
- Chirag Patel's Group:
  - Main page: http://www.chiragjpgroup.org
  - Software to conduct EWAS: https://github.com/chiragjp/xwas
  - Correlation globes: http://www.chiragjpgroup.org/exposome_correlation/html/
- NIH/NIEHS Children's Health Exposure Analysis Resource (CHEAR) program:
  - http://www.niehs.nih.gov/research/supported/dert/programs/chear/

References

Bartell, S.M., W.C. Griffith, and E.M. Faustman. 2004. Temporal error in biomarker-based mean exposure estimates for individuals. *Journal of Exposure Analysis and Environmental Epidemiology* 14: 173–179.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological* 57: 289–300.

Borrell, B. 2011. Epidemiology: Every bite you take. *Nature* 470: 320–322.

Buck Louis, G.M., and R. Sundaram. 2012. Exposome: Time for transformative research. *Statistics in Medicine* 31: 2569–2575.

Burdett, T., P.N. Hall, E. Hastings, L.A. Hindorff, H.A. Junkins, A.K. Klemm, J. Macarthur, T.A. Manolio, J. Morales, H. Parkinson and D. Welter. 2015. The NHGRI-EBI Catalog of published genome-wide association studies. Available at: www.ebi.ac.uk/gwas. Accessed 20 January 2018 version 1.

Butte, A.J., and I.S. Kohane. 2000. Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. *Pacific Symposium on Biocomputing*: 418–429.

Butte, A.J., P. Tamayo, D. Slonim, T.R. Golub, and I.S. Kohane. 2000. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proceedings of the National Academy of Sciences of the United States of America* 97: 12182–12186.

Centers for disease control and prevention (CDC). 2013. *National Health and Nutrition Examination Survey* [Online]. Atlanta: CDC. Available: http://www.cdc.gov/nchs/nhanes/. Accessed 1/1/2013 2013.

Davey Smith, G., and S. Ebrahim. 2003. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 32: 1–22.

Efron, B. 2010. *Large-Scale Inference*. Cambridge: Cambridge University Press.

Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* 95: 14863–14868.

Fallin, M.D., and W.H.L. Kao. 2011. Is 'X'-WAS the future for all of epidemiology? *Epidemiology* 22: 457–459.

Frayling, T.M., N.J. Timpson, M.N. Weedon, E. Zeggini, R.M. Freathy, C.M. Lindgren, J.R. Perry, K.S. Elliott, H. Lango, N.W. Rayner, B. Shields, L.W. Harries, J.C. Barrett, S. Ellard, C.J. Groves, B. Knight, A.M. Patch, A.R. Ness, S. Ebrahim, D.A. Lawlor, S.M. Ring, Y. - Ben-Shlomo, M.R. Jarvelin, U. Sovio, A.J. Bennett, D. Melzer, L. Ferrucci, R.J. Loos, I. Barroso, N.J. Wareham, F. Karpe, K.R. Owen, L.R. Cardon, M. Walker, G.A. Hitman, C.N. Palmer, A.S. Doney, A.D. Morris, G.D. Smith, A.T. Hattersley, and M.I. Mccarthy. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316: 889–894.

Goldstein, D.B. 2009. Common genetic variation and human traits. *The New England Journal of Medicine* 360: 1696–1698.

Greenland, S. 1990. Randomization, statistics, and causal inference. *Epidemiology* 1: 421–429.

Greenland, S., and H. Morgenstern. 2001. Confounding in health research. *Annual Review of Public Health* 22: 189–212.

Hardy, J., and A. Singleton. 2009. Genomewide association studies and human disease. *The New England Journal of Medicine* 360: 1759–1768.

Hastie, T., R. Tibshirani, and J.H. Friedman. 2009. *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Heard, E., S. Tishkoff, J.A. Todd, M. Vidal, G.N.P. Wagner, J. Wang, D. Weigel, and R. Young. 2010. Ten years of genetics and genomics: What have we achieved and where are we heading? *Nature Reviews. Genetics* 11: 723–733.

Hooper, L., A.R. Ness, and G.D. Smith. 2001. Antioxidant strategy for cardiovascular diseases. *Lancet* 357: 1705–1706.

Horvath, S. 2011. *Weighted network analysis: Applications in genomics and systems biology*. New York: Springer.

International Hapmap Consortium. 2005. A haplotype map of the human genome. *Nature* 437: 1299–1320.

Ioannidis, J.P., E.E. Ntzani, T.A. Trikalinos, and D.G. Contopoulos-Ioannidis. 2001. Replication validity of genetic association studies. *Nature Genetics* 29: 306–309.

Ioannidis, J.P.A., E.Y. Loy, R. Poulton, and K.S. Chia. 2009. Researching genetic versus nongenetic determinants of disease: A comparison and proposed unification. *Science Translational Medicine* 1: 8.

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, and M.A. Marra. 2009. Circos: An information aesthetic for comparative genomics. *Genome Research* 19: 1639–1645.

Mak, H.C. 2011. Trends in computational biology – 2010. *Nature Biotechnology* 29: 45–45.

Manly, B.F. 2007. *Randomization, bootstrap and monte carlo methods in biology*. Boca Raton: Chapman and Hall/CRC.

Manolio, T.A., L.D. Brooks, and F.S. Collins. 2008. A HapMap harvest of insights into the genetics of common disease. *The Journal of Clinical Investigation* 118: 1590–1605.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I. Mccarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C.N. Rotimi, M. Slatkin, D. Valle, A.S. Whittemore, M. Boehnke, A.G. Clark, E.E. Eichler, G. Gibson, J.L. Haines, T.F.C. Mackay, S.A. Mccarroll, and P.M. Visscher. 2009. Finding the missing heritability of complex diseases. *Nature* 461: 747–753.

Mccarthy, M.I., G.R. Abecasis, L.R. Cardon, D.B. Goldstein, J. Little, J.P.A. Ioannidis, and J.N. Hirschhorn. 2008. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Reviews. Genetics* 9: 356–369.

Miller, G.W., and D.P. Jones. 2014. The nature of nurture: Refining the definition of the exposome. *Toxicological Sciences* 137: 1–2.

NCI-NHGRI working group on replication in association studies. 2007. Replicating genotype-phenotype associations. *Nature* 447: 655–660.

Noble, W.S. 2009. How does multiple testing correction work? *Nature Biotechnology* 27: 1135–1137.

Omenn, G.S., G.E. Goodman, M.D. Thornquist, J. Balmes, M.R. Cullen, A. Glass, J.P. Keogh, F.L. Meyskens, B. Valanis, J.H. Williams, S. Barnhart, and S. Hammar. 1996. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *The New England Journal of Medicine* 334: 1150–1155.

Patel, C.J. 2017. Analytic complexity and challenges in identifying mixtures of exposures associated with phenotypes in the exposome era. Current Epidemiology Reports, January. Springer International Publishing, 4(1): 22–30.

Patel, C.J., J. Bhattacharya, and A.J. Butte. 2010. An Environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS One* 5: e10746.

Patel, C.J., M.R. Cullen, J.P. Ioannidis, and A.J. Butte. 2012. Systematic evaluation of environmental factors: Persistent pollutants and nutrients correlated with serum lipid levels. *International Journal of Epidemiology* 41: 828–843.

Patel, C.J., M.R. Cullen, J.P.A. Ioannidis, and D.H. Rehkopf. 2014. Systematic assessment of the correlation of household income with infectious, biochemical, physiological factors in the United States, 1999-2006. *American Journal of Epidemiology* 181: 171–179.

Patel, C.J., and J.P. Ioannidis. 2014a. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. *Journal of Epidemiology and Community Health* 68: 1096–1100.

———. 2014b. Studying the elusive environment in large scale. *Journal of the American Medical Association* 311: 2173–2174.

Patel, C.J., and A.K. Manrai. 2015. Development of exposome globes to map out environment-wide associations. *Pacific Symposium on Biocomputing*: 231–242.

Patel, Chirag J., Arjun K. Manrai, Erik Corona, and Isaac S. Kohane. 2016. Systematic correlation of environmental exposure and physiological and self-reported behaviour factors with leukocyte telomere length. International Journal of Epidemiology. https://doi.org/10.1093/ije/dyw043.

Patel, C.J., D.H. Rehkopf, J.T. Leppert, W.M. Bortz, M.R. Cullen, G. Chertow, and J.P. Ioannidis. 2013a. Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the United States National Health and Nutrition Examination Survey. *International Journal of Epidemiology* 42: 1795–1810.

Patel, C.J., T. Yang, Z. Hu, Q. Wen, J. Sung, Y. El-Sayed, H. Cohen, J. Gould, D. Stevenson, G. Shaw, X. Ling, and A.J. Butte. 2013b. Investigation of maternal environmental exposures in association with self-reported preterm birth. *Reproductive Toxicology* 45C: 1–7.

Pearson, T.A., and T.A. Manolio. 2008. How to interpret a genome-wide association study. *Journal of the American Medical Association* 299: 1335–1344.

Westfall, Peter H., and S.S. Young. 1993. *Resampling-based Multiple Testing*. New York: Wiley.

Peto, R., R. Doll, J.D. Buckley, and M.B. Sporn. 1981. Can dietary beta-carotene materially reduce human cancer rates? *Nature* 290: 201–208.

Rappaport, S.M. 2012. Discovering environmental causes of disease. *Journal of Epidemiology and Community Health* 66: 99–102.

Rappaport, S.M., D.K. Barupal, D. Wishart, P. Vineis, and A. Scalbert. 2014. The blood exposome and its role in discovering causes of disease. *Environmental Health Perspectives* 122 (8): 769–774.

Rappaport, S.M., and M.T. Smith. 2010. Environment and disease risks. *Science* 330: 460–461.

Rothman, K., S. Greenland, and T. Lash, eds. 2008. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins.

Salonen, J.T., P. Uimari, J.M. Aalto, M. Pirskanen, J. Kaikkonen, B. Todorova, J. Hypponen, V.P. Korhonen, J. Asikainen, C. Devine, T.P. Tuomainen, J. Luedemann, M. Nauck, W. Kerner, R.H. Stephens, J.P. New, W.E. Ollier, J.M. Gibson, A. Payton, M.A. Horan, N. Pendleton, W. Mahoney, D. Meyre, J. Delplanque, P. Froguel, O. Luzzatto, B. Yakir, and A. Darvasi. 2007. Type 2 diabetes whole-genome association study in four populations: The DiaGen consortium. *American Journal of Human Genetics* 81: 338–345.

Saxena, R., B.F. Voight, V. Lyssenko, N.P. Burtt, P.I. De Bakker, H. Chen, J.J. Roix, S. Kathiresan, J.N. Hirschhorn, M.J. Daly, T.E. Hughes, L. Groop, D. Altshuler, P. Almgren, J.C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Bostrom, B. Isomaa, G. Lettre, U. Lindblad, H.N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Rastam, E.K. Speliotes, M.R. Taskinen, T. Tuomi, C. Guiducci, A. Berglund, J. Carlson, L. Gianniny, R. Hackett, L. Hall, J. Holmkvist, E. Laurila, M. Sjogren, M. Sterner, A. Surti, M. Svensson, R. Tewhey, B. Blumenstiel, M. Parkin, M. Defelice, R. Barry, W. Brodeur, J. Camarata, N. Chia, M. Fava, J. Gibbons, B. Handsaker, C. Healy, K. Nguyen, C. Gates, C. Sougnez, D. Gage, M. Nizzari, S.B. Gabriel, G.W. Chirn, Q. Ma, H. Parikh, D. Richardson, D. Ricke, and S. Purcell. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316: 1331–1336.

Schwartz, D., and F. Collins. 2007. Medicine. environmental biology and human disease. *Science* 316: 695–696.

Sladek, R., G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre, P. Boutin, D. Vincent, A. Belisle, S. Hadjadj, B. Balkau, B. Heude, G. Charpentier, T.J. Hudson, A. Montpetit, A.V. Pshezhetsky, M. Prentki, B.I. Posner, D.J. Balding, D. Meyre, C. Polychronakos, and P. Froguel. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445: 881–885.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B: Methodological* 58: 267–288.

Todd, J.A. 2010. D'oh! genes and environment cause crohn's disease. *Cell* 141: 1114–1116.

Tzoulaki, I., C.J. Patel, T. Okamura, Q. Chan, I.J. Brown, K. Miura, H. Ueshima, L. Zhao, L. Van Horn, M.L. Daviglus, J. Stamler, A.J. Butte, J.P. Ioannidis, and P. Elliott. 2012. A nutrient-wide association study on blood pressure. *Circulation* 126: 2456–2464.

Vittinghoff, E., D. Glidden, S. Shiboski, and C. Mcculloch. 2005. *Regression methods in biostatistics: Linear, logistic, survival, and repeated measures models*. New York: Springer.

Wetterstrand, K. 2011. *DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program* [Online]. Available: http://www.genome.gov/sequencingcosts. [Accessed 2011/08/12].

Wild, C.P. 2005. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology, Biomarkers & Prevention* 14: 1847–1850.

———. 2012. The exposome: From concept to utility. *International Journal of Epidemiology* 41: 24–32.

Wild, C.P., A. Scalbert, and Z. Herceg. 2013. Measuring the exposome: A powerful basis for evaluating environmental exposures and cancer risk. *Environmental and Molecular Mutagenesis* 54: 480–499.

Witten, D.M., and R. Tibshirani. 2010. Survival analysis with high-dimensional covariates. *Statistical Methods in Medical Research* 19: 29–51.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 67: 301–320.

Chapter 6
Ecological Assembly of Chemical Mixtures

Rogelio Tornero-Velez and Peter P. Egeghy

Abstract Human-environment interactions have a significant role in the formation of chemical mixtures in the environment and by extension in human tissues and fluids. These interactions, which include decisions to purchase and use products containing chemicals as well as behaviors and activities that explain the uptake and absorption of chemicals, may be viewed as an ecological relationship between humans and their environments. Methods with origins in community ecology for evaluating structure in assemblages of flora and fauna are applied to investigate the nonrandom assembly of chemical species. Presence-absence matrix-based techniques are used to elaborate co-occurrence patterns with the aim of identifying the principal chemicals which tend to co-occur. This ecological premise is expanded by drawing on consumer market basket analysis techniques to show how this approach may help identify robust co-occurrence patterns.

Keywords Co-occurrence patterns · Community ecology · Chemical assembly · Human-environment interaction · Frequent itemset mining · Null model analysis

6.1 The Chemical Landscape and Emerging Technologies

Manufactured chemicals are integrated into nearly all industrial processes, building materials, and commercial goods such as furnishings, clothing, electronic equipment, cleaning products, and cosmetics (Weschler 2009; Wilson and Schwarzman 2009). Tens of thousands of anthropogenic chemicals are believed to be in wide commercial use, and several hundred additional chemicals are introduced into the market every year (U.S. EPA 2014). Despite a pervasive concern about the health consequences of exposure to ubiquitous synthetic chemicals (Wilson and Schwarzman 2009), particularly with respect to their role in the etiology of

R. Tornero-Velez (✉) · P. P. Egeghy
National Exposure Research Laboratory, U.S. Environmental Protection Agency, Washington, DC, USA
e-mail: tornero-velez.rogelio@epa.gov; egeghy.peter@epa.gov

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_6

increasingly prevalent diseases, such as autism, asthma, and childhood leukemia (Hertz-Picciotto and Delwiche 2009; Meeker 2012; Perrin et al. 2007), the hazard and exposure data necessary to assess risk adequately are unavailable for the vast majority of chemicals in commerce, even among those produced at a quantity greater than one million pounds per year (Judson et al. 2009).

In response to the need for accelerated risk assessment to substantially expand the number of chemicals evaluated, large-scale efforts are underway to apply high-throughput screening (HTS) approaches for toxicity testing (Kavlock et al. 2012; Tice et al. 2013). These approaches employ in vitro assays of toxicity pathway perturbations to predict the potential toxicity of chemicals as a cost-effective means of prioritizing candidates for more extensive toxicological evaluation. Although high throughput in nature, to date, these systems have been applied mainly to individual compounds and not to mixtures. Given the infinite possible ways chemicals may combine to form mixtures, information on the co-occurrence of chemicals in environmental and biological media is needed to inform the selection of mixtures for testing. Unfortunately, traditional approaches for measuring and monitoring chemicals in environmental and biological media have left us without even the most basic information on the occurrence, much less the co-occurrence, for the vast majority of man-made chemicals (Muir and Howard 2006; Schwarzman and Wilson 2009). Only a small fraction of all chemicals in commercial use have been measured in our environment or in our bodies largely due to the high cost of developing and applying targeted methods for the analysis of chemicals at trace levels in complex environmental matrices (Egeghy et al. 2012; McLachlan et al. 2014). Over the past two decades, and particularly over the past few years, new chromatography methods that allow screening for a large number of compounds in a single analysis have been developed to efficiently address the extensive number of chemicals that remain largely unmeasured (Cherta et al. 2015; Croley et al. 2012). These screening and nontargeted measurement methods are typically based on high-resolution mass spectrometry platforms, often with Orbitrap[1] or time-of-flight instrumentation, and are applicable to a variety of environmental and biological media. The approach relies on the accurate determination of the mass of the molecular ion and subsequent inference of its elemental composition (mass formula) but remains limited by difficulties in processing and filtering the large amounts of resulting data and by the absence of sufficient reference resources and optimization algorithms to facilitate the assignment of specific molecular structures from accurate mass (McLachlan et al. 2014). To compensate for the paucity of available measurements, approaches for modeling the presence of chemicals in the environment from information on sources and transportation processes based on linked environmental fate and bioaccumulation models have been developed (Arnot et al. 2010, 2012, 2014). Such modeling has been successful at predicting

[1] An Orbitrap (Orbitrap™, Thermo Scientific™) mass analyzer traps ions in an electrostatic field where the oscillations of the ions vary in accordance with their mass to charge ratio (Hu et al. 2005).

presence and relative concentrations of individual chemicals (Arnot et al. 2012) but has not been evaluated for the ability to correctly predict observed co-occurrence.

Biomonitoring studies show that humans carry a body burden of large numbers of chemicals, yet representatives of different chemical classes are rarely studied together (Meek et al. 2011; Yorita Christensen et al. 2013). In their prescient commentary which set forth a vision for collaboration between the National Institute of Environmental Health Sciences (NIEHS) and National Center for Environmental Health (NCEH) and Centers for Disease Control and Prevention (CDC), melding advances in toxicological testing (transgenic mouse, estrogen screen) and data on biomarkers in the population, Bucher and Lucier (1998) discussed use of large population-based surveys such as the National Health and Nutrition Examination Surveys (NHANES), (1) to determine whether there are common subsets of chemicals that persist in large segments of the population (and proposed that such subsets provide a logical starting point from which mixture studies can be designed and tested in rodents) and (2) to evaluate persistent chemicals found in the population for common biological activity.

The recent study by Qian et al. (2015) exemplifies the first aim of Bucher and Lucier (1998). In their evaluation of urinary metabolite data for six phthalates, Qian et al. (2015) observed that none of the individuals in the NHANES survey had a concentration at or above the 95th percentile for all six phthalates measured. However, at least 40% of individuals were exposed to 3–6 phthalates above their 50th percentiles. Thus, characteristic phthalate subsets were identified in population samples, and these data provide a basis to inform toxicological evaluations in rodent studies or streamlined high-throughput methods.

The NHANES National Report on Human Exposure to Environmental Chemicals is the most comprehensive repository of nationally representative chemical biomonitoring data (NHANES 2015a). However, Sobus et al. (2015) suggest that the potential of NHANES in this regard has not been fully realized as only a small percentage of the sampled NHANES-related publications reported on chemical biomarkers (8% yearly average). Indeed, the balance of publications focus on growth (obesity), physical activity and fitness surveys, diet behavior and nutritional status, and various health measures (including blood pressure, cholesterol, hepatitis, herpes, HIV status, osteoporosis, and cardiovascular health) (NHANES 2015b). To be sure, these metrics are of fundamental importance, yet by themselves they do not necessarily unravel the role of environmental chemicals in disease etiology.

A number of recent publications illustrate how both NHANES health and exposure data may be mutually informative. Yorita Christensen et al. (2013) used data from the 2003–2004 cycle of NHANES to evaluate the relationship between alanine aminotransferase (ALT) and 37 environmental contaminants and observed significant associations between elevated ALT levels and levels of heavy metals, non-dioxin-like polychlorinated biphenyls (PCBs), and dioxin-like compounds. Furthermore, those authors reported interactions between these exposure classes and ALT and attributed 78% of this interaction to mercury, PCB 180, and 3,3′,4,4′,5-pentachlorobiphenyl. Patel et al. (2010) used 1999–2006 NHANES to investigate the relationship between 266 environmental factors and a marker of type

2 diabetes (fasting sugar $\geq$ 126 mg/dl), employing a novel methodology which they denote environment-wide association study (EWAS), modeled after the genome-wide association study (GWAS) in controlling for multiple comparisons (Chap. 5). The authors observed increased odds ratios (OR) for exposure to heptachlor epoxide (OR 1.7), the vitamin ɣ-tocopherol (OR 1.5), and exposure to high concentrations of PCBs (OR 2.2), as well as a protective effect for beta-carotenes (OR 0.6). Bell and Edwards (2014, 2015) applied a methodology with origins in market basket analysis known as *frequent itemset mining*[2] to discover relationships between 219 chemicals and 93 health outcomes/biomarkers in the 1999–2010 NHANES. The authors not only confirmed chemicals implicated in type 2 diabetes by Patel et al. (2010) but also confirmed several findings of the "C8 Science Panel" (Fletcher et al. 2005), which was commissioned to evaluate links between exposure to perfluorooctanoic acid (PFOA or "C8") and several health outcomes as part of a class action lawsuit.

Bucher and Lucier (1998) noted the potential of NHANES in guiding toxicological assessments of complex exposures. They proposed that NHANES may be used to identify *common subsets* of chemicals which persist in populations and that the identified subsets serve as a template to recreate mixtures for testing purposes. Accordingly, understanding the underlying exposures which may explain chemical co-occurrence in biological media continues to be a driving goal. The number of ways that even a relatively small number of chemicals may combine to form unique chemical combinations can be astronomically large and is equal to 2^r for a pool of r chemicals; for example, with the 212 chemical biomarkers measured in the 2003–2004 NHANES cycle, this works out to 6.58×10^{63} combinations. However, it is imperative to recognize that not all combinations are equally plausible (Tornero-Velez et al. 2012). For that reason, an understanding of the generative processes that lead to plausible mixtures is sought, that is, an understanding of which single chemicals tend to combine to form common subsets and which of these subsets tend to combine to form higher-order mixtures.

This work explores the hypothesis that chemical mixtures in the environment, and by extension in human tissues and fluids, are assembled through human-environment interactions. Methods that were originally developed for understanding ecological assembly of communities are applied to investigate the nonrandom assembly of chemical species using presence-absence matrix-based techniques. Machine learning tools are then used to investigate and identify co-occurrence patterns and their nested structure, thereby revealing which chemicals tend to combine. The result is an identification of specific chemicals that are likely to have high probability of co-occurrence at measurable concentrations. These observed combinations can inform the intelligent toxicity testing of mixtures,

[2]Now used for the identification of sets of a variety of items in large databases (e.g., a specific set of symptoms characteristic of a rare disease observed in a medical records database), the technique originally was developed to examine customer behavior with respect to consumer products purchased.

whether for high-throughput screening or more traditional in vivo evaluations (Kapraun et al., 2017). Similarly, these approaches can inform exposure monitoring and biomonitoring studies by identifying targets for measurement to optimize monitoring strategies for collecting data of highest value in understanding chemical exposure.

6.2 Human-Environment Interaction

An *ecological assembly* of chemical mixtures (*ecological* in this context refers to the shaping of chemical mixtures through human-environment interaction) is considered. The environment refers to modern civilization, not devoid of forests and savannahs and other natural spaces, yet clearly marked by the city and its urban landscape, in a state of perpetual growth and flux. Modern civilization is inseparable from technology and anthropogenic chemicals. Humans are in perpetual contact with anthropogenic chemicals, in the air breathed, the food and water consumed, the daily commute, the workplace, and most notably in the ever-expanding portfolio of consumer products being brought to market. Chemicals released into the environment from distant, "far-field" sources (e.g., factories, refineries) are ultimately dispersed away from the source and among various environmental compartments (e.g., air, water, soil, vegetation). Those that are more persistent in the environment can be distributed over great distances and can come into contact with humans through a series of complex environmental pathways that include transport through several environmental media and food web bioaccumulation (Arnot et al. 2012; NRC 2000). Persistence (i.e., resistance to environmental degradation) is often identified as the most important factor affecting such indirect exposure via the environment as it, along with release rate, controls the amount present in the environment. Furthermore, chemicals that degrade slowly may travel greater distances by long-range transport in air or water and may affect a larger population (Mackay et al. 2014). Although wind and water serve to disperse chemicals, modern society also distributes chemicals in a nonrandom fashion, deliberate in its infrastructure and in the production and consumption of consumer goods. Indeed, it is the proximal, "near-field" sources of chemicals in the residential environment (e.g., consumer goods, household articles, building materials) that are believed to produce the highest exposure for the vast majority of chemicals, particularly those that are nonpersistent (Jayjock et al., Jayjock et al. 2009; Mitchell et al. 2013; Wambaugh et al. 2014). Consequently, chemical co-occurrence is structured through these societal processes. This chapter focuses on the *assembly* and *identification* of chemical mixtures that arise through human-environment interaction.

6.2.1 Community Ecology

Human ecology is the study of the relationship between humans and their natural, social, and built environment. This discipline addresses the process of human-environment interaction, which encompasses the distribution and use of chemicals (and the products made from them) and facilitates the creation and maintenance of one's immediate surroundings as well as one's own lifestyle choices. While acknowledging the broad and diverse scope of human ecology (McKenzie 1924; Mumford 1938; Hawley 1986; Urban Land Institute 2013), the focus here is on interactions that involve consumer products. Consumer products contain multiple chemicals, and thus they introduce chemical mixtures into the environment. However, these mixtures are also dependent on the combinations of products that consumers choose to acquire and use. An ecological paradigm is demonstrated as a useful construct to examine how these choices may impact chemical co-occurrence in our immediate environment.

In ecology, a community is an assemblage of two or more different species occupying the same geographical area. The question of whether ecological communities (species assemblages) are structured (nonrandom) and, if so, whether the structure derives from competitive interaction has been a central question in community ecology (Diamond 1975; Connor and Simberloff 1979; Sepkoski 1996; Gotelli 2000; Connor et al. 2013). Diamond (1975) hypothesized that interspecific competition (*wherein different species compete for the same resources*) leads to a lack of co-occurrence of species. In describing the distribution of birds in the Bismarck Archipelago, Diamond noted "checkerboard" patterns and proposed that these patterns signify community assembly rules. These *assembly rules* are as follows:

a. If one considers all the combinations that can be formed from a group of related species, only certain ones of these combinations exist in nature.
b. Permissible combinations resist invaders that would transform them into forbidden combinations.
c. A combination that is stable on a large or species-rich island may be unstable on a small or species poor island.
d. On a small or species-poor island, a combination may resist invaders that would be incorporated on a larger or more species-rich island.
e. Some pairs of species never co-exist, either by themselves or as part of a larger combination.
f. Some pairs of species that form an unstable combination by themselves, may form part of a stable larger combination.
g. Conversely, some potential larger combinations that are composed entirely of stable sub-combinations are themselves unstable.

6.2.2 Parallels Between Chemical and Biotic Species

As divorced as these rules may appear to be from chemical mixtures, consider that chemicals in commerce are subject to market pressures. Borrowing from the community ecology model, Tornero-Velez et al. (2012) likened avian species to chemical species to investigate the co-occurrence patterns of pyrethroid pesticide residues in 168 child care centers across the country (Tulve et al. 2006). The authors observed nonrandom co-occurrence patterns of pyrethroids reminiscent of patterns observed for bird species by Diamond (1975) and concluded that studies of species co-occurrence parallel the issue of chemical co-occurrence at specific locations. Both are driven by processes that introduce structure in the pattern of co-occurrence. Although Tornero-Velez et al. (2012) described a parallel specifically with avian co-occurrence patterns (i.e., the West Indian Finch matrix), similar patterns are evident in assemblages of ants (Gotelli and Ellison 2002), reptiles (França and Araújo 2007), plants (Götzenberger et al. 2012), fungi (Horner-Devine et al. 2007), plant-animal combinations (Peng et al. 2008), and even parasites on marine fish (Gotelli and Rohde 2002). The critical implication of community structure is that there are fewer realized species combinations than under the purely random case (where the theoretical tally of unique combinations is 2^r for a pool of r species). But how does structure manifest?

If the idea of replacing *avian* species with *chemical* species is entertained, in particular for chemicals with a social context (consumer products), *assembly rules* may be inferred from observed chemical combinations. Diamond's (1975) assembly rules are not presented here as orthodoxy – indeed, some rules appear as tautologies of others (Connor and Simberloff 1979). The aim is to repurpose methods by which community ecologists contemplated a "mixtures problem" (identification of a community) as one of *assembly* – assembly of chemical mixtures. This is evidently germane to ecology – to delineate an apparent chance assemblage of flora/fauna species from a real community and to question the basis for species assembly whether via immigration or competition or both. It is proposed that the discourse concerning *ecological assembly* of species is relevant to the toxicology of *chemical mixtures*. Mixtures toxicology has developed the methods to assess departures from *additivity*,[3] though one of the obstacles has been in identifying specific mixtures to test. If mixture assembly can be understood as a generative process, it is possible to anticipate which mixtures are likely to form. In parallel with ecology, chemical assemblages (mixtures) are assumed relevant to an *ecological niche* (e.g., licensed child care centers on the national scale).

In a critical examination of Diamond's (1975) assembly rules, Connor and Simberloff (1979) accepted some and dismissed others, but what drew most contention was the notion that competitive segregation is the underlying basis for community structure (e.g., Brooks 1985; Bartha et al. 1995; Krüger et al. 2010).

[3]See Gennings et al. (2005) for discussion of a method based on consideration of changes in the slope of the dose-response curve of one chemical produced by the presence of other chemicals.

The purpose here is not to delve into this controversy – one which has endured 40 years – but rather to say that the controversy drew attention to an elegant hypothesis (competitive segregation) and a useful methodology (namely, *null model analysis*) for identifying species co-occurrence patterns which can also be applied to the problem of identifying plausible chemical combinations. It is this discourse and inquiry that is relevant to mixture toxicology. Popularized by this debate, the presence-absence matrix has become a fundamental unit of analysis in community ecology and biogeography (Connor and Simberloff 1979; Diamond and Gilpin 1982; Gilpin and Diamond 1982; Manly 1995; Jackson et al. 1992; Kelt et al. 1995; Gotelli 2000; Miklós and Podani 2004; Ulrich and Gotelli 2013). In preparing a site-by-species (0,1) matrix from raw data, information on species abundance (population size) is discarded beyond registering the presence (1) or absence (0) of the species. Although this represents a loss of information, the problem is recast as one of co-occurrence or mutual exclusion (*anti* co-occurrence) of species.

Figure 6.1 is an example of presence-absence matrix derived from the distribution of 17 species of finches from 19 of the largest islands in the West Indies (West Indian Finch matrix (WIF); Gotelli and Abele 1982). Each row i represents a species ($i = 1$ to R rows), and each column j represents a site ($j = 1$ to C columns). The entries X_{ij} in the matrix represent the presence (1) or absence (0) of species i at site j.

Note the widespread occurrence of *Loxigilla noctis* and *Tiara bicolor* and, on the other hand, the rare occurrence of other species such as *Loxigilla violacea* and *Tiara canora* that find habitat in just one or a few islands. This is not unlike the market presence of some anthropogenic chemicals. Note that in the West Indian Finch matrix in Fig. 6.1, Hispaniola (the second largest island) exhibits the greatest species diversity.[4] In Fig. 6.1, the total number of species occurrences is 55. Two 2-by-2 submatrices ("checkerboards") each showing mutual exclusion for a species pair are emphasized, one with adjacent cells (pair, *T. bicolor* and *T. canora*), the other with nonadjacent cells (pair, *T. olivacea* and *L. noctis*). Attention is called to these checkerboards because when this pattern is sustained across all sites of occupancy for the pair, it is the basis of Diamond's (1975) assembly rules. Overall, of the 136 possible species pairings,[5] 91 pairs form perfect checkerboard distributions in the WIF archipelago (species pairs that never co-occur on the same islands; Gotelli 2000). This is more than the 55 pairs expected (through simulation) had the occurrences been randomly cast on a "blank" matrix of similar dimensions (Gotelli 2000). Under the model of Diamond (1975), this difference between observed and expected patterns implies *structure* derived from competitive segregation.

[4]All other factors held constant; the number of species on an island is generally observed to increase with increasing area (Connor and McCoy 1979).

[5]For N species, there are $N(N–1)/2$ possible species pairs or 136 pairs for $N = 17$ species.

	Cuba	Hispanolia	Jamaica	Puerto Rico	Guadeloupe	Martinique	Dominica	St Lucia	Barbados	St Vincent	Grenada	Antigua	St Croix	Grand Cayman	St Kitts	Barbuda	Montserrat	St Martin	St Thomas	row sums
Carduelis dominicensis	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Loxia leucoptera	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Volatina jacarina	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Sprophilia nigricolis	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Melopyrha nigra	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
Loxigilla portoricensis	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Loxigilla violacea	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Loxigilla noctis	0	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	12
Melanospiza richardsoni	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Tiara olivacea	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	5
Tiara bicolor	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	17
Tiara canora	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Loxipasser anoxanthus	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Saltator albicollis	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4
Torreornis inexpectata	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Ammodramus savannarum	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Zonotrichia capensis	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
column sums	4	7	5	4	3	3	3	4	2	2	4	2	1	2	2	2	2	2	1	55

Fig. 6.1 West Indian Finch (WIF) matrix as an example of a presence-absence matrix. Two submatrices ("checkerboards") are shown indicating mutual exclusion of species pairs in two islands. The left one has adjacent cells; the right one has nonadjacent cells (Data from Gotelli and Abele 1982)

6.3 Null Model Analysis of Pyrethroid Mixtures

Null model analysis has been a way to test hypothesized species interaction, in particular competitive segregation, in presence-absence matrices. Null model analysis involves developing a set of randomized matrices and comparing the randomized set with the observed matrix to accept or reject a hypothesis of structure (via the observation of patterns). As defined by Gotelli and Graves (1996):

> *A null model is a pattern-generating model that is based on randomization of ecological data or random sampling from a known or specified distribution. The null model is designed with respect to some ecological or evolutionary process of interest. Certain elements of the data are held constant, and others are allowed to vary stochastically to create new assemblage patterns. The randomization is designed to produce a pattern that would be expected in the absence of a particular ecological mechanism.*

To examine co-occurrence patterns of pesticides in the child care center study reported by Tulve et al. (2006), Tornero-Velez et al. (2012) applied null model analyses (as prescribed by Gotelli 2000). The Tulve et al. (2006) study aimed to characterize the environments of young children in a randomly selected, nationally representative sample of licensed institutional child care centers. Multistage sampling with clustering was used to select 168 child care centers in 30 primary sampling units in the United States. Pesticides, lead, and allergens were measured at multiple locations in each center. The authors reported floor surface loadings

(ng/cm^2) of 15 different pyrethroid or pyrethrin pesticides (hereafter, "pyrethroids") in each of the centers. A pyrethroid was considered present (1) if its surface loading at a particular site (child care center) exceeded the respective method of detection limit (ranging from 0.002 to 0.016 ng/cm^2), and absent (0) otherwise. In this way, a 15-by-168 presence-absence matrix was derived for 15 distinct pyrethroids in 168 centers. In the context of null model analysis, this matrix represents the *observed matrix*. An important assumption/parallel in the null model analysis approach was that the 168 licensed child care centers were considered analogous to the archipelagic collection of islands/sites in that both represent a broad ecological niche. Inferences on chemical co-occurrence patterns correspond to the domain of the niche (i.e., licensed child care centers). Furthermore, co-occurrence was assessed among pyrethroids in the same way that ecologists constrain the analysis to a guild; that is, a group of species competing for the same environmental resources is analogous to a class of chemicals (pyrethroids) competing for market share.

Gotelli (2000) reviewed the null model analysis literature and developed a 3-by-3 matrix locating nine historical null models used for randomizing and testing (Fig. 6.2). This summary representation provides an internally consistent interpretation of the methodology. A balance between randomness and structure is achieved by constraining the randomization according to parameters of the observed matrix. Thus, a null (1 of 9) can be chosen, and the species-site probabilities calculated based on the parameters, producing randomized matrices in accordance with the chosen null.

The analysis works by constraining the null according to beliefs of structure (based on the marginal row and column totals of the data matrix). More specifically, the structural hypothesis is derived from parameters of the observed matrix (N, S_i, T_j). Let S_i be the total number of occurrences of species i across all sites (row sums). Define T_j as the total number of species occurring in site j (column sums). And, let N equal the total number of all species occurrences in the matrix. Allowing the row and column margin totals to vary randomly with either equal probability ("equally likely") or with probabilities proportional to the margin totals in the original matrix ("proportional") or to be identical to the margin totals in the original matrix ("fixed") leads to nine null models (Fig. 6.2).

With the exception of Null 9 (see Fig. 6.2), the randomization procedure is the same for all null models, a bootstrapping of the observed matrix. Thus, the N occurrences are placed back on a blank matrix of the same dimensions as the observed matrix according to the probability specified by a formula of N, S_i, and T_j (Fig. 6.2). This is performed multiple times, say 5000 times, and a set of matrices consistent with the null is produced. A metric of choice is then computed for the observed matrix and compared with a distribution of that metric across the randomized set. If the frequency of the metric in the observed matrix resides in the tails of the distribution in the randomized set, the null is rejected. The metric at the center of the assembly rules debate has been the CHECKER (pairs of species which are mutually exclusive; Gotelli 2000). Tornero-Velez et al. (2012) examined the CHECKER and considered a metric directly based on *association*, both in the total

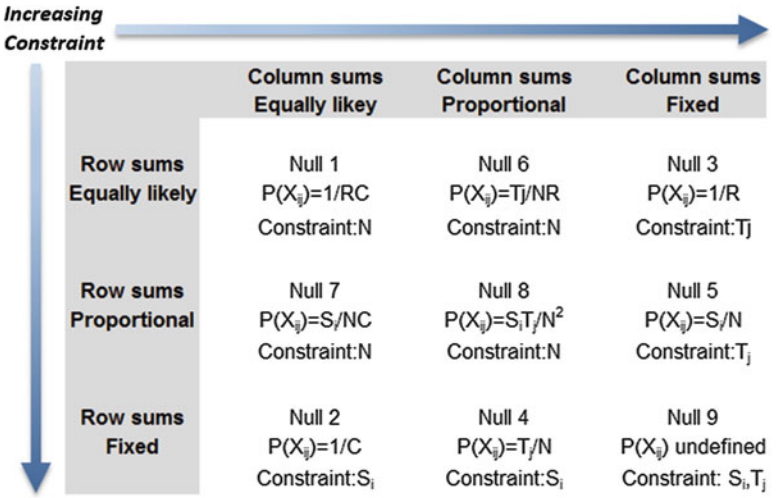

Fig. 6.2 Null models used to evaluate structure in presence-absence matrices. Each null has a unique formula for calculating the probability of occupancy for the first cell in the matrix [$P(X_{ij})$], where N = matrix total, R=number of rows, C =number of columns, S_i = total for row i, and T = total for column j. "Proportional" denotes that margin totals are allowed to vary randomly but with probabilities proportional to the margin totals in the original matrix (Adapted from Gotelli 2000)

number of unique combinations (COMBO metric; Pielou and Pielou 1968) and, in particular, k-way combinations. This does not imply that competitive segregation was rejected as operative in chemical mixture assembly.

In investigation of pyrethroid co-occurrence in child care centers, a larger CHECKER score (quantifying the number of pyrethroid pairs forming perfect checkerboard distributions in the presence-absence matrix) was observed compared with a chance process (Tornero-Velez et al. 2012). A lower COMBO score (the total number of unique combinations) was observed compared with purely chance process. The trend in both these indices provides evidence of structure in co-occurrence patterns. Table 6.1 shows the parallel between the CCC (child care centers) matrix and the WIF (West Indian Finch) matrix – both supporting competitive segregation (CHECKER score). For both WIF and CCC matrices, Null 1 (random) simulated too low of a CHECKER score compared with the observed matrix and was thus rejected (p-value < 0.001). For both WIF and CCC, Null 9 simulated the CHECKER score in accordance with the observed data. For both WIF and CCC, the COMBO score, the number of unique combinations, tended to decrease with increasing structure (Null 1→ Null 8 → Null 9)[6]; however, Null 9 tended to overpredict the COMBO score in both the CCC and WIF matrix. This suggested that competitive segregation provides a preferable explanation for

[6]Evaluations of CHECKER and COMBO indices were conducted only for null models 1, 8, and 9 (Tornero-Velez et al. 2012).

Table 6.1 Null model analysis on the child care centers (CCC) and West Indian Finch (WIF) matrices

	CCC matrix 15 species × 168 site		CCC matrix, lumped 12 species × 168 sites		WIF 17 species × 19 sites	
	CHECKER	**COMBO**	**CHECKER**	**COMBO**	**CHECKER**	**COMBO**
Observed	34	39	20	35	91	10
Null 1	7.2**	101.5**	8.8**	65.6**	70.8**	18.6**
Null 8	35.9	61.9**	21.7	46.3**	51.0**	17.2**
Null 9	34.9	42.8*	19.7	35.4	89.4	15.1**

Adapted from Tornero-Velez et al. (2012)

Note: CHECKER index is the number of species pairs forming checker patterns. Segregation increases as CHECKER increases. COMBO index is the total number of unique combinations. Structure (nonrandomness) decreases as COMBO decreases. Expected indices under the null model are based on 5000 simulations, $^{**}p$-value < 0.001, $^{*}p$-value < 0.05. To remove the effect of structuring due to co-occurring isomers produced during synthesis, "CCC matrix, lumped" lumps isomers together, resulting in 12 distinct species of pyrethroids

structure. However, for the CCC matrix, the structural isomers of permethrin (e.g., cis-permethrin and trans-permethrin) were posited to inflate the number of unique combinations. After lumping isomers in the CCC matrix into 12 distinct species, a COMBO score statistically consistent with the observed matrix (Null 9) was achieved (Table 6.1). Thus, the take-home message from these analyses was that null model 9, the most structured null, was most consistent with both species and chemical co-occurrence patterns.

6.3.1 Specific Combinations and Nested Behavior

Tornero-Velez et al. (2012) found that observed combinations were generally of low order (i.e., five-way or less). Among the 168 centers, there was only one observation each of seven-, eight-, and nine-way combinations. Combinations higher than nine-way (considering 12 or 15 distinct chemical species) were not observed. Four unique five-way combinations were observed; in particular, {cyfluthrin, cypermethrin, esfenvalerate, cis-permethrin, trans-permethrin} and {cyhalothrin, cypermethrin, esfenvalerate, cis-permethrin, trans-permethrin} were consistent with nulls requiring some form of species proportionality. Thus, the prominence of some species/chemicals across multiple sites suggests they serve as *generators* for forming combinations. Null models predicated on equal probability species (Null 1, 2, and 7) were rejected for combinations of order three-way and higher; however, these nulls were not rejected for some one-way and two-way combinations. Furthermore, for combinations three-way and higher, nulls permitting equal probability of sites were tolerated (not rejected). It is unlikely that sites are equally likely to "host" chemical species. This may indicate that null model analysis does not have sufficient power to determine the robustness of individual

combinations. Indeed, it has been used by ecologists only in conjunction with general measures of co-occurrence (Gotelli 2000). What was striking – though not systematically evaluated by Tornero-Velez et al. (2012) – was a pattern of nested subsets. For example, the most frequent four-way combinations, {cyfluthrin, cypermethrin, cis-permethrin, trans-permethrin} and {esfenvalerate, cypermethrin, cis-permethrin, trans-permethrin}, were each observed three times in the 168 centers, all of which are subsets of the two five-way combinations. And, of the three-way combinations, {cypermethrin, cis-permethrin, trans-permethrin} was observed 14 times and {cis-permethrin, trans-permethrin} was observed 46 times by itself. Thus, a nested pattern was apparent in the data.

Nested species patterns are observed in ecological systems (Patterson 1990). The classic pattern in island geography consists of a cluster of islands off the mainland. Species richness is greater in the mainland and becomes more fractured on islands farther from the mainland. The overall pattern is one of nested subsets; as distance from the mainland increases subsets of species drop-off. The pattern is thought to result from the predictable sequence of species undergoing extinction from inhabiting fragmented land (Patterson 1990). The causality of nestedness in insular communities, whether from selective immigrations or extinctions, is the subject of debate in ecology (Lomolino 1996). Methods to assess nestedness in ecology tend toward general measures, some, for example, focusing on the entropy of the ecological niche (Atmar and Patterson 1993; Rodríguez-Gironés and Santamaría 2006). A matrix exhibiting nestedness has a "low temperature" indicating low entropy/structure, while a matrix devoid of nestedness has a "high temperature" indicating high entropy/randomness. Given the potential for many nested patterns, it is understandable that ecologists have gravitated toward general measures of co-occurrence and nestedness in characterizing communities. For purposes of examining co-occurrence of chemical mixtures, a promising methodology to organize and make use of nested behavior is the technique of *frequent itemset mining*.

6.4 Frequent Itemset Mining and Nested Behavior

Community ecology and market basket analysis are of course very distinct and separate endeavors; however an interesting parallel is that both are concerned with nested behavior of species (or products). The nested behavior of items (consumer products, environmental chemical residues) reveals the generative processes that lead to chemical co-occurrence. The market basket analysis method of *frequent itemset mining* exploits the nested behavior of consumer product transactions to determine if consumers are more or less likely to buy specific items based on their other purchases. If, for example, an online merchant is aware of the items in a consumer's electronic shopping cart, other items are suggested to the consumer based on the *nested* behavior of past transactions by a larger population of consumers. Frequent itemset mining relates transactions (T) or "baskets" to items (I), allowing for mining of *association rules* (Agrawal et al. 1993; Hahsler et al. 2005).

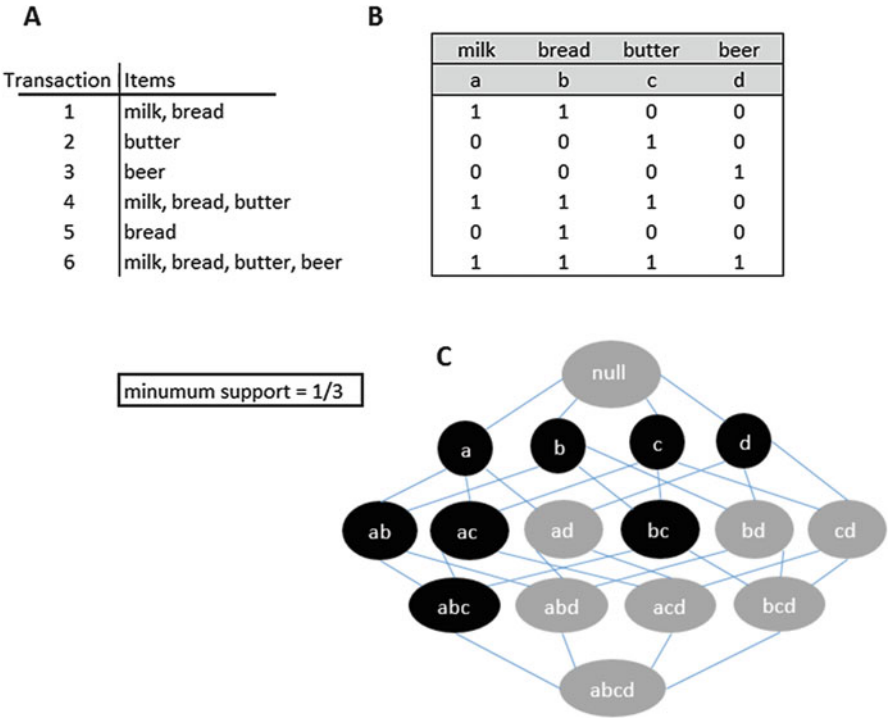

Fig. 6.3 Example database with six transactions and four items. (**a**) Transaction ID list. (**b**) Presence-absence representation. (**c**) Lattice with frequent itemsets (black) satisfying minimum support count of 1/3 (Example abstracted from Hahsler and Hornik 2007)

In the ecological context, transactions or baskets may refer to islands (sites) and items to species. In the environmental chemical context, baskets remain sites, and items become the chemicals of interest. The structure is as follows (Agrawal et al. 1993):

$$\begin{aligned}&\text{Set of items} : I = \{I_1, I_2, \ldots, I_m\}\\ &\text{Set of transactions} : T = \{t_1, t_2, \ldots, t_{\mathrm{n}}\}, t_j \subseteq I\\ &\text{Any subset} B \subseteq I \text{ is called an itemset}\end{aligned}$$

More concretely, Fig. 6.3 shows transactions for grocery store purchases. Here, the items, $I = \{\text{milk, bread, butter, beer}\}$. There are six transactions; the fourth transaction is $t_4 = \{\text{milk, bread, butter}\}$. Any subset B of I is an itemset; however, here the only concern will be with frequent itemsets.

Frequent itemset mining identifies the frequent itemsets among the infrequent or spurious itemsets. This is achieved by requiring a global threshold level of occurrence (minimum support count) for itemsets. In this way, the major associations are uncovered when the minimum support count is set high, and rarer more obscure associations are discovered when the minimum support count is decreased. The

space of all itemsets can be elaborated as a network (Fig. 6.3) of all possible subsets of I starting with the empty set (null) and working downward, ending with the set of all members {abcd}. The nodes represent the universe of unique combinations. In total, there are $2^R - 1$ unique combinations (excluding the null set) where R is the number of items/species/chemicals (4, in the current example).

The *Apriori* algorithm (Agrawal et al. 1993) is used to trim the exponential search space (2^R) based on *support* (occurrence). The algorithm proceeds by elaborating the network from *null* down through all k-order combinations and determines if such combinations are observed in the transaction database T at or above a minimum support count (minimum occurrence threshold) set by the investigator. Nodes satisfying this criterion are defined *frequent itemsets*. The set properties of *Apriori* are as follows (Agrawal et al. 1993):

No superset of an infrequent itemset can be frequent.
All subsets of a frequent itemset are frequent.

Thus, in Fig. 6.3C since {abc} is frequent, all of its subsets are frequent. Conversely, since item {ad} is infrequent, all of its *supersets* are infrequent. Once frequent itemsets are identified, *association rules* are developed. A *rule* is defined as an implication of the form (Agrawal et al. 1993):

$$X \rightarrow Y$$
$$\text{where} X, Y \subseteq I \text{ and} XY = \phi$$

In the expression "$X \rightarrow Y$," consider that X is {milk, bread} and Y is {butter}. Depending on the number of baskets containing {milk, bread, butter}, the rule $X \rightarrow Y$ is supported with support $(X \cup Y)$. Only if $X \rightarrow Y$ is realized do X and Y co-occur, {X, Y} (Agrawal et al. 1993). X is called the left-hand side (LHS) of the implication and Y the consequent right-hand side (RHS). The *Apriori* algorithm requires that the support count (o) (occurrence in the transaction database) of X and Y be at least equal to the minimum support count or greater. Support (S) is expressed as an occurrence fraction (o/N), where N is the number of transactions (examples). Thus, if the minimum support is set to 0.001 for a database of 50,000 consumer transactions, the minimum support count is 50 occurrences.

The confidence of a rule, conf$(X \rightarrow Y)$, is defined as the conditional probability of observing the RHS given the LHS, $P(Y|X)$. In other words, conf$(X \rightarrow Y)$ is the proportion of transactions that contain X which also contain Y (Hahsler et al. 2005):

$$\text{conf}(X \rightarrow Y) = S(X \cup Y)/S(X)$$

In this example of six transactions (Fig. 6.3), three of the six (3/6) involve milk and bread, $S(X)$ is 3/6, but only two of *those* six (2/6) also involve butter, an $S(X \cup Y)$ of 2/6. The confidence of the rule: conf$(\{milk, bread\} \rightarrow \{butter\}) = (2/6)/(3/6) = 2/3$, meaning that 2/3 of transactions involving {milk, bread} have the co-occurrence of interest {milk, bread, butter}. $S(X \cup Y)$ can be considered in two important ways, as the support of the rule (1/3) and as the co-occurrence (1/3).

Normalizing the confidence by the support of Y gives the lift measure. The lift is the ratio of observed support (for co-occurrence) to that expected if X and Y were independent (Hahsler et al. 2005):

$$\text{lift}(X \rightarrow Y) = S(X \cup Y)/(S(X) \times S(Y))$$

In the example, lift *({milk, bread} → {butter})* = (2/6)/(3/6 × 3/6) = 1.33. If independence is assumed between {milk, bread} and {butter}, a frequency of co-occurrence of 1/4 is estimated. But the observed co-occurrence of {milk, bread, butter} is 1/3, so the estimate would be off by a factor of 1.33. The lower bound of lift is one (independence), and an increase in lift suggests a stronger dependent association.

6.4.1 Association Rules Mining of the CCC Study

As discussed earlier in this chapter, Tornero-Velez et al. (2012) examined co-occurrence of pyrethroid residues in child care centers. Co-occurrence patterns were reexamined using *frequent itemset mining* to further investigate nested patterns. Minimum support count was set to 5%, thus requiring occurrences in at least 5% of centers. A total of 29 rules were found, and these involve only five pyrethroids *cis-/trans*-permethrin, cypermethrin, cyhalothrin, esfenvalerate, and cyfluthrin. Figure 6.4 shows that all the pyrethroids are networked with one another except cyfluthrin. While cyfluthrin co-occurs with other pyrethroids, those co-occurrences do not meet the minimum support threshold (5%).

At 5% minimum support, the highest-order combinations are three-way occurrences: {cis-permethrin, trans-permethrin, cypermethrin} in 18% of centers, {cis-permethrin, trans-permethrin, cyhalothrin} in 6% of centers, and {cis-permethrin, trans-permethrin, esfenvalerate} in 6% of centers (Table 6.2). Rules of the form {} → {Y} indicate % occurrence of Y.

To harvest the remaining co-occurrence patterns, the minimum support was lowered from 5% (8/168) to 3/167. To set minimum support lower than 3/167 would be to rely on 1 or 2 counts. At minimum support of 3/167, the same actors are present, yet now with additional pyrethroids: cis-/trans-allethrin, bifenthrin, deltamethrin, pyrethrin II, sumithrin, tetramethrin, and resmethrin. Apart from bifenthrin and deltamethrin, these additional pyrethroids were found to only co-occur with cis-/trans-permethrin. Among these, cis-/trans-allethrin exhibited the strongest association with cis-/trans-permethrin. Figure 6.5 provides a graphical interpretation for the top 50 of 197 rules, sorted by lift. More edges are observed with the previously identified pyrethroids (trans-permethrin, cis-permethrin, cypermethrin, esfenvalerate, cyhalothrin, and cyfluthrin). By comparison, it is evident that cis-/trans-allethrin, bifenthrin, and deltamethrin have less linkages. On the threshold of minimum support (3/167), deltamethrin and allethrin co-occur exclusively with cyhalothrin and permethrin, respectively. These findings are

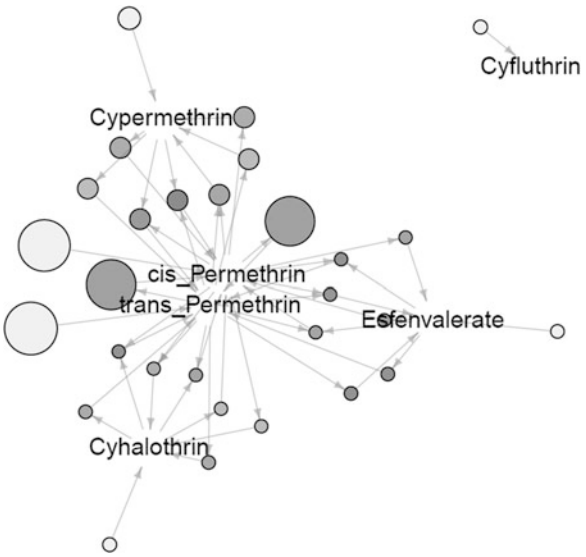

Fig. 6.4 Association rules mining of the CCC matrix. Minimum support count set to 5% (29 rules). Support (circle size) ranged from 6% to 71%. Lift (shading) ranged from 1 to 1.45. Analysis with R (R Core Team 2013) and R package arules (Hahsler et al., 2011) (Data from Tulve et al. 2006)

Table 6.2 Association rules (9 of 29) for floor wipe pyrethroid residues in the CCC[a] study (minimum support = 5%, $N = 167$[b])

Rule	Support	Confidence	Lift
{*cis*-permethrin, *trans*-permethrin} → {cypermethrin}	0.18	0.28	1.3
{cis-permethrin, trans-permethrin} → {cyhalothrin}	0.06	0.09	1.3
{cis-permethrin, trans-permethrin} → {esfenvalerate}	0.06	0.09	1.4
{ } → cis-permethrin	0.69	0.69	1.0
{ } → trans-permethrin	0.71	0.71	1.0
{ } → cypermethrin	0.21	0.21	1.0
{ } → cyfluthrin	0.07	0.07	1.0
{ } → cyhalothrin	0.07	0.07	1.0
{ } → esfenvalerate	0.07	0.07	1.0

[a]Tulve et al. 2006 applied population weights in summary statistics, including % detect. In contrast, Tornero-Velez et al. 2012 did not apply weights. Thus, presence/absence was determined by an unweighted comparison of the floor sample loading and the corresponding limit of detection
[b]One of the 168 centers (ID = 326) had missing values for *cis*- and *tran*-permethrin, and all other pyrethroid values for this center were below the limit of detection. This center did not impact analysis in Tornero-Velez et al. 2012 but should have been recognized as a center to be removed, thus $N = 167$

determined with ease using *frequent itemset mining* but may be difficult/tedious to make otherwise. In Table 6.3, a portion of the 197 rules are highlighted. For the three rules involving deltamethrin and cyhalothrin, note that co-occurrence $S(X \cup Y) = 3/167$ is the same regardless of the directionality of the rule. On the other hand, the confidence depends on the directionality of the rule. Although bifenthrin is on the periphery in Fig. 6.5, it is involved in a four-way co-occurrence {bifenthrin, trans-permethrin, cis-permethrin, cypermethrin} with support 5/167.

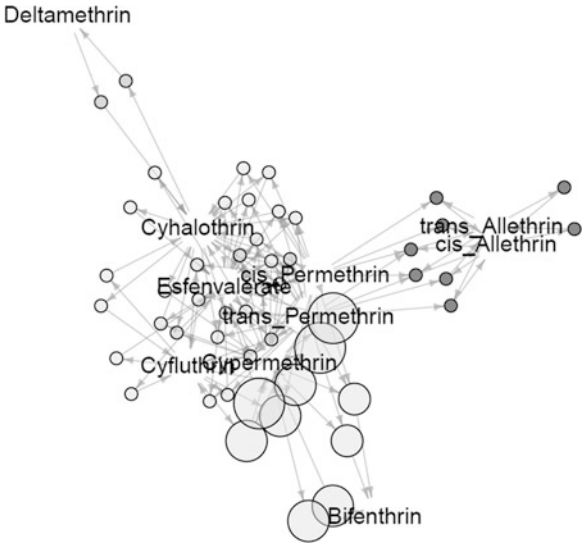

Fig. 6.5 Association rules mining of the CCC matrix. Showing for 50/197 rules with a minimum support count set to 3/167. Support (circle size) ranged from 1.8% to 4.2%. Lift (shading) ranged from 3.4 to 55.7. Analysis with R (R Core Team 2013) and R package arules (Hahsler et al. 2011) (Data from Tulve et al. 2006)

Table 6.3 Association rules (10 of 197) for floor wipe pyrethroid residues in the CCC study (minimum support = 3/167, $N = 167$)

Rule	Support	Confidence	Lift
All Rules involving deltamethrin			
{} → deltamethrin	3/167	3/167	1
{deltamethrin} → {cyhalothrin}	3/167	1.0	13.9
{cyhalothrin} → {deltamethrin}	3/167	0.25	13.9
Highlighted rules involving bifenthrin			
{} → {bifenthrin}	8/167	8/167	1
{bifenthrin} → {cypermethrin}	6/167	0.75	3.6
{cypermethrin, cis-permethrin, trans-permethrin} → {bifenthrin}	5/167	0.17	3.5
Highlighted rules involving cis-/trans-allethrin			
{cis-allethrin} → {trans-allethrin}	3/167	1.0	55.7
{cis-permethrin, trans-permethrin, cis-allethrin}→ {trans-allethrin}	3/167	1.0	55.7
Highlighted rules involving center cluster			
cypermethrin, cis-permethrin, trans-permethrin} → {esfenvalerate}	7/167	0.23	3.5
{cypermethrin, cis-permethrin, trans-permethrin} → {cyfluthrin}	6/167	0.2	2.8
{cypermethrin, cis-permethrin, trans-permethrin} → {cyhalothrin}	4/167	0.13	1.9
{esfenvalerate, cypermethrin, cis-permethrin, trans-permethrin} → {cyhalothrin}	3/167	0.43	5.9

In Table 6.3, the 3 of 20 rules associated with bifenthrin are highlighted. Importantly, rules not shown are permutations involving subsets of the superset {bifenthrin, trans-permethrin, cis-permethrin, cypermethrin}. Next, exploring the center region of Fig. 6.5, the most frequent observed four-way combinations involve the nested three-way combination {cis-permethin, trans-permethrin, cypermethrin} and either {esfenvalerate}, {cyfluthrin}, or {cyhalothrin}. Finally, the highest-order superset is a five-way {cis-permethrin, trans-permethrin, cypermethrin, esfenvalerate, cyhalothrin} with a support of 3/167. In Tornero-Velez et al. (2012), this set was observed once as a stand-alone five-way combination and embedded in the single eight-way and single nine-way, for a total of three occurrences.

6.5 A Narrative of Ecological Assembly

The toxicological profile of a single chemical, that is, its inherent properties, its mechanism of action, its potency, its absorption, and other pharmacokinetic characteristics, invariably extends to whether or not exposure is at all possible, and under which circumstances the chemical agent is presented to a living organism as an exposure. Basic toxicological parameters such as dose administered, route of exposure, and dose rate are germane to toxicology and informed by exposure. The same narrative holds for mixtures, be they PCBs, brominated flame retardants, diesel exhaust, water disinfection byproducts, or phthalates. Mixtures recognized in toxicology for testing have an accompanying exposure narrative. An important part of this narrative is *who are the actors* and *how do the actors manifest?* Most often, the chemical constituents of a tested mixture represent the target or the by-products of a chemical process or manufacturing process. By relating the actors to a specific process (e.g., chlorination, combustion, formulation), an "organizing principle" is established. Without the notion of an organizing principle, there may be reluctance to test an environmental mixture simply because it may be a spurious occurrence. If the origin of the mixture is not understood, there is doubt as to the value of the acquired information – doubt as to which population will encounter the mixture and under what circumstances. It's useful to examine the idea of "organizing principle":

> *An organizing principle is a core assumption from which everything else by proximity can derive a classification or a value. It is like a central reference point that allows all other objects to be located, often used in a conceptual map. (NCHRP 2014)*

Mixtures that originate from chemical or manufacturing processes undergo further mixing in the environment owing to their dispersal by humans and nature. Human-environment interaction is less explored as an organizing principle for mixtures, and yet there is recognition that disease may cluster by region for certain populations, motivating investigation of ecologic factors (Kellen et al. 2008; Savitz 1993; Stang et al. 2016; Holowaty et al. 2010). Non-alcohol fatty liver disease, cardiovascular disease, and developmental effects have been investigated in

northern populations by testing mixtures of various environmental contaminants unique to the Arctic, using the so-called northern contaminant mixtures (NCM) (Mailloux et al. 2014; Elabbas et al. 2011, 2014; Florian et al. 2013). Derived from concentrations of organic and inorganic chemicals found in the blood of Inuit populations during a 2004–2005 assessment, there is a supposition that inferences can be made to the Inuit population from the NCM.

Climate, rate and transport processes, and Inuit customs shape the NCM. Because the climate of the Arctic is ill-suited for agriculture and lacks plant matter that may be foraged for much of the year, the traditional Inuit diet is higher in fat and animal protein compared to the global average (Searles 2002). The fat-based food web of the Arctic favors the accumulation of persistent organic pollutants which are subject to long-range transport. This combination serves as an organizing principle to make the NCM composition an environmentally relevant mixture for the Inuit.

Clearly though, organizing principles may not be as evident for human populations engaged in an array of modern life activities comprising consumerism and interaction with the built environment. However, it is reasonable to assume that biogeographical and market basket techniques can be applied to more complex scenarios to elucidate new organizing principles for contaminant subsets (Bucher and Lucier 1998). By analogy, although the *Island Theory of Biogeography* was developed by ecologists to explain uneven distributions of animal species based on observations made on islands (isolated habitats), it was intended also to have explanatory power for community structure on the mainland (non-island habitats) where demarcations are less sharply defined (MacArthur and Wilson 1967). As these techniques are based on 0's and 1's, they easily accommodate the presence/absence of stressors, membership to a demographic group, socioeconomic class, age grouping, or behavior pattern (Huang et al., 2017). Consideration of such factors may be needed to elucidate an organizing principle driving an environmentally relevant mixture of chemicals.

The examples provided in this chapter illustrate the application of methods developed for understanding ecological assembly of communities using presence-absence matrix-based techniques and methods developed for understanding consumer-purchasing behavior using frequent itemset mining to investigate the nonrandom assembly of chemicals in mixtures and to identify co-occurrence patterns. Moreover, these techniques are useful for identifying those structuring factors that produce specific, environmentally relevant mixtures. Explicit knowledge of these factors, be they related to natural processes, economic drivers, or human decision-making, can facilitate the intelligent toxicity testing of mixtures, whether for high-throughput screening or more traditional in vivo evaluations. Similarly, this knowledge can inform exposure monitoring and biomonitoring studies by identifying targets for measurement to optimize monitoring strategies for collecting data of highest value in understanding chemical exposure.

Acknowledgments We are grateful to the editors, Cynthia Rider and Jane Ellen Simmons, for their guidance and constructive feedback. Special thanks to Hongtai Huang and Dustin F. Kapraun for their thoughtful comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily reflect views or policies of the U.S. Environmental Protection Agency.

References

Agrawal, R., T. Imieliński, and A. Swami. 1993. Mining association rules between sets of items in large databases. *ACM SIGMOD Record* 22 (2): 207–216.

Arnot, J.A., D. Mackay, T.F. Parkerton, R.T. Zaleski, and C.S. Warren. 2010. Multimedia modeling of human exposure to chemical substances: The roles of food web biomagnification and biotransformation. *Environmental Toxicology and Chemistry* 29 (1): 45–55.

Arnot, J.A., T.N. Brown, F. Wania, K. Breivik, and M.S. McLachlan. 2012. Prioritizing chemicals and data requirements for screening-level exposure and risk assessment. *Environ Health Persp* 120: 1565−1570.

Arnot, J.A., T.N. Brown, and F. Wania. 2014. Estimating screening-level organic chemical half-lives in humans. *Environmental Science & Technology* 48 (1): 723–730.

Atmar, W., and B. Patterson. 1993. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* 96 (3): 373–382.

Bartha, S., T. Czárán, and B. Oborny. 1995. Spatial constraints masking community assembly rules: A simulation study. *Folia Geobotanica* 30: 471–482.

Bell SM, S.W. Edwards. 2014. Building associations between markers of environmental stressors and adverse human health impacts using frequent itemset mining. In: Proc 2014 SIAM Int Conf Data Min, 551–559.

Bell, S.M., and S.W. Edwards. 2015. Identification and prioritization of relationships between environmental stressors and adverse human health impacts. *Environmental Health Perspectives* 123 (11): 1193–1199.

Brooks, D.R. 1985. Historical ecology: A new approach to studying the evolution of ecological associations. *Annals of the Missouri Botanical Garden* 72: 660–680.

Bucher, J.R., and G. Lucier. 1998. Current approaches toward chemical mixture studies at the National Institute of Environmental Health Sciences and the U.S. National Toxicology Program. *Environmental Health Perspectives* 106 (Suppl 6): 1295–1298. Review.

Cherta, L., T. Portolés, E. Pitarch, J. Beltran, F.J. López, C. Calatayud, et al. 2015. Analytical strategy based on the combination of gas chromatography coupled to time-of-flight and hybrid quadrupole time-of-flight mass analyzers for non-target analysis in food packaging. *Food Chemistry* 188: 301–308.

Connor, E.F., and E.D. McCoy. 1979. The statistics and biology of the species-area relationship. *The American Naturalist* 113 (6): 791–833.

Connor, E.F., and D. Simberloff. 1979. The assembly of species communities: Chance or competition? *Ecology* 60: 1132–1140.

Connor, E.F., M.D. Collins, and D. Simberloff. 2013. The checkered history of checkerboard distributions. *Ecology* 94 (11): 2403–2414.

Croley, T.R., K.D. White, J.H. Callahan, and S.M. Musser. 2012. The chromatographic role in high resolution mass spectrometry for non-targeted analysis. *Journal of the American Society for Mass Spectrometry* 23 (9): 1569–1578.

Diamond, J.M. 1975. Assembly of species communities. In *Ecology and evolution of communities*, ed. M.L. Cody and J.M. Diamond, 342–444. Cambridge, Massachusetts: Belknap Press.

Diamond, J.M., and M.E. Gilpin. 1982. Examination of the "null" model of Connor and Simberloff for species co-occurrences on islands. *Oecologia* 52 (1): 64–74.

Egeghy, P.P., R. Judson, S. Gangwal, S. Mosher, D. Smith, J. Vail, et al. 2012. The exposure data landscape for manufactured chemicals. *Sci Total Environ* 414: 159–166.

Elabbas, L.E., M.A. Finnilä, M. Herlin, N. Stern, C. Trossvik, W.J. Bowers, J. Nakai, J. Tuukkanen, R.A. Heimeier, A. Åkesson, and H. Håkansson. 2011. Perinatal exposure to environmental contaminants detected in Canadian Arctic human populations changes bone geometry and biomechanical properties in rat offspring. *Journal of Toxicology and Environmental Health, Part A* 74 (19): 1304–1318.

Elabbas, L.E., J. Esteban, X. Barber, G. Hamscher, H. Nau, W.J. Bowers, J.S. Nakai, M. Herlin, A. Åkesson, M. Viluksela, and D. Borg. 2014. In utero and lactational exposure to a mixture of environmental contaminants detected in Canadian Arctic human populations alters retinoid levels in rat offspring with low margins of exposure. *Journal of Toxicology and Environmental Health, Part A* 77 (5): 223–245.

Fletcher T, Savitz D, Steenland K. 2005. C8 science panel. http://www.c8sciencepanel.org/index.html. Accessed 20 January 2018.

Florian, M., J. Yan, S. Ulhaq, M. Coughlan, M. Laziyan, W. Willmore, and X. Jin. 2013. Northern contaminant mixtures induced morphological and functional changes in human coronary artery endothelial cells under culture conditions typifying high fat/sugar diet and ethanol exposure. *Toxicology* 313 (2): 103–112.

França, F.G.R., and A.F.B. Araújo. 2007. Are there co-occurrence patterns that structure snake communities in central Brazil? *Brazilian Journal of Biology* 67 (1): 33–40.

Gennings, C., W.H. Carter Jr., R.A. Carchman, L.K. Teuschler, J.E. Simmons, and E.W. Carney. 2005. A unifying concept for assessing toxicological interactions: Changes in slope. *Toxicological Sciences* 88 (2): 287–297.

Gilpin, M.E., and J.M. Diamond. 1982. Factors contributing to non-randomness in species co-occurrences on islands. *Oecologia* 52 (1): 75–84.

Gotelli, N.J. 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81: 2606–2621.

Gotelli, N.J., and L.G. Abele. 1982. Statistical distributions of West Indian land bird families. *Journal of Biogeography* 9: 421–435.

Gotelli, N.J., and A.M. Ellison. 2002. Assembly rules for New England ant assemblages. *Oikos* 99 (3): 591–599.

Gotelli, N.J., and G.R. Graves. 1996. *Null models in ecology*. Washington, DC: Smithsonian Institution Press. http://vertebrates.si.edu/birds/birds_staff_pages/Garygraves_Nullmodelsbook.html. Accessed 20 January 2018.

Gotelli, N.J., and K. Rohde. 2002. Co-occurrence of ectoparasites of marine fishes: A null model analysis. *Ecology Letters* 5 (1): 86–94.

Götzenberger, L., F. de Bello, K.A. Bråthen, J. Davison, A. Dubuis, A. Guisan, J. Lepš, R. Lindborg, M. Moora, M. Pärtel, and L. Pellissier. 2012. Ecological assembly rules in plant communities—approaches, patterns and prospects. *Biological Reviews* 87 (1): 111–127.

Hahsler, M., B. Grün, and K. Hornik. 2005. arules - A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software* 14 (15).

Hahsler, M., and K. Hornik. 2007. Building on the arules infrastructure for analyzing transaction data with R. In *Advances in Data Analysis*, 449–456. Heidelberg: Springer Berlin.

Hahsler, M., S. Chelluboina, K. Hornik, and C. Buchta. 2011. The arules R-package ecosystem: analyzing interesting patterns from large transaction data sets. *Journal of Machine Learning Research* 12 (Jun): 2021–2025.

Hawley, A.H. 1986. *Human ecology: A theoretical essay*. Chicago: The University of Chicago Press.

Hertz-Picciotto, I., and L. Delwiche. 2009. The rise in autism and the role of age at diagnosis. *Epidemiology* 20: 84–90.

Holowaty, E.J., T.A. Norwood, S. Wanigaratne, J.J. Abellan, and L. Beale. 2010. Feasibility and utility of mapping disease risk at the neighbourhood level within a Canadian public health unit: An ecological study. *International Journal of Health Geographics* 10 (9): 21.

Horner-Devine, M.C., J.M. Silver, M.A. Leibold, B.J. Bohannan, R.K. Colwell, J.A. Fuhrman, J.L. Green, C.R. Kuske, J.B. Martiny, G. Muyzer, and L. Øvreås. 2007. A comparison of taxon co-occurrence patterns for macro-and microorganisms. *Ecology* 88 (6): 1345–1353.

Huang, H., R. Tornero-Velez, and T.M. Barzyk. 2017. Associations between socio-demographic characteristics and chemical concentrations contributing to cumulative exposures in the United States. *Journal of Exposure Science & Environmental Epidemiology* 27 (6): 544–550.

Hu, Q., R.J. Noll, H. Li, A. Makarov, M. Hardman, and R. Graham Cooks. 2005. The Orbitrap: A new mass spectrometer. *Journal of Mass Spectrometry* 40 (4): 430–443.

Jackson, D.A., K.M. Somers, and H.H. Harvey. 1992. Null models and fish communities: Evidence of nonrandom patterns. *American Naturalist* 139: 930–951.

Jayjock, M.A., C.F. Chaisson, C.A. Franklin, S. Arnold, and P.S. Price. 2009. Using publicly available information to create exposure and risk-based ranking of chemicals used in the workplace and consumer products. *Journal of Exposure Science & Environmental Epidemiology* 19 (5): 515–524.

Judson, R., A. Richard, D.J. Dix, K. Houck, M. Martin, R. Kavlock, et al. 2009. The toxicity data landscape for environmental chemicals. *Environmental Health Perspectives* 117 (5): 685–695.

Kapraun, D.F., J.F. Wambaugh, C.L. Ring, R. Tornero-Velez, and R.W. Setzer. 2017. A method for identifying prevalent chemical combinations in the U.S. population. *Environmental Health Perspectives* 125 (8): 087017.

Kavlock, R., K. Chandler, K. Houck, S. Hunter, R. Judson, N. Kleinstreuer, T. Knudsen, M. Martin, S. Padilla, D. Reif, A. Richard, D. Rotroff, N. Sipes, and D. Dix. 2012. Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chemical Research in Toxicology* 25 (7): 1287–1302.

Kellen, E., M.P. Zeegers, L. Bruckers, and F. Buntinx. 2008. The investigation of a geographical cluster of bladder cancer. *Acta Clinica Belgica* 63 (5): 313–320.

Kelt, D.A., M.L. Taper, and P.L. Meserve. 1995. Assessing the impact of competition on community assembly: A case study using small mammals. *Ecology* 76: 1283–1296.

Krüger, R.F., C.J. de Carvalho, and P.B. Ribeiro. 2010. Assembly rules in muscid fly assemblages in the grasslands biome of southern Brazil. *Neotropical Entomology* 39 (3): 345–353.

Lomolino, M.V. 1996. Investigating causality of nestedness of insular communities: Selective immigrations or extinctions? *Journal of Biogeography* 23 (5): 699–703.

MacArthur, R.H., and E.O. Wilson. 1967. *The theory of island biogeography*. Princeton: Princeton University Press.

Mackay, D., D.M. Hughes, M.L. Romano, and M. Bonnell. 2014. The role of persistence in chemical evaluations. *Integrated Environmental Assessment and Management* 10: 588–594.

Mailloux, R.J., M. Florian, Q. Chen, J. Yan, I. Petrov, M.C. Coughlan, M. Laziyan, D. Caldwell, M. Lalande, D. Patry, C. Gagnon, K. Sarafin, J. Truong, H.M. Chan, N. Ratnayake, N. Li, W.G. Willmore, and X. Jin. 2014. Exposure to a Northern Contaminant Mixture (NCM) alters hepatic energy and lipid metabolism exacerbating hepatic steatosis in obese JCR rats. *PLoS One* 9 (9): e106832.

Manly, B.F. 1995. A note on the analysis of species co-occurrences. *Ecology* 76: 1109–1115.

McKenzie, R.D. 1924. The Ecological approach to the study of the human community. *American Journal of Sociology.* 30 (3): 287–301.

McLachlan, M.S., A. Kierkegaard, M. Radke, A. Sobek, A. Malmvärn, T. Alsberg, et al. 2014. Using model-based screening to help discover unknown environmental contaminants. *Environmental Science & Technology* 48: 7264–7271.

Meek, M., A.R. Boobis, K.M. Crofton, G. Heinemeyer, M. Van Raaij, and C. Vickers. 2011. Risk assessment of combined exposure to multiple chemicals: A WHO/IPCS framework. *Regulatory Toxicology and Pharmacology* 60: S1–S14.

Meeker, J.D. 2012. Exposure to environmental endocrine disruptors and child development. *Archives of Pediatrics & Adolescent Medicine* 166: 952–958.

Miklós, I., and J. Podani. 2004. Randomization of presence-absence matrices: Comments and new algorithms. *Ecology* 85 (1): 86–92.

Mitchell, J., J.A. Arnot, O. Jolliet, P.G. Georgopoulos, S. Isukapalli, S. Dasgupta, et al. 2013. Comparison of modeling approaches to prioritize chemicals based on estimates of exposure and exposure potential. *Science of the Total Environment* 458–460: 555–567.

Muir, D.C., and P.H. Howard. 2006. Are there other persistent organic pollutants? A challenge for environmental chemists. *Environmental Science & Technology* 40: 7157–7166.

Mumford, L. 1938. *The culture of cities*. Orlando: Harcourt Brace & Company.

National Cooperative Highway Research Program (NCHRP). 2014. Strategic Issues Facing Transportation (Volume 4: Sustainability as an Organizing Principle). In *NCHRP Report 750*. Washington, DC: Transportation Research Board of the National Academies.

National Health and Nutrition Examination Survey (NHANES). 2015a. http://www.cdc.gov/nchs/tutorials/environmental/orientation/overview/index.htm. Accessed 20 January 2018.

———. 2015b. http://www.cdc.gov/nchs/nhanes/new_nhanes.htm. Accessed 20 January 2018.

National Research Council (NRC). 2000. *Waste incineration and public health*. Washington, DC: National Academy Press.

Patel, C.J., J. Bhattacharya, and A.J. Butte. 2010. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PlosS one* 5 (5): e10746.

Patterson, B.D. 1990. On the temporal development of nested subset patterns of species composition. *Oikos* 59: 330–342.

Peng, Y.Q., Z.B. Duan, D.R. Yang, and J.Y. Rasplus. 2008. Co-occurrence of two Eupristina species on Ficus altissima in Xishuangbanna, SW China. *Symbiosis (Rehovot)* 45 (1): 9.

Perrin, J.M., S.R. Bloom, and S.L. Gortmaker. 2007. The increase of childhood chronic conditions in the United States. *JAMA* 297 (24): 2755–2759.

Pielou, D.P., and E.C. Pielou. 1968. Association among species of infrequent occurrence: the insect and spider fauna of Polyporus betulinus (Bulliard) Fries. *Journal of Theoretical Biology* 21 (2): 202–216.

Qian, H., M. Chen, K.M. Kransler, and R.T. Zaleski. 2015. Assessment of chemical coexposure patterns based upon phthalate biomonitoring data within the 2007/2008 national health and nutrition examination survey. *Journal of Exposure Science and Environmental Epidemiology* 25: 249–255.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rodríguez-Gironés, M.A., and L. Santamaría. 2006. A new algorithm to calculate the nestedness temperature of presence–absence matrices. *Journal of Biogeography* 33 (5): 924–935.

Savitz, D.A. 1993. Epidemiologic studies of electric and magnetic fields and cancer: Strategies for extending knowledge. *Environmental Health Perspectives* 101 (Suppl 4): 83–91.

Schwarzman, M.R., and M.P. Wilson. 2009. Science and regulation. New science for chemicals policy. *Science* 326: 1065–1066.

Searles, E. 2002. Food and the making of modern inuit identities. *Food & Foodways: History & Culture of Human Nourishment* 10: 55–78.

Sepkoski, J.J. 1996. Competition in macroevolution: The double wedge revisited. In *Evolutionary paleobiology*, ed. D. Jablonski, D.H. Erwin, and J.H. Lipps, 211–255. Chicago, IL: Chicago University Press.

Sobus, J.R., R.S. DeWoskin, Y.M. Tan, J.D. Pleil, M.B. Phillips, B.J. George, K. Christensen, D.M. Schreinemachers, M.A. Williams, E.A. Hubal, and S.W. Edwards. 2015. Uses of NHANES biomarker Data for chemical risk assessment: Trends, challenges, and opportunities. *Environmental Health Perspectives* 123 (10): 919–927.

Stang, A., B. Kowall, C. Rusner, B. Trabert, F. Bray, J. Schüz, K.A. McGlynn, and O. Kuss. 2016. A novel method for identifying settings for well-motivated ecologic studies of cancer. *International Journal of Cancer* 138 (8): 1887–1893.

Tice, R.R., C.P. Austin, R.J. Kavlock, and J.R. Bucher. 2013. Improving the human hazardcharacterization of chemicals: A Tox21 update. *Environmental Health Perspectives* 121 (7): 756–765.

Tornero-Velez, R., P.P. Egeghy, and E.A. Cohen Hubal. 2012. Biogeographical analysis of chemical co-occurrence data to identify priorities for mixtures research. *Risk Analysis* 32 (2): 224–236.

Tulve, N.S., P.A. Jones, M.G. Nishioka, R.C. Fortmann, C.W. Croghan, J.Y. Zhou, A. Fraser, C. Cavel, and W. Friedman. 2006. Pesticide measurements from the first national environmental health survey of child care centers using a multi-residue GC/MS analysis method. *Environmental Science & Technology* 40 (20): 6269–6274.

Ulrich, W., and N.J. Gotelli. 2013. Pattern detection in null model analysis. *Oikos* 122: 2–18.

Urban Land Institute. 2013. *Intersections: Health and the built environment*. Washington, D.C.: Urban Land Institute.

U.S. EPA. 2014. T.S.C.A. Chemical substance inventory. Available: https://www.epa.gov/tsca-inventory. Accessed 20 January 2018.

Wambaugh, J.F., A. Wang, K.L. Dionisio, A. Frame, P. Egeghy, R. Judson, et al. 2014. High throughput heuristics for prioritizing human exposure to environmental chemicals. *Environmental Science & Technology* 48: 12760–12767.

Weschler, C.J. 2009. Changes in indoor pollutants since the 1950s. *Atmospheric Environment* 43: 153–169.

Wilson, M.P., and M.R. Schwarzman. 2009. Toward a new U.S. chemicals policy: Rebuilding the foundation to advance new science, green chemistry, and environmental health. *Environmental Health Perspectives* 117 (8): 1202–1209.

Yorita Christensen, K.L., C.K. Carrico, A.J. Sanyal, and C. Gennings. 2013. Multiple classes of environmental chemicals are associated with liver disease: Nhanes 2003-2004. *International Journal of Hygiene and Environmental Health* 216: 703–709.

Chapter 7
Adverse Outcome Pathways to Support the Assessment of Chemical Mixtures

Mark D. Nelms, Jane Ellen Simmons, and Stephen W. Edwards

Abstract Due to the ever-increasing number of chemicals coming to market, and the cost of performing traditional in vivo studies, there has been a shift toward the use of less costly alternative techniques. The adverse outcome pathway (AOP) concept has emerged as a scaffold for organizing mechanistic information from these methods. Two main elements – key events (KEs) and key event relationships (KERs) – are utilized to describe the underlying mechanism outlined by the AOP. Each KE depicts the measureable changes in the state of the biological system at each level of organization that are essential for the progression along the pathway. The KERs, meanwhile, contain the biological information that connects each of the KEs. This chapter covers some of the potential applications for AOPs when performing risk assessment of chemical mixtures. The structure of the AOP provides much more precision when considering mechanistic data in a mixtures assessment. The use of this concept provides a means to allow more specificity when deciding whether to use dose addition, independent action or integrated addition risk assessment methodologies. Furthermore, AOPs enable novel approaches for determining chemical groups and how they may be utilized within mixtures risk assessment.

Disclaimer The views expressed in this chapter are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

M. D. Nelms
Oak Ridge Institute for Science and Education, Oak Ridge, TN, USA

Integrated Systems Toxicology Division, National Health and Environmental Effects Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

J. E. Simmons (✉) · S. W. Edwards
Integrated Systems Toxicology Division, National Health and Environmental Effects Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA
e-mail: simmons.jane@epa.gov

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_7

Keywords Chemical grouping · Adverse outcome pathway · Key event · Key event relationship · Molecular initiating event · Dose additivity · Independent action · AOP network

Abbreviations

ADME	Absorption, distribution, metabolism, and elimination
AO	Adverse outcome
AOP	Adverse outcome pathway
KE	Key event
KER	Key event relationship
MIE	Molecular initiating event

7.1 Introduction

One goal of research to support hazard/risk assessment is to generate data upon which to make an informed decision concerning a chemical's (potential) toxicity. Traditionally, risk assessments have been performed utilizing data from in vivo methods to identify the apical adverse effect(s) observed upon exposure to a chemical. These in vivo tests are expensive to perform not only in terms of their monetary cost but also with respect to the amount of time and the number of animals required. For example, a single 90-day oral repeat dose rodent study can cost an estimated €116,000 (~$125,000) and requires approximately 80 animals (10 animals/sex/group and at least four dose groups, including a control) per chemical (Taylor et al. 2014; Fleischer 2007). Thus, this limits the number of hazard assessment studies that can be conducted each year for any novel, or commercially available, chemicals.

In 2007, the National Research Council of the U.S. National Academies of Sciences published a report entitled *Toxicity Testing in the 21st Century: A Vision and a Strategy*, in which a paradigm shift in toxicity testing was suggested (NRC 2007). This shift was away from expensive in vivo tests that typically identify apical effects toward a testing strategy that utilizes various alternative techniques that are time- and cost-effective and rely on an understanding of the underlying toxicity pathway(s). A toxicity pathway is defined by the NRC as "a cellular response pathway that would result in an adverse health effect when sufficiently perturbed" (NRC 2007). In addition to the 2007 NRC report, legislation has been passed and a number of initiatives implemented, in both the United States and Europe that suggest advancing science through development and application of alternative toxicity testing data (EC 2003, 2006a, b). These programs act to either directly (i.e., the 7th Amendment to the Cosmetics Directive) or indirectly (i.e., REACH) promote the development and use of alternative testing methods to assess the hazard/risk of the ever-increasing number of chemicals introduced into commerce. In addition, the increase in animal welfare campaigns that call for in vivo testing to occur only as a last resort has

provided extra impetus for development of alternative testing methods, especially for the testing of cosmetic ingredients (Stoddart and Brown 2014).

Since the 2007 NRC report, there has been an increased effort to develop alternative methods for chemical hazard/risk assessment. These alternative methods employ in vitro, in chemico, in silico, and "omics" techniques. In vitro-based methods, such as the Ames test, use isolated biological organisms (e.g., bacteria, yeast), cell lines, or subcellular components to identify the effects of xenobiotics on the biological system. In chemico-based methods, such as the glutathione assay, are abiotic assays that use macromolecules (e.g., proteins/DNA) to measure the level of reactivity between a chemical and a surrogate biological macromolecule (Gerberick et al. 2008; Asturiol and Worth 2011). In silico-based methods, such as (Quantitative) Structure-Activity Relationships ((Q)SARs), utilize computational systems to enable predictions regarding various end points based on chemical structure. Meanwhile, "omics" approaches utilize multiple technologies, such as microarray, mass spectrometry, or chromatography, which may be able to detect changes across all of the genes, mRNA, proteins, or metabolites present within a biological sample. However, these types of data require an organizational scaffold to provide the level of mechanistic detail needed to link the data to traditional regulatory end points. To address this problem, the adverse outcome pathway (AOP) concept was introduced (Ankley et al. 2010). The intention behind the AOP concept is to utilize the multitude of mechanistic data that are generated by these alternative techniques in conjunction with advances in systems biology and bioinformatics (Garcia-Reyero 2015). As such, AOPs provide a framework for establishing a mechanistic connection between a molecular initiating event and a downstream adverse outcome relevant for risk assessment. This is accomplished via a number of measurable intermediate key events that traverse different levels of biological organization (Fig. 7.1) (Villeneuve et al. 2014a; Ankley et al. 2010). Ideally, an individual AOP should have a single molecular initiating event associated with a single adverse outcome, while the amount of intermediate key events required to link these together will vary based upon the number of biological processes that may be disrupted (Villeneuve et al. 2014a; OECD 2014). An exception to this rule, however, is when multiple late-stage key events could all be considered adverse under different decision contexts. Furthermore, it is important to note that a molecular initiating event, intermediate key event, and/or an adverse outcome may be shared by multiple AOPs. Additionally, a single molecular initiating event may be capable of inducing multiple adverse outcomes, and a single adverse outcome may be induced by multiple molecular initiating events. The latter scenario is important when considering chemical mixtures, as described later in this chapter. This process – i.e., the identification of measurable key events that describe the mechanistic pathway leading to an adverse effect – is similar to the Mode of Action (MoA) framework that was developed by the World Health Organization's International Programme on Chemical Safety for use in human health hazard/risk assessment (Meek et al. 2014). There is, however, at least one subtle difference between the MoA and AOP frameworks: the primary purpose behind their identification and development. MoA-based pathways are developed for individual chemicals to perform regulatory

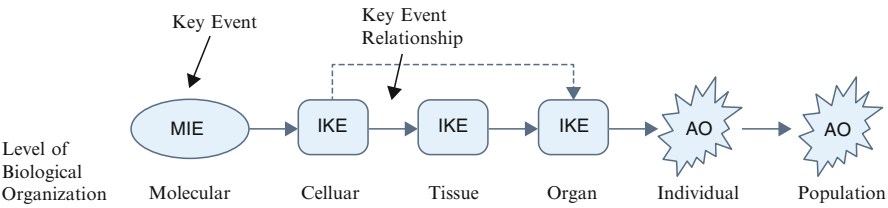

Fig. 7.1 Diagram of the components that comprise the AOP concept. *MIE* molecular initiating event, *IKE* intermediate key event, *AO* adverse outcome, and arrows – key event relationship. The direct key event relationships are illustrated by arrows with solid lines, while indirect key event relationships are illustrated by arrows with dotted lines

risk assessments. In contrast, AOPs are chemical agnostic pathways that describe the biological processes that are perturbed leading to an adverse outcome of regulatory significance (Edwards et al. 2016).

AOPs are typically depicted as linear constructs with one molecular initiating event linked to one adverse outcome via an unbranched set of intermediate key events (Fig. 7.1). However, this is an oversimplification used to better enable the development and evaluation of individual AOPs. Therefore, to be of most use when making regulatory decisions, and to better encapsulate the inherent complexities of toxicological processes within biological systems, it is likely that multiple AOPs will need to be used (Villeneuve et al. 2014a; Knapen et al. 2015). This is due to the potential for one-to-many, and many-to-one, relationships being present throughout an AOP. For example, one chemical may perturb many molecular initiating events, or many upstream events may induce a single downstream event. When several AOPs share at least one common key event, or key event relationship, an AOP network may be produced. Thus, these AOP networks provide a more realistic representation as to the underlying processes that are perturbed between a (set of) given molecular initiating event(s) and a (set of) given adverse outcome (s) (Villeneuve et al. 2014a; Knapen et al. 2015). An example of such an AOP network is depicted in Fig. 7.2, whereby the gradient-filled nodes highlight the initial key event that connects two individual AOPs together (Knapen et al. 2015).

7.1.1 Constituent Components of an AOP

There are two major components required to describe an AOP: key events (KE) and key event relationships (KER). Within the AOP diagram (Fig. 7.1), nodes depict the KEs, while the KERs are depicted by edges (i.e., visual representation of relation among nodes – represented by arrows in Fig. 7.1) that connect nodes together.

Each KE within an AOP will fall into one of three categories: a molecular initiating event (MIE), an intermediate key event, or an adverse outcome (AO) (Villeneuve et al. 2014b); each of which is discussed in more detail below. Overall, these KEs represent a change in biological state at varying levels of organization from the molecular level

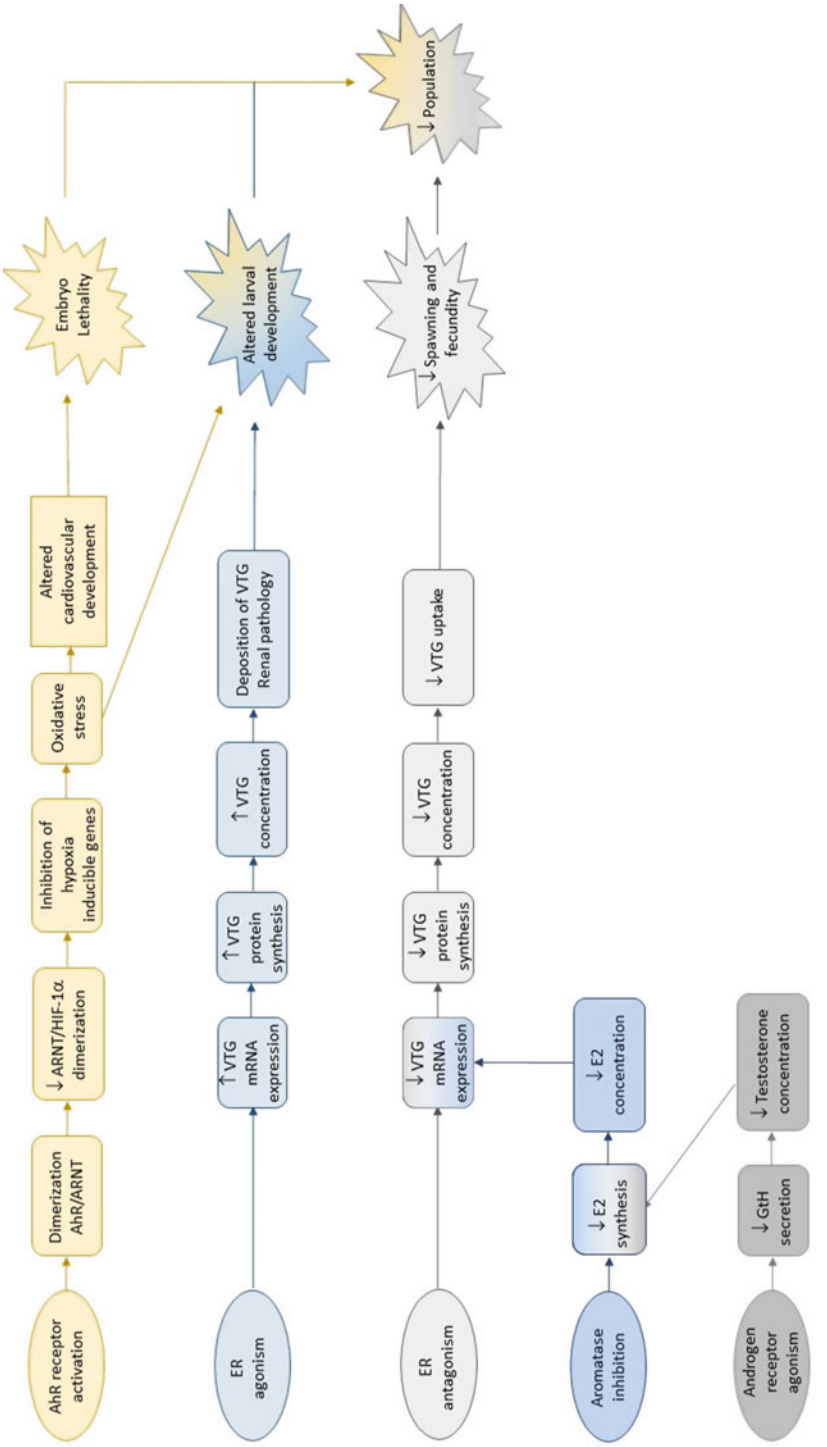

Fig. 7.2 Example AOP network adapted from Knapen et al. 2015 pertaining to the AOPs for reproductive and developmental toxicities in fish present in the AOP-Wiki. Nodes that contain the colors of two AOPs represent the initial common key events that generate the AOP network

to effects at the individual or population levels. Additionally, an event is considered "key" if (1) the change in state can be measured by use of one, or more, testing strategies and (2) the event is essential for the advancement along the pathway toward the AO (Villeneuve et al. 2014b; Groh et al. 2015a). However, even though KEs are essential in the progression of an AOP, the observation of a single KE, in isolation, does not necessarily mean the AO will be observed; it merely implies the adjacent downstream KE has the potential to be perturbed.

7.1.2 Molecular Initiating Event

Perturbation of the MIE is a prerequisite for the initiation of an AOP; without the perturbation of the MIE, the subsequent downstream KEs cannot commence. In the context of AOPs, a perturbation is any change in state of the normal biological processes occurring within an organism due to the interaction with an exogenous stressor. For example, either agonism or antagonism of a receptor by a chemical would be considered a perturbation. Within the AOP framework, the MIE is the first event that triggers the progression of the pathway toward the AO; as such, it also is seen as the upstream anchor within the AOP (Ankley et al. 2010; Villeneuve et al. 2014a). The MIE is unique compared to the other KEs in that it is the only point in the AOP where a chemical or stressor has a direct interaction with a biological (macro)molecule (Villeneuve et al. 2014b). Therefore, knowledge regarding the MIE provides mechanistic information pertaining to the initial interaction between the chemical/stressor and a biological system at the molecular level. Developing an understanding of the mechanisms underlying the MIE can be useful in determining the structural and/or physicochemical properties required by a (group of) chemical (s) to perturb the same MIE (Przybylak and Schultz 2013; Enoch and Roberts 2013; Enoch et al. 2013). For example, chemicals capable of inducing mitochondrial toxicity via uncoupling of oxidative phosphorylation (i.e., the futile translocation of protons across the inner mitochondrial membrane with no ATP production) are typically lipophilic (log P 1.5–5), weak acids (pKa 3–6) with an aromatic moiety capable of stabilizing a delocalized electron. Local anesthetics and nitroaromatic chemicals would fit within this mechanism-based chemical category (Kadenbach 2003; Nelms et al. 2015a, b; Spycher et al. 2008; Naven et al. 2013; OECD 2011; Przybylak and Schultz 2013).

A number of different MIEs have been identified as trigger points in a variety of AOPs, including reproductive toxicity, skin sensitization, cholestasis, weak acid respiratory uncoupling, and neural toxicity (Ankley et al. 2010; Schultz 2010; Landesmann et al. 2012; OECD 2011; Vinken et al. 2013; OECD 2012). A common type of MIE is (ant)agonistic binding of chemicals to specific molecular receptors or enzymes. For example, agonistic binding of chemicals to the androgen receptor induced reproductive toxicity in female fish (Ankley et al. 2010) (https://aopwiki.org/aops/23, accessed January 6th, 2018). Alternatively, there are MIEs that perturb less specific biomolecules; such MIEs include those constrained by the electrophilic

reactivity of the chemical. One example of this is the covalent interaction between the perturbant and the cysteine/lysine residues present in the proteins of epidermal cells within the skin sensitization AOP (OECD 2012; Enoch et al. 2008; Roberts et al. 2006). However, it should be noted that a molecular interaction described as the MIE within one AOP may, potentially, be defined as an intermediate KE within another AOP if the interaction between the chemical and biological system occurs upstream (Villeneuve et al. 2014b). An example is activation of the estrogen receptor, which could be either an MIE or intermediate KE depending upon the ligand. If activation of the estrogen receptor occurs via direct binding of a xenobiotic to the receptor, it would, in this instance, be considered an MIE. In contrast, if upstream KEs initiate the release of an endogenous ligand that subsequently binds to, and activates, the estrogen receptor, it would be considered an intermediate KE (Villeneuve et al. 2014b).

7.1.3 Intermediate Key Events

Upon sufficient perturbation of the MIE, the intermediate KEs are the transitional stages by which the AOP progresses toward the AO. The preferred convention when developing an AOP is to include an intermediate key event from each level of biological organization between the molecular initiating event and the expected adverse outcome (Villeneuve et al. 2014b; OECD 2014). Obviously, due to the complex and multifaceted nature of biological systems and the incomplete mechanistic knowledge currently available for many end points, this is not always feasible. For an event to be considered an intermediate KE, it has to be (1) measureable within a testing method and (2) essential to the progression of the AOP toward the AO, i.e., without the perturbation of each intermediate KE within the AOP, the AO cannot be induced. The former is required so that hypotheses surrounding an AOP, and each of the steps within it, can be experimentally tested to verify the mechanism(s) by which the MIE may induce the AO (Przybylak and Schultz 2013). Once the AOP is established, the associated testing methods could be used during hazard identification and risk assessment to identify whether a (set of) chemical(s) has the potential to perturb each level of organization and therefore has the capability to induce the AO. Additionally, the association of an alternative testing method with an intermediate KE (that has been causally linked to the AO) will minimize the number of animal studies needed to identify the potential hazards/risks posed by a chemical.

7.1.4 Adverse Outcome

The terminal stage of an AOP is the observation of an AO. As the ultimate objective of AOPs is for use within regulatory decision-making, the downstream AO(s) relate to the apical effects observed when undertaking in vivo experimentation and should,

therefore, be relevant for performing hazard/risk assessment (OECD 2013). As such, the level of biological organization at which a KE is considered to be the AO can vary (Ankley et al. 2010; OECD 2013, 2014; Vinken et al. 2013; Villeneuve et al. 2014b). Similar to the MIE, the presence of an AO is a prerequisite of AOP development as it acts to anchor the AOP to an observable downstream adverse effect (OECD 2013; Ankley et al. 2010). As with the other KEs in an AOP, clearly defining the AO affects the upstream KEs that may be present within the AOP. For example, if skin sensitization were the regulatory end point under consideration (i.e., the AO), this would thus limit the upstream KEs to only those that culminate in the observation of a sensitized individual. In addition, clearly defining the AO under consideration drives the number and type of individual AOPs that may comprise an AOP network.

7.1.5 Key Event Relationships

The information relating to how pairs of adjoining KEs are connected is provided by the KERs, i.e., the edges (arrows) in Fig. 7.1. The KERs act to identify one KE as the upstream event and the event it induces as the downstream event, thereby providing the AOP with directional information. The primary purpose of these KERs is to document the evidence that supports these causal relationships between the upstream and downstream KEs (Villeneuve et al. 2014b). Each KER present within an AOP can describe one of two types of relationship connecting KEs, either (1) a direct connection between two adjacent KEs (solid arrows), e.g., AhR agonism inducing a change in gene expression (Ankley et al. 2010), or (2) an indirect connection between two KEs separated by one, or more, KEs (dotted arrow), e.g., sustained AhR activation inducing hepatopathy (Becker et al. 2015b; OECD 2014; Villeneuve et al. 2014a). The indirect KER can allow evidence supporting nonadjacent KEs to be included. This is important because in many cases the majority of the evidence for an AOP will be concentrated on a subset of the KEs. By including these indirect relationships, one may achieve a higher degree of confidence in the AOP even with more limited support for individual KE pairs. For example, the amount and support present for an indirect KER linking KE_n and KE_{n+2} (dotted arrow in Fig. 7.1) within an AOP may be much higher than the confidence in the direct KER(s) linking KE_n to KE_{n+1} and/or KE_{n+1} to KE_{n+2}. This may be due to a number of factors such as the assay used to measure KE_{n+1} may be inherently more error prone than that used to measure either of the other KEs; therefore, the level of confidence in the indirect KER (i.e., between KE_n and KE_{n+2}) would be higher.

The data utilized in support of the scientific explanation linking two KEs can come from a variety of testing methods, as well as knowledge present in the available literature. Depending upon the level of understanding of the relationship between the two KEs and the type of assay(s) available for testing, these data may be quantitative, semiquantitative, or qualitative in nature. Additionally, these data are used to support, or refute, the KER by using a weight of evidence approach to assess the

biological plausibility and empirical support for the KER (Villeneuve et al. 2014a; OECD 2014; Becker et al. 2015a). In general, current understanding of the biological system's functioning under "normal" (homeostatic) conditions is used to inform the biological plausibility of the KER, i.e., under "normal" circumstances what is the relationship between the two KEs? In comparison, empirical support for the KER is typically founded upon (1) temporal concordance, i.e., observation of the upstream KE occurring prior to the downstream KE; (2) response-response concordance, i.e., a lesser, or equal, concentration of stressor is required to see perturbations in the upstream KE compared to the downstream KE; and (3) incidence concordance, i.e., the upstream KE is as, if not more, prevalent than the downstream KE. In practice, there are many cases where empirical data are inconsistent relative to these guidelines even when the KER is real. For example, the assay used to test the downstream KE may be more sensitive than that for the upstream KE. Therefore, it may identify perturbation of the downstream KE at a lower concentration of test chemical or a higher prevalence for the downstream KE even though the perturbation of that KE is mediated by the upstream KE. This is why biological plausibility carries a higher weight than empirical support when evaluating the weight of evidence. Furthermore, the KERs allow a single KE to connect to different downstream KEs based on the context of the perturbation. For example, sustained activation of the aryl hydrocarbon receptor (AhR) is required for the changes in cellular homeostasis and apoptosis that are the next step leading to rodent liver tumors (Becker et al. 2015b) (https://aopwiki.org/aops/41, accessed January 6th, 2018). Capturing this information in the KER, rather than having separate KEs for sustained vs. acute activation of the receptor, ensures that all chemicals that interact with the AhR can be considered for their potential to cause rodent liver tumors with the understanding that it would only occur under circumstances where sustained activation was present.

7.2 AOP Development

In 2012, the Organisation for Economic Co-operation and Development (OECD) introduced the AOP development program to promote the development and use of AOPs (Villeneuve et al. 2014a). The AOP-Knowledge Base (AOP-KB, https://aopkb.oecd.org/, accessed January 6th, 2018) has emerged as part of this program. The AOP-KB consists of four discrete modules designed to assist with the different aspects of the AOP development process: AOPXplorer (currently under development) will enable users to view networks of interrelated AOPs; Effectopedia (http://www.effectopedia.org/, accessed January 6th, 2018) contains quantitative KER data in addition to information pertaining to the available assays and biomarkers for each KE; and Intermediate Effects Database (IEDB) will provide information on chemicals that perturb an AOP and can be used in a regulatory setting utilizing data from Europe's International Uniform Chemical Information Database (IUCLID) database (Oki et al. 2016). The fourth module, the AOP-Wiki (http://aopwiki.org, accessed January 6th, 2018), provides a template that AOP developers

can use to capture each of the KEs within an AOP and evaluate the supporting scientific evidence for an AOP. The AOP development handbook, written by the OECD, acts as the blueprint upon which the AOP-Wiki is based (OECD 2014). In addition to the handbook, a number of articles have been published that provide further advice and guidance for developing and assessing AOPs for their reliability and robustness (Villeneuve et al. 2014a, b; Groh et al. 2015a, b; Vinken 2013).

Five core concepts have been identified to aid researchers when developing AOPs (Villeneuve et al. 2014a):

- AOPs are chemical agnostic – any chemical that is sufficiently able to perturb the MIE may start a cascade, whereby each upstream KE perturbs the neighboring downstream KE up to and including the apical AO.
- AOPs are modular – this is accomplished by ensuring the two constituent components of AOPs (i.e., the KE and KER) are not necessarily specific to a single AOP. The reusability of these building blocks provides the AOP with this modular structure, allowing for a single KE or KER to be incorporated into multiple AOPs.
- An individual AOP is a pragmatic unit of development and evaluation – the single chain of KEs organized from one MIE to one AO provides a simplified structure that assists with the development and evaluation of AOPs.
- AOP networks are the functional unit of prediction – AOP networks will build up naturally over time due to shared KEs and KERs appearing across multiple individual AOPs. Typically, organisms are exposed to complex chemical mixtures. Therefore, AOP networks better illustrate the complex biology at play, enabling more accurate predictions to be made with regard to the most likely AO.
- AOPs are living documents – the KEs and KERs are supported by data from various testing methods. Thus, as these methods develop and their accuracy increases, so too will the information for the KEs and KERs. Additionally, the AOPs that are comprised of these modular components will improve over time as our understanding of the underlying biology improves.

The first step in AOP development is assembly of the AOP scaffold by identification of KEs in the AOP, including the MIE and AO, and defining the relationships among the KEs. Several strategies have been discussed that may be employed when developing an AOP: (1) bottom-up, where data pertaining to the MIE are the starting point and mechanistic information linking the MIE to KEs at higher levels of biological organization is required; (2) middle-out, where data are available for an intermediate KE in an AOP and mechanistic information is needed to anchor the intermediate KE by connecting to both the upstream MIE and downstream AO; and (3) top-down, where data surrounding an observable AO are present; therefore, mechanistic information is needed that links the AO to the upstream KEs (Villeneuve et al. 2014a; Groh et al. 2015b). The mechanistic knowledge relating to each information block within the AOP can be derived from a variety of testing methods, including in silico techniques such as the identification of chemical structures (or fragments) that are associated with inducing an MIE, in chemico methods that can be used to measure the ability of a chemical to covalently bind to important

biological macromolecules, in vitro assays that utilize cultured bacterial/mammalian cells to observe the (sub)cellular responses initiated by a chemical, and in vivo tests that enable observations to be made for regulatory relevant end points, e.g., at the organ, individual, or population level (OECD 2013, 2014). In addition to performing one of these methods, evidence that mechanistically connects one KE to another may be available within the current literature and should be considered in support of the KEs within the AOP.

The second step utilizes the AOP scaffold set out in the first step to assemble the scientific evidence in support of each KE and KER in preparation for the evaluation step (discussed below). The AOP-Wiki can be utilized at this stage, in conjunction with the associated handbook (OECD 2014), as it provides a common platform for aggregating and displaying data that supports the inclusion of each KE and KER in the AOP. The summarized data pertaining to each KE in an AOP should (1) have a description of the KE's role under normal physiological conditions and how this is perturbed during the course of eliciting an AO and (2) have a description of the assay (s) that can be utilized to measure the KE. When performing the overall assessment of the AOP, an evaluation of the evidence regarding the essentiality of the KEs for the progression of the AOP is required. In addition, the KER descriptions should outline the weight of evidence present in the primary literature and cover (1) the biological plausibility of the connection, based upon the relationship between the two KEs under normal physiological conditions, (2) the empirical evidence that supports the association between the perturbation of the upstream KE and the subsequent initiation of the downstream KE, and (3) any uncertainty that may be present within the literature surrounding the relationship between the KEs, i.e., is there conflicting evidence present that "disputes" the connection between the KEs?

Once the evidence has been assembled, it is possible to assess the support for the AOP by systematically evaluating the evidence for each component. This information can be used as a guide to the most appropriate use of the AOP within a regulatory setting, as well as identifying any data gaps present within the AOP. Considerations based on the Bradford Hill criteria, commonly applied to the analysis of toxicological and epidemiological data, form the basis for assessing AOPs as they can help to determine the relevance of the supporting information identified during the data summation phase (OECD 2014; Przybylak and Schultz 2013; Becker et al. 2015a; Meek et al. 2014; Bradford Hill 1965). Together these considerations enable evaluations to be made about the biological plausibility and empirical support for each individual KER, as well as the essentiality of each KE, present in the AOP (OECD 2014). Assessment of KER biological plausibility uses knowledge surrounding the biological processes in question to ascertain whether the KER is credible, i.e., is there mechanistic evidence supporting the relationship between the upstream and downstream KEs?

Assessing the empirical support for each KER within an AOP requires information regarding the response-response (change in the response for the downstream KE based upon a change in response for the upstream KE), temporal, and incidence concordance. Such an assessment enables an evaluation to determine whether a change in an upstream KE is succeeded by an appropriate change in the downstream

KE. Importantly, while dose-response, time-course, and incidence data from specific chemicals are used to evaluate the KERs, the comparison is between the response of the upstream KE and the response of the downstream KE. The KER defines a causal connection between the upstream and downstream KEs. Therefore, the perturbation of the downstream KE by the chemical is mediated by the upstream KE except for the MIE. If this is true, the perturbation of the upstream KE should be seen at lower doses, earlier times, and with increased incidence compared with the downstream KE. In practice, this is not always true due to technical considerations with the methods used to measure the perturbations of the KEs, so care must be taken when assembling the weight of evidence based on the empirical support.

The assessments into essentiality of each KE within an AOP use information from studies that block the perturbation of an upstream KE to determine if there is a concomitant inhibition/reduction in the downstream KEs. The assembled evidence also allows the domain of applicability to be determined for the AOP and its components. Typically, the applicability domain is restricted to only the species, life stages, and sexes of the organism(s) exhibiting the relevant AO. However, in cases where the AOP has been developed to assess human health end points, information from other species may be used if there is concordance among the species, i.e., the underlying mechanism is conserved across taxa.

When developing an AOP, it is important to work through each of these steps in turn from the identification of the KEs and KERs relevant for the AOP, to assembling the supporting evidence for each of the KE/KERs and culminating in the evaluation of the collected evidence. In cases where quantitative precision is needed to predict adverse outcomes at the individual or population level based upon assays for KEs, the development of a quantitative AOP (qAOP) is needed. This requires quantitative response-response information for each pair of KEs so that the level of change in the upstream KE required to induce the downstream KE can be elucidated (Villeneuve et al. 2014a). Subsequently, this quantitative information can be used to identify and develop in silico models (e.g., Quantitative Structure-Activity Relationships) that may be used to predict the likelihood that a chemical would initiate the AOP via direct perturbation of the MIE, thereby enabling a prediction to be made as to whether the AO will be observed given a measured change in an assay or biomarker for an early KE.

7.3 Use of AOPs

An AOP may lie anywhere along a broad continuum with respect to its completeness and associated uncertainty, which affects the applications for that AOP in a hazard/risk assessment setting. For example, when used for screening and prioritization purposes, an AOP with limited confidence may still provide valuable information when prioritizing based on high-throughput toxicity assays or predictions based on chemical structure. In contrast, to use an AOP as the basis for replacing an established in vivo test with (likely multiple) high-throughput in vitro tests would

require a higher level of confidence in the AOP itself. However, since confidence is measured for each KER within the AOP as well, the overall confidence in the AOP may not necessarily matter for use. For example, an AOP with a relatively low confidence score (due to a high degree of uncertainty at downstream KEs/KERs) may still be of use for performing read-across or prioritization if there is a high level of confidence in the upstream KE(R)s and their associated assays (Becker et al. 2015a).

One of the broadest applications of the AOP framework is for use in chemical category formation and subsequent read-across analysis. Chemical category formation is an approach whereby a set of chemicals with similar properties are grouped together into a category (Enoch 2010; Enoch and Roberts 2013; ECHA 2009; OECD 2011; ECHA 2008). Chemical categories can be developed via a variety of approaches, one of the most powerful being a common mechanism of action, typically relating to the MIE (Enoch and Roberts 2013; Przybylak and Schultz 2013). This approach relies on identifying key structural features within a group of chemicals (e.g., substructures, fragments, chemical classes) that are associated with inducing toxicity (e.g., skin sensitization) through the same MIE (e.g., covalent protein binding). These features can then be utilized as the basis for developing mechanism-based structural alerts. Additionally, if certain toxicokinetic parameters are known to be required for a given end point (e.g., a molecular weight less than 700), these can be associated with the structural alert as a method of refinement. Thus, structural alerts can lie anywhere along a continuum from only having knowledge of a conserved structural fragment to having physicochemical parameters associated with the structural fragment that are required to be present. Where along this continuum a specific structural alert lies depends on the level of information and support that was available during its development. Once a chemical category has been developed, for example, based upon the presence of one (or more) structural alert(s), category members with existing toxicity data (called source chemicals) can be used to make predictions via interpolation/extrapolation for the target chemical (s) for which no toxicity data exists for the end point under consideration. This interpolation of missing data between source and target chemicals within a chemical category is called read-across (Enoch et al. 2013; Enoch 2010; Cronin 2013; ECHA 2009). These read-across predictions rely on the premise that similar chemicals are expected to have similar biological/chemical activities (Jaworska and Nikolova-Jeliazkova 2007; Cronin 2013; ECHA 2008). Importantly, the confidence in the information pertaining to the AOP (and the relevant KE(s)) will affect the level of confidence in the read-across predictions, i.e., a higher confidence in the connection between the MIE and AO within an AOP will lead to a higher confidence in the resulting prediction.

Profiling a large dataset or inventory to prioritize chemicals for further testing within an alternative testing method can be carried out utilizing mechanism-based structural alerts developed using AOPs (Przybylak and Schultz 2013; Enoch and Roberts 2013; Perkins et al. 2015; Gutsell and Russell 2013). Inventory screening uses a set of structural alerts relating to the same outcome (e.g., protein or DNA binding) for rapid identification of chemicals with the potential to induce toxicity. In

this instance, as further testing will be performed, nonmechanistic structural alerts (i.e., structural fragments associated with inducing toxicity but lacking mechanistic information) may be used. However, the use of mechanism-based structural alerts is preferable as these better guide the end points that should be addressed by the alternative methods.

As the level of mechanistic understanding pertaining to the KEs, and the relationships between them, increases, so too does their use within hazard/risk assessment. For example, an AOP with a moderate level of understanding and confidence may be useful in building Integrated Approaches to Testing and Assessment (IATA) (Tollefsen et al. 2014; Perkins et al. 2015; OECD 2015). In this instance, the AOP acts as the anchor point upon which chemical-specific testing data for each KE can be assembled and evaluated. Once this process is complete, an assessment can be made as to the suitability of the current data and the uncertainty associated with the AOP. If more information is required, the AOP can inform the amount and type of data that are required. Finally, an AOP that contains a high level of quantitatively defined causal relationships between the MIE, intermediate KEs, and the AO along with dose-response and absorption, distribution, metabolism, and excretion (ADME) information for the chemical in question can be utilized to perform quantitative risk assessment or predictive toxicology (Leonard et al. 2016).

While these applications can be used for individual AOPs, it is likely that most regulatory decisions will be made using AOP networks connected via common KEs (Fig. 7.2). This is not only because typical exposure scenarios are to mixtures of chemicals that may act on different biological targets to impact a common AO but also because individual chemicals have the potential to trigger multiple MIEs depending upon concentration, length of exposure, and route of administration and thereby impact more than one AO. In addition, the AOP networks allow for additional modifying factors such as genetic susceptibility or preexisting disease to be incorporated within the proper mechanistic context.

7.4 AOP-Informed Risk Assessment for Mixtures

While others have briefly mentioned the potential utilization of AOPs for chemical mixtures (Ankley et al. 2010; NRC 2007; Groh et al. 2015a, b), discussions surrounding the use of AOPs within hazard/risk assessment are limited typically to the exposure of one chemical in isolation. However, within a real-world scenario, humans are exposed continually to varying numbers of chemicals present as mixtures in the environment. These mixtures might result either from sequential exposure to individual chemicals (i.e., where the chemicals are “ingested” separately but, due to their toxicokinetic/toxicodynamic profiles, the chemicals or the biological consequences of chemical exposure or both appear within the body at the same time) or from concurrent exposure to multiple chemicals (i.e., where the chemicals are “ingested” simultaneously). In addition, chemicals can enter the body via either the same route or different routes (i.e., oral, dermal, or inhalation). Therefore, it is

important to consider how AOPs can be utilized for performing hazard/risk assessments for defined chemical mixtures. In this instance, a defined mixture is one where all of the chemical constituents present are known. Furthermore, it should be noted that in the context of this book chapter, the focus is on low doses of chemicals likely to be present within the environment.

When performing risk assessment on a chemical mixture, the ideal scenario is to have data pertaining to the mixture itself; however, these data are rarely available. In the absence of suitable data on the mixture itself, data for a sufficiently similar mixture, if available, are preferred. The main limitation of sufficient similarity is that methods to determine what constitutes a sufficiently similar mixture have not been widely applied or validated (Rider and Simmons 2015). If neither of these data types are available, risk assessment is performed using single-chemical data for chemicals within the mixture considered important or for which concentration levels and point of departure values are available, i.e., component-based analysis. Due to the scarcity of data available for mixtures, component-based analysis is used in the majority of cases when undertaking mixtures risk assessment.

Currently, one of the first stages in mixtures risk assessment is to identify when (i.e., sequential or concurrent) and how (i.e., oral, dermal, or inhalation) exposure to a chemical within the mixture occurs. Additionally, the toxicokinetic/toxicodynamic profiles of each constituent chemical need to be taken into consideration. This is because toxicokinetics will determine how persistent the chemicals are within the body based upon their respective ADME properties. In comparison, toxicodynamics will determine how persistent the biological effects (if any) are within the environment/organism. In this book chapter, chemicals are deemed to co-occur if their toxicokinetic/toxicodynamic profiles are such that they or their biological effects are present within an organism at the same time. For a screening-level assessment, co-occurrence may be sufficient for inclusion of the chemical in the mixtures risk assessment (Chap. 14). In subsequent steps, or as the initial stage, relevant toxicity data for the individual chemicals are used to subcategorize the broader grouping based on co-occurrence. In both cases, the toxicity data are used to identify the tissue(s)/organ(s) in which adverse effects are observed following exposure.

In the future, as the number and scope of AOPs increase, it is envisioned that the use of the AOP construct within mixtures risk assessment will also increase. One of the fundamental applications of the AOP framework is providing mixtures risk assessment with the mechanistic understanding of the sequence of events that occur between the MIE and AO. This sequence of events forms a scaffold onto which toxicological information from in silico models, in chemico, in vitro, and/or in vivo assays, together with information from the available literature, can be placed to help make more informed decisions. Therefore, once an AOP has been developed (and verified), it may be of use when performing mixtures risk assessment for a variety of scenarios. Here, in no particular order, we will discuss some of the mixtures risk assessment scenarios for which AOPs could be of use. For the purpose of clarity, each of the scenarios below will discuss how AOP(s), and structural alerts, for a specific end point may be used in selection of the most appropriate mixtures risk assessment method for use.

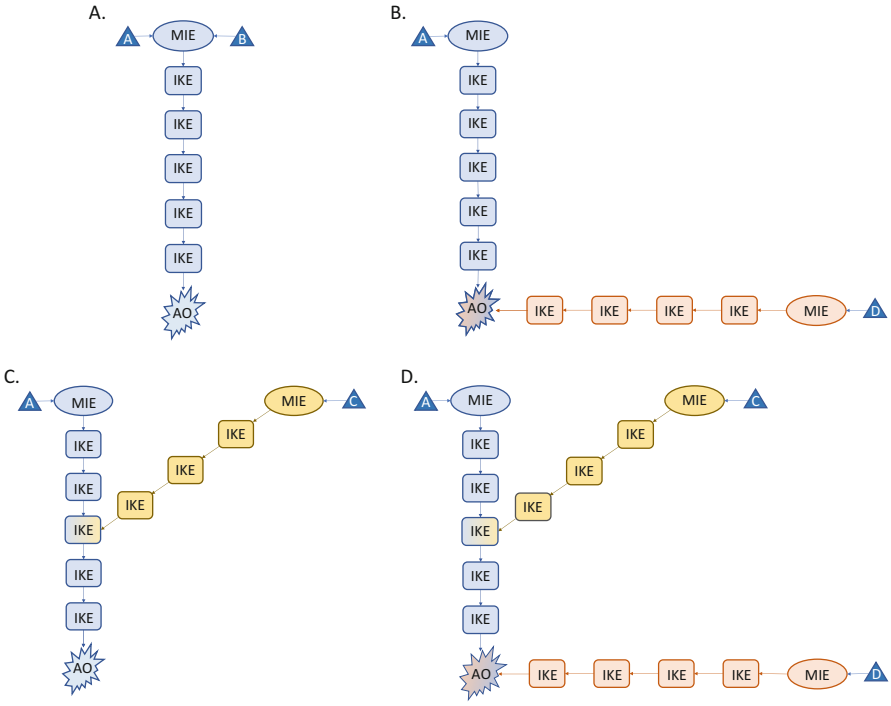

Fig. 7.3 Graphical representations of how AOP networks may be used to influence the decision as to whether dose addition, independent action, or integrated addition would be most appropriate to undertake a mixtures risk assessment depending upon the chemicals present in the mixture. (**a**) Chemicals A and B perturb the same AOP; therefore, dose addition would be an appropriate strategy to perform. (**b**) Chemicals A and D only converge at a common AO; therefore, independent action would be an appropriate strategy. (**c**) Chemicals A and C perturb separate MIEs but converge at an intermediate KE; therefore dose addition would be an appropriate strategy. (**d**) Chemicals A and C converge at an intermediate KE, while Chemical D only converges at the AO; therefore integrated addition would be an appropriate strategy. *MIE* molecular initiating event, *IKE* intermediate key event, *AO* adverse outcome, *arrows* key event relationships, *triangles* chemicals. Nodes that contain the colors of two AOPs represent the initial common key events that generate the AOP network

The first scenario (Fig. 7.3a) involves utilizing an AOP related to an end point of concern, to identify mixtures of chemicals that a risk assessor could expect to be dose additive, under a) the assumption that the chemicals are from the same (structural/mechanistic) group(s) which perturb the same AOP and b) that the chemicals co-occur. In this respect, a risk assessor would likely use structural alerts, which have been associated with perturbing the MIE and inducing the AO, to develop chemical categories specific for that AOP. Thus, any chemicals that co-occur and contain one, or more, of the associated structural alerts would form a logical grouping. Subsequently, the available LOAEL/BMD data for the analogues may be used to perform read-across. Thus, this may enable a prediction to be made (via

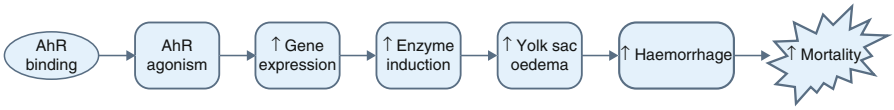

Fig. 7.4 AOP concerning larval fish death initiated by binding to the aryl hydrocarbon receptor (AhR) (Adapted from Ankley et al. 2010)

interpolation or extrapolation) regarding the concentration at which the MIE would be sufficiently perturbed to induce the downstream KEs for the target chemical(s), i.e., those chemicals that do not have associated toxicity data. This in silico prediction could then be verified by performing in vitro or in chemico assays that are associated with the MIE and/or other KEs in the AOP. With the assumption that each co-occurring chemical initiates the same AOP via the same MIE, dose addition would be assumed. Subsequently, a dose-additive risk assessment approach may be employed to assess the combined risk of the mixture of chemicals toward the end point in question.

For example, a risk assessor may be concerned about the potential for larval fish death to be initiated after ligand binding to the aryl hydrocarbon receptor (AhR). The pathway for this AO has been set forth by Ankley and colleagues (Ankley et al. 2010) and is summarized in Fig. 7.4. Briefly stated, AhR agonists bind to the receptor inducing dimerization with aryl hydrocarbon receptor nuclear translocator, thereby increasing the expression of various genes, including those that encode for enzymes involved in xenobiotic metabolism (such as CYP1A1 and CYP1B1) (Ankley et al. 2010; Peterson et al. 1993; Henry et al. 1997; Walker and Peterson 1994). Together these changes initiate a variety of other KEs that culminate in the death of fish at the larval life stage. A number of polychlorinated dibenzodioxins (PCDD), polychlorinated dibenzofurans (PCDF), and other similar halogenated planar aromatic chemicals have the ability to bind to AhR (Belair et al. 2001; Henry et al. 1997; Peterson et al. 1993; King-Heiden et al. 2012). Therefore, these chemical classes could be used to produce separate mechanism-based structural alerts, i.e., each alert is associated with the ability to bind to AhR. If multiple chemicals from any one or more of these chemical categories were to co-occur, AhR binding would be expected based upon read-across predictions. As such, at environmentally relevant levels, these chemicals would be expected to exhibit dose additivity, and mixtures risk assessment using dose addition methodologies would be undertaken (Ankley et al. 2010).

The assumption here of low dose additivity is entirely consistent with the acknowledgment in the "Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures" (U.S. EPA, 2000) that the majority of component-based procedures have an underlying assumption that interaction effects either do not occur at low doses or are sufficiently small to be inconsequential to the risk estimate. The AOP framework may also provide an in vitro assay that could be performed to verify the results from the in silico predictions. Furthermore, the results from these assays could potentially be used to rank order and calculate relative

potency factors for the chemicals within the same category (i.e., the chemicals that perturb the same MIE).

The second scenario involves having two, or more, AOPs that only share a common AO (i.e., no overlap in MIE or intermediate KEs). This knowledge may be used to identify chemicals that would likely be expected to act via independent action if they were to co-occur (Fig. 7.3b). In this instance, it is assumed that knowledge pertaining to the AO of concern has been utilized to identify an AOP network, whereby the constituent AOPs only share a common AO. Upon development of this AOP network, the structural alerts associated with each of the MIEs may be used to group the co-occurring chemicals based upon the MIE and thus the AOP; they are likely to perturb. As with the first scenario, other available testing methods may be utilized to verify that the individual chemicals induce the KEs within the relevant AOP, including the MIE and AO. For example, available assays may be used to verify that (1) chemical A can perturb the KEs within the blue AOP but does not perturb (within the limits of the assay(s)) the KEs within the orange AOP and (2) chemical D can perturb the KEs within the orange AOP but does not perturb (within the limits of the assay(s)) the KEs with the blue AOP (Fig. 7.3b). In this instance, an independent action mixtures risk assessment method would be assumed to be appropriate as the co-occurring chemicals (chemicals A and D) initiate separate AOPs, by perturbing different MIEs, while eliciting the same AO, i.e., the same response (Fig. 7.3b).

The third scenario involves having knowledge of two, or more, AOPs that have separate MIEs but converge at an intermediate KE and initiate the same downstream KEs, up to and including the AO (Fig. 7.3c). As with the previous scenarios, the assessment starts by identifying the structural alerts associated with the MIEs of the individual AOPs that comprise the AOP network, enabling chemicals to be grouped based upon the MIE perturbed, which may in turn be utilized to facilitate read-across. Within this scenario as long as at least one of the intermediate KEs upstream of the convergence point between the AOP (i.e., the common intermediate KE) is known, then a dose-additive mixtures risk assessment strategy may be adopted. A dose-additive strategy appears appropriate in this context because even though chemicals A and C perturb separate MIEs, the AOPs converge at an intermediate KE upstream of the AO (Fig. 7.3c). Therefore, downstream of the initial common intermediate KE (shaded blue and yellow Fig. 7.3c), the AOPs can be considered the same. As such, the effect at the AO can be estimated from the sum of the scaled concentrations of the individual chemicals that perturb the MIEs consistent with the recommendation that dose-additive methods for toxicologically similar chemicals (Rider and Simmons 2015; NRC 2008). One example where this scenario may be of use is for performing a mixtures risk assessment with respect to male reproductive toxicity by androgen antagonism. At present, no formal AOPs have been fully developed for this end point. However, having tested a variety of phthalates and other antiandrogenic chemicals, a number of studies and review articles discuss the biological mechanisms leading to the AO (NRC 2008; Mylchreest et al. 1998; Fisher 2004; Wolf et al. 1999). The results from these studies seem to suggest that, if AOPs were developed, while there may be separate MIEs (e.g., androgen receptor

antagonism or 5α-reductase inhibition), these separate pathways would converge into an AOP network upstream of the AO, in a manner similar to that illustrated in Fig. 7.3c. Therefore, dose additivity may be assumed an appropriate strategy to use for a mixtures risk assessment for androgen antagonists associated with male reproductive toxicity. This is backed up by the findings within the "Phthalates and Cumulative Risk Assessment: The Task Ahead" report by the National Research Council in 2008; this report summarizes studies that tested combinations of different antiandrogens where the experimentally observed effects of the mixture were consistent with those predicted under an assumption of dose addition (NRC 2008).

Finally, the fourth scenario pertains to instances where the AOPs within the AOP network merge at multiple points. A generic example of this is illustrated in Fig. 7.3d, whereby three separate AOPs merge at different points: with the blue and yellow AOPs converging at a common intermediate KE and the orange AOP converging at the common AO. Due to the complexity of this AOP network, i.e., individual AOPs merging at both the intermediate KEs and the AO within the pathway, neither dose addition nor independent action in isolation is likely to give the most accurate prediction. However, a combination of these two strategies, termed integrated addition, may be suitable (Rider and Simmons 2015; Rider and LeBlanc 2005; Olmstead and LeBlanc 2005). The integrated addition approach first groups chemicals based upon a shared "mechanism of action," and a prediction is made for each group separately using dose addition. Subsequently, the results for each of these mechanism-based groups are merged via an overall prediction made using independent action (Rider and Simmons 2015; Rider and LeBlanc 2005). In this instance, the AOP paradigm would provide the rationales to be used to more accurately form both the chemical- and mechanism-based categories. The initial chemical categories could be developed based upon their ability to perturb the same MIE, i.e., grouped by conserved mechanism. Subsequently, these mechanism-based categories would be grouped together based upon whether or not the AOPs shared common KEs upstream of the AO.

Even though each of the above scenarios has focused on only one use case, i.e., identifying the more appropriate risk assessment strategy, other use cases exist that AOPs could be utilized for; these include (1) having the information regarding the presence of a disease incidence within the community (either human or environmental) and identification of the chemical mixture(s) that would be of potential concern or (2) having knowledge that chemicals co-occur within the environment and utilizing this information to elucidate which AOP(s) will likely be triggered. A disease incidence would act as the AO within an AOP, thereby enabling the AOP (s) that culminate in the shared outcome to be elucidated. Upon identification of the AOPs, the appropriate MIE(s) and their associated structural alert(s) can be discerned. These alerts can then be utilized to screen and categorize the groups of chemicals of most concern if they were to co-occur. Alternatively, knowledge of co-occurring chemicals would require the chemicals to be categorized according to the presence of a structural alert. Subsequently, this information regarding the alert (s) present within the chemicals could be related to the MIE(s) they perturb. In turn, this information can be used to identify the AOP(s) that are most likely to be initiated

and thereby the AOs of most concern. In addition, the information within these AOPs would act as a guide as to the assays to perform to refine the predictions. Each of these scenarios illustrate an added benefit the AOP paradigm may provide risk assessors when conducting a mixtures risk assessment as an aid to define the toxicological (and mechanistic) similarity of chemicals within the mixture. Consequently, this information can be utilized in better determining the most appropriate mixtures risk assessment strategy to undertake for a given scenario.

7.5 Benefits of Using AOPs in Mixtures Risk Assessment

In cases where assessments of chemical mixtures are needed, AOP networks can be utilized to decide whether a dose additive, independent action, or integrated addition method is more appropriate to use when performing the mixtures risk assessment for a specific chemical mixture and AO. The initial step is to organize the individual AOPs that relate to the AO of concern into an AOP network. The AOP network is developed by collapsing individual AOPs together based upon the presence of one or more common KEs across each of the AOPs. An example of such an AOP network is shown in Fig. 7.2. Subsequently, the AOP network can be annotated with chemical information. This annotation involves matching each of the co-occurring chemicals to their respective MIE, either through the use of structural alerts, high-throughput screening assays, or prior knowledge. Thereafter, a decision can be made as to whether dose addition, independent action, or integrated addition is the most appropriate mixtures risk assessment technique to perform. The decision will be made based upon where in the AOP network the chemicals overlap, i.e., at the MIE, an intermediate KE, or the AO. Consequently, there are four situations that are the most conceivable through which a decision could be made, as shown in Fig. 7.3.

The additional structure provided by the AOP construct when describing the mechanistic basis for toxicity can support an increase in the confidence both in selection of the appropriate mixtures risk assessment strategy and in use of the selected methodology. By defining AOPs in terms of common KEs as shown in Fig. 7.3, we can now specifically define at what point in the pathway any two chemical perturbations will intersect. As an example, chemicals A and C both have three to four unique KEs (including the MIE) as well as three common intermediate KEs leading to the common AO. Currently, a determination would be made regarding whether these chemicals share a "common mode of action" to decide between dose addition, independent action, and integrated addition. The additional specificity of the AOP definition allows us to evaluate this assumption more carefully. In the example shown in Fig. 7.3c, the key question is whether the KEs that are unique to the two different AOPs could collectively produce dose-response curves of different shapes when considering perturbation of the common KEs including the final AO. This could be due to differences in the types of biological processes involved in the unique KEs such as receptor signaling cascades vs. enzymatic

processes, or it could result from differences in homeostatic processes such as an effective threshold response in one AOP vs. a continuous response in another. Since these differences can be hypothesized based on the AOP description alone, the default assumption of dose addition could be evaluated. In cases where the confidence in this assumption is not sufficient for the given decision context, the AOP framework highlights the experiments needed to confirm or refute the assumption of dose addition. Furthermore, since AOPs are chemical agnostic, as these questions are addressed for a given set of chemicals, the information can then be used for any other chemicals that operate through that set of KEs.

The AOP also provides more options with regard to the response used for independent action of a chemical mixture. Using the example of chemicals A and C, if the first shared intermediate KE is more easily measured than later KEs or the AO, tests could be run for the two chemicals to refine the response at this KE. If needed, based on the decision context, the combined response at the KE could be estimated by these data using the independent action approach without requiring data for the apical outcome from all chemicals. The predicted response at this KE from the mixture in question could then be used to predict the overall AO based on knowledge of the downstream events in the AOP. Alternatively, if the response-response characteristics of the unique KEs in the two AOPs are well characterized, the predicted response at the first common KE could potentially be computationally predicted based on the perturbation of the MIE, even in cases where differences in the dose-response curves across the unique KEs are distinct between the two AOPs. All chemicals like A could then be combined via dose addition, similarly for all chemicals like C, and then the combined response at the first shared KE could be calculated from the predicted responses for each group. Since these earlier key events can be investigated via in vitro and less time- and resource-intensive assays than in vivo studies, it should be much easier to get models that predict an early stage KE than models that extend to the AO.

The examples above are intended to show the theoretical potential for AOPs to impact mixtures assessments. They are by no means comprehensive and will not be applicable in all cases. We expect them to be revised, refined, and expanded based on the valuable insights that will be gained from attempts to apply these strategies to actual data and by comparative analysis of mixtures risk assessments performed via more traditional methods to those that take full advantage of the AOP concept. However, as the examples show, the AOP framework should provide a means to think more precisely about the question of dose addition and independent action in the future. Because AOPs are chemical agnostic, both the number of AOPs available for decision-making and the quality of those AOPs will increase over time. In addition, as the AOP construct is used to support mixtures assessments, experimental evaluation will determine the best methods for using AOPs in this context. For this reason, AOPs should play an ever-increasing role in supporting risk assessment of chemical mixtures.

Acknowledgments M.D.N. was supported by an appointment to the Research Participation Program of the U.S. Environmental Protection Agency, Office of Research and Development, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. EPA.

References

Ankley, G.T., R.S. Bennett, R.J. Erickson, D.J. Hoff, M.W. Hornung, R.D. Johnson, D.R. Mount, J.W. Nichols, C.L. Russom, P.K. Schmieder, J.A. Serrrano, J.E. Tietge, and D.L. Villeneuve. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry* 29: 730–741.

Asturiol, D., and A. Worth. 2011. The use of chemical reactivity assays in toxicity prediction. *JRC Scientific and Technical Reports*.

Becker, R.A., G.T. Ankley, S.W. Edwards, S.W. Kennedy, I. Linkov, B. Meek, M. Sachana, H. Segner, B. Van Der Burg, D.L. Villeneuve, H. Watanabe, and T.S. Barton-Maclaren. 2015a. Increasing scientific confidence in adverse outcome pathways: Application of tailored Bradford-Hill considerations for evaluating weight of evidence. *Regulatory Toxicology and Pharmacology* 72: 514–537.

Becker, R.A., G. Patlewicz, T.W. Simon, J.C. Rowlands, and R.A. Budinsky. 2015b. The adverse outcome pathway for rodent liver tumor promotion by sustained activation of the aryl hydrocarbon receptor. *Regulatory Toxicology and Pharmacology* 73: 172–190.

Belair, C.D., R.E. Peterson, and W. Heideman. 2001. Disruption of erythropoiesis by dioxin in the zebrafish. *Developmental Dynamics* 222: 581–594.

Bradford Hill, A. 1965. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine* 58: 295–300.

Cronin, M.T.D. 2013. An introduction to chemical grouping, categories and read-across to predict toxicity. In *Chemical toxicity prediction: Category formation and read-across*, ed. M.T.D. Cronin, J.C. Madden, S.J. Enoch, and D.W. Roberts. Cambridge, UK: Royal Society of Chemistry.

EC. 2003. Directive 2003/15/EC of the European Parliament and of the Council (Cosmetics Directive 7th Amendment). *Official Journal of the European Union*.

———. 2006a. Corrigenda of Regulation (EC) No 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Official Journal of the European Union*.

———. 2006b. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Official Journal of the European Union*.

ECHA. 2008. Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals.

———. 2009. *Practical guide 6: How to report read-across and categories*. Helsinki: ECHA.

Edwards, S.W., T. Yu-Mei, D.L. Villeneuve, M.E. Meek, and C.A. Mcqueen. 2016. Adverse outcome pathways – organizing toxicological information to improve decision making. *The Journal of Pharmacology and Experimental Therapeutics* 356: 170–181.

Enoch, S.J. 2010. Chemical category formation and read-across for the prediction of toxicity. In *Recent advances in QSAR studies: Methods and applications*, ed. T. Puzyn, J. Leszczynski, and M.T.D. Cronin. Dordrecht/Heidelberg/London/New York: Springer.

Enoch, S.J., and D.W. Roberts. 2013. Approaches for grouping chemicals into categories. In *Chemical toxicity prediction: Category formation and read-across*, ed. M.T.D. Cronin. Cambridge, UK: Royal Society of Chemistry.

Enoch, S.J., J.C. Madden, and M.T.D. Cronin. 2008. Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR and QSAR in Environmental Research* 19: 555–578.

Enoch, S.J., K.R. Przybylak, and M.T.D. Cronin. 2013. Category formation case studies. In *Chemical toxicity prediction: Category formation and read-across*, ed. M.T.D. Cronin, J.C. Madden, S.J. Enoch, and D.W. Roberts. Cambridge, UK: Royal Society of Chemistry.

Fisher, J.S. 2004. Environmental anti-androgens and male reproductive health: Focus on phthalates and testicular dysgenesis syndrome. *Reproduction* 127: 305–315.

Fleischer, M. 2007. Testing costs and testing capacity according to the REACH requirements – results of a survey of independent and corporate GLP laboratories in the EU and Switzerland. *Journal of Business Chemistry* 4: 96–114.

Garcia-Reyero, N. 2015. Are adverse outcome pathways here to stay? *Environmental Science and Technology* 49: 3–9.

Gerberick, F., M. Aleksic, D. Basketter, S. Casati, A.-T. Karlberg, P. Kern, I. Kimber, J.P. Lepottevin, A. Natsch, J.M. Ovigne, C. Rovida, H. Sakaguchi, and T.W. Schultz. 2008. Chemical reactivity measurement and the predictive identification of skin sentisers. *ALTA* 36: 215–242.

Groh, K.J., R.N. Carvalho, J.K. Chipman, N.D. Denslow, M. Halder, C.A. Murphy, D. Roelofs, A. Rolaki, K. Schirmer, and K.H. Watanabe. 2015a. Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology. *Chemosphere* 120: 764–777.

———. 2015b. Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: II. A focus on growth impairment in fish. *Chemosphere* 120: 778–792.

Gutsell, S., and P. Russell. 2013. The role of chemistry in developing understanding of adverse outcome pathways and their application in risk assessment. *Toxicology Research* 2: 299.

Henry, T.R., J.M. Spitsbergen, M.W. HornunG, C.C. Abnet, and R.E. Peterson. 1997. Early life stage toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin in zebrafish (Danrio rerio). *Toxicology and Applied Pharmacology* 142: 56–68.

Jaworska, J., and N. Nikolova-Jeliazkova. 2007. How can structural similarity analysis help in category formation? *SAR and QSAR in Environmental Research* 18: 195–207.

Kadenbach, B. 2003. Intrinsic and extrinsic uncoupling of oxidative phosphorylation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1604: 77–94.

King-Heiden, T.C., V. Mehta, K.M. Xiong, K.A. Lanham, D.S. Antkiewicz, A. Ganser, W. Heideman, and R.E. PetersoN. 2012. Reproductive and developmental toxicity of dioxin in fish. *Molecular and Cellular Endocrinology* 354: 121–138.

Knapen, D., L. Vergauwen, D.L. Villeneuve, and G.T. Ankley. 2015. The potential of AOP networks for reproductive and developmental toxicity assay development. *Reproductive Toxicology* 56: 52–55.

Landesmann, B., M. Goumenou, S. Munn, and M. Whelan. 2012. Description of prototype modes-of-action related to repeated dose toxicity. *JRC Scientific and Policy Report 75689*.

Leonard, J.A., Y.-M. Tan, M. Gilbert, K. Isaacs, and H. El-Masri. 2016. Estimating margin of exposure to thyroid peroxidase inhibitors using high-throughput in vitro data, high-throughput exposure modeling, and physiologically-based pharmacokinetic/Pharmacodynamic modeling. *Toxicological Sciences* 151(1): 57–70.

Meek, M.E.B., C.M. Palermo, A.N. Bachman, C.M. North, and R.J. Lewis. 2014. Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. *Journal of Applied Toxicology* 34: 595–606.

Mylchreest, E., R.C. Cattley, and P.M.D. Foster. 1998. Male reproductive tract malformations in rats following gestational and lactational exposure to Di(n-butyl) phthalate: An antiandrogenic mechanism? *Toxicological Sciences* 43: 47–60.

Naven, R.T., R. Swiss, J. Klug-Mcleod, Y. Will, and N. GreenE. 2013. The development of structure-activity relationships for mitochondrial dysfunction: Uncoupling of oxidative phosphorylation. *Toxicological Sciences* 131: 271–278.

Nelms, M.D., G. Ates, J.C. Madden, M. Vinken, M.T. Cronin, V. Rogiers, and S.J. Enoch. 2015a. Proposal of an in silico profiler for categorisation of repeat dose toxicity data of hair dyes. *Archives of Toxicology* 89: 733–741.

Nelms, M.D., C.L. Mellor, M.T. Cronin, J.C. Madden, and S.J. Enoch. 2015b. Development of an in Silico Profiler for Mitochondrial Toxicity. *Chemical Research in Toxicology* 28: 1891–1902.

NRC. 2007. *Toxicity testing in the 21st century: A vision and a strategy*. Washington, DC: The National Academies Press.

———. 2008. *Phthalates and cumulative risk assessment: The tasks ahead*. Washington D.C: The National Academies Press.

OECD. 2011. Report of the workshop on using mechanistic information on forming chemical categories. Series on testing and assessment No. 138. ENV/JM/MONO(2011)8.

———. 2012. The Adverse Outcome Pathway for skin sensitisation initiated by covalent binding to proteins part I: scientific evidence. Series on testing and assessment No. 168. ENV/JM/MONO (2012)10/PART1.

———. 2013. Guidance document on developing and assessing Adverse Outcome Pathways. Series on testing and assessment No. 184. ENV/JM/MONO(2013)6.

———. 2014. Users' handbook supplement to the guidance document for developing and assessing AOPs.

———. 2015. Report of the workshop on a framework for the development and use of Intergrated Approaches to Testing and Assessment. *Environment Directorate Joint Meeting of the Chemicals Committee and the Working Party on Chemicals.* Paris: Organisation for Economic Co-operation and Development.

Oki, N.O., M.D. Nelms, S.M. Bell, H.M. Mortensen, and S.W. Edwards. 2016. Accelerating adverse outcome pathway development using publicly available data sources. *Current Environmental Health Reports* 3(1): 53–63.

Olmstead, A.W., and G.A. Leblanc. 2005. Toxicity assessment of environmentally relevant pollutant mixtures using a heuristic model. *Integrated Environmental Assessment and Management* 1: 114–122.

Perkins, E.J., P. Antczak, L. Burgoon, F. Falciani, S. Gutsell, G. Hodges, A. Kienzler, D. Knapen, M. Mcbride, and C. Willett. 2015. Adverse outcome pathways for regulatory applications: Examination of four case studies with different degrees of completeness and scientific confidence. *Toxicological Sciences* 148: 14–25.

Peterson, R.E., H.M. Theobald, and G.L. Kimmel. 1993. Developmental and reproductive toxicity of dioxins and related compounds: Cross-species comparisons. *Critical Reviews in Toxicology* 23: 283–335.

Przybylak, K.R., and T.W. Schultz. 2013. Informing chemical categories through the development of adverse outcome pathways. In *Chemical toxicity prediction: Category formation and read-across*, ed. M.T.D. Cronin, J.C. Madden, S.J. Enoch, and D.W. Roberts. Cambridge, UK: Royal Society of Chemistry.

Rider, C.V., and G.A. Leblanc. 2005. An integrated addition and interaction model for assessing toxicity of chemical mixtures. *Toxicological Sciences* 87: 520–528.

Rider, C.V., and J.E. Simmons. 2015. Risk assessment strategies and techniques for combined exposures. In *Toxicological risk assessment for beginners*, ed. J.A. Torres and S. BobsT. Heidelberg/New York/Dordrecht/London: Springer International Publishing.

Roberts, D.W., A.O. Aptula, and G. PatlewicZ. 2006. Mechanistic applicability domains for non-animal based predictions of toxicological endpoints. QSAR analysis of the Schiff base applicability domain for skin sensitisation. *Chemical Research in Toxicology* 19: 1228–1233.

Schultz, T.W. 2010. Adverse outcome pathways: A way of linking chemical structure to in vivo toxicological hazards. In *In silico toxicology: Principles and applications*, ed. M.T.D. Cronin and J.C. Madden. Cambridge, UK: Royal Society of Chemistry.

Spycher, S., P. Smejtek, T.I. Netzeva, and B.I. Escher. 2008. Towards a class-independent quantitative structure-activity relationship model for uncouplers of oxidative phosphorylation. *Chemical Research in Toxicology* 21: 911–927.

Stoddart, G., and J. Brown. 2014. A campaign to end animal testing: Introducing the PETA international science consortium ltd. *ATLA* 42: 387–393.

Taylor, K., D.J. Andrew, and L. Rego. 2014. The added value of the 90-day repeated dose oral toxicity test for industrial chemicals with a low (sub)acute toxicity profile in a high quality dataset. *Regulatory Toxicology and Pharmacology* 69: 320–332.

Tollefsen, K.E., S. Scholz, M.T. Cronin, S.W. Edwards, J. De Knecht, K. Crofton, N. Garcia-Reyero, T. Hartung, A. Worth, and G. Patlewicz. 2014. Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regulatory Toxicology and Pharmacology* 70: 629–640.

U.S. EPA. 2000. Supplementary guidance for conducting health risk assessment of chemical mixtures. Risk Assessment Forum. EPA/630/R-00/002.

Villeneuve, D.L., D. Crump, N. Garcia-Reyero, M. Hecker, T.H. HutchinsoN, C.A. Lalone, B. Landesmann, T. Lettieri, S. Munn, M. NepelskA, M.A. Ottinger, L. VergauweN, and M. Whelan. 2014a. Adverse outcome pathway (AOP) development I: Strategies and principles. *Toxicological Sciences* 142: 312–320.

———. 2014b. Adverse outcome pathway development II: Best practices. *Toxicological Sciences* 142: 321–330.

Vinken, M. 2013. The adverse outcome pathway concept: A pragmatic tool in toxicology. *Toxicology* 312: 158–165.

Vinken, M., B. Landesmann, M. Goumenou, S. Vinken, I. Shah, H. Jaeschke, C. Willett, M. Whelan, and V. Rogiers. 2013. Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury. *Toxicological Sciences* 136: 97–106.

Walker, M.K., and R.E. Peterson. 1994. Aquatic toxicity of dioxins and related chemicals. In *Dioxins and Health*, ed. A. Schecter. New York: Plenum.

Wolf, C.J., C. Lambright, P. Mann, M. Price, R.L. Cooper, J. Ostby, and L.E.J. Gray. 1999. Administration of potentially antiandrogenic pesticides (procymidone, linuron, iprodione, chlozolinate, p,p'-DDE, and ketoconazole) and toxic substances (dibutyl- and diethylhexyl phthalate, PCB 169, and ethane dimethane sulphonate) during sexual differentiation produces diverse profiles of reproductive malformations in the male rat. *Toxicology and Industrial Health* 15: 94–118.

Part III
Toxicology

Chapter 8
Dose-Response Modeling

Gregg E. Dinse and David M. Umbach

Abstract For any definition of additivity, evaluating whether an organism's response to a mixture is additive depends on the dose-response relationships for each of the mixture's component chemicals. Consequently, the statistical analysis of dose-response relationships is fundamental to mixture toxicology – as well as to other areas of toxicology. This chapter offers a broad overview of dose-response modeling and an introduction to some statistical issues that arise in the use of dose-response models – with an eye to evaluating additivity. It does not, however, attempt to be a handbook or guide to the use of any specific models; instead, it tries to make readers aware of issues that need attention to achieve efficient and valid inference. The chapter mentions features of study design and describes how they can influence both aspects of model fitting and the quality of results. It considers the choice of functional form used to describe how the mean response changes as dose increases as well as the evaluation of how well the chosen form fits the data at hand. The chapter also points out that proper modeling of the variability inherent in the structure of the data is crucial to efficient statistical inference. Finally, because many dose-response models require iterative numerical methods, it offers a few pointers to help overcome problems when these methods fail to converge. Dose-response modeling is an essential tool in mixture toxicology but one that demands careful application to achieve the best results.

Keywords Experimental design · Hill model · Model assessment · Nonlinear regression · Statistical model · Variance structure

G. E. Dinse (✉)
Public Health Sciences, Social & Scientific Systems, Inc, Durham, NC, USA
e-mail: dinse@niehs.nih.gov

D. M. Umbach (✉)
Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA
e-mail: umbach@niehs.nih.gov

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_8

8.1 Introduction

One fundamental project in mixture toxicology is the evaluation of whether a mixture obeys an additive or no-interaction null model, whereby the typical response of an organism (or an isolated biological system) to different concentrations of the mixture can be predicted based on the organism's (or system's) typical response to different concentrations of each component of the mixture individually. This project has three essential elements: (1) a suitable definition of the additive state, formulated in a way that is amenable to constructing quantitative predictions (see Chap. 9); (2) an appropriate collection of data from experiments that examined dose-response relationships both for the individual component chemicals and for the mixture; and (3) an array of statistical techniques that use these data to make inferences about the question of additivity (see Chap. 11). Shortcomings regarding any of these three elements can impinge on the value of mixture experiments for risk assessment.

Because prediction of the response of an organism to a mixture under additivity – regardless of the precise definition of additivity invoked – is rooted in the dose-response relationships for each of the mixture's component chemicals, statistical analysis of dose-response curves for individual chemicals becomes a key building block for evaluating hypotheses about departures from additivity. In fact, the statistical properties of predicted responses to a mixture under an additivity assumption depend crucially on how well the statistical models used to analyze data from the individual component chemicals represent the underlying true dose-response relationships. If dose-response relationships for the component chemicals in the mixture are estimated with bias, the prediction for the mixture would likely be estimated with bias as well. Correspondingly, if the estimated dose-response relationships for the component chemicals are highly variable, the prediction for the mixture would likely be highly variable, so that tests of the additivity null hypothesis would have low statistical power. In addition, if a statistical dose-response model represents the typical dose-response trajectory accurately but fails to properly account for multiple sources of variation in the data, the precision attributed by the data analysis to the dose-response relationships might be larger or smaller than it actually is. Such errors in modeling variability in dose-response relationships can lead to incorrect conclusions about additivity, including false positive declarations of departures from additivity or false negative declarations of no departure from additivity. Analogous kinds of statistical issues centered on adequate specification and fitting of a dose-response relationship also arise when specifying models that are intended to reflect departures from additivity and fitting those models to data from mixtures of chemicals. Sound statistical analysis is essential for valid quantification of dose-response relationships; consequently, it is important throughout toxicology – including mixture toxicology.

The purpose of this chapter is to acquaint toxicologists who are now (or intend to be) working in the area of mixtures with some statistical issues that are ubiquitous in studying dose-response relationships. Because toxicologists conduct an exceedingly broad range of experiments with various outcomes and study designs, the variety of

statistical techniques needed to analyze the range of toxicologic data is comparably broad. Any attempt to even touch on all possibilities would require a book-length treatment, not simply a chapter. Instead, this chapter highlights certain issues that are common across many statistical techniques and that can have a distinct impact on inferences about dose-response relationships. First, it defines some terminology and notation that will be used throughout the chapter. Next, it discusses statistical issues related to study design and describes how they can impinge on the choice of statistical models employed and on the quality of the results. Then, it goes on to talk about modeling strategies, including the choice of functional form for fitting the mean dose-response trajectory. The chapter also draws attention to the importance of specifying a variance structure that properly reflects sources of variation in the data. In addition, it points out techniques for evaluating the adequacy of the specified dose-response model for the data at hand. Finally, because many dose-response models useful in toxicology require iterative numerical methods for fitting, it mentions some ways to cope with failure of these methods to converge to a unique best fit to the data. The chapter closes with a summary.

8.2 A Statistical Perspective on Dose-Response Modeling

To establish some definitions and basic notation, consider as a template a simple dose-response study involving a single chemical and a single response or outcome. Assume that the study involves a total number N of experimental units, where "experimental unit" is a generic term that statisticians use to denote the entity that is assigned at random to receive a particular treatment or condition in an experiment. Thus, the experimental units could be male mice in a rodent carcinogenicity study, pregnant rats in a teratogenicity study, petri plates containing a *Salmonella* tester strain in an Ames mutagenicity assay, culture dishes growing a given cell line in in vitro studies, and so on. In a dose-response study, the treatments to which the experimental units are assigned at random are the particular dose or concentration levels of the compound being investigated. Let D denote the number of dose levels and $d_1, d_2, d_3, \cdots, d_D$ denote the D particular dose levels used in the study. Because these dose levels are under the control of the experimenter, one usually regards them as known constants and not subject to errors of measurement. For simplicity, assume that the total number of experimental units N is a multiple of the number of dose levels D, such that $N/D = n$ and that the same number, n, of units are assigned to each dose level d_i. Here the index i is one of the numbers $\{1, 2, 3, \cdots, D\}$. At each dose level, each experimental unit is labeled by a second index j that is one of the numbers $\{1, 2, 3, \cdots, n\}$. Consequently, in this study, each experimental unit is uniquely identified by two indices, i indicating dose group and j indicating experimental unit within dose group. Let Y_{ij} or y_{ij} represent the response of the j^{th} experimental unit at the i^{th} dose level; uppercase Y indicates that one is thinking of the response as a random quantity that has probabilistic properties (but lacks a known numerical value), whereas lowercase y indicates that one is referring to an

actual value that would be observed in a study. Thus, y_{ij} is the observed value of the random variable Y_{ij}. Also, one omits the subscripts when referring to a response in general or uses only subscript i to emphasize the role of dose level when the particular experimental unit is inconsequential. Typically, both subscripts are needed only when referring to the response of a particular experimental unit assigned to a given dose level. Also, though i indexes the dose levels d_i used in the experiment, the notation d is used for dose level more generically – including dose levels not included in the experiment.

Depending on the response measured, data analysis for toxicologic studies utilizes a range of statistical distributions. The response of interest in a mouse carcinogenicity study might be the presence/absence of a certain type of tumor in each mouse at death. Then, Y (viewed as 1/0 for presence/absence) might be modeled statistically as having a Bernoulli distribution, and a dose-response analysis might focus on how steeply the probability of the tumor being present increases with increasing dose level. In a teratogenicity study, if the chemical under study was known to have no effect on implantation, the response of interest might be the number of rat pups in each litter born alive and without any malformations. For each litter, the count Y (which could range from 0 to the number of zygotes implanted) might be modeled statistically as having a binomial distribution, and a dose-response analysis might focus on how steeply the probability of being born alive and without malformation decreases with increasing dose level. In an Ames assay for mutagenicity, the response of interest is the number of revertant colonies on the plate; then, for each plate, the count Y (which could range from 0 upward) might be modeled statistically as having a Poisson distribution, and a dose-response analysis might focus on the rate of increase in the expected number of revertant colonies with increasing dose level. In studies that use cell lines to assess toxic effects, the response of interest might be the level of a particular enzyme. The enzyme level Y (whose value could be any non-negative number) might be modeled as having a log-normal distribution, and a dose-response analysis might examine how the typical enzyme level changes with increasing dose level.

Although the details of an appropriate data analysis would differ for each of these examples, they share some fundamental commonalities: each asks how some characteristic of the response's statistical distribution, a characteristic whose true value is unknown, changes as the dose level changes; and each acknowledges inherent variability of the response around its true value. One can think of the characteristic of interest as the signal and the inherent variability (random error) as the noise – so the goal of statistical dose-response analysis is to uncover the signal in noisy data.

From a statistical perspective, several decisions must be made in carrying out a dose-response study. These decisions fall into two phases: the design phase and the data analysis phase. First is the design phase: one must have a plan for gathering the needed dose-response data. Having a statistically efficient design for an experiment enhances validity and cost-effectiveness. Although many aspects of study design are the purview of the toxicologist, statisticians can help address issues like choosing the number and location of dose levels or setting the number of experimental units – both overall and at each dose level – to achieve acceptable statistical performance. In

complex experiments where, for example, the experiment might need to be conducted over several days and involve multiple batches of experimental units, statistical design is indispensable for allocating experimental units from different batches to different days to ensure that the dose effects of interest can be estimated without bias and that uncertainty attributable to days, batches, and experimental units can properly be assessed. The statistical design of experiments is a broad and challenging field, beyond the scope of this chapter, but the next section (Sect. 8.3) considers a few aspects of statistical design and how they impinge on dose-response data analysis.

The second phase of a dose-response study is, of course, the analysis of the data. Here the decisions have to do with constructing a statistical model that describes the data and allows estimation of the dose-response relationship of interest and quantification of uncertainty in the estimate. As indicated in the examples presented earlier in this section, one must choose a statistical or probability distribution that describes the random behavior of the experimental units at each dose level. Usually, with knowledge of the type of response being measured and personal experience or guidance from the literature regarding similar responses, the analyst will quickly nominate a small set of candidate distributions that can be refined as the analysis proceeds.

A second early decision is what feature of the statistical distribution to focus on for describing the dose-response relationship. The distributional feature used most often in the dose-response context is the mean of the dose-level-specific distribution of responses. A mean is a common way to assess central tendency in a statistical distribution, not only for continuous responses but more generally. For example, for a list of presence/absence responses (coded as 1/0), their average (i.e., the number of ones divided by the number of experimental units) gives the proportion of experimental units with 1 as a response, and such a proportion can often be interpreted as an estimate of the probability of the characteristic being present. Though features other than the mean response might sometimes be useful in dose-response analysis, this chapter considers only the mean.

Another crucial aspect of the analysis phase is specifying a model or algebraic expression that describes how the mean response changes with dose level. For example, one can represent each observation using the mathematical formula:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \tag{8.1}$$

where μ_i is the unknown true mean response at dose level d_i from the experiment and ε_{ij} is an error term that is conceptualized as a random deviation from that mean for experimental unit j at dose level d_i, one that corresponds to the particular statistical distribution under consideration. This representation embodies the common notion that an observed value is an imperfect or noisy reflection of a true but unknown "typical" value for any experimental units that experience a particular dose level. The collection of μ_i for i in the set $\{1, 2, 3, \cdots, D\}$ represents a model for the mean as a function of dose in that the μ_i values describe how the mean changes with applied

dose level. The goal of the statistical analysis might be obtaining point and confidence interval estimates for each μ_i or for differences among them.

If the investigator were only interested in the mean response at the pre-selected dose levels tested in the experiment, estimation of those particular mean responses via Eq. 8.1 would be useful. More commonly, however, dose-response studies are conducted to make inferences about dose levels in addition to those actually tested in the experiment; perhaps an investigator seeks to characterize an entire dose-response curve or to make inference about mean responses at arbitrary dose levels either between or beyond (e.g., low-dose extrapolation) those studied experimentally. In this case, one could replace Eq. 8.1 with a possibly nonlinear regression model:

$$Y_{ij} = f(d_i|\boldsymbol{\theta}) + \varepsilon_{ij}, \tag{8.2}$$

where $f(d_i|\boldsymbol{\theta})$ is a prespecified regression function that relates dose level d_i to the unknown true dose-specific mean response μ_i and that depends on a vector of unknown parameters $\boldsymbol{\theta}$, and ε_{ij} is an error term as in Eq. 8.1. For example, in the Ames assay for mutagenicity, one often specifies $f(d|\boldsymbol{\theta}) = \alpha + \beta d$ (here, $\boldsymbol{\theta} = (\alpha, \beta)$), where α reflects the background mutant yield and β represents the mutagenic potency (Bernstein et al. 1982). A second example might involve enzyme concentrations measured in such a way that the response ranges from 0 to 1 (or, equivalently, from 0% to 100%) and one might specify $f(d|\boldsymbol{\theta})$ using the Hill model (Hill 1910), namely, $f(d|\boldsymbol{\theta}) = d^{\gamma}/(d^{\gamma} + \delta^{\gamma})$ (here, $\boldsymbol{\theta} = (\delta, \gamma)$), where δ represents the ED_{50} (or median effective dose) and γ is known as the Hill coefficient. Because Eq. 8.2 allows inferences for any d, a key difference between Eqs. 8.1 and 8.2 is the latter's capacity for allowing inference to dose levels that were not observed in the experiment. Of course, there is a trade-off: proper inferences with Eq. 8.2 rely, in part, on the assumption that $f(d|\boldsymbol{\theta})$ is correctly specified for (or at least a close approximation to) the true dose-response relationship under study. Further consideration of the mean function $f(d|\boldsymbol{\theta})$ and consequences of specifying it incorrectly appear later in the chapter.

The presentation of Eqs. 8.1 and 8.2 has so far focused on modeling the relationship between dose level and mean response as embodied in the μ_i or in $f(d_i|\boldsymbol{\theta})$; but both models also involve random errors, as embodied in the set of ε_{ij}. For now, assume that the mean response model is correctly specified so that the ε_{ij}, averaged across experimental units, have mean zero at each dose level. The important remaining properties of the ε_{ij} are their variances and their covariances. Variance may be the same for every experimental unit or may change across dose levels. A covariance is zero when two experimental units are independent and non-zero when they are correlated. In fitting Eqs. 8.1 and 8.2 to data, correct specification of variances and covariances for all experimental units is critical. This specification is done via a variance-covariance matrix for the vector $\boldsymbol{\varepsilon}$ (whose N entries are the ε_{ij}). Let $\boldsymbol{\Sigma}$ denote the $N \times N$ variance-covariance matrix for $\boldsymbol{\varepsilon}$ (or, more generally, for the vector of observations $\mathbf{Y}$ given their true means); the diagonal elements of $\boldsymbol{\Sigma}$ are the variances of the ε_{ij} and the off-diagonal elements are their pairwise covariances. The matrix $\boldsymbol{\Sigma}$ can depend on one or more unknown parameters (denoted by vector $\boldsymbol{\omega}$);

one can write $\boldsymbol{\Sigma}(\boldsymbol{\omega})$ to emphasize that dependence and regard $\boldsymbol{\Sigma}(\boldsymbol{\omega})$ as representing a model for the variance-covariance matrix in terms of unknown parameters to be estimated.

For example, when the dose-level-specific response distribution is normal, a typical default assumption is that its variance is unknown but constant across dose levels and that the individual ε_{ij} are independent (their covariances are zero) both within dose level and across dose levels (in this example, the off-diagonal elements of $\boldsymbol{\Sigma}$ are all zero, and the diagonal elements are all the same unknown constant often denoted σ^2). The assumption of constant variance may not always hold, however. Some probability distributions have the property that their variance and their mean are related in a defined way. For example, for the Poisson distribution, the mean and the variance are equal; for the Bernoulli and binomial distributions, both the mean and the variance depend on the success probability; for the log-normal distribution, the variance increases as the mean increases in a prescribed way. Thus, certain forms of heteroskedasticity (nonconstant variance) are built into the data analysis via the statistical distribution chosen for the analysis. Sometimes, however, the mean-variance dependence that is built in by the chosen distribution does not adequately accommodate the observed degree of heteroskedasticity, so that the data analyst must incorporate additional parameters to accommodate extra variability (Breslow 1984; Williams 1982). For example, models that incorporate extra-binomial variation are commonly applied in studies of possible teratogens where a dam is the treated experimental unit, but a presence/absence response is assessed on each dam's individual pups (Haseman and Hogan 1975; Piegorsch and Haseman 1991; Zorrilla 1997) and summarized into a single Y_{ij} for each dam. Even for distributions like the normal that accommodate constant variance, the underlying data-generating mechanism may deliver heteroskedasticity across dose levels; and such heteroskedasticity must be properly taken into account to achieve efficient and valid statistical inference.

Another consideration when modeling the variance-covariance structure of the data is whether the observations can be modeled as independent, which implies that covariances are zero. When experimental units that are homogeneous (representing a single unstructured population of units) are assigned at random to dose levels, the homogeneity of the units together with the randomization process strongly supports that the observations would be independent. If the experimental units are not homogeneous to begin with, however, but instead represent multiple subgroups, then non-zero covariances can arise even with randomization. Say, for example, the experimental units are mice and the mice needed for the study were accumulated by taking multiple littermates from several different litters, then arguably the responses of two littermates may well be more similar to each other than the responses of two mice from different litters – leading to non-zero covariances for certain pairs of units that should be acknowledged in the data analysis. Aspects of the way an experiment is conducted can also lead to non-zero covariances between experimental units. For example, if an experiment is so large that it must be carried out in multiple runs and each run involves several experimental units and is accomplished on a different day, or if a procedure involves incubating treated plates (the experimental units) at a

controlled temperature for a certain period and the large number of plates involved necessitates the use of multiple incubators, then two units from the same run or in the same incubator could arguably be more similar in response than two units from different runs or in different incubators. In this way, details of the conduct of an experiment can have a strong bearing on the covariance structure that may exist among the observations.

Thus, an overall statistical model for the data from a dose-response study typically consists of three components: (1) a probability distribution appropriate for the response under study; (2) a mathematical model, denoted here by $f(d|\boldsymbol{\theta})$, that describes the relationship between mean response and dose; and (3) a model, denoted here by $\boldsymbol{\Sigma}(\boldsymbol{\omega})$, for the variance-covariance matrix of the data. After the data analyst has established such a statistical model for the dose-response data, the task is to estimate the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ and to quantify the uncertainty in those estimates. For example, the model $f(d|\boldsymbol{\theta}) = \alpha + \beta d$ describes an entire family of possible straight lines because both α and β can take on values ranging from $-\infty$ to ∞. Estimating the unknown parameter $\boldsymbol{\theta} = (\alpha, \beta)$ amounts to choosing the specific values of α and β that produce the straight line that fits the data best.

The process of fitting a model to data – that is, of estimating the particular values of the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ that provide the best fit to the data at hand – is, of course, a key step in data analysis. Statisticians have devised various criteria to operationalize the concept of "best fit." For regression models for normally distributed observations, estimates of $\boldsymbol{\theta}$, denoted $\widehat{\boldsymbol{\theta}}$, are typically derived using the principle of least squares, where the best estimates are those that minimize the sum of squared differences between observed data and model predictions (Seber and Wild 1989). Variance parameters $\boldsymbol{\omega}$ for normally distributed observations can be estimated by equating observed average squared deviations to their expected values expressed as functions of $\boldsymbol{\omega}$ (Searle et al. 2006). Another widely used criterion for "best fit" is based on the principle of maximum likelihood. Generally, the likelihood is an algebraic expression of the joint probability of the observed data regarded as a function of the unknown parameters. The value of the unknown parameter declared to fit best is the one that makes this joint probability as large as possible – hence, maximum likelihood. Maximum likelihood delivers estimates of both $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ directly. For normally distributed data, the least squares and the maximum likelihood estimates of $\boldsymbol{\theta}$ coincide – a relationship that may not hold for other distributions. Maximum likelihood estimates are widely used because they have desirable statistical properties that hold for many distributions (Mood and Graybill 1963). Of course, other ways of defining "best fit" could also be used: some are versions of maximum likelihood such as restricted maximum likelihood or penalized maximum likelihood; others are developed in a Bayesian framework (Box and Tiao 1992; Carlin and Louis 2000). Maximum likelihood receives the most attention in this chapter.

8.3 Design Considerations

Proper experimental design can have a substantial impact on the precision and power of statistical inferences from a dose-response study. A study's design can also influence the range of models that can be successfully fitted to the eventual data. Designing an experiment always entails trade-offs. In a study with multiple goals, a good design for addressing one goal may be less useful for addressing another; in a study with multiple endpoints, an ideal design for one endpoint may be suboptimal for another. Consider a study where the investigator anticipates that the appropriate dose-response function is a straight line between the lowest and highest dose levels of interest. Assume further that the responses will be normally distributed and have constant variance across all dose levels. Then, for the goal of having the most precise estimate of the slope of the line, the optimal design is to allocate half the experimental units to the lowest dose of interest and the other half to the highest (Seber 1977). If the investigator entertained any doubts that a straight line was the correct dose-response model, this design could be catastrophic, as it is completely unable to detect any curvature in dose-response trajectory. Allocating experiment units to intermediate dose levels would allow one to check whether, in fact, the straight line was an appropriate dose-response model, though at the cost of some loss of precision for slope estimation, and would enable the investigator to fit a more appropriate model if necessary. When selecting dose levels for an Ames mutagenicity assay, the investigator is faced with just this kind of trade-off; at higher dose levels, toxicity begins to dominate mutagenicity, and an increasing dose-response curve tends to bend downward. Possible goals for a dose-response study for an individual endpoint include choosing a model that aptly describes the dose-response trajectory, estimating the unknown parameters of the selected model, assessing risk through quantities such as the median effective dose or some other benchmark dose, and predicting expected response at any desired dose level – and seeking the most accurate and precise estimation possible for these last three. A chosen design will often have to compromise among such competing goals – while simultaneously honoring any constraints imposed by budgets, facilities, and available time.

Consider first a simple setting, known to statisticians as a completely randomized design. In that setting all the experimental units are viewed as homogeneous, and all are allocated at random to the chosen dose levels; such a design incorporates no structure arising from different batches of experimental units or from different technicians or different runs or other restrictions on randomization. In that setting, the basic quantities contributing to the statistical design are the number of dose levels in the study, the actual dose levels employed (think of their placement – location and spacing – on the dose axis of a graph), and the number of experimental units assigned to each dose level. The choice of each of these quantities is informed both by the nature of the dose-response models to be fitted to the data, namely, the set of functions $f(d|\boldsymbol{\theta})$ under consideration, and by the number of experimental units that the investigator can afford to obtain and carry through the study. Generally speaking, more flexible models can be fitted when more dose levels are studied and,

for a fixed number of dose levels, estimates become more precise as the number of replicate experimental units allocated to each dose level increases – but costs and other practicalities typically limit these numbers.

A fixed number N of experimental units could be allocated to particular dose levels in various ways, ranging from placing all of them at a single dose to placing a single observation at N distinct dose levels. Statisticians often think in terms of an optimal experimental design – one where the placement of dose levels and the proportion of the total number of experimental units allocated to each dose level is guaranteed to meet some desirable property (often to minimize the variances of the parameter estimates) for any total number of experimental units. An optimal design for minimizing the variance of an estimated slope for a straight line was mentioned earlier. Such optimal designs are a prominent area of theoretical statistics, but most of the work and the results are applicable when the model used to analyze the data is linear in the unknown parameters, for example, a straight line or many models used in analysis of variance. In these settings, often there will be a unique optimal design regardless of the true values of the unknown parameters. A majority of dose-response models used in toxicology, however, are nonlinear in the unknown parameters (e.g., the Hill model mentioned earlier). Though optimal designs may still be found, finding them is much more difficult for nonlinear models; and, when found, they often have the unfortunate property that the placement of the dose levels for the optimal design changes depending on the true values of the unknown parameters (Seber and Wild 1989) – which means that the investigator has to know a good deal about the underlying truth before a dose-response experiment can be designed optimally. Consequently, focusing on optimal designs is not practical in the present context, and instead the focus will be on heuristics for good design.

In general, the number of dose levels (including dose zero) in the study must, at minimum, equal the number of individual parameters in the vector $\boldsymbol{\theta}$. Under ideal circumstances, that minimal number of dose levels ensures all individual parameters can be estimated uniquely (if several different models are being considered, then use the $\boldsymbol{\theta}$ with the most elements to determine the minimum). For a simplistic example illustrating the problem with having too few dose levels, consider fitting a straight line (where $\boldsymbol{\theta}$ has two individual parameters) to a design where all the replicates were allocated to a single dose level. Although one could certainly estimate a mean response at the single dose level, one could not estimate an intercept and slope uniquely: there is no best choice among the infinitely many lines that can be fitted through a single point. To estimate both parameters, a design must employ at least two dose levels. Analogously, at least two dose levels would be required to estimate the parameters γ and δ in the Hill model mentioned earlier. Similarly, if $\boldsymbol{\theta}$ contained four individual parameters, then a study would need at least four distinct dose levels to have any chance of obtaining unique estimates of the four parameters.

While the number of parameters in the selected model provides a minimum for the number of dose levels required, usually it would be unwise to implement a design with such a limited number of dose levels – for several reasons. One has already been mentioned in connection with the optimal design for a straight line: if an investigator wants to check whether a more complex model, one with more parameters to allow a

richer variety of shapes, might fit better than the one used to determine the minimal number of dose levels, the design needs to have additional dose levels to accommodate the more complex alternative model. A second reason has to do with the placement of the dose levels.

How the placement of dose levels can influence the ability of a design to estimate unknown parameters can be illustrated by an example using a sigmoid dose-response function, a shape often relevant in toxicology. Consider an extended version of the Hill model mentioned previously, namely,

$$f(d|\boldsymbol{\theta}) = L + (U - L) \times [d^{\gamma}/(d^{\gamma} + \delta^{\gamma})] \tag{8.3}$$

(here, $\boldsymbol{\theta} = (L, U, \delta, \gamma)$), where L is the lower response limit, U is the upper response limit, δ represents the ED_{50} (or median effective dose), and γ is known as the Hill coefficient and is related to how sharply the curve rises (or falls). Whereas in the original version of the Hill model (the portion enclosed in brackets in Eq. 8.3), the possible response values ranged from 0 to 1 (assuming an increasing function, i.e., $\gamma > 0$); in this version, response values range from L to U, and both of these limits are to be estimated. If $f(d|\boldsymbol{\theta})$ is plotted against the logarithm of dose, the resulting graph is sigmoid, with a long shallow rise from an asymptote at L with small dose levels, eventually rising more quickly, reaching its steepest slope at the ED_{50}, then gradually rising more slowly as it increases toward an upper asymptote at U with large dose levels.

For the four-parameter model of Eq. 8.3, a design must incorporate at least four distinct dose levels. Suppose an experimental design assigned four non-zero dose levels to be equally spaced on a logarithmic scale (often a convenient spacing for a laboratory because of the ease of serial dilutions). If the four dose levels selected all corresponded to the early part of the curve when the increase was shallow (say, all were well below the ED_{50}), then one gets a lot of information about response levels close to L but little information about other response levels (Fig. 8.1a). Thus, it is likely that only the parameter L would be estimated satisfactorily; the data from such a design would not serve to estimate the other parameters well at all. Similarly, if the four dose levels selected for the design were all above the ED_{50}, then data resulting from the design might estimate U well but not the other three parameters (Fig. 8.1b). Alternatively, if the four doses were widely spaced on the log scale but the true curve rose rapidly over a narrow intermediate dose range, it is possible that the two lower doses would correspond to the early relatively flat part of the curve and the two higher doses would correspond to the late relatively flat part of the curve, with no data collected in the region where the dose-response function changed rapidly (Fig. 8.1c). In that situation, data arising from the design might estimate L and U well but not the ED_{50} or Hill coefficient. Of course, estimating all the model parameters well simultaneously is important if one seeks a reliable estimate of the entire dose-response curve, which requires appropriate placement of the four dose levels (Fig. 8.1d). Even the placement of doses d_1–d_4 in Fig. 8.1d, though better than the placements in the other panels, may not sufficiently capture information about the flat parts of the curve. The use of two additional dose levels (d_0, d_5) to extend the

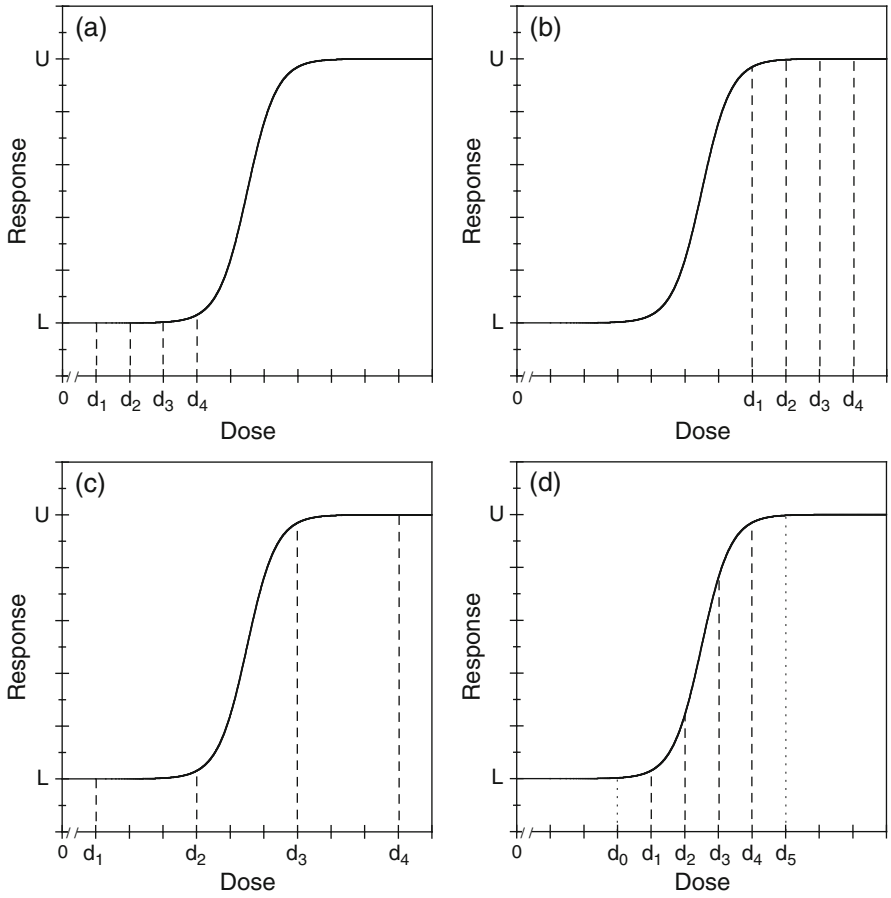

Fig. 8.1 Illustration of how dose-level placement may influence estimation of dose-response relationships. Each panel shows a dose-response function generated from the four-parameter Hill model in Eq. 8.3 and four evenly spaced dose levels (d_1, d_2, d_3, d_4). Panel (**a**): all dose levels on the lower flat portion of the curve provide information mostly about parameter L. Panel (**b**): all dose levels on the upper flat portion of the curve provide information mostly about parameter U. Panel (**c**): failure to include dose levels in the region where the curve rises provides information mostly about parameters L and U. Panel (**d**): dose-level placement that includes the region where the curve rises as well as the shoulders where the curve is nearly level provides information about all four parameters; placing additional dose levels at d_0 and d_5 (alternately, replacing dose levels d_1 and d_4 with d_0 and d_5) would provide better information about the flat parts of the curve and enhance estimation of all parameters

range of dose levels would likely improve estimation of all parameters, pointing to the value of using more dose levels than the minimum required number.

Another way to think about why the dose level placements mentioned in the previous paragraph are problematic is by relating the dose levels in the design to the expected response levels. The set of dose levels included in a design should

correspond to a set of true unknown response levels that are representative of the entire range of possible underlying response values. Thus, a design whose dose levels only capture the low end but not the middle or high end of the response range – like the design mentioned earlier with all dose levels below the ED_{50} – is unlikely to be a good design for a sigmoid curve. Similarly, a design whose dose levels capture only the highest and lowest response levels but none in the middle – because the dose spacing is too wide to have doses corresponding to intermediate response levels – is unlikely to be a good design. The placement of dose levels is important so that the design captures a range of responses at the collection of chosen doses.

This discussion serves to illustrate a point made earlier – that an optimal design for nonlinear models depends on the unknown true values of the parameters. One must have some idea about the shape and location of the dose-response curve one is trying to estimate if one hopes to design an experiment to estimate it well. Although equally spaced dose levels, either on an additive or logarithmic scale, are common default choices, irregular spacing of dose levels – farther apart where the dose-response curve is expected to be flat and closer together where the dose-response is expected to change rapidly – can be a useful strategy, but one that demands more extensive prior knowledge of the dose-response curve. If an investigator does have a good sense of an appropriate range of dose levels to represent the full range of response levels, then choosing a number of dose levels nearer the minimum required by the number of model parameters seems reasonable – though including several additional dose levels that will allow flexibility for fitting more complex models and will accommodate any lingering uncertainty as to the appropriate dose range would be a prudent strategy. On the other hand, if the investigator is very uncertain of the nature of the dose-response curve, other strategies are needed. Most important would be to employ a pilot dose-finding study to help home in on an appropriate dose range. In addition, when the dose-response shape is uncertain, designs using a relatively large number of dose levels (and necessarily fewer experimental units at each dose) would increase the chances for avoiding some of the aforementioned problems with poor dose placement.

Usually designs assign an equal number of observations at each dose level – this strategy is just simple to execute. If the investigator expects that variability in response is greater at some dose levels than others – say, variability is greater at higher dose levels – allocating more experimental units at those dose levels expected to be most variable can be an efficient strategy but one that is not widely used.

The design considerations to this point have focused on a completely randomized design; but such simple designs are not the best approach in every study. Many dose-response experiments must accommodate distinct batches of experimental units or involve procedures that must be carried out on different days or using several instruments of the same type to accommodate the throughput. In such experiments, statistical efficiency demands that one account for the batches or days or instruments at the data analysis stage – because they contribute batch-to-batch or day-to-day or instrument-to-instrument variability. The statistical design goal is to be able to remove, when possible, such unavoidable but attributable variability from the variances of dose-level comparisons and the variances of parameter estimates. One

statistically useful but relatively simple design in this context is a randomized complete block design, where "block" is a generic statistical term for a set of experimental units that have some feature in common. That feature might be a common source (e.g., mice from different litters could constitute separate blocks, or mice from different strains could constitute separate blocks), or that feature might be some commonality in the way units are handled during the conduct of the experiment (e.g., if multiple incubators are used in an experiment, the set of experimental units assigned to each incubator would be considered a block). A randomized complete block design assigns the experimental units within each block at random to the set of dose levels under study. Thus, each block is a dose-response study with one experimental unit allocated to each dose level; in a sense, each block is a separate mini-dose-response study. This design is useful when the number of experimental units in a block (the block size) is large enough to handle the entire set of dose levels under study. If the block size is smaller than the number of distinct dose levels, then more complicated designs known as randomized incomplete block designs may be appropriate. Statisticians have devised many sorts of statistical designs to handle various sorts of restrictions on randomization imposed by the way experiments must be conducted. The intent here is to make readers aware of these issues and to point out that consultation with a statistician at the design stage can help an investigator make the best use of resources in settings where complex blocking may be needed.

8.4 The Model Describing How Mean Response Depends on Dose: $f(d|\theta)$

Toxicologists typically consider dose-response curves that are continuous (i.e., without jumps) and where the mean response either increases across the entire range of doses or decreases across that entire range. Dose-response curves that never change directions are called "monotone." A monotone dose-response model can have one or more flat regions; however, if a monotone model has no flat regions, it is called "strictly monotone." A model is called "non-monotone" when it exhibits any change in direction (i.e., the response increases over some dose ranges but decreases over others).

Non-monotone dose-response models are not widely used in toxicology, but they have some applications. For example, in the Ames assay, dose-response curves may turn downward at the highest dose levels when cell toxicity dominates mutagenicity (Margolin et al. 1981); others have allowed the possibility of non-monotonicity when fitting flexible curves for relative potency estimation (Guardabasso et al. 1987, 1988). Consideration of hormesis also leads to non-monotonicity in dose-response models (Hunt and Bowman 2004; Kim et al. 2016). As with the toxicity-mutagenicity competition just mentioned, models constructed to reflect an increase in response attributable to one process (say, mutagenicity) with a decrease in

response due to another (say, toxicity) have been applied to non-monotone dose-response data in other settings, such as high-throughput screening experiments (EPA 2016; Shockley 2016).

Monotone models with flat regions are used more often, however. A typical example is a threshold model where the response is initially flat from dose zero up to some critical dose where the monotone increase or decrease in response begins (e.g., Casey et al. 2004). The initial flatness of a threshold model is interpreted as a continuation of the response in the absence of chemical exposure until the dose becomes sufficiently high to elicit a measureable response. Though such models are useful, using noisy data to distinguish a flat region (where the slope is zero) from a region with a very shallow positive or negative slope is difficult; consequently, inference about the critical dose (join point) is difficult in the sense that estimates of the join point often have large standard errors. Also, from a curve-fitting perspective, a flexible and carefully chosen strictly monotone model can often closely mimic and be difficult to distinguish statistically from a threshold model. Consequently, this chapter will focus primarily on strictly monotone dose-response models.

Another reason to favor strictly monotone dose-response models in the context of mixtures is that methods for constructing the dose-response curve for a mixture under an assumption of additivity (e.g., Berenbaum's definition of dose additivity; Berenbaum 1985) become problematic without strict monotonicity. A dose-response model maps a given dose to the corresponding expected response level; but use of Berenbaum's definition requires mapping an observed response level to the dose of each component chemical that separately would induce that response level. What is required is a mapping from response to dose – the inverse of the dose-response function. If a dose-response curve has a flat section between two dose levels, the response level of the flat portion does not correspond to a unique dose – but to any dose between those two dose levels. A strictly monotone dose-response curve, however, has a unique response corresponding to each dose and can be inverted to provide a unique dose corresponding to each response. Thus, strict monotonicity of the dose-response model is desirable in connection with mixtures.

Luckily, the class of strictly monotone mathematical functions to use as dose-response models is large and can accommodate a wide variety of curve shapes. Many dose-response models that toxicologists use routinely are nonlinear, often sigmoid. In Eq. 8.3, an extended version of the Hill model was introduced; this model is strictly monotone so long as $\gamma \neq 0$, and it is increasing or decreasing depending on the sign of γ. Consider replacing the Hill function, $d^{\gamma}/(d^{\gamma} + \delta^{\gamma})$, in Eq. 8.3 by a general strictly monotone function $g(d|\boldsymbol{\theta}^*)$ that is governed by a vector of parameters $\boldsymbol{\theta}^*$ and that takes values between 0 and 1 as d increases from 0 to ∞ (or, equivalently, as the logarithm of d increases from $-\infty$ to ∞). One can write the resulting more general dose-response model as:

$$f(d|\boldsymbol{\theta}) = L + (U - L) \times g(d|\boldsymbol{\theta}^*) \tag{8.4}$$

(here, $\boldsymbol{\theta} = (L, U, \boldsymbol{\theta}^*)$). This formulation allows $f(d|\boldsymbol{\theta})$ to increase from L to U or to decrease from U to L, depending on whether $g(d|\boldsymbol{\theta}^*)$ is monotone increasing or

decreasing in d. For concreteness, the mean response in this chapter will be assumed to increase monotonically with dose, unless otherwise stated. Therefore, the lower response limit L is the value of $f(d|\boldsymbol{\theta})$ when $d = 0$ and the upper response limit U is the value of $f(d|\boldsymbol{\theta})$ as dose d gets arbitrarily large.

Generally speaking, the response limits in Eq. 8.4 can take any values such that L is smaller than U. Some experiments may involve natural boundaries for the mean response, in which case L and U might be assigned fixed values a priori. For example, if the response is a percentage of experimental units responding, one might specify $L = 0$ and $U = 100$ or in fact any pair of intermediate values that satisfy $0 \leq L < U \leq 100$. Alternatively, one or both of the response limits can be treated as unknown and estimated from data in the current experiment (Dinse and Umbach 2011) or from data on negative and positive controls in historical studies. Even if the mean response is a proportion, the lower limit could exceed 0% if there were a background rate, which could occur if a fraction of the population showed an effect even without chemical exposure. Similarly, the upper limit could be less than 100% if a fraction of the population did not show an effect regardless of how great the dose became.

The family of curves described by the dose-response model of Eq. 8.4 changes for different specifications of the function $g(d|\boldsymbol{\theta}^*)$. Of course, the Hill model of Eq. 8.3 is one family where $g(d|\boldsymbol{\theta}^*) = d^\gamma/(d^\gamma + \delta^\gamma)$ with $\boldsymbol{\theta}^* = (\delta, \gamma)$**.** For statisticians, one convenient way to specify a monotone increasing curve for $g(d|\boldsymbol{\theta}^*)$, which has a lower bound of 0 and an upper bound of 1, is to use a cumulative distribution function for a known statistical distribution, which by definition increases monotonically from 0 to 1. (If a monotone decreasing $g(d|\boldsymbol{\theta}^*)$ is desired, subtract the cumulative distribution function from 1 and use that difference as $g(d|\boldsymbol{\theta}^*)$.) Some common statistical distributions used in toxicology include the logistic (Reeve and Turner 2013); the normal, which leads to the well-known probit model (Finney 1971); and the Weibull (Christensen and Nyholm 1984). Many other choices are possible, of course; but usually choices are restricted to familiar statistical distributions.

Statistical cumulative distribution functions have rigidly defined shapes. Investigators looking for more flexibility in the shape of a dose-response model to better fit the data at hand have several options. One is to create a new dose-response model by replacing dose d in an existing model with a transformed version of dose d; a second is to use other kinds of models that allow more flexible dose-response shapes, such as regression splines or smoothing splines.

Perhaps the most frequently used transformation of dose is the natural logarithm of dose, in symbols, $t(d) = ln\,(d)$. Here, $t(\cdot)$ is notation for a generic transformation, and $ln(\cdot)$ is the natural (base e) logarithm function. If one starts with a function $f(d|\boldsymbol{\theta})$ and substitutes $t(d)$ for d, one gets a new dose-response function $f^*(d|\boldsymbol{\theta})$. Consider the linear dose-response model $f(d|\boldsymbol{\theta}) = \alpha + \beta d$. If one replaces d with $\ln(d)$, the new dose-response model becomes $f^*(d|\boldsymbol{\theta}) = \alpha + \beta \ln(d)$. The latter model describes a different family of curves than the original model. The same kind of manipulation is possible starting from any model $f(d|\boldsymbol{\theta})$; of course, one has no guarantee that the resulting model $f^*(d|\boldsymbol{\theta})$ will be better suited than the original for the data at hand.

The distinction between creating a new dose-response model and simply reparameterizing the original model is important and is often a point of confusion. Often a model expressed as a function of d is reexpressed as a function of $ln(d)$ for mathematical convenience or computational stability. This reexpression takes advantage of the mathematical identity: $d \equiv \exp(\ln(d))$. For example, the popular Hill model (limited here to the response range 0–1) is frequently written as a function of dose d:

$$g(d|\boldsymbol{\theta}) = d^{\gamma}/(d^{\gamma} + \delta^{\gamma}). \tag{8.5}$$

Alternatively, Eq. 8.5 can be algebraically rearranged and reparameterized to give the following logistic model, which is expressed as a function of log dose:

$$g(d|\boldsymbol{\theta}) = 1/[1 + \exp(-\alpha - \beta \ln(d))], \tag{8.6}$$

where parameters α and β have different interpretations from parameters δ and γ. Often α is referred to as an intercept and β as a slope because Eq. 8.6 can be rewritten as a linear function of $\ln(d)$, namely, $\ln\{g(d|\boldsymbol{\theta})/[1 - g(d|\boldsymbol{\theta})]\} = \alpha + \beta \ln(d)$. If one sets $\alpha = -\gamma \ln(\delta)$ and $\beta = \gamma$, then Eqs. 8.5 and 8.6 coincide and both specify exactly the same mean response for any particular dose. Thus, simply because two dose-response models look distinct algebraically does not mean that they must specify two different families of dose-response relationships – sometimes two models are exactly the same even though the functional forms appear to be different.

For some models, one parameterization may offer computational or interpretational advantages over another. For example, when summarizing results in terms of the ED_{50}, one might prefer a model that incorporates the ED_{50} directly as a parameter, such as parameter δ in Eq. 8.5, rather than calculating it indirectly from other parameters. Alternatively, a model expressed in terms of $\ln(d)$, such as Eq. 8.6, has properties that make it less susceptible to numerical problems with model fitting: it minimizes curvature and thus reduces bias by more closely mimicking a linear model (Bates and Watts 1988; Reeve and Turner 2013).

Rather than selecting a fixed dose transformation in advance, one can write the transformation function $t(\cdot)$ as a function of unknown parameters and build those additional parameters into an original dose-response model, in essence estimating a dose transformation that enhances model fit. Perhaps the most common transformation function with an adjustable parameter is the Box-Cox transformation (Box and Cox 1964): $t(d) = (d^{\lambda} - 1)/\lambda$, where the parameter λ governs the shape of the dose metric, with a continuum of dose transformations for the range of λ values between $-\infty$ and ∞. This family of transformations includes (after changing multiplicative and additive constants to 1 and 0, respectively) $t(d) = \sqrt{d}$ if $\lambda = \frac{1}{2}$, $t(d) = 1/d$ if $\lambda = -1$, and $t(d) = ln(d)$ in the limit as $\lambda \to 0$. For example, to obtain a linear dose-response model with an arbitrary dose metric, one could substitute $(d^{\lambda} - 1)/\lambda$ for d in $f(d|\boldsymbol{\theta}) = \alpha + \beta d$ to create a new model $f^{*}(d|\boldsymbol{\theta}) = \alpha + \beta[(d^{\lambda} - 1)/\lambda]$ and then estimate λ together with α and β using software for nonlinear regression. The same general procedure could be applied to almost any dose-response model. For instance,

Altenburger et al. (2000) considered several models that included a Box-Cox transformation of dose.

One feature to be aware of when applying the Box-Cox approach is that the mean response at dose 0 can differ from the response limit expected under the original model. For example, consider a dose-response model $f(d|\boldsymbol{\theta})$ as in Eq. 8.4, with $g(d|\boldsymbol{\theta}^*)$ having the form shown in Eq. 8.6, except that $(d^\lambda - 1)/\lambda$ is substituted for $\ln(d)$. An estimate of λ close to 1 would suggest using $d - 1$ (or, equivalently, d) as the dose metric. By definition, at dose 0 the values of $g(d|\boldsymbol{\theta}^*)$ and $f(d|\boldsymbol{\theta})$ should be 0 and L, respectively. For an increasing dose-response curve ($\beta > 0$), however, substituting d for $\ln(d)$ in Eq. 8.6 gives $g(0|\boldsymbol{\theta}^*) > 0$, and thus $f(0|\boldsymbol{\theta}) > L$. Therefore, use of the Box-Cox approach can force a non-zero background rate even if $L = 0$. A similar problem occurs for a decreasing dose-response curve ($\beta < 0$), where the mean response at dose 0 would remain below the upper limit, i.e., $f(0|\boldsymbol{\theta}) < U$, thereby precluding 100% response even if $U = 100$. In other words, employing the Box-Cox transformation approach can imply a redefinition of parameters L and U in models having the basic structure of Eq. 8.4.

Up to this point, the presentation has focused on parametric dose-response models. Each of these models has a family of curves associated with it, and each expresses the mean response as a smooth and strictly monotone function of dose. These models often have parameters with useful interpretations. The shapes of the curves within each family are rigid in certain ways, however. Despite one's ability to adjust a model's parameters to achieve a best fit within that model or family of curves, one may not always be able to find a model that adequately fits the data at hand. Certain desired inferences, such as extrapolation below the lowest doses in the experiment, rely heavily on the dose-response shape determined by the parametric model.

As an alternative, one might seek approaches that also produce a smooth and strictly monotone curve for the dose-response function but alleviate some of the rigidity in shape of particular parametric models. To achieve greater flexibility in curve shape, one generally has to sacrifice some interpretability of model parameters. Still, for mixture applications, the goal is often fitting a smooth monotone dose-response curve. From such a curve, one can estimate any quantities of interest such as the ED_{50} or other benchmark doses even when a single parameter identified with the quantity of interest is not part of the model. On the other hand, low-dose extrapolation is even more uncertain with such models because their flexibility precludes using the rigid parametric model structure to help make inferences beyond the range of the available data. Various kinds of flexible smoothing models could be applied for dose-response modeling with choices dictated to some extent by the nature of the available data. Splines, which are piecewise polynomial models, could be useful in many dose-response settings (Harrell 2001; Ramsay 1988). The analyst chooses the number and location of knots (the dose levels where the polynomial pieces join), the degree of polynomial to employ, and sometimes a constraint on the function beyond the lowest and highest knots. For dose-response modeling, constraining the fitted spline to be monotone is an important consideration. For example, with respect to evaluating mixture data for departures from additivity,

Kelly and Rice (1990) used a monotone hybrid of smoothing and least-squares splines, where monotonicity is achieved by constraining coefficients and smoothing is controlled by a penalty parameter and by the number of knots. For data with a binary response, Dette et al. (2005) proposed a nonparametric method for obtaining monotone estimates of effective dose without requiring constrained optimization or function inversion. In a relative potency setting, Guardabasso et al. (1987) assumed that multiple chemicals share a common but arbitrarily shaped dose-response curve, which can be shifted or stretched along the log-dose scale to give chemical-specific curves; they model the common curve by a cubic spline (though without requiring monotonicity) and then apply chemical-specific shift and scale parameters. Nottingham and Birch (2000) combined parametric and nonparametric estimates of a dose-response function, with a mixing parameter that adjusts the relative weight given to each component based on how well it individually fits the data.

Another modeling strategy, one that is capable of generating quite flexible predictive models but one that has largely been unexplored for dose-response modeling, is model averaging. The idea is that one postulates a list of K possible dose-response model families, say $f_1(d|\boldsymbol{\theta_1})$, $f_2(d|\boldsymbol{\theta_2})$, $f_3(d|\boldsymbol{\theta_3})$, $\cdots$, $f_K(d|\boldsymbol{\theta_K})$, and fits each of the K models to the available data. The final predictive model is a weighted average of the best-fitting models, one from each family. Typically, model fitting is accomplished using a Bayesian paradigm where the weights are also estimated. Although we have not yet seen model averaging used in the context of mixture models, model averaging has been proposed by several authors as a way to estimate a benchmark dose that is not tied to any single parametric model family (Fang et al. 2015; Simmons et al. 2015; Wheeler and Bailer 2007, 2008, 2009) and to detect hormesis (Kim et al. 2016).

When examining whether a mixture is obeying some definition of additivity, there is a premium on accurate and precise estimation of the dose-response curves of the component chemicals because the predictions from those curves are combined to calculate the expected dose-response curve for the mixture under additivity. Hertzberg et al. (2013) proposed an approach based on guidance from the U.S. Environmental Protection Agency that assumes that all component chemicals are toxicologically similar and that the specific dose-response curve for each chemical comes from a common family of models; that is, the component-specific curves differ only in their specific parameter values, but not the form of $f(d|\boldsymbol{\theta})$. Those authors use model selection criteria (see Sect. 8.5) to select the simplest model family that still provides an adequate fit to the data for each component chemical. This approach is straightforward and likely sound if all component chemicals are indeed toxicologically similar. There is, however, no guarantee that a single model family will provide the best fit to the data on each component chemical. The Hertzberg et al. approach could suffer if the dose-response relationships of certain component chemicals were poorly fit by the common model family. In a less restrictive approach, Altenburger et al. (2000) specified a list of model families (such as described above for model averaging) and selected a best-fitting model for each component chemical across the list of model families. Thus, each component chemical could have a dose-response curve from a different model family, a strategy

which improves overall fit at the expense of greater complexity. A model averaging approach would, in principle, offer better response prediction for each component chemical than could any single model – but that potential advantage would come at the expense of a vastly more computationally intensive fitting procedure.

8.5 Analysis Considerations

As mentioned previously, a dose-response model is specified via three components: a probability distribution, a model $f(d|\boldsymbol{\theta})$ that describes the relationship between mean response and dose, and a model $\boldsymbol{\Sigma}(\boldsymbol{\omega})$ for the variance-covariance matrix of the data. The process of fitting a dose-response model to data is the process of estimating the values for the unknown parameters $\boldsymbol{\theta}$ in the mean model and $\boldsymbol{\omega}$ in the variance-covariance model that best fit the data. Denote the estimated values by $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\omega}}$, respectively**.** Although the details of the calculations differ depending on the assumed probability distribution for the data and the criterion employed in defining "best fit," statistical software will provide the estimates $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\omega}}$ together with estimates of the variances and covariances of those parameter estimates – as well as related quantities such as confidence intervals and test statistics. After fitting a particular model $f(d|\boldsymbol{\theta})$ to data from an experiment and having the estimate $\widehat{\boldsymbol{\theta}}$ available, one can estimate the true mean response under the model at any specified dose level d^* by calculating $f(d^*|\widehat{\boldsymbol{\theta}})$, that is, by substituting d^* and $\widehat{\boldsymbol{\theta}}$ into the relevant equation. The value $f(d^*|\widehat{\boldsymbol{\theta}})$ is called the predicted response (or predicted value) at dose d^*. The variance of these predicted responses can be estimated using $\widehat{\boldsymbol{\omega}}$**,** but details of the calculation differ depending on the particular dose-response model assumed. Again, many statistical software tools will provide these predicted values and their variances (or standard errors). Thus, fitting a dose-response model allows one to plot the estimated mean response trajectory as a function of dose and to construct confidence bands for and test hypotheses about that trajectory.

A well-known aphorism attributed to statistician GEP Box is "All models are wrong; but some are useful." With any regression models, and dose-response models are no exception, one seeks a model where the estimated dose-response trajectory closely mimics the trajectory of the observed data, and the assumed variance model and probability distribution are faithfully reflected by the observed data. If the assumed model does not satisfactorily reflect the data, then inferences based on that model are questionable and conclusions are potentially misleading. The model must be a sufficiently good approximation to be useful. Thus, a critical part of any dose-response analysis involves assessing the aptness of the model for the data at hand. Essentially, this process is one of model criticism – what are the good and bad aspects of the model in terms of being in accord with the data.

Inspecting residual plots is often a sensible first step in model criticism. Procedures for examining model assumptions through residual plots are commonplace for models that are linear in the parameters; for such linear models, the residuals have

two desirable features: (1) the variability in the residuals is a straightforward reflection of the variability in the data; and (2) the residuals and the predicted values are uncorrelated. A so-called raw residual, $\hat{\varepsilon}_{ij}$, is the difference between an observed response and its predicted response under the fitted model; thus, $\hat{\varepsilon}_{ij} = y_{ij} - f(d_i|\widehat{\boldsymbol{\theta}})$; sometimes these raw residuals are rescaled to have variance 1 by dividing each by some measure of variability. The raw residuals, and their various rescaled counterparts, are commonly used with linear models. Residuals with names such as Pearson residuals or deviance residuals are often used with models involving binomial or Poisson distributional assumptions (Agresti 2013), but the fundamental idea remains the same – a residual assesses how far an observed response is from the value expected under the fitted model.

What is less often recognized is that, for the nonlinear models common in dose-response modeling, raw residuals do not necessarily retain the features that make them useful for checking aptness of linear models via plotting. (Problems arise when a nonlinear model has high intrinsic curvature, a concept beyond the scope of the current presentation.) Instead, another concept of residuals, called projected residuals (Cook and Tsai 1985), can be used with nonlinear models to recover the desirable properties needed for using simple plots of residuals as diagnostic tools. Statistical software packages such as SAS provide projected residuals for plotting.

Consider modeling a continuous response, like enzyme activity, with $n > 1$ experimental units per dose level under the assumption that the response at each dose level is normally distributed about its mean with constant variance across dose levels. Aspects of model aptness can be examined by plotting the residuals appropriate for the type of model fitted against dose level d_i or against predicted response $f(d_i|\widehat{\boldsymbol{\theta}})$. If the assumed model for the mean fits well, the n residuals at each dose level should be centered near zero at every dose (Fig. 8.2a); departures where the residuals are centered above zero for some dose levels and below zero for others suggest that a different mean function may provide a better fit (Fig. 8.2b). If, in addition, the variance is constant as assumed, residuals at every dose level will exhibit approximately equal spreads (Fig. 8.2a); patterns where the spread in the residuals grows larger or smaller with increasing dose level provide evidence for heteroskedasticity (Fig. 8.2c) and may indicate that a transformation of the response or a more complex model that incorporates heteroskedasticity is needed. Moreover, a histogram of the residuals should reveal a symmetric distribution if the data are distributed normally.

Another common graphical diagnostic approach applied with linear regression is to look for influential observations – those that have an unusually large influence on parameter estimates or on predicted values when they are deleted from the data set – by plotting various statistics known collectively as influence diagnostics against a variable that identifies each observation. Again, as with residuals, inherent characteristics of nonlinear regression models that differ from those of linear models imply that some concepts used for influence diagnostics must be reinterpreted for use with nonlinear models (St. Laurent and Cook 1992, 1993).

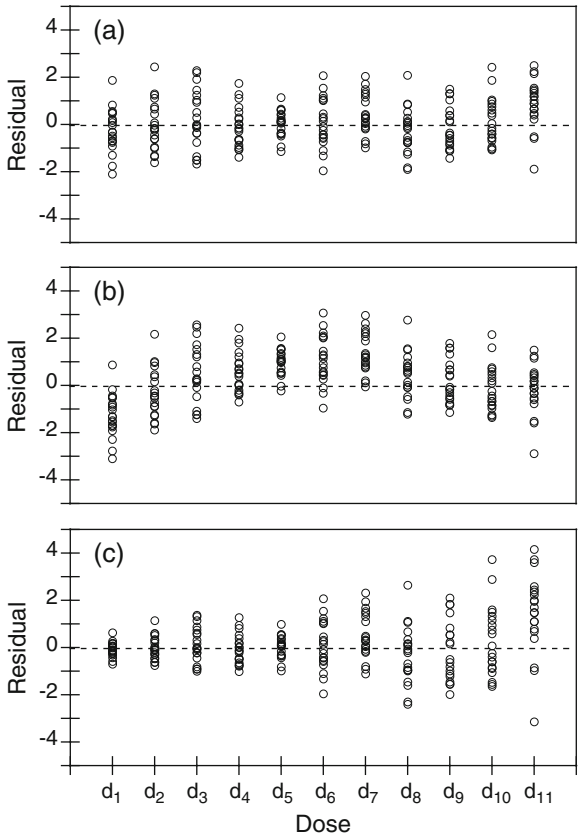

Fig. 8.2 Characteristic residual plots for a dose-response model whose errors ε_{ij} have a normal distribution. Underlying data have responses measured on 20 experimental units at each of 11 dose levels (d_1, d_2, d_3,. . ., d_{11}). Panel (**a**): data-generating model has homoskedastic errors and the fitted $f(d\,|\,\boldsymbol{\theta})$ was correctly specified; residuals have similar spreads and are vertically centered near zero at all dose levels. Panel (**b**): data-generating model has homoskedastic errors, but the fitted $f(d\,|\,\boldsymbol{\theta})$ was incorrectly specified; residuals have similar spreads at all dose levels, but their centers exhibit a non-horizontal trajectory (above zero at intermediate dose levels and below zero at lower or higher dose levels). Panel (**c**): data-generating model has heteroskedastic errors, but the fitted $f(d\,|\,\boldsymbol{\theta})$ was correctly specified; residuals are vertically centered near zero at all dose levels but have spreads that increase with dose level

In addition to informal visual inspections of residual plots, formal statistical tests can be applied as well. Perhaps the most basic test for aptness of the regression model $f(d|\boldsymbol{\theta})$ arises from comparing the fit of Eqs. 8.1 and 8.2. When each of the D distinct dose levels in a dose-response experiment has $n > 1$ experimental units assigned, the estimates of the set of μ_i in Eq. 8.1, denoted $\hat{\mu}_i$, are the average values of the observations at each dose level. Those values should be unbiased estimates of the true unknown responses at those dose levels (most accurate estimates available with the data at hand); the larger n, the more precise the estimates. The use of regression

model $f(d|\boldsymbol{\theta})$ to provide inference about responses at untested doses proceeds under the belief that it also provides unbiased estimates of the true unknown response at each tested dose level. On the other hand, if the mean responses at tested dose levels based on fitting $f(d|\boldsymbol{\theta})$ appear biased, then it may not be a useful model. Thus, one regards the regression model $f(d|\boldsymbol{\theta})$ as fitting the data well when its predicted responses at tested doses closely match the corresponding dose-specific means, i.e., when $f(d_i|\widehat{\boldsymbol{\theta}}) \approx \hat{\mu}_i$. This idea is the basis of a test of model fit that can be applied whenever the vector $\boldsymbol{\theta}$ contains fewer than D parameters. The general procedure is to fit both Eqs. 8.1 and 8.2 via maximum likelihood and construct a likelihood ratio test (Seber and Wild 1989) comparing the fit of Eq. 8.2 to that of Eq. 8.1. Rejection of the null hypothesis that both models fit equally well implies that the regression model $f(d|\boldsymbol{\theta})$ was unsuccessful in estimating the observed dose-specific mean responses and should be replaced with a different model. Of course, details of constructing the likelihood ratio test depend on the statistical distribution assumed for the responses.

For example, consider a study involving D distinct dose levels with n experimental units assigned to each dose and assume that the response has a normal distribution at each dose level. Then, the fitting of Eq. 8.1 amounts to conducting a one-way analysis of variance with the dose levels as the treatment groups. Taking $f(d|\boldsymbol{\theta})$ to be the Hill model of Eq. 8.3 where $\boldsymbol{\theta}$ contains four parameters, a least-squares-based test that is essentially equivalent to the likelihood ratio test comparing Eq. 8.2 to 8.1 would involve an F-statistic with $D - 4$ degrees of freedom in the numerator and $D \times (n - 1)$ in the denominator (Seber and Wild 1989).

The underlying principle used in the goodness-of-fit test described above for comparing Eqs. 8.1 and 8.2 can be applied to any pair of nested models to decide whether the smaller model fits as well as the larger. Consider two regression models, with $f_2(\cdot|\boldsymbol{\theta_2})$ being a nested sub-model of $f_1(\cdot|\boldsymbol{\theta_1})$. One way to think of a nested sub-model is that the parameter vector $\boldsymbol{\theta_2}$ of the sub-model contains the same parameters as $\boldsymbol{\theta_1}$ except that, in the sub-model, some of the parameters are fixed at specified values and do not need to be estimated. For example, consider the Hill model of Eq. 8.3 with parameter vector $\boldsymbol{\theta_1} = (L, U, \delta, \gamma)$. Another Hill model with the lower and upper asymptotes fixed at 0 and 1, respectively, would have parameter vector $\boldsymbol{\theta_2} = (0, 1, \delta, \gamma)$ and be nested within the first because the parameters to be estimated in $\boldsymbol{\theta_2}$ are a subset of those to be estimated in $\boldsymbol{\theta_1}$. When one model is a special case of another, the one with more parameters is more flexible and will necessarily fit at least as well as the one with fewer parameters; however, if the difference in fit is negligible, the simpler model with fewer parameters would typically be preferred based on parsimony. Formal statistical tests such as likelihood ratio tests can be used to decide whether or not fixing a subset of the parameters at specified values degrades model fit (Seber and Wild 1989).

Comparing two regression models that are not nested requires a different strategy. Formal tests to compare non-nested models are rarely used in toxicology; instead, one chooses the "better" model by using a model selection criterion. Because models

with more parameters might be expected to fit better than models with fewer parameters, model selection criteria typically make adjustments for the number of parameters in the model. The general procedure is to choose a model selection criterion and calculate its value for each candidate model under consideration. Then select as the best model the one with the largest value (or sometimes smallest, depending on the particular criterion used) of the criterion. A great variety of such criteria are in use. The coefficient of determination (R^2), or a version of it adjusted for the number of parameters in the model, selects according to the proportion of variation in the data accounted for by the fitted model. Other commonly used criteria, such as the Akaike information criterion (AIC) (Akaike 1974) or the Bayesian information criterion (BIC) (Schwarz 1978; Montgomery et al. 2012), evaluate goodness of fit through the likelihood of the observed data but impose a penalty that increases with the number of model parameters. Thus, if two models produced the same likelihood, these criteria would favor the one with fewer parameters. In experiments that involve very few observations, one might prefer the AICc (Burnham and Anderson 2002), a version of the AIC with a correction for small sample sizes. Model selection criteria such as these are useful adjuncts to strategies for modeling mixture components, such as those of Hertzberg et al. (2013) and Altenburger et al. (2000), which involve selecting best-fitting models. On the other hand, even the best-fitting among a list of candidate regression models may exhibit important lack of fit when compared to fitting the dose-specific mean responses via Eq. 8.1.

When the model $f(d|\boldsymbol{\theta})$ for the mean response shows evident lack of fit, how should it be remediated? An obvious answer is to choose a different model with superior fit – but that may be easier said than done. Experience with the particular response or assay may suggest alternative models to try; similarly, a literature search or reaching out to colleagues might turn up alternatives. A plot with the fitted dose-response model overlaying the observed data sometimes reveals the main discrepancies between model and data, thereby suggesting modifications to improve the model – including perhaps changing the dose metric. Such a plot or residual plots may instead reveal "unusual" or "influential" data values that adversely affect model fit, initiating careful scrutiny of the validity or correctness of those data points. If efforts to find a better-fitting parametric model fail, one could consider more flexible modeling approaches such as splines or model averaging, as mentioned earlier. Another consideration is whether the model is useful for its intended purpose despite some lack of fit. For example, suppose model fit suffers mainly at high doses but is satisfactory at lower doses. If inferences at lower doses are the primary use for the model, perhaps the formally ill-fitting model will yield useful information – interpretation should be cautious, however: lack of fit in the mean model can distort variance estimates so that, for example, confidence interval coverage may suffer even at the low doses where the mean model fits well. Although the use of a model that exhibits substantial lack of fit is undesirable, it is occasionally unavoidable; in those unavoidable cases, one should clearly acknowledge evident lack of fit in reporting the results of the data analysis.

Although the primary focus is often on the mean response, a full dose-response model must also properly model the variability of the data around the mean. This variability is partly described by the nature of the probability distribution employed and partly by the model $\boldsymbol{\Sigma}(\boldsymbol{\omega})$ for the variance-covariance matrix.

As mentioned previously, the nature of the response (binary, count, or continuous) and experience often is sufficient for properly specifying an appropriate probability distribution. Nevertheless checking distributional assumptions is always useful. Histograms or *Q*-*Q* plots (Wilk and Gnanadesikan 1968) of residuals can reveal deviations from an assumed distributional shape, particularly for continuous responses. Formal procedures for testing distributional assumptions are available. Some, such as the Kolmogorov-Smirnov test or the Anderson-Darling test, are general purpose (Stephens 1974); others, such as the Shapiro-Wilks normality test, are directed toward particular distributions (Stephens 1974).

Building models for the variance is often a complex undertaking and well beyond the scope of this chapter – so the remarks here merely scratch the surface. As mentioned earlier, for distributions like the binomial or Poisson, the variance is a function of the mean, so the variance model is partly preset. Additional variance parameters are introduced only when the preset model proves inadequate for the data. When the normal distribution is the relevant probability model, often the default assumption is that the variance is constant and governed by a single parameter. Nonconstant variance is possible; it can sometimes be stabilized by a transformation of the response variable to achieve constant variance (Bates and Watts 1988; Bickel and Doksum 1977) or, alternatively, be modeled as a parametric function of dose. For complex experimental designs that involve more structure than a completely randomized design – such as multiple batches of experimental units or an experiment that is carried out in blocks over several days – modeling the variance often requires the use of several variance parameters. For example, the model might need a parameter for variance in response among units in a single batch and one for variance among batches. If the overall variability can be partitioned into such components, statistical analysis implements a so-called mixed model approach that allows simultaneous estimation of mean parameters and multiple variance parameters (Searle et al. 2006).

Another kind of departure from assumptions arises if observations are correlated instead of being independent as many models assume. Such correlations typically arise from block structure in the experimental design or from certain pre-analysis data manipulations. One common practice is to rescale a continuous response so that its limits are 0 and 1 on a probability scale or 0 and 100 on a percent scale. If the largest responses occur at dose zero, one might rescale by dividing all responses by the mean response among the negative controls so that on average the responses have an upper limit of 100% (Crofton et al. 2005; Hertzberg et al. 2013). Although such rescaling has intuitive appeal, in principle, dividing several responses by the same random quantity (average of negative control responses) induces correlations among them that contradict independence – because the rescaled responses all depend on the same mean response among controls. Particularly when the rescaling is done separately for different runs that are part of the same experiment (using

run-specific control means), the analysis should arguably account for this dependence. Again, mixed models can be constructed to account for correlations. Alternatively, with likelihood-based estimation methods, statisticians have developed alternative estimators for the standard errors of parameter estimates that adjust for such correlations. These estimators, known as "sandwich estimators" (Kauermann and Carroll 2001; Freedman 2006), are also useful in settings with heteroskedasticity. In situations where appropriate variance estimators are difficult to derive theoretically, bootstrap methods or other resampling methods can be used to estimate standard errors appropriately (Efron and Tibshirani 1993).

In evaluating model fit, it is important to recognize that all three components of the dose-response model play off one another. If the probability model assumes a symmetric distribution but the actual data distribution is skewed, the variance model may be (or appear to be) misspecified. If the mean function fails to fit the data well, that could also impinge on the diagnostics for evaluating the variance model or the distribution. In evaluating the aptness of a model, one must keep in mind that all three components work together.

8.6 Computational Issues

Many dose-response models used in toxicology require iterative computational methods for model fitting. Whether the criterion for best fit is least squares or maximum likelihood, estimation for models that are nonlinear in the parameters uses computational algorithms that approach the best fit incrementally through successive cycles of computation. The iterations stop when an additional cycle fails to improve the fitting criterion by a preset amount – in which case the algorithm is said to have converged. For maximum likelihood, one can think of the process as analogous to finding the highest point in a landscape (for least squares, the lowest); only the dimension of the "landscape" – which depends on the number of parameters in the model – is often higher than three. Also, the algorithms cannot just look around and see the highest point and head toward it; they must use clues available at the current location, such as steepness and uphill direction, to determine which way and how far to go for the next iteration. When the topology is complicated, the algorithms have trouble finding the maximum. A long and nearly flat ridge makes finding the maximum difficult. Sometimes an algorithm "falls off a cliff" and has difficulty climbing back. Multiple nearby peaks of different heights also make finding the unique maximum difficult. All these issues may lead to failure of an algorithm to converge.

One common cause for failure to converge is a design that is not well suited to the model at hand, as was more fully discussed earlier. Thus, an appropriate choice of the number and spacing of dose levels goes a long way toward avoiding convergence problems in model fitting. Even with a sound design, however, nonlinear models can be quirky to fit. Sometimes a model can be reparameterized so that the numerical properties of the fitting algorithms are improved while the trajectory of predicted

responses remains unchanged. Although the best way to parameterize a given model is not always obvious, some choices work better than others. A reparameterization of the Hill model to improve numerical performance was described earlier (see Sect. 8.4). In addition to the mean response, the overall model may involve one or more variance parameters, and reparameterizing the variance may also help in some situations. For example, rather than directly estimating the variance or even the standard deviation, the logarithm of the standard deviation may work better, possibly due to its scale being more similar to that of the mean parameters. Sometimes centering dose levels can enhance convergence and lessen variability of parameter estimates by reducing multicollinearity (Reeve and Turner 2013). All iterative algorithms must start at some initial set of guesses at the parameter values and proceed iteratively to improve those estimates – but convergence can be highly dependent on the choice of starting values. Ill-chosen values can lead to non-convergence. Moreover, in situations where the likelihood surface has multiple peaks, different starting values can lead to seemingly successful convergence to estimates that represent different local maxima – but the goal is to find the global maximum. Even if convergence appears to have been achieved, it is good practice to try multiple distinct starting values and confirm that all produce the same ultimate estimates. When fitting closely related dose-response curves simultaneously to multiple chemicals that differ widely in potency, one simple but effective step that frequently helps with convergence issues is to rescale the doses of each chemical, so corresponding parameter estimates will have similar magnitudes across chemicals. For example, in simultaneously fitting Hill models to two chemicals, the first with an ED_{50} near 0.005 mg/kg and the second with an ED_{50} near 5.0 mg/kg, convergence might be improved if the dose levels of the first chemical were rescaled to μg/kg, so both ED_{50} values were near 5.0 in their respective units. Of course, the resulting estimates and confidence limits could be reexpressed in any common units desired. Finally, there are usually several different computational algorithms that can be used to estimate the parameters of a given model. For example, the NLIN procedure in SAS offers four choices: steepest descent (or gradient), Newton, modified Gauss-Newton, and Marquardt. Some methods may work better than others for a given set of data, so if one algorithm fails to converge, try another. Also, most algorithms involve preset constants that control aspects of the algorithm; sometimes adjusting these "tuning parameters" helps with computational issues.

8.7 Summary

Dose-response modeling is an important data analysis tool throughout toxicology, particularly so in evaluating chemical mixtures. This chapter provides a broad introductory overview to statistical issues that arise in studying dose-response relationships. Statistical dose-response models consist of a probability distribution for the response, a function $f(d|\boldsymbol{\theta})$ to describe the relationship between the mean response and dose, and a model $\boldsymbol{\Sigma}(\boldsymbol{\omega})$ for the variance-covariance matrix of the data.

The chapter discusses the role of statistical study design, including the roles of dose placement and of properly accounting for multiple sources of variation or correlations among observations, in achieving accurate and precise parameter estimation and efficient hypothesis testing. It discusses the choice of a functional form for $f(d|\boldsymbol{\theta})$ and describes strategies for examining the adequacy of the proposed dose-response model. Finally, the chapter considers some ways to cope with the failure of iterative model-fitting algorithms to converge to a unique solution. Careful attention to study design and the use of statistical models that are appropriate for the data at hand are critical for achieving the best possible results from dose-response studies.

Acknowledgments This work was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

References

Agresti, A. 2013. *Categorical data analysis*. 3rd ed. Hoboken: Wiley.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6): 716–723.

Altenburger, R., T. Backhaus, W. Boedeker, M. Faust, M. Scholze, and L.H. Grimme. 2000. Predictability of the toxicity of multiple chemical mixtures to vibrio fischeri: Mixtures composed of similarly acting chemicals. *Environmental Toxicology and Chemistry* 19 (9): 2341–2347.

Bates, D.M., and D.G. Watts. 1988. *Nonlinear regression analysis and its applications*. New York: Wiley.

Berenbaum, M.C. 1985. The expected effect of a combination of agents: The general solution. *Journal of Theoretical Biology* 114: 413–431.

Bernstein, L., J. Kaldor, J. McCann, and M.C. Pike. 1982. An empirical approach to the statistical analysis of mutagenesis data from the Salmonella test. *Mutation Research* 97: 267–281.

Bickel, P.J., and K.A. Doksum. 1977. *Mathematical statistics: Basic ideas and selected topics*. San Francisco: Holden-Day.

Box, G.E.P., and D.R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society, Series B* 26 (2): 211–252.

Box, G.E.P., and G.C. Tiao. 1992. *Bayesian inference in statistical analysis*. New York: Wiley.

Breslow, N.E. 1984. Extra-poisson variation in log-linear models. *Applied Statistics* 33 (1): 38–44.

Burnham, K.P., and D.R. Anderson. 2002. *Model selection and multimodel inference: A practical information-theoretic approach*. 2nd ed. New York: Springer.

Carlin, B.P., and T.A. Louis. 2000. *Bayes and empirical Bayes methods for data analysis*. 2nd ed. Boca Raton: Chapman & Hall.

Casey, M., C. Gennings, W.H. Carter, V.C. Moser, and J.E. Simmons. 2004. Detecting interaction(s) and assessing the impact of component subsets in a chemical mixture using fixed-ratio mixture ray designs. *Journal of Agricultural, Biological, and Environmental Statistics* 9 (3): 339–361.

Christensen, E.R., and N. Nyholm. 1984. Ecotoxicological assays with algae: Weibull dose-response curves. *Environmental Science & Technology* 18 (9): 713–718.

Cook, R.D., and C.L. Tsai. 1985. Residuals in nonlinear regression. *Biometrika* 72 (1): 23–29.

Crofton, K.M., E.S. Craft, J.M. Hedge, C. Gennings, J.E. Simmons, R.A. Carchman, W.H. Carter, and M.J. DeVito. 2005. Thyroid-hormone-disrupting chemicals: Evidence for dose-dependent additivity or synergism. *Environmental Health Perspectives* 113 (11): 1549–1554.

Dette, H., N. Neumeyer, and K.F. Pilz. 2005. A note on nonparametric estimation of the effective dose in quantal bioassay. *Journal of the American Statistical Association* 100 (470): 503–510.
Dinse, G.E., and D.M. Umbach. 2011. Characterizing non-constant relative potency. *Regulatory Toxicology and Pharmacology* 60: 342–353.
Efron, B., and R. Tibshirani. 1993. *An introduction to the bootstrap*. Boca Raton: CRC Press.
EPA (Environmental Protection Agency). 2016. The ToxCast analysis pipeline: An R package for processing and modeling chemical screening data. Date of access: 12 December 2017. https://www.epa.gov/sites/production/files/2015-08/documents/pipeline_overview.pdf.
Fang, Q., W.W. Piegorsch, S.J. Simmons, X. Li, C. Chen, and Y. Wang. 2015. Bayesian model-averaged benchmark dose analysis via reparameterized quantal-response models. *Biometrics* 71 (4): 1168–1175.
Finney, D.J. 1971. *Probit analysis*. Cambridge: Cambridge University Press.
Freedman, D.A. 2006. On the so-called "Huber sandwich estimator" and "robust standard errors". *The American Statistician* 60 (4): 299–302.
Guardabasso, V., D. Rodbard, and P.J. Munson. 1987. A model-free approach to estimation of relative potency in dose-response curve analysis. *The American Journal of Physiology* 252 (3): E357–E364.
Guardabasso, V., P.J. Munson, and D. Rodbard. 1988. A versatile method for simultaneous analysis of families of curves. *The FASEB Journal* 2 (3): 209–215.
Harrell, F.E. 2001. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer.
Haseman, J.K., and M.D. Hogan. 1975. Selection of the experimental unit in teratology studies. *Teratology* 12 (2): 165–171.
Hertzberg, R.C., Y. Pan, R. Li, L.T. Haber, R.H. Lyles, D.W. Herr, V.C. Moser, and J.E. Simmons. 2013. A four-step approach to evaluate mixtures for consistency with dose addition. *Toxicology* 313: 134–144.
Hill, A.V. 1910. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *The Journal of Physiology* 40 (Suppl): iv–vii.
Hunt, D.L., and D. Bowman. 2004. A parametric model for detecting hormetic effects in developmental toxicity studies. *Risk Analysis* 24 (1): 65–72.
Kauermann, G., and R.J. Carroll. 2001. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 96 (456): 1387–1396.
Kelly, C., and J. Rice. 1990. Monotone smoothing with application to dose-response curves and assessment of synergism. *Biometrics* 46 (4): 1071–1085.
Kim, S.B., S.M. Bartell, and D.L. Gillen. 2016. Inference for the existence of hormetic dose-response relationships in toxicology studies. *Biostatistics* 17 (3): 523–536.
Margolin, B.H., N. Kaplan, and E. Zeiger. 1981. Statistical analysis of the Ames Salmonella/microsome test. *Proceedings of the National Academy of Sciences of the United States of America* 78 (6): 3779–3783.
Montgomery, D.C., E.A. Peck, and G.G. Vining. 2012. *Introduction to linear regression analysis*. 5th ed. Hoboken: Wiley.
Mood, A.M., and F.A. Graybill. 1963. *Introduction to the theory of statistics*. 2nd ed. New York: McGraw Hill.
Nottingham, Q.J., and J.B. Birch. 2000. A semiparametric approach to analyzing dose-response data. *Statistics in Medicine* 19: 389–404.
Piegorsch, W.W., and J.K. Haseman. 1991. Statistical methods for analyzing developmental toxicity data. *Teratogenesis, Carcinogenesis, and Mutagenesis* 11: 115–133.
Ramsay, J.O. 1988. Monotone regression splines in action. *Statistical Science* 3 (4): 425–461.
Reeve, R., and J.R. Turner. 2013. Pharmacodynamic models: Parameterizing the Hill equation, Michaelis-Menten, the logistic curve, and relationships among these models. *Journal of Biopharmaceutical Statistics* 23: 648–661.
Schwarz, G.E. 1978. Estimating the dimension of a model. *Annals of Statistics* 6 (2): 461–464.
Searle, S.R., G. Casella, and C.E. McCulloch. 2006. *Variance components*. New York: Wiley.

Seber, G.A.F. 1977. *Linear regression analysis*. New York: Wiley.
Seber, G., and C. Wild. 1989. *Nonlinear regression*. New York: Wiley.
Shockley, K.R. 2016. Estimating potency in high-throughput screening experiments by maximizing the rate of change in weighted Shannon entropy. *Scientific Reports* 6: 27897. https://doi.org/10.1038/srep27897.
Simmons, S.J., C. Chen, X. Li, Y. Wang, W.W. Piegorsch, Q. Fang, B. Hu, and G.E. Dunn. 2015. Bayesian model averaging for benchmark dose estimation. *Environmental and Ecological Statistics* 22: 5–16.
St. Laurent, R.T., and R.D. Cook. 1992. Leverage and superleverage in nonlinear regression. *Journal of the American Statistical Association* 87 (420): 985–990.
St. Laurent, R.T., and R.D. Cook. 1993. Leverage, local influence and curvature in nonlinear regression. *Biometrika* 80 (1): 99–106.
Stephens, M.A. 1974. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association* 69 (347): 730–737.
Wheeler, M.W., and A.J. Bailer. 2007. Properties of model-averaged BMDLs: A study of model averaging in dichotomous response risk estimation. *Risk Analysis* 27 (3): 659–670.
———. 2008. Model averaging software for dichotomous dose response risk estimation. *Journal of Statistical Software* 26 (5): 1–15. https://doi.org/10.18637/jss.v026.i05.
———. 2009. Comparing model averaging with other model selection strategies for benchmark dose estimation. *Environmental and Ecological Statistics* 16: 37–51.
Wilk, M.B., and R. Gnanadesikan. 1968. Probability plotting methods for the analysis of data. *Biometrika* 55 (1): 1–17.
Williams, D.A. 1982. Extra-binomial variation in logistic linear models. *Applied Statistics* 31 (2): 144–148.
Zorrilla, E.P. 1997. Multiparous species present problems (and possibilities) to developmentalists. *Developmental Psychobiology* 30 (2): 141–150.

Chapter 9
Predicting Mixture Toxicity with Models of Additivity

Cynthia V. Rider, Gregg E. Dinse, David M. Umbach, Jane Ellen Simmons, and Richard C. Hertzberg

Abstract Researchers in numerous fields (e.g., pharmacology, entomology, toxicology, and epidemiology) have attempted to model the joint action of chemicals using simple formulas based only on knowledge of individual chemical toxicity or pharmacological effect (i.e., dose-response relationships). Collectively, these formulas are referred to as "additivity models," and they are based on concepts of additivity that include dose addition, independent action, integrated addition, and effect summation. In toxicology, additivity-based predictions are often compared to observed mixture data to assess the presence and magnitude of interactions (greater-than-additive or less-than-additive) among chemicals. These models can also be used to estimate the toxicity of a defined mixture for comparison to the observed toxicity of a related, but more complex, mixture. Alternatively, additivity models have been used to explore mechanisms of joint action. In general, the steps for investigating joint toxicity using additivity models include (1) deciding on which additivity model(s) to apply (e.g., dose addition, independent action, or

Disclaimer The views expressed in this chapter are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

C. V. Rider (✉)
National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA
e-mail: cynthia.rider@nih.gov

G. E. Dinse
Public Health Sciences, Social & Scientific Systems, Inc., Durham, NC, USA

D. M. Umbach
Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA

J. E. Simmons
National Health and Environmental Effects Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

R. C. Hertzberg
Emory University, Atlanta, GA, USA

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_9

both), (2) collecting dose-response data on individual chemicals and the mixture, (3) incorporating individual chemical data in an additivity model to generate predictions, and (4) comparing predicted to observed mixture responses. Many of the additivity models have a long and sometimes controversial history. This chapter provides background on several of the common additivity models, illustrates their application with examples, and discusses their advantages and limitations.

Keywords Joint action · Dose addition · Loewe additivity · Independent action · Response addition · Bliss independence · Integrated addition · Effect summation · Component-based approaches

9.1 Introduction

The toxicity of a mixture can be evaluated by treating the whole mixture as a single chemical or by investigating how components in the mixture contribute to the mixture's toxicity. Each approach has advantages and disadvantages. While whole mixture testing can offer a straightforward experimental design (e.g., one test article), the major impediment lies in relating the toxicity results to real-world exposures or to other mixtures of interest. Specifically, challenges include selecting the whole mixture for study, generating the test article in concentrated form in a volume sufficient for testing, and relating findings to the vast number of potentially similar whole mixtures. Since the actual toxicological evaluation of the whole mixture is the same as for single chemicals, the whole mixture approach will not be discussed in this chapter. Challenges in whole mixture research relating to test article selection and generation of test material can be found in the literature (Simmons et al. 2008; Pressman et al. 2010), and approaches for comparing across whole mixtures are discussed in a risk assessment context in Chap. 15. The focus of this chapter is on understanding mixture toxicity through evaluation of the components of the mixture. This chapter discusses approaches for using individual chemical data in simple mathematical models to predict mixture effects.

The first step in understanding the joint action of two or more chemicals (i.e., how they behave when they co-occur in an organism) is to formulate a testable null hypothesis based on what is known about the individual chemicals and to evaluate it empirically. In studying additivity, this hypothesis typically involves a key assumption that the chemicals will abide by a defined model of joint action (e.g., dose addition, independent action) and do not interact (pharmacokinetically or pharmacodynamically) to alter the response expected under the model. In other words, the null hypothesis of additivity describes the "no-interaction," "zero-interaction," or "baseline" situation.

Empirical evaluation of a null hypothesis of additivity requires data. For the additivity models considered here, the required data usually consist of dose-response data for each individual chemical in the mixture as well as dose-response data for one or more mixtures, each defined by specified proportions of the individual chemicals. Also, the null hypothesis of additivity is generally evaluated

for a single specified response; or, when several specific responses are of simultaneous interest, additivity is evaluated for each response separately.

Of course, the totality of an organism's responses to a tested mixture might include characteristics that are biologically distinct from those seen with any of the component chemicals. Such qualitatively different responses could be caused when the components combine chemically to form a new chemical that differs from all of the components, for example, the potent carcinogen *N*-nitrosomorpholine can be formed from coadministration of nitrogen dioxide and morpholine (Van Stee et al. 1995). Alternatively, qualitatively different responses can occur when a pharmacokinetic interaction among chemicals leads to the formation of a toxic metabolite that would not otherwise be formed by any of the components (Dobrev et al. 2002). Toxicodynamic interactions could also alter physiological behavior or histological structure so that new effects become prominent. Although the absence of dose-response data on these responses for the individual chemicals makes formal evaluation of the additivity null hypothesis problematic, the appearance of those novel responses only with exposure to the mixture and not to the individual chemicals would provide prima facie evidence of departure from additivity.

In general, evaluation of the null hypothesis of additivity proceeds by (1) using the individual component chemical dose-response data together with a specific additivity model to predict the expected response at various doses of the mixture and (2) comparing those predictions under additivity to observed dose-response data for the mixture. At least four kinds of outcomes are possible (Fig. 9.1):

- The responses of the tested mixture are close enough to predictions under additivity that they support the stated additivity hypothesis.
- The responses of the tested mixture occur at lower doses than predicted under additivity, contradict the stated additivity hypothesis, and indicate a potential greater-than-additive interaction among the component chemicals.

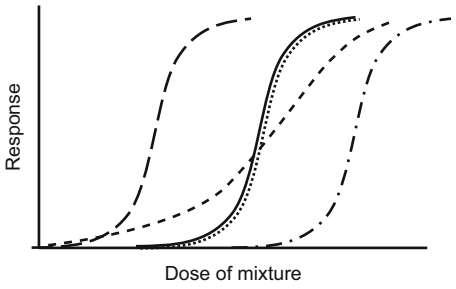

Fig. 9.1 Comparison of predicted and observed mixture responses. The solid black line represents predicted effects based on an assumption of additivity (a.k.a. "no-interaction" scenario). The dotted line represents observed mixture data that conform to additivity-based predictions. The long-dashed line represents mixture data that deviate from additivity, displaying greater-than-additive toxicity. The dash-dot line represents mixture data that deviate from additivity, displaying less-than-additive toxicity. Finally, the short-dashed line represents mixture data that deviate from additivity, displaying greater-than-additive toxicity in the low-dose range and less-than-additive toxicity in the high-dose range

- The responses of the tested mixture occur at higher doses than predicted under additivity, contradict the stated additivity hypothesis, and indicate a potential less-than-additive interaction among the component chemicals.
- The responses of the tested mixture occur at lower doses than predicted under additivity at some dose levels and higher doses than predicted under additivity at other dose levels, contradict the stated additivity hypothesis, and indicate potential interactions that are either greater-than-additive or less-than-additive, depending on the dose level.

Selection of an appropriate model is complicated by the fact that there are multiple additivity models, often with numerous versions, available in the mixture literature. Furthermore, it is important to note that deviation from an additivity prediction is not enough to confirm the presence of an interaction among mixture constituents. For example, an inappropriate additivity model may have been applied – a dose-addition model was used when the chemicals displayed independence. Conversely, consistency with model predictions does not necessarily rule out the presence of greater-than-additive or less-than-additive interactions. One can imagine, for example, that the presence of greater-than-additive interactions among some mixture components and less-than-additive interactions among others could counteract each other in some sense and result in responses that approximate additivity. Also, as determination of consistency with or deviation from a model is based on statistical comparisons between the predicted and observed responses, some real deviations from additivity may not be detected due to inadequacy of experimental design, including insufficient power (See Chap. 13), and some detected deviations may represent statistical false positives.

As suggested above, no definition of additivity is universally accepted. Instead, there are multiple definitions of additivity that can be separated into two basic categories: (1) additivity defined through formulas involving dose levels (e.g., dose additivity) and (2) additivity defined through formulas involving response levels (e.g., independent action and effect summation). There are also integrated-addition models, which combine the concepts of dose addition and independent action. For dose addition, multiple mathematical approaches are available for calculating the predicted response of the mixture from dose-response data on individual chemical constituents. This chapter will focus on the different concepts of additivity and the varied methods that have been used to predict mixture toxicity under an assumption of additivity. It will present the history, assumptions, current applications, and relevance of four general types of additivity (viz., dose addition, independent action, effect summation, and integrated addition).

Many factors must be considered when tested mixtures result in responses that deviate from those predicted under additivity. The first set of factors involves the series of decisions made in the modeling process. Typical examples include the quality of the individual chemical data, the dose-response modeling of the individual chemical data (Chap. 8), and the selection of the additivity model (e.g., was independent action applied, when a dose-addition model would have been more appropriate). The second group of factors relates to the simplicity of the models

versus the complexity of biology. The models described in this chapter are simple mathematical models, whereas the biological responses that are being evaluated often involve complex, interdependent signaling networks. The additivity models represent a reductionist approach that may provide different results depending on the model system and endpoints being evaluated. Additionally, chemicals can interact with biological systems in complex ways that are not captured by the models (e.g., a single chemical can activate multiple adverse outcome pathways).

Despite these complications, the models of additivity discussed in this chapter offer a starting place for understanding the joint action of chemicals. They can be used to address important issues in toxicology ranging from limiting the potential toxicity of combination therapies to informing the process of cumulative risk assessment. Sometimes, simple models of additivity will be adequate for describing mixture responses and will support continued application of the models for predicting the toxicological consequences of exposure to mixtures. Other times, information gained from applying different additivity models, such as patterns of deviation from the model(s), could inform the development of more sophisticated models of mixture toxicity that would better integrate component chemical dose-response data. Regardless, targeted mechanistic studies can be designed based on results from comparing observed mixture responses to predictions under additivity to further characterize and understand the joint action of chemical mixtures.

Displaying dose-response relationships for mixtures graphically can be useful for visualizing some key mixture concepts (Greco et al. 1995). For a single chemical, a dose-response relationship is typically depicted graphically as a curve with responses on the vertical axis and dose levels on the horizontal axis. With binary mixtures, graphs of the dose-response relationship that depict the responses and the dose levels of both chemicals are called response surfaces and require three dimensions, a vertical axis for response and two perpendicular "horizontal" axes for the dose levels (dose plane). (This idea generalizes mathematically to higher-order mixtures, but visualization is inconvenient at best.) Two-dimensional representations of features of response surfaces are often used. For binary mixtures with a fixed ratio of the two constituent chemicals, increasing doses of the mixture extend along a ray that lies in the dose plane and starts at the origin (zero dose of the mixture). A plane that is perpendicular to the dose plane and that contains the fixed-ratio ray cuts through the response surface; the set of points at the intersection of that perpendicular plane and the response surface is the dose-response curve for that fixed-ratio mixture. Treating the mixture as a single "chemical," the mixture's dose-response curve can be displayed in two dimensions just like any dose-response curve for a single chemical. One can also cut the dose-response surface with a plane parallel to the dose plane at any selected response level; this parallel plane also has the dose levels as its axes. The intersection of such a parallel plane with the dose-response surface is called an isobole: a line or curve that connects equi-effective doses on a graph whose axes are the dose levels of each component chemical. Thus, an isobole is always associated with a specified response level. A graph of isoboles is called an isobologram. See Greco et al. (1995) for a more thorough explanation and illustration of these ideas. As described later, isobolograms have been used widely for evaluating the additivity of binary mixtures.

9.2 Toxicological Similarity and Additivity Model Selection

The degree of similarity in toxicological responses to chemicals has generally served as the basis for selection of an appropriate additivity model, with dose additivity traditionally being applied to chemicals that elicit "similar" responses and independent action being applied to chemicals that elicit "dissimilar" responses. However, the terms "similar" and "dissimilar" have had varied interpretations. The spectrum of toxicological similarity among chemicals, along with examples of chemicals that correspond to each level, is presented in Fig. 9.2. As chemicals move down the spectrum of similarity, scientific support for applying the concept of dose addition to predict combined effects becomes more tenuous.

As noted in Fig. 9.2, chemicals with the highest degree of similarity share a common active metabolite (assuming that the parent compounds require metabolic activation). In this case, the metabolite responsible for initiating the cascade of responses is identical among chemicals. It follows that the sequence of biochemical steps resulting from the molecular initiating event is identical between all chemicals that share the active metabolite. Any difference in potency among chemicals could be attributed to differences in the rate and extent to which the chemicals are converted to the active metabolite and would be reflected in the individual chemical dose-response relationships.

The next level of similarity involves chemicals that elicit toxicity through a common molecular initiating event. There are many classes of chemicals that meet this criterion, for example, organophosphate pesticides that share inhibition of acetylcholinesterase as the molecular initiating event that leads to downstream adverse nervous system effects (Mileson et al. 1998). Toxicological similarity at the level of shared molecular initiating event has provided the conceptual basis for dose addition. The concept holds that chemicals displaying similarity at the molecular initiating event level act as dilutions or concentrations of each other, thus

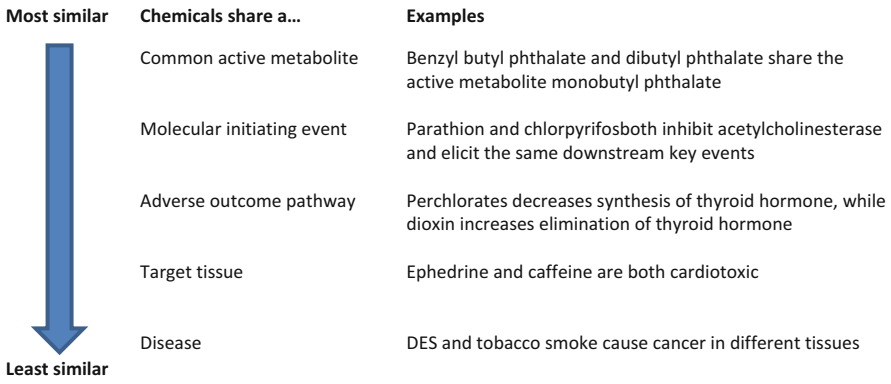

Fig. 9.2 Continuum of toxicological similarity

eliciting the same toxic effects by the same toxicological pathways, once dose is scaled for that "dilution" or potency factor (Cedergreen 2014; Hertzberg et al. 2013; Loewe and Muischnek 1926). The dilution concept, first described by Bliss (1939), can be empirically tested by determining whether the chemical components and the mixture share a common dose-response model, except for scaling the dose (Meadows et al. 2002), and in the most complete dilution concept, by determining whether the response variances depend only on the response mean and otherwise are the same across all chemical components and the mixture (Hertzberg et al. 2013).

It is at the level of adverse outcome pathway that the concepts of toxicological similarity and independence begin to blur. Chemicals can have different molecular initiating events that converge at any of a number of key events in their adverse outcome pathways (Chap. 7). Moving one step further, chemicals can act through different signaling pathways that intersect at the target tissue. Although a common target tissue among chemicals can be one of a number of factors used to build a "weight of evidence" case for toxicological similarity, there is continued debate on using shared target tissue alone as the basis for applying dose addition. Many chemicals have notably different adverse outcome pathways yet display toxicity at the same tissue (see Sect. 3.3). Determining how the degree of toxicological similarity of chemicals in a mixture relates to predicted mixture toxicity from available additivity models is an active area of research and has important implications for cumulative risk assessment (Chap. 14).

Regarding the concept of toxicological similarity, it is important to note that evidence for a concept is not the same as the lack of evidence against a concept. Evidence supporting independent action has often involved demonstration of differing adverse outcome pathways for the most important effects. Evidence supporting similar adverse outcome pathways has usually been interpreted as support for toxicological similarity in general and thus for the dose additivity model. For weakly studied mechanisms, a similar conclusion has been drawn based on the lack of evidence for multiple differing adverse outcome pathways, i.e., inability to reject toxicological similarity. Similarly shaped dose-response curves (generalized parallelism) have also been interpreted as indicating similar adverse outcome pathways, and thus dose addition, yet counterexamples exist (Hertzberg et al. 2013). Specifically, a "lack of a different slope cannot be taken as a proof for a similar mode of action" (Altenburger et al. 2005). The decision about sufficient toxicological similarity for inclusion in a similarity group is a judgment and thus requires involvement of experts in toxicology and data analysis.

Finally, evaluating the evidence for toxicological similarity necessarily involves consideration of data quality. Different evidence streams can support determination of similarity, including structural similarity among chemicals, consistency of toxicological profile, and similar findings in mechanistic studies used to build adverse outcome pathways. For risk assessment purposes, frameworks have been developed to make decisions on including chemicals into a toxicological similarity group (U.S. EPA 1999). Structural comparisons (e.g., quantitative structure activity relationships) have been used for a long time, but many counterexamples exist

regarding toxic effect, so structure alone is insufficient for defining a similarity group. For any chemical's toxicological profile, key elements include the target organ(s), adverse effects/apical outcome(s), and pharmacokinetic properties, including whether the parent or metabolite is the toxic form. The preferred information on a toxicological profile involves effects related to specific molecular targets (e.g., particular enzyme or other protein, hormone) or target tissues (e.g., thyroid, blood) as similar specific effects (e.g., receptor-mediated effects) are more likely to follow a common toxicity pathway than nonspecific toxicity (e.g., narcosis). Knowledge of the adverse outcome pathway and related pharmacokinetics is the best information for judging toxicological similarity; however, these data are often unavailable or incomplete.

In contrast to toxicological similarity and dose addition, most of the classical literature that presents component-based formulas for joint toxicity includes minimal biological justification for independent-action models. The most common justification is evidence of important differences in toxicological mode of action or mechanism; in some cases, that evidence is only of effects occurring in different organs. In a review comparing concentration-addition (CA) and independent-action (IA) formulas, the conclusion was that little biological clarity has been proposed: "The CA model has some theoretical underpinnings and some experimental data supporting its use for mixtures of chemicals with the same site of action. In contrast, IA is not as well grounded in theory, nor is it unambiguously supported by data" (Cedergreen et al. 2008).

The mechanistic or functional argument for independence should be based on data and concepts showing that an individual chemical's toxic function, dose-response relationship, and adverse outcome pathway are not altered by co-exposure to the other chemical(s) in the mixture. Because toxicological interactions are known to involve one or more key event(s) from exposure to apical (adverse) effect, an evaluation of independence should include as many of those steps as possible. For chemicals, the following processes have been identified as involved in toxicological interaction: contact, uptake, in vivo chemical reactions, absorption, distribution, excretion, metabolism, receptor site (antagonism), receptor function (antagonism), and DNA binding (Mumtaz and Hertzberg 1993; Mumtaz et al. 1993). Evidence only of a difference in toxicodynamics, i.e., in toxicological mode of action, thus provides fairly weak support for toxicological independence. For example, independence might not be plausible when adverse outcome pathways overlap for secondary effects: "commonality of biological/biochemical events that may not be part of the recognized toxic mode or mechanism of action of one mixture component can lead to unanticipated interactions and consequences" (Lambert and Lipscomb 2007). Such information is not commonly available. Usually independence and the independent-action formula are assumed when there is insufficient evidence to establish toxicological similarity.

Finally, the experimental endpoint that is selected in a mixture evaluation (i.e., measured response) can also influence the determination of similarity versus independence. The specificity of the endpoint measured could increase the likelihood of classifying chemicals as toxicologically independent. Consider the

difference between an in vitro assay measuring binding to a hormone receptor and a downstream functional measure of the system in which the hormone operates as a signaling molecule. In the case of the in vitro binding assay, there is a limited capacity for interactions between the chemicals and the system. Instead, there is a focus on a single, clearly defined mechanism. In contrast, the functional measure incorporates multiple mechanisms and potential avenues for interaction among chemicals and complex biological systems.

9.3 Dose Addition

The phrases "dose addition" and "concentration addition" represent the same concept, with "dose addition" used more frequently in mammalian toxicology and "concentration addition" commonly associated with ecotoxicology. For simplicity, the term "dose" will be used throughout this chapter. As indicated by its name, dose addition is a concept that defines or predicts the additive joint action of chemical mixtures by imposing a constraint on a weighted sum of the component-specific dose levels (see below). Dose addition is a single concept, but it has numerous mathematical representations. A general description of the history of dose addition is presented below, followed by highlights of recent advances to dose-addition modeling. For a more thorough review of work mentioned in the historical section and other methods developed prior to 1995, see the review by Greco et al. (1995).

9.3.1 Background

The first scientific publications on the effects of chemical combinations come from the field of pharmacology and its consideration of the joint action of drugs. Although often cited as providing the first description of the concept of dose additivity, Loewe and Muischnek (1926) refer to earlier work by Emil Bürgi and interpret the "Bürgi rule" as a reference to joint chemical effects described by an isobole (Bürgi 1912). Loewe and Muischnek also point out the inconsistent use of mixture-related terminology including synergism, antagonism, addition, and potentiation in the literature of the time; disagreement regarding these terms continues today (Kodell and Pounds 1991; Hertzberg and MacDonell 2002; Greco et al. 1992; Simmons 2013). In their seminal paper, Loewe and Muischnek (1926) present a spectrum of possible effects of a binary combination of chemicals through the use of isobolograms; in particular, they illustrate what we now call dose addition (then referred to as non-varying joint effects). This classic work presents a conceptual framework for addressing chemical combinations, but it does not describe applications of these concepts to experimental data.

The isobologram is a simple but effective graph for illustrating the concept of dose addition. Suppose that the horizontal and vertical axes of a graph represent the component doses of a binary mixture. Loewe and Muischnek proposed that the two chemicals are dose-additive if the isobole for a given response level is a diagonal line (with a negative slope) connecting the dose levels on the two axes that elicit that given response. For example, if chemical A elicits a 50% response at dose *a*, chemical B elicits a 50% response at dose *b*, and the straight line connecting points (*a*, 0) and (0, *b*) covers every dose combination of chemicals A and B that elicit a 50% response, then A and B are dose-additive at the 50% response level (Fig. 9.3). Loewe and Muischnek compared dose addition to other concepts: (1) deviation from dose additivity (e.g., greater than dose-additive, less than dose-additive); (2) independence of the mechanism that results in the same measured effect, i.e., exposure to each chemical results in the same apical outcome (e.g., death), but in each case, the outcome is achieved by a different underlying causative series of events (see Sect. 4.1); and (3) joint effects that are not elicited by either chemical alone but only by the combination (Fig. 9.3). Throughout their manuscript, Loewe and Muischnek described many key issues that continue to plague the field of mixture toxicology. They touched on the difficulty in distinguishing independent action of chemicals from "antagonism" (less than dose-additive interaction). They

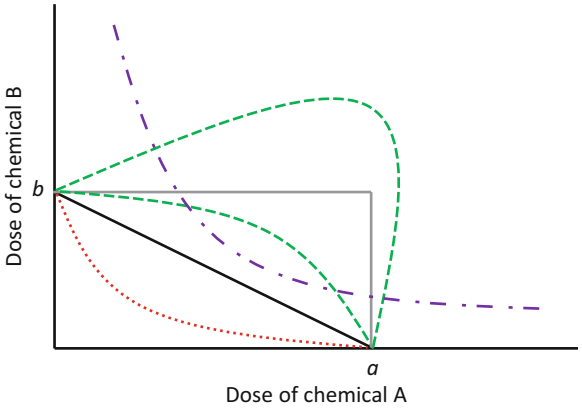

Fig. 9.3 Isobologram illustrating possible effects of a binary combination of chemicals (adapted from Loewe and Muischnek (1926)). Points *a* and *b* (with coordinates (*a*,0) and (0,*b*)) represent doses of chemicals A and B, respectively, that elicit equivalent effect levels (e.g., doses that elicit an effect that is 50% of the maximum response or the ED50). Each line/curve in the figure is an isobole, meaning that any combination of the concentrations of the two chemicals on that curve will result in the same effect level. The straight (black) diagonal line connecting points *a* and *b* is an isobole for two chemicals that are dose-additive. The dotted (red) curve represents a scenario where the combination of chemicals A and B results in effects that are greater-than dose-additive, and the two dashed (green) curves provide examples of less-than dose-additive scenarios. The dash-dot (purple) curve is an example of a combination of chemicals A and B eliciting effects not produced by any dose of either chemical alone. The angular isobole (perpendicular gray line segments) corresponds to a mixture where chemicals A and B act independently, and susceptibility to A is perfectly positively correlated with susceptibility to B (see Sect. 4.1)

also highlighted the difficulties in describing complex biological processes using simple mathematics, stating eloquently that "Biology does not know such simple and constant relationships." Furthermore, they acknowledged the need to incorporate experimental variability into evaluations of mixtures. Finally, Loewe and Muischnek suggested ways to quantify deviations from additivity. It is not surprising that this early, insightful work on mixtures has been so influential in shaping our current understanding and practice of mixture toxicology.

Bliss (1939), emerging from the fields of entomology and statistics and focusing on pesticide efficacy, provided another significant contribution to the field. Bliss described three possible joint action scenarios: "independent joint action" (a.k.a., independent action or response addition), "similar joint action" (dose addition), and "synergistic action" (encompassing greater-than- or less-than-additive scenarios). Although Bliss was primarily cited in reference to the origins of independent action (discussed in greater detail later in this chapter), his discussion of the joint action of chemicals was wide-ranging with many important observations. Regarding dose addition, he articulated more specific requirements for chemicals expected to exhibit dose-additive toxicity. In particular, Bliss argued that they should have the same mechanism of action ("act upon the same system of receptors" in producing the outcome of interest), identical susceptibilities (in a population context) to each chemical, and parallel dose-response curves, so that the only difference between chemicals would be in their potencies. Bliss noted that mixtures that are dose-additive "have a greater expected potency than those" that are response-additive. In other words, the dose-addition model predicts greater mixture toxicity than the independent-action model for the same component doses. In studies that compare results from dose-addition and independent-action models, there is a general trend of dose addition providing a higher predicted toxicity than independent action (Belden et al. 2007; Altenburger et al. 1996). However, there are exceptions to this observed trend (Cedergreen et al. 2008), and it has been demonstrated that the shapes of the individual chemical dose-response curves are critical in determining whether dose addition or independent action results in greater predicted toxicity (Christensen and Chen 1985; Drescher and Boedeker; 1995). To facilitate visualization of combination effects, Bliss proposed linearizing the dose-response data using a log-dose probit function and provided an example using data from experiments measuring mortality in house flies exposed to combinations of rotenone and pyrethrin (Bliss 1939). In that paper, Bliss focused on the "dosage-mortality curve" and thus defined toxic response only as the fraction of the exposed population dying. Other types of toxic response, such as continuous measurements of physiological function, are considered in later sections of this chapter.

Finney (1942) further developed many of the concepts proposed by Bliss (1939). In his work, Finney sought to more clearly define statistical methods for analyzing mixture data to obtain predictions based on an assumption of either independent action or dose addition. Finney demonstrated that the proposed methods can be applied to mixtures which contain more than two chemicals (Finney 1942). Drawing notable distinctions from Bliss (1939) and Finney (1942), Hewlett and Plackett published a body of work on models to describe joint action (Hewlett and Plackett

1957, 1959; Plackett and Hewlett 1952, 1963). In particular, they pointed out that individual chemicals need not have parallel dose-response relationships to be consistent with dose addition (Hewlett and Plackett 1959).

The next major contribution to the dose-addition literature was offered by Chou and Talalay in a series of highly cited papers (Chou and Talalay 1977, 1981, 1983, 1984). Their approach is based on the median effect principle, which was derived from the mass action law and has been applied mechanistically in biochemistry and pharmacology. For any given dose and corresponding fraction responding, the median effect principle focuses on two ratios: the ratio of the fraction responding to the fraction not responding and the ratio of the specified dose to the median effective dose (i.e., the ED50 or dose that elicits a 50% response). The first ratio is assumed to equal the second ratio raised to some power, where that power reflects the shape of the dose-response curve. The method relies on calculation of a combination index (CI), where CI $= 1$ indicates dose-additive effects, CI < 1 indicates greater-than-additive effects, and CI > 1 indicates less-than-additive effects. The CI is calculated by first fitting data from each chemical alone and a fixed ratio combination of the chemicals to the "median-effect equation" derived by Chou (1976), which is a rearrangement of the Hill equation, of which the Michaelis-Menten equation is a special case (Chou 2010). The CI value is then calculated from the resulting parameter estimates. Software was developed to facilitate the use of the approach, with the currently available CompuSyn (http://www.combosyn.com/index.html) representing the third generation of the software (Chou 2010). Greco et al. (1995) provide a thorough review and critique of the CI of Chou and Talalay, while Chou has more recently (2010) provided a review with recommendations for successful application of the method.

Contemporaneously with Chou and Talalay, Berenbaum presented another model for dose additivity (Berenbaum 1977, 1985, 1989). Berenbaum employed hypothetical "sham combinations" – combining a chemical with dilutions of itself – to articulate the dose-addition concept (Berenbaum 1989). In contrast to work by Loewe and Muischnek (1926) and Bliss (1939) described above, Berenbaum proposed that dose addition should be used as a general empirical method to describe the joint action of noninteracting chemicals, without requiring toxicological similarity among mixture constituents. Notably, Berenbaum stated that the only requirement is that the dose-response relationships for the individual chemicals are known over an "adequate range." The functions describing those relationships are unconstrained and are not required to be consistent across mixture constituents. On the topic of non-monotonic dose-response curves, Berenbaum posited that the proposed method can be used in cases of non-monotonicity observed in individual chemical dose-response curves while acknowledging that additional uncertainty may be associated with the analysis (Berenbaum 1985). He claimed that the proposed "general solution" put forth in his 1985 paper is a major departure from alternative models for describing combination effects, which require that "all the agents in the combination show similar dose-effect relations of the appropriate type" in the sense of similar functions and/or similarly shaped dose-response

curves. Many of those other approaches are based on or motivated by the interpretation of toxicological similarity as chemicals that are dilutions (concentrations) of each other and hence have the characteristics Berenbaum describes in his sham mixture example. He also contrasted his general solution with postulated requirements of sigmoidal dose-response curves for the median-effect equation, linear dose-response curves for effect summation, and simple exponential dose-response curves for independent action. While appropriate for the literature of 1985, those latter "requirements" are not considered necessary in current formulas based on independence (see Chap. 14).

The dose-addition model proposed by Berenbaum (1985) is based on the concept of linear isoboles and can be described mathematically by Eq. 9.1 for a given response level y:

$$\sum_{j=1}^{J} \frac{d_{j,y}}{\mathrm{ED}_{j,y}} = 1 \tag{9.1}$$

where J is the number of chemicals in the mixture, $d_{j,y}$ is the dose of chemical j in the mixture that produces response y, and $\mathrm{ED}_{j,y}$ is the dose of j alone that will produce response y (as estimated, e.g., from dose-response data for each individual chemical). For a binary combination, the equation simplifies to:

$$\frac{d_{\mathrm{A},y}}{\mathrm{ED}_{\mathrm{A},y}} + \frac{d_{\mathrm{B},y}}{\mathrm{ED}_{\mathrm{B},y}} = 1 \tag{9.2}$$

where $d_{\mathrm{A},y}$ and $d_{\mathrm{B},y}$ are the doses of chemicals A and B, respectively, in the mixture that elicit response y and $\mathrm{ED}_{\mathrm{A},y}$ and $\mathrm{ED}_{\mathrm{B},y}$ are the respective doses of chemicals A and B alone that elicit response y. Berenbaum referred to this equation as the "hyperplane theorem." Chemicals that individually do not produce an effect but that could increase or decrease the effect of other chemicals in a mixture can be incorporated into Eq. 9.1, as Berenbaum pointed out, by assuming that ED_j is infinite so that $\frac{d_j}{\mathrm{ED}_j}$ for that chemical is 0. At least one chemical must produce a nonzero response for Eq. 9.1 to hold. While chemicals that are dose-additive obey Eq. 9.1, a sum less than 1 indicates greater than dose-additive effects of the mixture, and a sum greater than 1 indicates less than dose-additive effects of the mixture. Berenbaum went on to demonstrate that the equation for dose addition (Eq. 9.1) can be rearranged to solve for the predicted mixture response *provided that all constituent chemicals display similar dose-response relationships* (Berenbaum 1985) – a restriction that is not needed for application of Eq. 9.1. Although Berenbaum's framework is relatively flexible compared to other approaches, limitations still exist. For example, mixture effects can only be calculated up to the maximum effect of the least effective constituent, thereby limiting the utility of this approach for mixtures containing partial agonists. Furthermore, Bosgra et al. have criticized the isobole approach advocated by Berenbaum and others based on the possibility of noninteracting chemicals that display nonlinear isoboles (Bosgra et al. 2009).

Nevertheless, the work of Berenbaum has provided the foundation for many subsequent efforts to refine predictive models for mixture toxicity based on an assumption of dose additivity.

9.3.2 Summary of Select Dose-Addition Models

This section discusses approaches that have been developed more recently. Altenburger and colleagues have made significant contributions to the dose-addition literature from the field of ecotoxicology (Altenburger et al. 1990, 2000, 2004; Faust et al. 2001). Their work systematically evaluated mixtures of chemicals with similar and dissimilar mechanisms of action in order to better understand the joint action of combinations of environmental chemicals. Their overarching goal appears to be the development of general principles that can improve our ability to predict mixture effects based on dose-response relationships for component chemicals. To this end, they have adapted established dose-addition (and independent-action) concepts and applied them to a number of different mixtures. Their work evaluates which additivity models are most appropriate for estimating the toxicity of different mixtures with known mechanisms of action. For example, they test the hypothesis that chemicals with the same mechanism of action will better conform to a model of dose addition than to a model of independent action. Their approach represents a practical application of the dose-addition principles described by Berenbaum. First, various dose-response functions (e.g., probit, logit, Weibull) are fit to data from each individual chemical to identify a separate "best fit" function for each chemical (i.e., there is no requirement for a shared function among constituents). This feature is relatively uncommon among dose-addition models, which typically use a common function to describe all individual chemical dose-response curves. Next, the proportion of each individual chemical in the mixture and the dose corresponding to a designated response level derived from the "best fit" function serve as input for calculating the predicted mixture dose corresponding to that effect level. For mixtures of known and fixed composition, Altenburger et al. (2000) rearrange the variables in the equation proposed by Berenbaum (Eq. 9.1) to solve for the dose of the mixture that would elicit a given effect level y:

$$\mathrm{ED}_{\mathrm{mix},y} = \left(\sum_{j=1}^{J} \frac{q_j}{\mathrm{ED}_{j,y}} \right)^{-1} \tag{9.3}$$

where $ED_{\mathrm{mix},y}$ and $ED_{j,y}$ are the doses of the mixture and individual chemical j, respectively, expected to elicit a designated effect level y, and q_j is the fraction of chemical j in the mixture. The mixture dose $ED_{\mathrm{mix},y}$ is calculated for a range of y values spanning approximately 1–99% of the maximal response. This results in articulation of the predicted dose-response relationship for the mixture. As with the

original Berenbaum approach, predictions can be made only up to the lowest maximum effect level among the individual constituents and down to the highest minimum effect level. This limitation is not ideal for mixtures containing constituents that display incomplete efficacy (i.e., partial agonists). Generally, Altenburger et al. compare predictions of mixture toxicity based on dose addition, along with predictions based on independent action as a second reference curve, to observed mixture toxicity over a range of doses.

The method first described by Altenburger et al. (2000) has been applied by many other groups to evaluate diverse environmental mixtures in a wide array of test systems from in vitro estrogenicity assays (Silva et al. 2002; Rajapakse et al. 2004; Payne et al. 2000) to in vivo rodent toxicity studies (Metzdorff et al. 2007). Although the essentials remain consistent, the approach has been applied with various modifications and increasing sophistication. For example, the collection of functions from which the "best fit" to individual chemicals is selected may differ among research groups. Additionally, bootstrap methods have been used to incorporate statistical uncertainty into model predictions to more rigorously evaluate the differences between predicted and observed dose-response relationships (Metzdorff et al. 2007). Finally, Scholze et al. have adapted the method to assess mixtures containing constituents that display partial efficacy (Scholze et al. 2014).

The work of Gennings and colleagues offers another significant body of literature providing methods for detecting departure from dose additivity of chemical mixtures (Casey et al. 2004; Gennings et al. 1997, 2004a, b). A defining hallmark of this work is in the statistical comparison of predicted to observed mixture effects using the whole dose-response curve instead of, or in addition to, a point-by-point comparison. Models developed by Gennings et al. have been applied to a diverse array of mixtures and testing paradigms. A detailed description of this and related statistical approaches for dose additivity can be found in Chap. 11.

Another example of adapting Berenbaum's dose-addition principles comes from Howard and Webster, who offer a novel approach to assessing mixtures containing constituents with partial efficacy (e.g., partial agonists) (Howard and Webster 2009). In contrast to the methods developed by Altenburger et al., this approach incorporates a single model (3-parameter Hill model) for all individual chemicals. The slope (i.e., Hill coefficient) for each chemical is assumed to be 1. Howard and Webster's equation for the "generalized concentration addition" model for a given response y is:

$$\sum_{j=1}^{J} \frac{d_{j,y}}{f_j^{-1}(y)} = 1 \tag{9.4}$$

where $d_{j,y}$ is the dose of chemical j in the mixture that produces response y and f_j^{-1} is the inverse of the dose-response function f_j for chemical j. Compared to the Berenbaum formula of Eq. 9.1, Eq. 9.4 is simply a change in notation: replacing $\mathrm{ED}_{j,y}$ by $f_j^{-1}(y)$. The novelty arises because Howard and Webster formally define the inverse function for partial agonists on the full range of responses possible for

the full agonist, revising the usual definition where each inverse function is defined only on the range of responses for that individual chemical. This revised definition eliminates some dose-response models from consideration. When the dose-response model is a Hill model with a slope equal to 1, this revised definition assigns "negative" doses of a partial agonist to response levels beyond that partial agonist's maximal response and, for binary mixtures of a partial and a full agonist, leads to linear isoboles with positive slope at those response levels. If the assumption about the Hill slope is tenable, this approach is particularly well-suited for assessing dose additivity for mixture data from receptor-based in vitro assays that frequently include partial agonists (Howard et al. 2010; Hadrup et al. 2012, 2013).

9.3.3 Application of Dose-Addition Modeling in Toxicology: Male Reproductive Tract Development

Endocrine-disrupting chemicals have been the focus of a large body of mixture research that uses dose-addition modeling as a tool to explore the joint action of potentially co-occurring chemicals. This line of research is motivated by multiple factors. First, there is concern that a number of reproductive tract malformations and pathologies (e.g., testicular dysgenesis syndrome) are increasing in certain populations; exposure to environmental contaminants has been implicated in this trend (Main et al. 2010). Second, the U.S. National Health and Nutrition Examination Survey (NHANES) (Buttke et al. 2012) and some international biomonitoring efforts (Frederiksen et al. 2014) have demonstrated that people are routinely exposed to numerous chemicals that have endocrine-disrupting potential. Third, the period of reproductive tract development represents a particularly sensitive window. Fourth, the joint effects of endocrine-disrupting chemicals that act at different points in complex signaling pathways are not fully understood. Fifth, knowledge of whether chemicals display dose-additive or greater than dose-additive toxicity provides information useful for the risk assessment of chemical mixtures.

Research groups at the U.S. EPA led by Gray (Rider et al. 2010, 2008) and in Europe led by Hass and Kortenkamp (Christiansen et al. 2009; Metzdorff et al. 2007; Ermler et al. 2011) have used dose-addition modeling to understand how individual endocrine-disrupting chemicals act jointly to disrupt male reproductive tract development. Both groups have identified individual environmental chemicals (e.g., pesticides, herbicides, plasticizers, personal care product ingredients) that disrupt male reproductive tract development through different mechanisms (e.g., androgen receptor antagonism, disruption of steroidogenesis). The general hypothesis developed by these researchers is that chemicals that share common key events in their respective adverse outcome pathways (see Chap. 7 for detailed information on using adverse outcome pathways to prioritize mixtures for study), or that simply share a common adverse outcome, will adhere to predictions based on dose

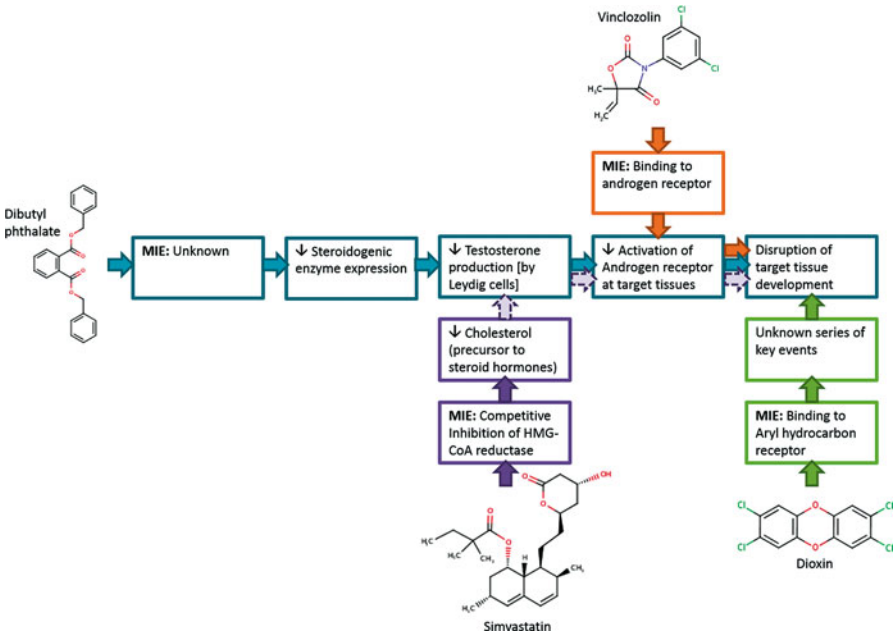

Fig. 9.4 Hypothetical adverse outcome pathway network for four chemicals that disrupt male reproductive tract development. Each chemical displays a unique molecular initiating event (MIE). Solid arrows represent pathways with substantial scientific support, while lighter arrows with dashed lines represent hypothesized pathway interactions. Some of the key events are unknown, including the molecular initiating event for phthalates and the intermediate steps for dioxin

addition. In other words, a convergence of pathways at or near the manifestation of the adverse outcome is hypothesized to result in combined effects that are dose-additive.

Figure 9.4 provides a visual representation of the general hypothesis. In this example, three chemicals (vinclozolin, dibutyl phthalate, and dioxin) elicit reproductive toxicity via different mechanisms of action (i.e., they activate different molecular initiating events). The fungicide vinclozolin binds to the androgen receptor thereby blocking the action of androgens (e.g., testosterone, dihydrotestosterone) responsible for normal development of male reproductive tissue. Dibutyl phthalate, a plasticizer, decreases the availability of testosterone. Vinclozolin and dibutyl phthalate share a common adverse outcome of disrupting the development of male reproductive organs in rats exposed in utero. Though not all of these chemicals elicit the same suite of effects, some effects are overlapping (e.g., disruption of epididymal development). The biological underpinning of applying the dose-addition hypothesis to binary combinations of vinclozolin and dibutyl phthalate is that the target tissue (i.e., epididymal tissue) does not recognize whether the decrease in activated androgen receptor is due to receptor antagonism

or decreased androgen; the tissue only recognizes the total decrease in activated receptors. Even though the specific mechanisms are different, their convergence supports the use of dose addition. Multiple studies with combinations of chemicals with the mechanisms described above have shown that predicted mixture responses based on dose addition are similar to observed mixture responses, supporting the hypothesis (Rider et al. 2008, 2010).

The hypothesis of dose addition is less clearly applied to the addition of dioxin to the binary mixture. Dioxin's mechanism of action is not known, and its effect on epididymal development is less well-defined. To expand the dose-addition hypothesis to the ternary mixture, dose-response data from the three chemicals that exhibit reproductive toxicity could be used to calculate expected effects of a three-chemical mixture based on an assumption of dose additivity. These predictions could then be compared to empirical data generated from testing the defined mixture.

A fourth chemical, simvastatin, is known to disrupt cholesterol synthesis, but it does not exhibit similar effects on reproductive endpoints. The hypothesis of dose addition could be used to investigate the joint action of simvastatin with the other three reproductive toxicants. Here, the assumption would be that simvastatin would not contribute to the toxicity of the three-chemical mixture (i.e., its contribution to dose-additive toxicity would be 0). A greater than dose-additive or less than dose-additive result would indicate that, although simvastatin does not directly induce the adverse outcome, it does alter the toxicity of one or more of the three chemicals that act through the androgen receptor-mediated adverse outcome pathway (Fig. 9.4). Mechanistic studies would then be required to investigate the hypothesis that decreased cholesterol could lead to a decrease in testosterone sufficient to contribute to downstream toxicity.

In addition to work aimed at exploring the dose additivity of antiandrogenic mixtures, similar work looked at estrogenic mixtures (Silva et al. 2002; Rajapakse et al. 2004) and at thyroid-disrupting mixtures (Crofton et al. 2005). Data from experiments like those described above help to inform cumulative risk assessment. For example, such studies contributed to the decision by a National Academy of Sciences panel to recommend to the U.S. EPA that reproductive toxicant phthalates, as well as other antiandrogenic chemicals, should be included in cumulative risk assessments (National Research Council 2008).

9.4 Independent Action

As with many terms in mixture toxicology, "independent action" is not used consistently throughout the literature. In particular, "independent action" has been used interchangeably with "independent joint action," "response addition," and "Bliss independence." Here, the term "independent action" was deliberately selected. Although "response addition" provides a convenient counterpart to the term "dose addition," it does not adequately convey the underlying concept. First,

the terms "response" and "addition" can be synonyms of "effect" and "summation," respectively. In this chapter, the concepts of "response addition" and "effect summation" are carefully defined so they are not equivalent and lead to different prediction formulas for the combination exposure (see detailed discussion later in chapter). Second, independent action does not necessarily lead to a simple addition of responses. Although independent action and independent joint action have both been used widely, a preference for simplicity favored the use of independent action herein.

9.4.1 Background

The first description of independent action is generally attributed to Bliss (1939). The foundational biological premise of independent action is that individual chemicals act through distinct, non-interfering pathways to arrive at an apical response (see Sect. 2 for further discussion). In a binary mixture with independent action, the presence of chemical B does not influence the dose of or the response to chemical A and vice versa. The work of Bliss is based on pesticides and refers specifically to quantal responses, those involving some measurable two-category event (generically, occurrence/nonoccurrence, presence/absence, or yes/no, e.g., death, presence of a particular tumor at sacrifice, or body weight less than 18 g). Quantal responses are characterized by the probability that the event occurs as estimated by the fraction of the population (or sample) that experiences the event. The equation for calculating mixture response under independent action most often cited in the toxicology literature is:

$$p_{\text{mix}} = 1 - \prod_{j=1}^{J} \left(1 - p_j\right) \tag{9.5}$$

where p_{mix} is the probability of the event occurring in response to a given dose of the mixture and p_j is the corresponding probability for chemical j alone at the same dose as it is present in the mixture. Equation 9.5 is a consequence of an elementary probability calculation that assumes that the occurrence of the event in response to an individual chemical is independent across the chemicals. The probability that the event does not occur in response to the mixture is $1–p_{\text{mix}}$. For the event not to occur in response to the mixture means that the event cannot occur in response to any of the individual chemicals (if one or more of the individual chemicals at their dose in the mixture had elicited the event, then the mixture dose would have elicited the event). The non-occurrence probability for chemical j is $1–p_j$. Under the independence assumption, the probability that no single chemical elicited the event is the product of their individual non-occurrence probabilities, namely, $\prod_{j=1}^{J} \left(1 - p_j\right)$. Equating $1–p_{\text{mix}}$ to this product and solving for p_{mix} yields Eq. 9.5.

For a binary combination of chemicals, the formula simplifies to:

$$p_{\text{mix}} = p_{\text{A}} + p_{\text{B}} - (p_{\text{A}} p_{\text{B}}) \tag{9.6}$$

where p_{A} and p_{B} are the event probabilities for individual chemicals A and B, respectively. Because the fraction of individuals who respond both to chemical A and to chemical B is included twice in the sum $p_{\text{A}} + p_{\text{B}}$, the term ($p_{\text{A}}$ p_{B}) must be subtracted to compensate for the double counting. Bliss noted that this formula corresponds to Abbott's formula, which is used in entomology to distinguish between mortality attributable to pesticide application or to natural causes (Abbott 1925).

Bliss pointed out that the underlying assumption about independent action leads to different predictions of the mixture response depending on the "correlation in susceptibility" to the distinct mixture constituents. Imagine a population in which individuals exhibit a range of susceptibilities or tolerances (i.e., the exposure level of a given chemical required to elicit the event differs among individuals). In addition, suppose that each individual has a separate susceptibility to each distinct chemical in a mixture. In a population of individuals, the susceptibilities to distinct chemicals may be correlated or not. Limiting himself to two chemicals, Bliss considered three possibilities: no correlation (no relationship between susceptibilities), perfect positive correlation (ordering of individuals' susceptibilities is the same for both chemicals), and some intermediate degree of positive correlation. Equations 9.5 and 9.6 correspond to independent action with uncorrelated susceptibilities.

In Fig. 9.3, the angular isobole (perpendicular gray line segments) corresponds to the case of independent action with perfect positive correlation among susceptibilities. That special case is likely to be extremely rare for real mixture exposures but merits inclusion here because it shows up in mixture literature, usually without explanation. When the susceptibility correlation is positive but less than perfect, the isobole would have a different shape that depended on the shapes of the dose-response functions for the two chemicals. The concept of independent action, regardless of the correlation among susceptibilities, properly applies only to quantal responses.

Plackett and Hewlett (1948) also contributed to describing the concept of independent action put forth by Bliss and went into further detail on potential correlation in susceptibilities. In addition to the options described by Bliss, Plackett and Hewlett considered perfect negative correlation (the ordering of individuals' susceptibilities to one chemical is exactly the reverse of their susceptibilities to the other), as well as intermediate degrees of negative correlation. For perfect negative correlation of susceptibilities, the probability of response to the mixture is given by:

$$p_{\text{mix}} = \min\,(p_A + p_B,\ \ 1), \tag{9.7}$$

$$p_{mix} = p_A + p_B - p_A p_B$$

.0199 = .01 + .01 = .02 .01 x .01 = .0001
.0494 = .02 + .03 = .05 .02 x .03 = .0006
.0976 = .06 + .04 = .10 .06 x .04 = .0024
.1904 = .08 + .12 = .20 .08 x .12 = .0096

Fig. 9.5 Example of calculating mixture effects at low individual effect levels (<15%) when assuming independent action. The product $p_A p_B$ is small and does not appreciably change the sum $p_A + p_B$

which reflects the constraint that a probability cannot exceed 1. In particular, the proportion responding to the mixture cannot exceed 1, regardless of the proportions responding to the individual constituents.

Effect summation is considered in the next section as a separate approach from independent action, though there is an overlap between the two concepts. The usual effect-summation formula for a binary mixture is:

$$p_{\text{mix}} = p_{\text{A}} + p_{\text{B}} \tag{9.8}$$

and corresponds to Eq. 9.7 (for independent action with perfect negative correlation of susceptibilities) if the sum of p_A and p_B does not exceed 1, and it approximates Eq. 9.6 (for independent action with no correlation of susceptibilities) if the response probabilities are small enough that the product of p_A and p_B is negligible (see, e.g., Fig. 9.5).

Assessing correlations among pairs of susceptibilities is not trivial with two chemicals and becomes unmanageable with more than two chemicals. Furthermore, if chemicals are indeed working through different (i.e., independent) mechanisms, perfect correlation (either positive or negative), though instructive for illustrating the limiting cases, seems unrealistic. The difficulty in assessing these correlations coupled with the implausibility of perfectly correlated susceptibilities could explain why correlation of susceptibilities is not discussed widely in current mixture literature and why most researchers default to Eq. 9.5, which corresponds to the assumption of uncorrelated susceptibilities. Unless specifically stated, this chapter will take "independent action" to mean "independent action with uncorrelated susceptibilities" as in Eqs. 9.5 and 9.6.

9.4.2 Application of Independent Action

As with many concepts in mixture toxicology, the application of independent action has evolved over time to scenarios beyond the original scope. Most importantly, independent action is currently used to predict a variety of biological responses that

do not reflect the initial definition, derived from probabilities of events. Equations 9.5 and 9.6 are specific calculations for the probability of co-occurrence of statistically independent events; consequently, they properly apply to the probability that the event occurs or the proportion of the population that experiences the event, quantities that must fall between zero and one. Quantal response data that meet the original definition of independent action include mortality and the proportion of animals within a dose group that displays a certain characteristic. Continuous endpoints (e.g., growth, organ weight, hormone levels, magnitude of gene expression changes) generally do not lie between zero and one and are not interpretable as probabilities; consequently, they fall outside of the scope of that definition. If the continuous response (y) has predefined minimal and maximal possible values (y_{min}, y_{max}), then it can be rescaled to lie between zero and one (equivalently, 0% and 100%) by transforming to $(y-y_{min})/(y_{max}-y_{min})$. Even such rescaled responses, however, are not typically interpretable as "probabilities." For example, knowing that an exposed rat lost 40% of its body weight says little about the probability of that response. Thus, the use of Eq. 9.5 with continuous responses, even after transformation to a scale that looks like a probability, is not supported by the underlying probability argument that justifies Eq. 9.5. Furthermore, while the $1–p$ term in Eq. 9.5 is clearly interpreted as the probability of non-occurrence (e.g., if the event is death, then non-occurrence is survival), we do not know of any analogous interpretation of the corresponding $1–y$ term. Consequently, application of Eqs. 9.5 or 9.6 with quantal responses has no similar theoretical basis as a benchmark for independent action when applied to continuous responses. In other words, there is no reason to believe that chemicals that act independently will obey Eqs. 9.5 or 9.6 when the ps are replaced by ys, even rescaled ys. Nevertheless, despite this serious theoretical shortcoming, using the independent-action formula as an empirical model with continuous responses is widely practiced and may have utility.

Early discussions of applying independent action for binary mixtures to continuous response data emphasized differences in response measures (Muska and Weber 1977). While quantal data show fractions of the population exceeding toxicity thresholds, continuous data show average measured response intensities. Thus, the concept of susceptibility correlation is not appropriate for continuous data and formulas like Eq. 9.7 cannot be used. Instead, estimates of combined responses are simple sums of component response intensities (see section 9.5 for more about effect summation). More recent instances of applying independent action to continuous endpoints employ Eq. 9.5. Examples in the literature of the application of independent action to continuous data include in vitro enzyme activity (Froment et al. 2016), degree of hypopigmentation in zebra fish (Schmidt et al. 2016), and organ weights in rat pups following in utero exposure (Rider et al. 2008), among others.

An example can be used to illustrate the different applications discussed above. Imagine a study where pregnant females are exposed to different doses of chemical A or chemical B alone and their pups are evaluated for responses. The two endpoints of interest are pup mortality (Table 9.1) and pup weight at birth (Table 9.2). For the mortality endpoint, the percent of pups who died is the response

Table 9.1 Hypothetical data to illustrate application of independent-action model

Chemical A		Chemical B		
Dose (mg/kg)	Mortality (%)	Dose (mg/kg)	Mortality (%)	Predicted mixture response under independent action (%)
0	0	0	0	0
2	3	50	5	8
4	10	100	20	28
8	50	200	45	73
16	95	400	75	99

used in the model. Therefore, Eq. 9.5 could be used to calculate the expected response of a binary combination of the two chemicals (at each of the dose pairings in a given row of the table) under an assumption of independent action as in the following example:

$$4\ \text{mg/kg A} + 100\ \text{mg/kg B} : 0.10 + 0.20 - (0.10^{*}0.20) = 0.30 - 0.02 = 0.28 = 28\%\text{mortality}$$

In contrast, pup weight does not easily fit into the probability-based equation. As opposed to mortality, mean pup weight represents an average value of a continuous variable. Before discussing the data transformation that is required to apply the model, it is important to address the rationale for applying the independent-action model to these data. There are two possible arguments. First, the model can be viewed as a strictly empirical tool, offering a point of comparison for dose-addition predictions and observed mixture data. Second, and more problematically, one could argue that individual chemicals are acting through different biological mechanisms to disrupt a common system (e.g., development) that has real biological limits, justifying a meaningful rescaling to 0–100%. In the current example, pup weight data could be converted to % decrease from control (but even this more meaningful rescaling is not interpretable as providing probabilities, so it should just be treated as an empirical model). The mean weight in the control group could serve as a maximum response level (y_{max}) – though using observed means can lead to anomalies (e.g., negative percentages which are not between 0 and 100%) (Table 9.2). As a minimum response level (y_{min}), 0 g could serve as a default minimum (pups cannot weigh less than 0 g), or, assuming pups are not viable below some nonzero weight, the average of "lowest pup weights" achieved in previous work could serve as a more biologically-based minimum. If all pup weights are rescaled using the control value (e.g., the average weight of the combined control animals from chemicals A and B, which is 6.3 g in this example) as y_{max} and the minimum viable weight (3.0 g in this example) as y_{min}, then the data can, in principle, be converted to a 0–100% response scale (Table 9.2). These converted data can then be used to make predictions based on the independent-action equation, as described above for mortality.

Table 9.2 Conversion of continuous data to a 0–100% scale for use in Eq. 9.5 to predict mixture responses assuming independent action

Chemical A			Chemical B			Predicted mixture response under independent action
Dose (mg/kg)	Mean pup weight at birth (g)	% Decrease from control	Dose (mg/kg)	Mean pup weight at birth (g)	% Decrease from control	% Decrease from control
0	6.2	3.0	0	6.4	−3.0	0.1
2	5.9	12.1	50	6.5	−6.1	6.7
4	5.7	18.2	100	5.7	18.2	33.1
8	4.5	54.5	200	5.1	36.4	71.1
16	3.9	72.7	400	4.2	63.6	90.1

[a]A minimum response of $y_{min} = 3.0$ g was assumed when calculating the % decrease from control, which is defined as $100 \times (y_{max} - y)/(y_{max} - y_{min})$, where y_{max} is the average control response of 6.3 g

In the past, debate has centered around which definition of additivity – dose addition or independent action – should serve as the default approach for describing the baseline prediction from which interactions should be measured (Greco et al. 1992). This argument has largely faded, with many researchers instead opting to include both dose-addition and independent-action predictions as referents for assessing tested mixtures (Rider et al. 2008; Olmstead and LeBlanc 2005; Christiansen et al. 2009; Gregorio et al. 2013; Altenburger et al. 2000; Qin et al. 2011). Including both formulas is usually done when there is inadequate biological information about the potential joint toxicity. It is typically held that independent action is appropriate for chemicals that act through different adverse outcome pathways, while dose addition is favored for chemicals that exhibit similar mechanisms of action. However, individual chemical adverse outcome pathways are not always known. Furthermore, instead of a priori classification of individual chemicals and selection of an appropriate model, the fit of observed mixture data to either dose-addition or independent-action models has been used to support arguments regarding similarity/dissimilarity of mechanism(s) among mixture constituents (Rider and LeBlanc 2006; Froment et al. 2016; Faria et al. 2016). In many cases, however, dose addition and independent action yield similar or even indistinguishable predicted mixture effects, particularly when the response (as a probability) is very small or when the dose-response curve is close to linear (Thienpont et al. 2013; Kortenkamp and Altenburger 1998; Cedergreen et al. 2008).

Several studies have addressed the hypothesis that chemicals with distinctly different mechanisms of action are better fit by a model of independent action than by one of dose addition (Backhaus et al. 2000, 2011; Ermler et al. 2014; Villa et al. 2012; Faust et al. 2003; Hermens and Leeuwangh 1982; Martin et al. 2009; Baylay et al. 2012; Cedergreen et al. 2008); however, few of these studies found that

independent action consistently provided a significantly better fit than dose addition (Backhaus et al. 2000; Faust et al. 2003).

Some of the most highly cited work on independent action comes from Backhaus and colleagues (Faust et al. 2003; Backhaus et al. 2000). These foundational studies set out to carefully investigate the joint action of chemicals with strictly dissimilar mechanisms of action. They tested mixtures of 14 and 16 constituents with the microtox assay (measuring disruption of the respiratory process of the marine bioluminescent bacteria *Vibrio fischeri*) and algal toxicity, respectively. In both studies, they found that independent action accurately predicted mixture toxicity, while dose addition overestimated the responses of the mixture, across the range of mixture dilutions along fixed-ratio rays with constituents present at equipotent concentrations (e.g., constituents present at the ratio of their concentrations eliciting a 50% effect when tested individually).

In the majority of papers evaluating chemicals with dissimilar mechanisms of action for adherence to independent-action predictions, the results are much less clear-cut. In a series of experiments assessing the effects of binary mixtures of dissimilar chemicals on *C. elegans* egg production, Martin et al. (2009) found some cases where independent action accurately predicted observed mixture effects and other cases where it over- or underestimated observed mixture effects. In another example, Ermler et al. (2014) found that a mixture of four chemicals with dissimilar mechanisms of action resulted in genotoxic effects that fell between independent-action and dose-addition predictions. Similarly, Petersen et al. (2014) observed a dose-dependent switch from independent action at low concentrations to dose addition at higher concentrations for a mixture of eight chemicals with diverse mechanisms of action on the growth rate of marine algae. Baylay et al. (2012) pursued a mechanistic understanding of the combination of two dissimilar chemicals using metabolomic profiling of earthworm tissue following treatment with nickel, chlorpyrifos, and a combination of the two. They concluded that while the measured effect of the mixture on reproduction was greater than that predicted by independent action, the hypothesis of dissimilar mechanisms of action was supported by the finding that the metabolomic profile for the mixture was intermediate between the unique profiles for nickel and chlorpyrifos alone, confirming their dissimilar activity and indicating that they both contributed to the mixture effect.

9.4.3 Challenges with Independent Action

As indicated by the name, independent action is predicated on the assumption that the individual chemicals within the mixture operate independently; however, the exact nature of the biological independence required to make this assumption plausible is usually not clearly articulated. Throughout the independent-action literature, there are references to independence at the level of mechanism of action (adverse outcome pathway), target tissue, or target system. There are two important points to keep in mind. First, considering the complexity of biological systems, it is

unlikely that the mechanisms of action will be strictly dissimilar. Second, many environmental chemicals can have more than one mechanism of action.

As alluded to above, independence of action is likely influenced by the chemical constituents, endpoint, and model system of interest. Therefore, one can hypothesize that experiments that include chemicals with specific and distinct mechanisms of action and simple model systems are more likely to produce observed results consistent with independent-action predictions. Conversely, chemicals with nonspecific mechanisms of action, common among environmental contaminants, and complex systems (e.g., carcinogenesis in a rodent) might be more likely to involve interactions among biological signaling pathways and less likely to conform to predictions based on an assumption of independent action. Hermens and Leeuwangh (1982) discuss some of these considerations in their studies with chemicals that display both specific and nonspecific mechanisms of toxicity. For example, both Hermens and Leeuwangh (1982) and Könemann (1981) discuss the concept that the primary mechanisms of constituent chemicals could act independently, while a lesser (secondary) narcotic mechanism present in all organic chemicals could contribute in a dose-additive manner to toxicity, resulting in observed mixture toxicity that exceeds independent-action predictions.

9.5 Effect Summation

As mentioned earlier (Sect. 4.1), under certain conditions (e.g., low effect levels), effect summation can be applied directly with quantal response data to approximate predictions under independent action. Nevertheless, some investigators have applied effect summation beyond quantal responses to continuous responses, even without prior conversion to a 0–100% scale. Avoiding such conversion is particularly useful when the responses being measured are not readily amenable to it. Of course, the sum of effects, whether rescaled or not, can exceed the 100% limit for probabilities or the experimentally determined maximum effect level for continuous responses – a characteristic that is often cited as a fatal flaw of the approach (see below for a discussion of the limited application of effect summation). The general equation for effect summation is:

$$Y_{\text{mix}} = \sum_{j=1}^{J} Y_j \tag{9.9}$$

where Y_j is the response elicited by chemical j alone. For a binary mixture, the equation simplifies to:

$$Y_{\text{mix}} = Y_{\text{A}} + Y_{\text{B}} \tag{9.10}$$

where Y_{A} and Y_{B} represent the responses elicited by chemicals A and B, respectively, when given alone.

9.5.1 Application and Challenges of Effect Summation

In toxicological studies, effect summation is often cited as an inappropriate approach for assessing mixtures because of a lack of biological plausibility at high effect levels (Boedeker and Backhaus 2010). A primary example of this implausibility is that an effect-summation model can produce response predictions beyond the natural response boundaries. In contrast, for independent action, the laws of probability ensure that the predicted probabilities satisfy the natural boundary constraints. Thus, effect summation is frequently discounted as an appropriate tool in toxicology because a biological plausibility threshold can be exceeded. Nevertheless, effect summation does have one important advantage over independent action: there is no requirement to convert the individual data, so they can be combined across chemicals in raw form. Therefore, effect summation could be used as a benchmark, if researchers acknowledge its limitations.

The example described in Sect. 4.2 can be used to illustrate the advantages and disadvantages of effect summation compared to independent action. In Table 9.3, it is apparent that the responses can be added together without converting the data. However, at the highest dose of the mixture (16 mg/kg chemical A + 400 mg/kg chemical B), the predicted effect is a 4.5 g decrease in pup weight. This decrease would result in pup weights around 2.0 g, well below the 3.0 g that was thought to be the minimum that is biologically plausible (see independent-action example above). This again emphasizes the need for a biologically-based limit on the predicted effect. Because most effect measures are surrogates for a complex physiological process, many effect limits are empirically derived, including limits on what values are considered "normal" or "healthy." At those highest doses in Table 9.3, the pups are of such low weight as to be biologically compromised. For example, at the highest dose of chemical A (16 mg/kg), the pup weight loss is 37%, well in excess of the common limit of 10% (Chapman et al. 2013). Furthermore,

Table 9.3 Predictions of mixture responses based on effect summation

Chemical A			Chemical B			Effect-summation prediction
Dose (mg/kg)	Mean pup weight (g)	Decrease in pup weight (g) from control[a]	Dose (mg/kg)	Mean pup weight (g)	Decrease in pup weight (g) from control[a]	Decrease in pup weight (g) from control[a]
0	6.2	0.1	0	6.4	−0.1	0
2	5.9	0.4	50	6.5	−0.2	0.2
4	5.7	0.6	100	5.7	0.6	1.2
8	4.5	1.8	200	5.1	1.2	3.0
16	3.9	2.4	400	4.2	2.1	4.5

[a]The control response for calculating decrease in weight was the average of the zero-dose responses for both chemicals, namely, 6.3 g

they likely have some physiological or biochemical processes that are no longer functioning normally, calling into question the purpose of the predictive model based on one measured effect.

9.6 Integrated Addition

Integrated addition is a more recent development in the mixture literature and represents a combining of concepts from dose addition and independent action to accommodate mixtures that contain constituents with similar and dissimilar adverse outcome pathways (Rider and LeBlanc 2005; Teuschler et al. 2004; Altenburger et al. 2005). First, chemicals are grouped into toxicologically similar groups (see discussion in Sect. 2). Next, the total response for each group is calculated using a dose-addition method. This response must be a measure of probability (see Sect. 4.2 for concerns). Finally, the responses from the different groups are combined using an independent-action approach (Figs. 9.6 and 9.7). The goal driving development of the integrated-addition approach was to refine predictions of mixture toxicity based on known mechanisms of action.

9.6.1 Application of Integrated Addition

In one of the first papers describing an integrated-addition approach, a ternary mixture of two organophosphate pesticides (malathion and parathion) and the

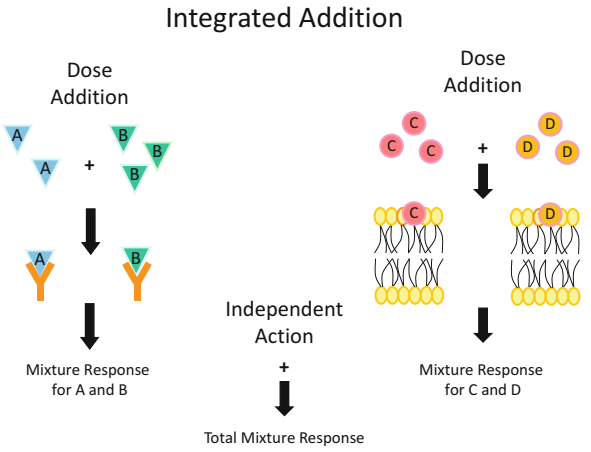

Fig. 9.6 Representation of integrated addition. In this example, chemicals A and B share a common mechanism of action (binding to a receptor to elicit a downstream effect), as do chemicals C and D (interfering with lipid membranes). A dose-addition model can be used to calculate expected mixture effects for each of the two mechanism-based groups. The responses of each of the mechanism-based groups can then be combined using an independent-action model

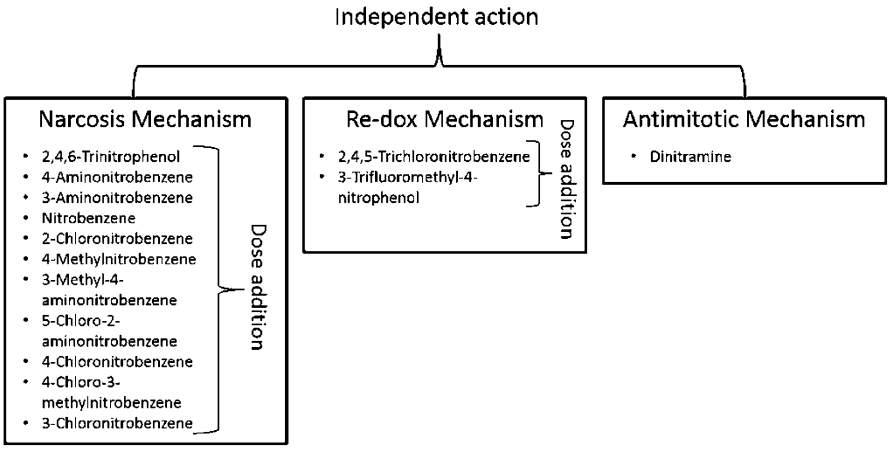

Fig. 9.7 Example of an application of the integrated-addition concept to a 14-chemical mixture of nitrobenzenes (Altenburger et al. 2005)

pesticide synergist piperonyl butoxide was assessed for its joint effects on immobilization of the crustacean *Daphnia magna* (Rider and LeBlanc 2005). Malathion and parathion inhibit acetylcholinesterase activity. Their joint action was estimated using dose addition. Next, piperonyl butoxide effects were combined with the dose-additive effects of malathion and parathion using independent action (Eq. 9.5). An interaction coefficient was also added to the equation to account for the interaction between organophosphates and piperonyl butoxide. (Quantitative interaction information is rarely available, and the modification of models to incorporate interactions is beyond the scope of this chapter.)

In another early effort to combine dose addition and independent action, Altenburger et al. (2005) used various mechanism-based grouping strategies to compare modeled predictions to observed algal toxicity of a 14-chemical nitrobenzene mixture. They used both individual chemical dose-response parameters and known mechanistic data to group chemicals. Based on this approach, they articulated three groups: 11 nitrobenzenes operating primarily through narcosis, two mononitrobenzenes acting through a redox-cycling mechanism, and a single chemical acting through an antimitotic pathway. Dose addition was used to calculate the expected joint effects from each of the two groups with more than one chemical. These two groups were then combined with the single chemical using independent action (Fig. 9.7).

Since its introduction, integrated addition has been used by various research groups to provide an additional point of comparison together with independent action and dose addition (Ra et al. 2006; Flippin et al. 2009). There have also been attempts to improve models based on integrated addition using more sophisticated statistical procedures. Mwense et al. (2004, 2006) used advanced approaches such as molecular modeling to derive chemical descriptors and fuzzy set theory to assign

chemicals to similarity groups. Qin et al. (2011, 2015) employed multiple linear regression techniques to combine predictions based on independent action and dose addition.

9.6.2 Challenges of Integrated Addition

The greatest challenge to applying integrated addition is a lack of information on the mechanisms of action or adverse outcome pathways associated with individual chemicals. Additionally, chemicals can induce more than one adverse outcome pathway (Chap. 7) and can activate different adverse outcomes depending on the species. Advances in model development address the need for knowing a priori the adverse outcome pathways associated with each individual chemical by relying on physicochemical properties and data from individual dose-response curves. These approaches, however, often require advanced statistical and toxicological understanding or access to larger datasets that may not be widely available.

Inclusion of both dose addition and independent action within a single approach necessarily includes all of the challenges associated with application of those models separately. Integrated addition does not specify a particular method for calculating predictions under dose addition. Thus, any of the dose-addition methods described above could be used in the integrated-addition framework; however, because the independent-action part of the framework requires probabilities or percentages, all of the issues associated with converting response data discussed previously also apply here.

9.7 Conclusions and Recommendations

This chapter has described the long history of using simple mathematical models to better understand the toxicity of mixtures. Over the course of the last 90 years, a lot of progress has been made in developing these tools. Despite the many overlapping approaches and the often confusing terminology, some basic principles have emerged.

- Simple mathematical models to predict mixture toxicity from individual chemical data provide a useful tool for exploring the joint action of chemicals. Most of the underlying toxicological concepts involve similarity or independence. In many cases, responses predicted using dose addition and/or independent action provide close approximations of observed mixture responses.
- Despite the many variations of dose addition available, the general concept remains intact, with specific changes resulting in potentially wider application (e.g., modification to accommodate partial agonists) without substantially changing the underlying null hypothesis.

- Deviations of observed mixture responses from predicted responses can be due to a number of factors including that the prediction model is too simple (it lacks biological complexity) or there are toxicological interactions between chemicals that are not captured by the models (e.g., greater than or less than dose-additive interactions). Significant deviations could signal that follow-up mechanistic studies are required to better understand potential interactions. Further work comparing biological models (e.g., toxicokinetic models) with these simple additivity models will be important in developing plausible interaction models as well as extensions of additivity models to reflect dose dependence or mixing ratio dependence.
- When biological evidence for similarity or for independence is weak, including more than one model (e.g., dose addition and independent action or effect summation) to compare observed and predicted responses is recommended. Since adverse outcome pathways associated with chemicals are often unknown or incomplete, including multiple models can better frame the potential range of predicted mixture responses without requiring a mechanism-based argument for application of a specific model. This range of predicted responses is not the expected range of true responses, but it characterizes responses under the null models most likely to be used.
- Much of the toxicology work on mixtures has focused on low numbers of chemicals (binary and ternary combinations), with less work addressing chemical mixtures containing 10–20 constituents. Work with higher-order mixtures will be important in determining the limitations of the models described here.

Acknowledgments This work was supported in part by the NIH, National Institute of Environmental Health Sciences. The authors would like to thank Dr. Kembra Howdeshell and Dr. Natasha Catlin for their thoughtful review of this chapter.

References

Abbott, W.S. 1925. A method of computing the effectiveness of an insecticide. *Journal of Economic Entomology* 18: 265–267.

Altenburger, R., T. Backhaus, W. Boedeker, M. Faust, M. Scholze, and L.H. Grimme. 2000. Predictability of the toxicity of multiple chemical mixtures to *Vibrio fischeri*: Mixtures composed of similarly acting chemicals. *Environmental Toxicology and Chemistry* 19: 2341–2347.

Altenburger, R., W. Bodeker, M. Faust, and L.H. Grimme. 1990. Evaluation of the Isobologram method for the assessment of mixtures of chemicals – combination effect studies with pesticides in algal biotests. *Ecotoxicology and Environmental Safety* 20: 98–114.

Altenburger, R., W. Boedeker, M. Faust, and L.H. Grimme. 1996. Regulations for combined effects of pollutants: Consequences from risk assessment in aquatic toxicology. *Food and Chemical Toxicology* 34: 1155–1157.

Altenburger, R., H. Schmitt, and G. Schuurmann. 2005. Algal toxicity of nitrobenzenes: Combined effect analysis as a pharmacological probe for similar modes of interaction. *Environmental Toxicology and Chemistry* 24: 324–333.

Altenburger, R., H. Walter, and M. Grote. 2004. What contributes to the combined effect of a complex mixture? *Environmental Science & Technology* 38: 6353–6362.

Backhaus, T., R. Altenburger, W. Boedeker, M. Faust, M. Scholze, and L.H. Grimme. 2000. Predictability of the toxicity of a multiple mixture of dissimilarly acting chemicals to *Vibrio fischeri*. *Environmental Toxicology and Chemistry* 19: 2348–2356.

Backhaus, T., T. Porsbring, A. Arrhenius, S. Brosche, P. Johansson, and H. Blanck. 2011. Single-substance and mixture toxicity of five pharmaceuticals and personal care products to marine periphyton communities. *Environmental Toxicology and Chemistry* 30: 2030–2040.

Baylay, A.J., D.J. Spurgeon, C. Svendsen, J.L. Griffin, S.C. Swain, S.R. Sturzenbaum, and O.A.H. Jones. 2012. A metabolomics based test of independent action and concentration addition using the earthworm *Lumbricus rubellus*. *Ecotoxicology* 21: 1436–1447.

Belden, J.B., R.J. Gilliom, and M.J. Lydy. 2007. How well can we predict the toxicity of pesticide mixtures to aquatic life? *Integrated Environmental Assessment and Management* 3: 364–372.

Berenbaum, M.C. 1977. Synergy, Additivism and antagonism in immunosuppression – critical-review. *Clinical and Experimental Immunology* 28: 1–18.

———. 1985. The expected effect of a combination of agents – the general-solution. *Journal of Theoretical Biology* 114: 413–431.

———. 1989. What is synergy. *Pharmacological Reviews* 41: 93–141.

Bliss, C.I. 1939. The toxicity of poisons applied jointly. *Annals of Applied Biology* 26: 585–615.

Boedeker, W., and T. Backhaus. 2010. The scientific assessment of combined effects of risk factors: Different approaches in experimental biosciences and epidemiology. *European Journal of Epidemiology* 25: 539–546.

Bosgra, S., J.C.H. Van Eijkeren, and W. Slob. 2009. Dose addition and the isobole method as approaches for predicting the cumulative effect of non-interacting chemicals: A critical evaluation. *Critical Reviews in Toxicology* 39: 418–426.

Bürgi, E. 1912. The synergism of medicines. *Pflugers Archiv Fur Die Gesamte Physiologie Des Menschen Und Der Tiere* 147: 275–278.

Buttke, D.E., K. Sircar, and C. Martin. 2012. Exposures to endocrine-disrupting chemicals and age of menarche in adolescent girls in NHANES (2003-2008). *Environmental Health Perspectives* 120: 1613–1618.

Casey, M., C. Gennings, W.H. Carter, V.C. Moser, and J.E. Simmons. 2004. Detecting interaction(s) and assessing the impact of component subsets in a chemical mixture using fixed-ratio mixture ray designs. *Journal of Agricultural Biological and Environmental Statistics* 9: 339–361.

Cedergreen, N. 2014. Quantifying synergy: A systematic review of mixture toxicity studies within environmental toxicology. *PLoS One* 9: e96580.

Cedergreen, N., A.M. Christensen, A. Kamper, P. Kudsk, S.K. Mathiassen, J.C. Streibig, and H. Sorensen. 2008. A review of independent action compared to concentration addition as reference models for mixtures of compounds with different molecular target sites. *Environmental Toxicology and Chemistry* 27: 1621–1632.

Chapman, K., F. Sewell, L. Allais, J.L. Delongeas, E. Donald, M. Festag, S. Kervyn, D. Ockert, V. Nogues, H. Palmer, M. Popovic, W. Roosen, A. Schoenmakers, K. Somers, C. Stark, P. Stei, and S. Robinson. 2013. A global pharmaceutical company initiative: An evidence-based approach to define the upper limit of body weight loss in short term toxicity studies. *Regulatory Toxicology and Pharmacology* 67: 27–38.

Chou, T.C. 1976. Derivation and properties of Michaelis-Menten type and hill type equations for reference ligands. *Journal of Theoretical Biology* 59: 253–276.

———. 2006. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacological Reviews* 58: 621–681.

———. 2010. Drug combination studies and their synergy quantification using the Chou-Talalay method. *Cancer Research* 70: 440–446.

Chou, T.C., and P. Talalay. 1977. Simple generalized equation for analysis of multiple inhibitions of Michaelis-Menten kinetic systems. *Journal of Biological Chemistry* 252: 6438–6442.

———. 1981. Generalized equations for the analysis of inhibitions of Michaelis-Menten and higher-order kinetic systems with 2 or more mutually exclusive and non-exclusive inhibitors. *European Journal of Biochemistry* 115: 207–216.

———. 1983. Analysis of combined drug effects – a new look at a very old problem. *Trends in Pharmacological Sciences* 4: 450–454.

———. 1984. Quantitative-analysis of dose-effect relationships – the combined effects of multiple-drugs or enzyme-inhibitors. *Advances in Enzyme Regulation* 22: 27–55.

Christensen, E.R., and C.Y. Chen. 1985. A general noninteractive multiple toxicity model including probit, logit, and Weibull transformations. *Biometrics* 41: 711–725.

Christiansen, S., M. Scholze, M. Dalgaard, A.M. Vinggaard, M. Axelstad, A. Kortenkamp, and U. Hass. 2009. Synergistic disruption of external male sex organ development by a mixture of four antiandrogens. *Environmental Health Perspectives* 117: 1839–1846.

Crofton, K.M., E.S. Craft, J.M. Hedge, C. Gennings, J.E. Simmons, R.A. Carchman, W.H. Carter, and M.J. Devito. 2005. Thyroid-hormone-disrupting chemicals: Evidence for dose-dependent additivity or synergism. *Environmental Health Perspectives* 113: 1549–1554.

Dobrev, I.D., M.E. Andersen, and R.S.H. Yang. 2002. In silico toxicology: Simulating interaction thresholds for human exposure to mixtures of trichloroethylene, tetrachloroethylene, and 1,1,1-trichloroethane. *Environmental Health Perspectives* 110: 1031–1039.

Drescher, K., and W. Boedeker. 1995. Assessment of the combined effects of substances – the relationship between concentration addition and independent action. *Biometrics* 51: 716–730.

Ermler, S., M. Scholze, and A. Kortenkamp. 2011. The suitability of concentration addition for predicting the effects of multi-component mixtures of up to 17 anti-androgens with varied structural features in an in vitro AR antagonist assay. *Toxicology and Applied Pharmacology* 257: 189–197.

———. 2014. Genotoxic mixtures and dissimilar action: Concepts for prediction and assessment. *Archives of Toxicology* 88: 799–814.

Faria, M., V. Pavlichenko, K. Burkhardt-Medicke, A.M.V.M. Soares, R. Altenburger, C. Barata, and T. Luckenbach. 2016. Use of a combined effect model approach for discriminating between ABCB1-and ABCC1-type efflux activities in native bivalve gill tissue. *Toxicology and Applied Pharmacology* 297: 56–67.

Faust, M., R. Altenburger, T. Backhaus, H. Blanck, W. Boedeker, P. Gramatica, V. Hamer, M. Scholze, M. Vighi, and L.H. Grimme. 2001. Predicting the joint algal toxicity of multi-component s-triazine mixtures at low-effect concentrations of individual toxicants. *Aquatic Toxicology* 56: 13–32.

———. 2003. Joint algal toxicity of 16 dissimilarly acting chemicals is predictable by the concept of independent action. *Aquatic Toxicology* 63: 43–63.

Finney, D.J. 1942. The analysis of toxicity tests on mixtures of poisons. *Annals of Applied Biology* 29: 82–94.

Flippin, J.L., J.M. Hedge, M.J. Devito, G.A. Leblanc, and K.M. Crofton. 2009. Predictive modeling of a mixture of thyroid hormone disrupting chemicals that affect production and clearance of thyroxine. *International Journal of Toxicology* 28: 368–381.

Frederiksen, H., T.K. Jensen, N. Jorgensen, H.B. Kyhl, S. Husby, N.E. Skakkebk, K.M. Main, A. Juul, and A.M. Andersson. 2014. Human urinary excretion of non-persistent environmental chemicals: An overview of Danish data collected between 2006 and 2012. *Reproduction* 147: 555–565.

Froment, J., K.V. Thomas, and K.E. Tollefsen. 2016. Automated high-throughput in vitro screening of the acetylcholine esterase inhibiting potential of environmental samples, mixtures and single compounds. *Ecotoxicology and Environmental Safety* 130: 74–80.

Gennings, C., W.H. Carter, E.W. Carney, G.D. Charles, B.B. Gollapudi, and R.A. Carchman. 2004a. A novel flexible approach for evaluating fixed ratio mixtures of full and partial agonists. *Toxicological Sciences* 80: 134–150.

Gennings, C., W.H. Carter, M. Casey, V. Moser, R. Carchman, and J.E. Simmons. 2004b. Analysis of functional effects of a mixture of five pesticides using a ray design. *Environmental Toxicology and Pharmacology* 18: 115–125.

Gennings, C., P. Schwartz, W. Carter, and J. Simmons. 1997. Detection of departures from additivity in mixtures of many chemicals with a threshold model. *Journal of Agricultural, Biological, and Environmental Statistics* 2: 198–211.

Greco, W., H.D. Unkelbach, G. Pöch, J. Sühnel, M. Kundi, and W. Bödeker. 1992. Consensus on concepts and terminology for combined-action assessment: The Saariselkä agreement. *Archives of Complex. Environmental Studies* 4: 65–69.

Greco, W.R., G. Bravo, and J.C. Parsons. 1995. The search for synergy – a critical-review from a response-surface perspective. *Pharmacological Reviews* 47: 331–385.

Gregorio, V., N. Chevre, and M. Junghans. 2013. Critical issues in using the common mixture toxicity models concentration addition or response addition on species sensitivity distributions: A theoretical approach. *Environmental Toxicology and Chemistry* 32: 2387–2395.

Hadrup, N., C. Taxvig, U. Hass, and A.M. Vinggaard. 2012. Applicability of the GCA model for effect prediction of endocrine disrupters in mixture. *Toxicology Letters* 211: S159–S159.

Hadrup, N., C. Taxvig, M. Pedersen, C. Nellemann, U. Hass, and A.M. Vinggaard. 2013. Concentration addition, independent action and generalized concentration addition models for mixture effect prediction of sex hormone synthesis in vitro. *PLoS One* 8: e70490.

Hermens, J., and P. Leeuwangh. 1982. Joint toxicity of mixtures of 8 and 24 chemicals to the guppy (*Poecilia reticulata*). *Ecotoxicology and Environmental Safety* 6: 302–310.

Hertzberg, R.C., and M.M. Macdonell. 2002. Synergy and other ineffective mixture risk definitions. *Science of the Total Environment* 288: 31–42.

Hertzberg, R.C., Y. Pan, R. Li, L.T. Haber, R.H. Lyles, D.W. Herr, V.C. Moser, and J.E. Simmons. 2013. A four-step approach to evaluate mixtures for consistency with dose addition. *Toxicology* 313: 134–144.

Hewlett, P.S., and R.L. Plackett. 1957. Quantal responses to mixtures of drugs. *Nature* 180: 712–713.

———. 1959. A unified theory for quantal responses to mixtures of drugs: Non-interactive action. *Biometrics* 15: 591–610.

Howard, G.J., J.J. Schlezinger, M.E. Hahn, and T.F. Webster. 2010. Generalized concentration addition predicts joint effects of aryl hydrocarbon receptor agonists with partial agonists and competitive antagonists. *Environmental Health Perspectives* 118: 666–672.

Howard, G.J., and T.F. Webster. 2009. Generalized concentration addition: A method for examining mixtures containing partial agonists. *Journal of Theoretical Biology* 259: 469–477.

Kodell, R.L., and J.G. Pounds. 1991. Assessing the toxicity of mixtures of chemicals. In *Statistics in toxicology*, ed. D. Krewski and C. Franklin. New York: Gordon and Breach.

Konemann, H. 1981. Fish toxicity tests with mixtures of more than 2 chemicals: A proposal for a quantitative approach and experimental results. *Toxicology* 19: 229–238.

Kortenkamp, A., and R. Altenburger. 1998. Synergisms with mixtures of xenoestrogens: A reevaluation using the method of isoboles. *Science of the Total Environment* 221: 59–73.

Lambert, J.C., and J.C. Lipscomb. 2007. Mode of action as a determining factor in additivity models for chemical mixture risk assessment. *Regulatory Toxicology and Pharmacology* 49: 183–194.

Loewe, S., and H. Muischnek. 1926. On combination effects (D. Doty trans.). *Archiv für Experimentalle Pathologie und Pharmakologie* 114: 313–326.

Main, K.M., N.E. Skakkebaek, H.E. Virtanen, and J. Toppari. 2010. Genital anomalies in boys and the environment. *Best Practice & Research Clinical Endocrinology & Metabolism* 24: 279–289.

Martin, H.L., C. Svendsen, L.J. Lister, J.L. Gomez-Eyles, and D.J. Spurgeon. 2009. Measurement and modeling of the toxicity of binary mixtures in the nematode *Caenorhabditis elegans*-a test of independent action. *Environmental Toxicology and Chemistry* 28: 97–104.

Meadows, S.L., C. Gennings, W.H. Carter Jr., and D.S. Bae. 2002. Experimental designs for mixtures of chemicals along fixed ratio rays. *Environmental Health Perspectives* 110 (Suppl 6): 979–983.

Metzdorff, S.B., M. Dalgaard, S. Christiansen, M. Axelstad, U. Hass, M.K. Kiersgaard, M. Scholze, A. Kortenkamp, and A.M. Vinggaard. 2007. Dysgenesis and histological changes of genitals and perturbations of gene expression in male rats after in utero exposure to antiandrogen mixtures. *Toxicological Sciences* 98: 87–98.

Mileson, B.E., J.E. Chambers, W.L. Chen, W. Dettbarn, M. Ehrich, A.T. Eldefrawi, D.W. Gaylor, K. Hamernik, E. Hodgson, A.G. Karczmar, S. Padilla, C.N. Pope, R.J. Richardson, D.R. Saunders, L.P. Sheets, L.G. Sultatos, and K.B. Wallace. 1998. Common mechanism of toxicity: A case study of organophosphorus pesticides. *Toxicological Sciences* 41: 8–20.

Mumtaz, M., and R.C. Hertzberg. 1993. The status of interaction data in risk assessment of chemical mixtures. In *Hazard assessment of chemicals*, ed. J. Saxena. Wahington DC: Hemisphere.

Mumtaz, M.M., I.G. Sipes, H.J. Clewell, and R.S. Yang. 1993. Risk assessment of chemical mixtures: Biologic and toxicologic issues. *Fundamental and Applied Toxicology* 21: 258–269.

Muska, C.F., and L.J. Weber. 1977. An approach for studying the effects of mixtures of environmental toxicants on whole organism performances. In *Recent advances in fish toxicology: A symposium*, ed. R.A. Tubb. Corvallis: U.S. EPA.

Mwense, M., X.Z. Wang, F.V. Buontempo, N. Horan, A. Young, and D. Osborn. 2004. Prediction of noninteractive mixture toxicity of organic compounds based on a fuzzy set method. *Journal of Chemical Information and Computer Sciences* 44: 1763–1773.

———. 2006. QSAR approach for mixture toxicity prediction using independent latent descriptors and fuzzy membership functions. *SAR and QSAR in Environmental Research* 17: 53–73.

National Research Council. 2008. *Phthalates and cumulative risk assessment: The task ahead*. Washington DC: National Academies Press.

Olmstead, A.W., and G.A. Leblanc. 2005. Joint action of polycyclic aromatic hydrocarbons: Predictive modeling of sublethal toxicity. *Aquatic Toxicology* 75: 253–262.

Payne, J., N. Rajapakse, M. Wilkins, and A. Kortenkamp. 2000. Prediction and assessment of the effects of mixtures of four xenoestrogens. *Environmental Health Perspectives* 108: 983–987.

Petersen, K., H.H. Heiaas, and K.E. Tollefsen. 2014. Combined effects of pharmaceuticals, personal care products, biocides and organic contaminants on the growth of *Skeletonema pseudocostatum*. *Aquatic Toxicology* 150: 45–54.

Plackett, R.L., and P.S. Hewlett. 1948. Statistical aspects of the independent joint action of poisons, particularly insecticides .1. The toxicity of a mixture of poisons. *Annals of Applied Biology* 35: 347–358.

———. 1952. Quantal responses to mixtures of poisons. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 14: 141–163.

———. 1963. A unified theory for Quantal responses to mixtures of drugs – fitting to data of certain models for 2 non-interactive drugs with complete positive correlation of tolerances. *Biometrics* 19: 517.

Pressman, J.G., S.D. Richardson, T.F. Speth, R.J. Miltner, M.G. Narotsky, E.S. Hunter, G.E. Rice, L.K. Teuschler, A. Mcdonald, S. Parvez, S.W. Krasner, H.S. Weinberg, A.B. Mckague, C.J. Parrett, N. Bodin, R. Chinn, C.F.T. Lee, and J.E. Simmons. 2010. Concentration, chlorination, and chemical analysis of drinking water for disinfection byproduct mixtures health effects research: US EPA's four lab study. *Environmental Science & Technology* 44: 7184–7192.

Qin, L.T., S.S. Liu, J. Zhang, and Q.F. Xiao. 2011. A novel model integrated concentration addition with independent action for the prediction of toxicity of multi-component mixture. *Toxicology* 280: 164–172.

Qin, L.T., J. Wu, L.Y. Mo, H.H. Zeng, and Y.P. Liang. 2015. Linear regression model for predicting interactive mixture toxicity of pesticide and ionic liquid. *Environmental Science and Pollution Research* 22: 12759–12768.

Ra, J.S., B.C. Lee, N.I. Chang, and S.D. Kim. 2006. Estimating the combined toxicity by two-step prediction model on the complicated chemical mixtures from wastewater treatment plant effluents. *Environmental Toxicology and Chemistry* 25: 2107–2113.

Rajapakse, N., E. Silva, M. Scholze, and A. Kortenkamp. 2004. Deviation from additivity with estrogenic mixtures containing 4-nonylphenol and 4-tert-octylphenol detected in the e-screen assay. *Environmental Science & Technology* 38: 6343–6352.

Rider, C.V., J. Furr, V.S. Wilson, and L.E. Gray. 2008. A mixture of seven antiandrogens induces reproductive malformations in rats. *International Journal of Andrology* 31: 249–262.

Rider, C.V., J.R. Furr, V.S. Wilson, and L.E. Gray. 2010. Cumulative effects of in utero administration of mixtures of reproductive toxicants that disrupt common target tissues via diverse mechanisms of toxicity. *International Journal of Andrology* 33: 443–462.

Rider, C.V., and G.A. Leblanc. 2005. An integrated addition and interaction model for assessing toxicity of chemical mixtures. *Toxicological Sciences* 87: 520–528.

———. 2006. Atrazine stimulates hemoglobin accumulation in *Daphnia magna*: Is it hormonal or hypoxic? *Toxicological Sciences* 93: 443–449.

Schmidt, S., W. Busch, R. Altenburger, and E. Kuster. 2016. Mixture toxicity of water contaminants-effect analysis using the zebrafish embryo assay (*Danio rerio*). *Chemosphere* 152: 503–512.

Scholze, M., E. Silva, and A. Kortenkamp. 2014. Extending the applicability of the dose addition model to the assessment of chemical mixtures of partial agonists by using a novel toxic unit extrapolation method. *PLoS One* 9: e88808.

Silva, E., N. Rajapakse, and A. Kortenkamp. 2002. Something from "nothing" – eight weak estrogenic chemicals combined at concentrations below NOECs produce significant mixture effects. *Environmental Science & Technology* 36: 1751–1756.

Simmons, J.E. 2013. ToxMix 2011: International toxicology of mixtures conference. A selection of papers. Preface. *Toxicology* 313: 71–72.

Simmons, J.E., S.D. Richardson, L.K. Teuschler, R.J. Miltner, T.F. Speth, K.M. Schenck, E.S. Hunter, and G. Rice. 2008. Research issues underlying the four-lab study: Integrated disinfection by-products mixtures research. *Journal of Toxicology and Environmental Health-Part a-Current Issues* 71: 1125–1132.

Teuschler, L.K., G.E. Rice, C.R. Wilkes, J.C. Lipscomb, and F.W. Power. 2004. Feasibility study of cumulative risk assessment methods for drinking water disinfection by-product mixtures. *Journal of Toxicology and Environmental Health-Part a-Current Issues* 67: 755–777.

Thienpont, B., C. Barata, and D. Raldua. 2013. Modeling mixtures of thyroid gland function disruptors in a vertebrate alternative model, the zebrafish eleutheroembryo. *Toxicology and Applied Pharmacology* 269: 169–175.

U.S. EPA. 1999. Guidance for identifying pesticide chemicals and other substances that have a common mechanism of toxicity. In *Programs, O. O. P.* Wahington, D.C: U.S. EPA.

Van Stee, E.W., R.A. Sloane, J.E. Simmons, M.P. Moorman, and K.D. Brunnemann. 1995. Endogenous formation of N-nitrosomorpholine in mice from (No2)-N-15 by inhalation and Morpholine by gavage. *Carcinogenesis* 16: 89–92.

Villa, S., S. Migliorati, G.S. Monti, and M. Vighi. 2012. Toxicity on the luminescent bacterium *Vibrio fischeri* (Beijerinck). II: Response to complex mixtures of heterogeneous chemicals at low levels of individual components. *Ecotoxicology and Environmental Safety* 86: 93–100.

Chapter 10
Mixtures: Contrasting Perspectives from Toxicology and Epidemiology

Thomas F. Webster

Abstract Investigation of the health effects of mixtures will benefit from the cooperation of toxicologists and epidemiologists. This chapter provides insight into the commonalities and differences in the viewpoint and methods that toxicologists and epidemiologists use to investigate health effects of mixtures. Ways in which these two important disciplines can work together are suggested.

Keywords Epidemiology · Toxicology · Mixtures · Interaction

10.1 Introduction

The field of mixtures has had a resurgence in the last decade, with increasing interest from researchers and governments. As a report from NIEHS on a 2011 workshop put it, "Traditionally, toxicological studies and human health risk assessments have focused primarily on single chemicals. However, people are exposed to a myriad of chemical and nonchemical stressors every day and throughout their lifetime. . .It is imperative to develop methods to assess the health effects associated with complex exposures in order to minimize their impact on the development of disease." Another conclusion was the "need for further collaboration among epidemiologists, toxicologists, and biostatisticians." (Carlin et al. 2013). For such collaborations to be most productive, researchers in all three fields need to be at least aware of each other's jargon, especially the definition of "interaction." Further, toxicologists and epidemiologists need to understand the viewpoints and methods that the two fields use to investigate health effects of mixtures (see also Boedeker and Backhaus 2010; Howard and Webster 2013).

T. F. Webster (✉)
Department of Environmental Health, Boston University School of Public Health, Boston, MA, USA
e-mail: twebster@bu.edu

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_10

10.1.1 The Mixtures Problem for Toxicologists

The mixtures problem typically faced by toxicologists and pharmacologists is to predict the effect of a combination of compounds based only on information for each compound individually, including toxicological mechanism of action. Put more concretely, when and how can the individual dose-response curves plus mechanistic information be used to predict the joint response across a range of doses, i.e., the *dose-response surface*. For simplicity, concepts will be illustrated with a mixture of two compounds; it is important to note that the ideas presented here can be generalized to more complicated situations. This problem can be visualized with a three-dimensional diagram: plot the dose of one compound on the X_1-axis, the dose of the second compound on the X_2-axis, and response on the Y-axis (Fig. 10.1, right). In some situations, the response surface can be accurately predicted using *component-based mixtures models*, as briefly discussed later in this chapter and elsewhere in this volume (Chaps. 9, 11 and 14). Sometimes toxicologists are faced with a well-defined mixture for which the composition is fixed or nearly so. For such multicomponent defined mixtures, direct toxicological testing of the complete mixture can be used. Toxicologists also directly test highly complex environmentally realistic mixtures (whole mixtures) where some part of the mixture mass is known and the rest consists of unidentified components (Narotsky et al. 2012, 2013).

10.1.2 The Mixtures Problem for Epidemiologists

Environmental epidemiologists face a related but complementary problem. For component-based mixtures, toxicologists can choose the compounds they are examining and the combinations of doses. Since epidemiologists cannot ethically expose people to toxic compounds, they look for natural experiments where such exposures occur. For each person, the investigator needs to know their exposure to each compound X_j (during the biologically relevant time period), the outcome Y, as well as potential confounders and effect measure modifiers. At the simplest level, a confounder is a third variable that is associated with the exposure *and* is an independent predictor of the outcome (for a more thorough definition and discussion of confounding, see Aschengrau and Seage 2013; Rothman et al. 2008). If not controlled in some way, confounding causes a distortion of the relationship between the exposure of interest and the outcome; this distortion can be in any direction, either diminishing/masking a true association or creating false associations. It is important to note that the omission of a risk factor for an outcome does not cause confounding if it is not also associated with the exposure of interest. Effect measure modification is discussed in more detail later in the chapter.

For now, two simplifying assumptions are made: the outcome Y is continuous, and there is no effect measure modification. The exposure of each person can be plotted as a point on the X_1-X_2 plane, generating the distribution of data in *exposure*

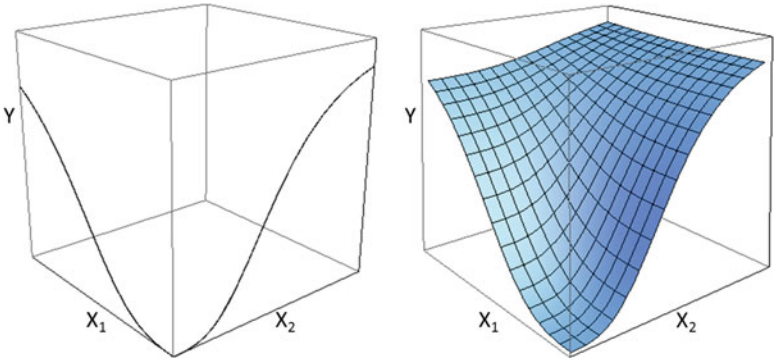

Fig. 10.1 An important mixtures problem in toxicology: When and how can dose-response curves and other information about individual components (left) predict the dose-response surface of the mixture (right)?

space (e.g., Fig. 10.2). Statisticians typically use regression modeling to estimate the associations between each exposure and the outcome (Chap. 8) as well as *statistical interaction*. For example, one might use a regression equation of the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_3 Z + \varepsilon \qquad (10.1)$$

where β_0 is a constant; β_1 and β_2 are the effect estimates for the two exposures X_1 and X_2 individually (called the main effects) (e.g., Harrell 2001). In statistics, $\beta_{12} X_1 X_2$ is called a multiplicative interaction term. As discussed later in this chapter, this is not necessarily the same as interaction from an epidemiologic or toxicologic point of view. Z is a confounder that requires adjustment, and ε is an error term, assumed to be random and unmeasured. Regression uses the data for each person (X_{ij}, Y_i, Z_i) to estimate the parameters β_0, β_1, β_2, β_{12}, and β_3. Equation 10.1 is quite simple. In addition to the two assumptions discussed above, it also assumes that individual dose-response curves are linear and there is only one confounder; these assumptions can be relaxed. More complicated interaction terms can also be used. While toxicologists can usually avoid confounding by experimental design (Chaps. 8 and 13), it is a major cause of concern in environmental epidemiology: this difference arises because of the uncontrolled quality of natural experiments. For example, only a limited number of exposure variables (e.g., different chemicals) are usually measured, potentially causing confounding.

While a component-based toxicology approach uses data on individual compounds to estimate the effect of the mixture, the epidemiologist tries to estimate the response surface *directly* from the data in exposure space: Equation 10.1 is a model of the dose-response surface (it also yields estimates of the individual dose-response curves, e.g., by setting the other compound equal to zero). Unlike toxicologists, epidemiologists often have neither a priori individual chemical dose-response information nor information about toxicological mechanism.

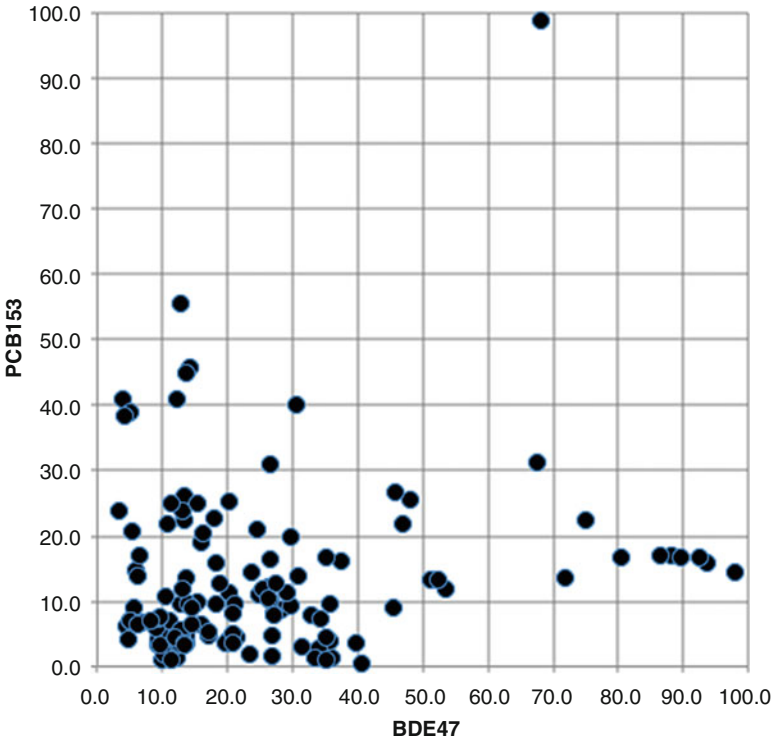

Fig. 10.2 Example of exposure space for two compounds BDE 47 and PCB153, as they occur in human serum (ng/g lipid weight, unpublished data). Each axis represents an exposure; each point represents a person

Epidemiologists have at least three questions in mind when studying exposure to mixtures (Braun et al. 2016): (1) *Variable selection*: Which components of the mixture contribute to the outcome? In our example, are both X_1 and X_2 associated with the effect? (2) Are there "interactions" (however defined) between the two exposures? (3) Can some kind of *summary measure* of the exposures be constructed (discussed later in this chapter)? Examining mixtures in epidemiology is a difficult problem, facing a number of challenges (e.g., Braun et al. 2016).

Several recent efforts have been aimed at developing methods to evaluate mixtures of chemicals in epidemiology: e.g., the EPA multipollutant workshop (Johns et al. 2012) and the NIEHS workshop *Statistical Approaches for Assessing Health Effects of Environmental Chemical Mixtures in Epidemiology* (NIEHS 2015). Novel methods that have been applied to assess mixtures in epidemiological settings include weighted quantile sum regression (Carrico et al. 2015), Bayesian kernel machine regression (Bobb et al. 2014), and exposure space smoothing (Webster and Vieira 2015). A series of publications are expected to come out of the NIEHS workshop comparing various methods (e.g., Taylor et al. 2016).

10.2 Toxicology and Epidemiology of Mixtures: The Importance of Exposure Space

To generalize, instead of two exposures, suppose there are *J* exposures. Thus, instead of the simple diagram in Fig. 10.2, the exposure space has *J* dimensions; the response variable adds an additional dimension. This space is potentially very large: it is estimated that there are somewhere between 25,000 and 84,000 chemicals in commerce in the USA (IOM 2014). To this one might add natural compounds, metabolites, pharmaceuticals, and nonchemical exposures. (However, we have still simplified, as considerations of time and exposure measurement error are omitted from the discussion.)

It is currently not possible to simply test our way out of the mixtures problem: the numbers are too big. For example, even if we examined only one dose of only three-way combinations of 25,000 chemicals, the number of toxicology experiments required is 2.6×10^{12}. We clearly need ways to reduce the number of combinations. This sobering fact provides a compelling rationale for the following two main toxicological approaches to mixtures. The component-based approach, when it is applicable, only requires data on individual compounds. When the relative composition of a mixture is fixed, then varying the dose of the whole mixture produces a ray in exposure space, a line from the origin outward (Chap. 13).

The distribution of points in exposure space is thus of key importance for both toxicology and epidemiology (e.g., Carlin et al. 2013). Large pieces of the space will be empty: such combinations of exposure do not occur and don't require investigation. Some exposures will be highly correlated, even forming rays amenable to whole mixture approaches. Understanding exposure space is also critical for environmental epidemiologists as they cannot control the distribution of exposures except to some degree by choice of populations. The good news is that epidemiologists study exposures as they actually occur—thus targeting important parts of exposure space—at least if they are measured. But this can also pose problems. For example, estimating both main effects and interactions is difficult unless the population under study is sufficiently large, and the data are to some degree spread across exposure space. For example, with two exposures, one would ideally have groups of people with exposure to both compounds, exposure to neither compound, and exposure to only one or the other. If two exposures are highly correlated, it is difficult to disentangle separate effects: putting both exposures in the regression model can produce unstable estimates, a problem called collinearity (note that if exposures are very highly correlated, the epidemiological problem is related to the whole mixture approaches of toxicology). Suppose components A and B are correlated because they come from a common source, but only A contributes to the effect. If only B is measured (or included in a model), it will incorrectly appear to be associated with the outcome, i.e., it is confounded by the missing exposure (Fig. 10.3). For all of these reasons, it is important to increase the number of exposures that are measured and examined: expanded targeted analysis, nontargeted analysis, and similar approaches are critical for a better understanding of what

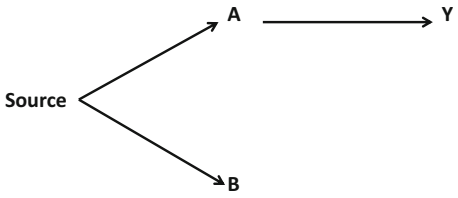

Fig. 10.3 Suppose that exposures A and B are correlated because they both arise from a common source. A, but not B, causes the outcome Y. If B is the only exposure measured, it will falsely appear to be associated with Y due to confounding

mixtures occur. Targeted analysis looks for specific compounds in a sample; nontargeted analysis is a screening approach that can identify previously unknown or uncharacterized exposures (e.g., Getzinger et al. 2015; Chaps. 3 and 4). Analysis of such expanded exposure data will also require larger sample sizes, e.g., to achieve desired levels of statistical power.

Methods for analyzing the information in exposure space are also important. For example, Fig. 10.4 illustrates the correlations of a set of persistent organic pollutants in human serum from the same cohort as Fig. 10.2, but with more compounds (unpublished data). The dendrogram, which can be interpreted similar to a family tree, was constructed using hierarchical clustering. Compounds joined closer to the bottom are more highly correlated (using Spearman's correlation coefficients of the serum concentrations), those joined at the top less so. For example, BDE47 and BDE99 are two highly correlated (and tightly clustered) compounds. They are two polybrominated diphenyl ether (PBDE) congeners that occur in the same commercial flame retardant and that have similar routes of exposure. The PBDEs and PCBs are in two different clusters because they are not very correlated with each other. For example, BDE47 is not well correlated with PCB153, as shown in Fig. 10.2. This suggests that PBDEs and PCBs are unlikely to confound each other (at least in this cohort); that does not preclude "additive" or "interactive" effects however.

10.3 Component-Based Mixture Methods in Toxicology

When mechanisms are sufficiently well understood, models of the effect of mixtures can sometimes be constructed. For example, biologically based mathematical models can be constructed of the effects on receptor activation by mixtures of ligands, one important mechanism for endocrine disruption (e.g., Weiss et al. 1996; Howard and Webster 2009; Webster 2013). However, toxicologists and risk assessors often need to make predictions without this kind of detail. Two main approaches are used. As these ideas are discussed in more detail elsewhere in this book (Chap. 9), these concepts will be only briefly reviewed.

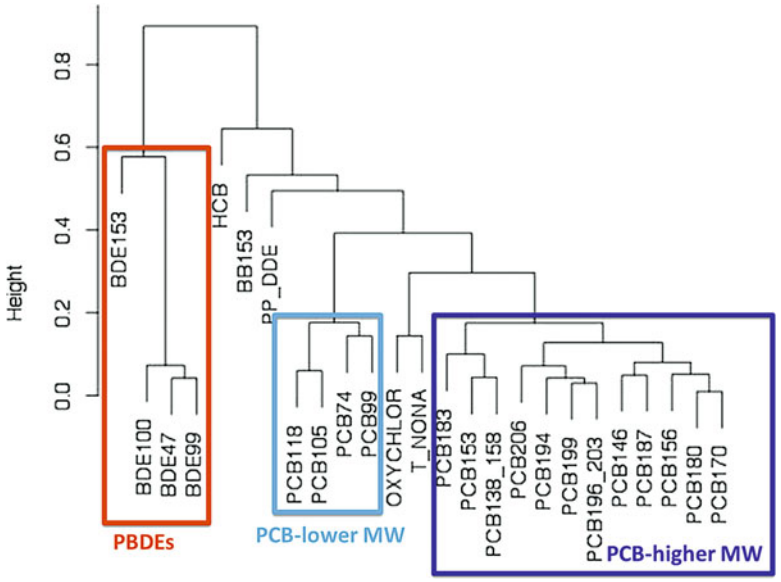

Fig. 10.4 Dendrogram showing the correlations between the concentrations of a number of persistent organic pollutants in human serum

For compounds that act via similar toxicological mechanisms, one approach assumes *dose addition*, also known as concentration addition. Compounds that are dose additive obey the following equation (see Chap. 9):

$$\sum_{j=1}^{J} \frac{d_j}{\mathrm{ED}_{j,y}} = 1 \tag{10.2}$$

where d_j is the dose of compound j and $\mathrm{ED}_{j,y}$ is the dose of compound j alone that causes response level y, e.g., the ED_{10} (Berenbaum 1989). Depending on the number of dimensions, Eq. 10.2 describes a line, plane, or hyperplane. As a result, the isoboles (contours) of the response surface form negatively sloped lines, planes, or hyperplanes when projected onto the exposure space (Fig. 10.5).

A simple version of dose addition is *toxic equivalence*: it can occur when the relative potency between compounds is the same at *all* response levels (Howard and Webster 2009; Chap. 14). The joint response of the mixture under toxic equivalence (f_{TE}) is a function of a linear combination of the component doses scaled by potency. If chemical 1 is selected as the reference compound, then the mixture response can be represented by the dose-response model for chemical 1 applied to the linear combination of the component doses

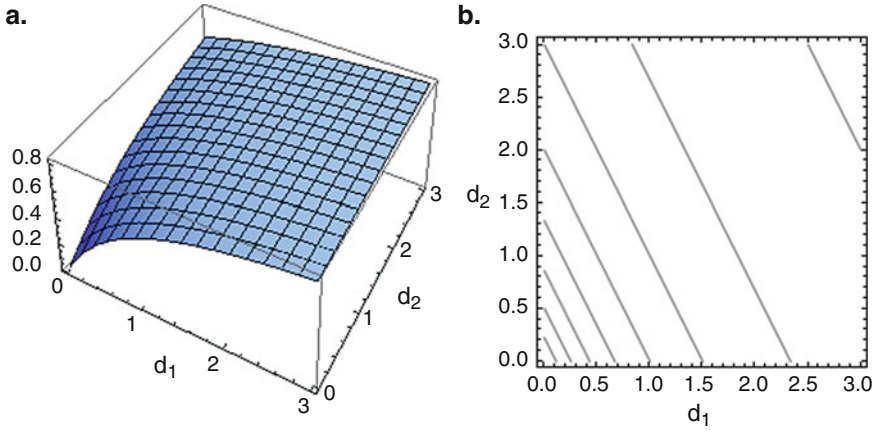

Fig. 10.5 Isoboles: The response surface on the left (**a**) yields the contours (isoboles) on the right (**b**). Compounds that are dose additive have isoboles that are negatively sloped straight lines. For compounds that follow the toxic equivalence model (a special case of dose addition with constant relative potency), the isoboles are also parallel

$$f_{\mathrm{TE}}[d_1, \ldots, d_J] = f_1\left[\sum_{j=1}^{J} \gamma_j d_j\right] \tag{10.3}$$

where $f_1[.]$ is the dose-response curve for the reference compound, compound 1. The γ_j are the relative potency factors (RPFs) compared to a reference compound. When the reference compound has the highest potency, the other compounds in the mixture act as if they are dilute versions of the reference compound (for a discussion of definitions of RPFs and TEFs, see Chap. 14 as well as USEPA 2008). The concept underlying dose addition is perhaps most easily seen for toxic equivalence: one first scales doses by their relative potencies (to give their equivalent doses as chemical 1) and then applies the dose-response function of chemical 1 to total equivalent dose. Under toxic equivalence, the isoboles are always parallel, negatively sloped straight lines (Fig. 10.5b).

We have proposed a modification of dose addition called generalized concentration addition (GCA) that can in principle handle mixtures of full and partial agonists, i.e., compounds with different maximal responses (Howard and Webster 2009). This class of models has been successfully applied to mixtures of full and partial agonists of the AhR and PPARγ receptors (Howard et al. 2010; Watt et al. 2016). Isoboles for mixtures obeying GCA are straight lines but can have negative or positive slopes: the latter implies that the partial agonist acts like a competitive antagonist at response levels above its maximum effect level (Howard and Webster 2009).

When compounds act via different mechanisms, many mixtures toxicologists use *independent action*. The mixture response expected under independent action (f_{IA}) is:

$$f_{\mathrm{IA}}[d_1,\ldots,d_J] = 1 - \prod_{j=1}^{J}\left(1 - f_j[d_j]\right) \tag{10.4}$$

Originally derived from independence in probability theory, independent action assumes that responses range between zero and one.

Effect summation (ES) is defined by

$$f_{\mathrm{ES}}[d_1,\ldots,d_J] = \sum_{j=1}^{J} f_j[d_j] \tag{10.5}$$

and describes the excess effect, above controls (Chap. 9) (note that at very low effect levels independent action is approximated by effect summation). It is worth emphasizing that effect summation has often been rejected as a general mixtures model by mixtures toxicologists, e.g., Howard and Webster (2009), but may be useful when dose-response curves are approximately linear or under other conditions (Chap. 14). Effect summation frequently appears in the toxicology literature as well as in textbooks.

Perhaps not surprisingly, there has been discussion in the toxicology literature about when one should use dose addition vs. independent action, in particular, how similar the toxicological mechanisms must be for dose addition to apply (Webster 2013). The choice of dose addition vs. independent action can have profound consequences; this is nicely illustrated in the "something from nothing" experiment of Silva et al. (Silva et al. 2002), where a mixture of xenoestrogens, each at doses less than their empirical no effect levels, produces a response in combination; independent action would predict no combination response. Further, the compounds exhibit toxic equivalence and thus dose addition well describes the mixture response. As the dose-response curves for these compounds are concave upward at low doses (Fig. 10.6a), the combined response exceeds that predicted by effect summation (Silva et al. 2002; Rajapakse et al. 2002). Some models (integrated addition models) combine features of both dose addition and independent action (e.g., Rider and LeBlanc 2005, Chap. 9)

In sum, mixtures toxicologists often use either dose addition or independent action to estimate the effect of mixtures from individual components. The choice often depends on the toxicological similarity of the components. Furthermore, dose addition (or GCA) and independent action can be considered as toxicologic definitions of *non-interaction/additivity*. Having specified such a definition, one can then determine if a mixture has an interaction, producing responses greater than additive or less than additive relative to the chosen definition.

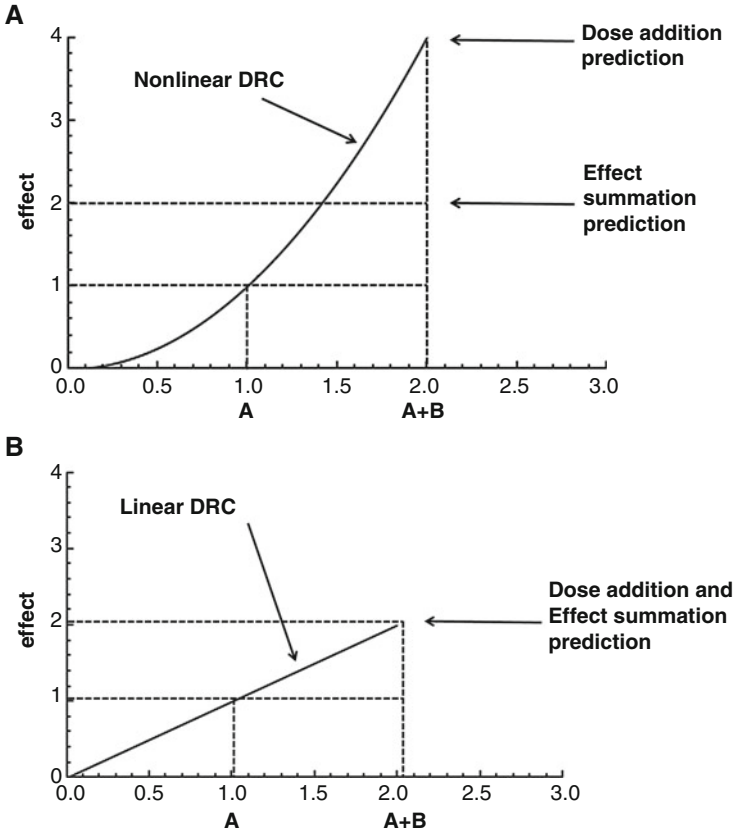

Fig. 10.6 (**a**) Suppose A and B follow the toxic equivalence model and have the same potency, with a nonlinear response curve that is concave up. The same dose of A alone or B alone would give the same response. For a mixture of these doses, effect summation would predict twice the response of either compound alone. Dose addition gives the correct, higher value. But from the point of view of epidemiology, the incremental effect of compound B depends on the amount of compound A. They interact, as defined by epidemiologists. (**b**) Suppose A and B follow the toxic equivalence model and have the same potency, with a linear response curve that is concave up. The same dose of A alone or B alone would give the same response. For a mixture of these doses, effect summation and dose addition gives the same, correct value. From the point of view of epidemiology, the incremental effect of compound B does not depend on the amount of compound A. They do not interact, as defined by epidemiologists

10.4 Additivity and Interaction in Epidemiology

To borrow a frequently used example (Rothman 1986), Table 10.1 shows hypothetical risks of lung cancer categorized by exposure to asbestos and smoking in a cohort study (rates of disease could also be used). Let's assume there is no confounding or

Table 10.1 Hypothetical risks of lung cancer categorized by exposure to asbestos and smoking (2 × 2 table). Risks and risk differences (RD) are expressed as cases per 100,000; relative risks (RR) are ratios

	Exposed to asbestos	Not exposed to asbestos	RR^a	RD^a
Smoker	**50**	**10**	*5*	*40*
Non-smoker	**5**	**1**	*5*	*4*
RR^b	*10*	*10*		
RD^b	*45*	*9*		

[a]Exposed to asbestos vs. unexposed to asbestos, stratified by smoking status
[b]Smokers vs. non-smokers, stratified by asbestos status

other biases. The highest risk is in the doubly exposed people. As will be seen, epidemiologists think about interaction in quite a different way from toxicologists.

Epidemiologists summarize the associations between exposure and disease using effect measures. For example, Table 10.1 shows the relative risks (RR) for the association between lung cancer and asbestos, holding smoking constant. In this case, the RR—the ratio of the risk in the asbestos exposed to the risk in the asbestos unexposed—equals 5 for smokers and also 5 for non-smokers. As the RRs are the same in both strata, there is no *effect measure modification* of the asbestos RR by smoking. Similarly, Table 10.1 shows that asbestos exposure does not modify the RR for smoking and lung cancer: it is 10 in both strata. On the other hand, suppose the investigator used another equally valid effect measure: the risk difference (RD), which equals the risk in the exposed minus the risk in the unexposed. The RD for the association between lung cancer and asbestos is 40/100,000 for smokers and 4/100,000 for non-smokers. Thus, the asbestos RD is modified by smoking (and vice versa). This phenomenon is called effect measure modification because it depends on the choice of effect measure, here either RD or RR (there are other possibilities such as odds ratios, typically used in case-control studies).

Epidemiologists consider effect measure modification to be descriptive. Interaction is a different concept. Like toxicologists, epidemiologists also use the term interaction to mean nonadditive, either greater than additive (called synergism by epidemiologists) or less than additive. Unlike toxicology, epidemiologists have a single definition of non-interaction (additivity), making no distinction for mechanism: it is based on additivity of risk differences. For the example in Table 10.1, epidemiologists would say that asbestos and smoking interact, having a greater than additive (synergistic) effect on lung cancer. Indeed, some epidemiologists call this "biologic interaction" (e.g., Ahlbom and Alfredsson 2005).

To see how epidemiologists judge Table 10.1 to display nonadditivity, let's briefly review one derivation of the epidemiologic definition of interaction (for a detailed explanation, see Howard and Webster 2013; Rothman et al. 2008). As in Table 10.1, epidemiologic examples traditionally use binary exposures and outcomes. One model for interaction in epidemiology relies on what are called counterfactual susceptibility types. For pairs of two binary exposures A and B, there are 16 possible patterns of exposures and outcomes, each of which can be considered a possible response type (Table 10.2). For some types, one needs to know the value of

Table 10.2 Counterfactual susceptibility type model for two exposures provides a basis for thinking about interaction for epidemiologists. The outcome (binary) depends on the combination of exposure (A, B)

Type	A = 1 B = 1	A = 0 B = 1	A = 1 B = 0	A = 0 B = 0	Description
1	1	1	1	1	Doomed (always develops outcome)
2[a]	1	1	1	0	A causal, B causal, A and B causal
3[a]	1	1	0	1	
4	1	1	0	0	A ineffective, B causal
5[a]	1	0	1	1	
6	1	0	1	0	A causal, B ineffective
7[a]	1	0	0	1	A preventative, B preventative, A and B antagonizes
8[a]	1	0	0	0	A and B causal
9[a]	0	1	1	1	A and B preventative
10[a]	0	1	1	0	A causal, B causal, A and B antagonizes
11	0	1	0	1	A preventative, B ineffective
12[a]	0	1	0	0	
13	0	0	1	1	A ineffective, B preventative
14[a]	0	0	1	0	
15[a]	0	0	0	1	A preventative, B preventative, A and B preventative
16	0	0	0	0	Immune (never develops outcome)

[a]Interdependent types

both exposures to know the outcome. For example, for people of type 8, the outcome occurs only if both exposures occur, i.e., A = B = 1. Epidemiologists call these types *interdependent*. For non-interdependent types, the effect of exposure to one compound does not depend on the other. For example, people of type 6 will have the outcome if A = 1, irrespective of the value of B. Risks associated with different exposure scenarios can be written as r_{AB}; e.g., r_{10} is the risk in the population exposed to A but not to B. Writing down the risks associated with only the non-interdependent types—types 1, 4, 6, 11, 13, and 16—(and assuming no confounding or bias) yields the following equation:

$$(r_{11} - r_{00}) = (r_{10} - r_{00}) + (r_{01} - r_{00}) \tag{10.6}$$

where r_{ij} denotes the risk for each exposure patterns (e.g., r_{10} means the risk in a population exposed to X_1 and unexposed to X_2). This equation means that the risk difference between the jointly exposed (r_{11}) and the jointly unexposed (r_{00}) is equal to the sum of the risk differences due to individual exposures, ($r_{10}-r_{00}$) and ($r_{01}-r_{00}$). Since this equation includes only non-interdependent types, deviation from this equation implies the presence of interdependent types. Thus risk difference additivity is used by epidemiologists as the criteria for interaction/interdependence. It is necessary but not sufficient, i.e., interaction may occur even if Eq. 10.6 holds, e.g., if risks associated with interdependent types cancel out (dividing Eq. 10.6 by r_{00}

provides equivalent criteria in terms of relative risks). Applying this equation to Table 10.1 shows that asbestos and smoking interact: indeed they have a greater than additive effect on lung cancer. Setting $r_{11} = 50$, $r_{10} = 10$, $r_{01} = 5$, $r_{00} = 1$ yields

$$(50 - 1) > (10 - 1) + (5 - 1) \tag{10.7}$$

For simplicity, the denominator of 100,000 was omitted for all of the numbers in Eq. 10.7.

This all might seem very reasonable until one realizes the following fact: *the epidemiologic definition of biological interaction is consistent with effect summation, the definition rejected by many mixtures toxicologists!* Equation 10.6 is a special case of effect summation, Eq. 10.5, where outcomes and exposures are binary (Howard and Webster 2013). To see this, recall that effect summation examines the excess effect above controls. For two exposures, Eq. 10.5 is equivalent to Eq. 10.6 with the background risk (r_{00}) subtracted.

Similar ideas about interaction are sometimes applied to continuous outcomes in epidemiology. For continuous outcomes Y (untransformed) and linear dose-response curves (or binary exposures), regression models such as Eq. 10.1 can test for statistical interaction (and control for confounding). The beta coefficients for main effects are the differences in outcome per unit of exposure. Results consistent with β_{12} meaningfully different from zero imply there is an interaction using the epidemiologic definition. This conclusion depends on the use of an additive scale. The presence of a non-zero interaction term in more general regression models does not necessarily imply interaction from the epidemiologic point of view, i.e., statistical interaction is not the same as epidemiologic interaction. For example, in a logistic model, as is commonly used for binary outcomes, one cannot simply examine an interaction term; there are, however, more complicated methods to assess epidemiologic interaction in such cases (Andersson et al. 2005).

10.5 Contrasting Interaction in Toxicology and Epidemiology

What would toxicologists and pharmacologists say about the data in Table 10.1? To highlight differences between epidemiology and toxicology, consider replacing asbestos and smoking by compounds A and B, about which little is known (the epidemiologic conclusion would remain the same). The answer would depend on whether one hypothesized that A and B worked by similar or different mechanisms. If the toxicologist thought they acted by "different" mechanisms, they might use independent action and conclude that it is greater than additive (since the risks are small, independent action is approximately equal to effect summation). Suppose they believed the chemicals acted by "similar" mechanisms? Unfortunately, Table 10.1 does not contain enough information to determine if A and B

(or smoking and asbestos) are dose additive. One would also need information or assumptions about the dose-response curves for each compound alone.

The contrasting ideas about interaction and additivity between epidemiology and toxicology are perhaps most stark when comparing effect summation with dose addition for a mixture of compounds that follow toxic equivalence (a special case of dose addition) and a dose-response curve that curves upward (e.g., Fig. 10.6a). The mixture is nonadditive from the epidemiologic point of view where one first applies the dose-response function to each compound separately and adds the results: the sum of the effects is less than the effect of the mixture. But toxicologists would define this mixture as additive (relative to dose addition). For toxic equivalence, one first adds component doses scaled by relative potency factors and then applies the dose-response function of the reference compound. The contrast derives from the underlying logic of the epidemiologic definition. As illustrated in Fig. 10.6a, the increase in the effect due to a dose of compound B *depends* on whether compound A is also present in the mixture. Hence, nonlinear dose-response curves will lead to interaction as defined by epidemiologists. The toxicologic and epidemiologic definitions coincide when the dose-response curve is linear (Fig. 10.6b).

Now suppose that A and B are merely different doses of the same substance, an idea called sham substitution (Berenbaum 1989). Although the definition of dose addition used here does not depend on this idea, it is sometimes used by toxicologists as a rationale for thinking of dose addition as noninteractive. According to this line of thought, a compound does not interact with itself. From the epidemiologic point of view, sham substitution implies that different doses of the same compound do interact when dose-response curves are nonlinear (Rothman 1974; Howard and Webster 2013).

Toxicologists and epidemiologists thus use the same terminology—additive, greater than additive, less than additive—but mean something quite different. Understanding this difference is important for interpreting mixtures studies that come out of the two fields. None of this discussion means that toxicology is correct and epidemiology is wrong or vice versa. Definitions cannot be "wrong" (at least, if used logically); the real test is whether they are useful. Research is needed comparing mixtures studies in the two fields. Epidemiology has the possible advantage of relying on one definition rather than two as in toxicology, where the choice depends on sometimes fuzzy distinctions about similarity of mechanism (Howard and Webster 2013). It is possible that the toxicologic definition sheds more light on biology, whereas the epidemiologic definition (despite being called biologic by some epidemiologists) might be more useful for thinking about intervention to protect public health (e.g., Rothman et al. 1980). As an example of the latter, consider two exposures that have a greater than additive effect from the epidemiologic point of view; this implies that reduction of either exposure may lead to a dramatic reduction of risk. Returning to our example of Table 10.1, preventing exposure to either smoking or asbestos would have a large impact, greatly reducing the risk of lung cancer in those who would otherwise have been exposed to both.

10.6 Combining Ideas from Toxicology and Epidemiology

Progress on mixtures would benefit from greater communication and collaboration between toxicologists, epidemiologists, statisticians, and exposure scientists. The mixtures problem can be thought of as having two sub-questions:

1. What are the patterns of co-exposure in real populations?
2. What are the health effects of the mixtures to which populations are exposed?

As discussed above, exposure science has much to contribute to the first question. The second can be investigated by the complementary approaches of toxicology and epidemiology.

Epidemiologists have used some of the toxicological ideas discussed in this chapter. Perhaps the best example is the use of TEFs when studying health effects in people of exposure to dioxin-like compounds (e.g., Korrick et al. 2011). For example, when exposure is measured using blood concentrations, one multiplies the concentrations by the appropriate TEFs and sums. The result is then used as a summary measure of exposure in a regression equation that gets around collinearity problems. With sufficient toxicological information to construct RPFs, this strategy could be applied to other classes of compounds. This approach could be a very fruitful line of collaboration between toxicologists and epidemiologists.

Let's now take a more general view of the mixtures problem for epidemiologists, one called *exposure space smoothing* (Webster and Vieira 2015). Important limitations of Eq. 10.1 include the assumption of linearity of dose-response functions and a particular mathematical form of the interaction term. These restrictions can be avoided by using a smoothing function $f(.)$ of the exposures

$$g[Y] = f[X_1, X_2] + \gamma' Z + \varepsilon \tag{10.8}$$

For simplicity, only two exposures (X_1, X_2) are shown, but higher dimensional smooths are possible with sufficient data. Equation 10.8 also includes a link function $g[Y]$ of the outcome, allowing the use of continuous, binary, and other types of outcome data. Equation 10.8 can be treated as a generalized additive model (gam) (Hastie and Tibshirani, 1990). Gams can also adjust for confounders, important for any epidemiologic analysis (Eq. 10.8 adjusts for a vector of confounders Z). Rather than impose a specific functional form (e.g., linearity), smoothing functions use the data to inform the shape. There are a number of ways to do smoothing, but one method estimates the value of the function at a particular point by using a weighted average of the outcomes at points that are nearby in exposure space (for details on how this works in two-dimensional geographic space, see Webster et al. 2006). The results of a two-dimensional smooth can be displayed in a number of ways, e.g., as color-coded maps or by using contours (e.g., Fig. 10.5b). For more dimensions, slices can be displayed. Such results can also be used as exploratory data analysis to inform additional modeling. The contours of the response surface, called isoboles, have a toxicologic interpretation. As discussed above, isoboles that are

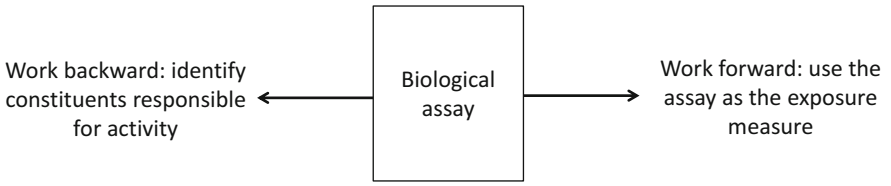

Fig. 10.7 Overview of EAMEDA: Exposomic Analysis of Mixtures via Effect Directed Analysis. A biological assay (e.g., luciferase reporter assay) is used to measure the integrated activity of the sample. The investigator uses these results as the exposure measure in an epidemiology study. One also uses effect-directed analysis (or some related technique) to determine which compounds in the sample account for the activity

approximately negatively sloped parallel straight lines suggest that a summary measure can be constructed using RPFs. If the RFPs are not known from toxicologic data, they might be estimated using other approaches, including methods such as weighted quantile sum regression (Carrico et al. 2015). Isoboles which curve toward the origin suggest a greater than dose-additive response; isoboles which curve away from the origin suggest a less than dose-additive response.

Another interesting potential approach that combines aspects of exposure science, toxicology, and epidemiology is EAMEDA: exposomic analysis of mixtures via effect-directed analysis (Fig. 10.7). Effect-directed analysis (EDA) uses a high-throughput response assay, e.g., reporter assays, to biologically compute the combined effect of a mixture (e.g., serum or dust) on that biological endpoint. Working backward, chemical fractionation and targeted and nontargeted analysis are used to identify the components of the mixture responsible for the result (e.g., Simon et al. 2013; Fang et al. 2015, Chap. 3). For example, Fang et al. (2015) measured total PPARγ activity of dust extracts using a reporter assay. The dust samples were then chemically fractionated with normal phase high-performance liquid chromatography. Each fraction was retested with the reporter assay. In fractions with significant activity, compounds were identified using targeted and nontargeted analysis. Fatty acids were determined to be a major contributor to the dust PPARγ activity. Simon et al. (2013) used a transthyretin-binding assay to examine an aspect of thyroid hormone disruption in polar bears. Nonylphenols and certain hydroxylated PCBs contributed to this activity in plasma. The biologically based measure of activity can be considered the central focus of the EAMEDA concept: (1) Investigators work backward using EDA to determine which compounds contribute to the activity. (2) They could also work forward, using the result of the assay—an integrated measure of the activity of the mixture—as the measure of exposure in an epidemiologic study. Clearly, the appropriateness of the assay and the samples would need to be carefully considered.

I look forward to greater synergy—or at least additivity—between toxicologists, epidemiologists, statisticians, and exposure scientists in investigations of the mixtures problem.

Acknowledgments I would like to thank the many colleagues whom I have had the pleasure to work with on mixtures, particularly Mingliang Fang, Greg Howard, Emma Virginia Preston, Jennifer Schlezinger, Heather Stapleton, Verónica Vieira, James Watt, and a number of others. Thanks also to NIEHS and the Superfund Research Program (grant P42ES007381).

References

Ahlbom, A., and L. Alfredsson. 2005. Interaction: A word with two meanings creates confusion. *European Journal of Epidemiology* 20 (7): 563–564.

Andersson, T., L. Alfredsson, H. Källberg, S. Zdravkovic, and A. Ahlbom. 2005. Calculating measures of biological interaction. *European Journal of Epidemiology* 20 (7): 575–579.

Aschengrau, A., and G. Seage. 2013. *Essentials of epidemiology in public health*. 3rd ed. Burlington: Jones & Bartlett Learning.

Berenbaum, M.C. 1989. What is synergy? *Pharmacological Reviews* 41 (2): 93–141.

Bobb, J.F., L. Valeri, B. Claus Henn, D.C. Christiani, R.O. Wright, M. Mazumdar, J.J. Godleski, and B.A. Coull. 2014. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics* 16: 1–16.

Boedeker, W., and T. Backhaus. 2010. The scientific assessment of combined effects of risk factors: different approaches in experimental biosciences and epidemiology. *European Journal of Epidemiology* 25: 539–546.

Braun, J.M., C. Gennings, R. Hauser, and T.F. Webster. 2016. What can epidemiological studies tell us about the impact of chemical mixtures on human health? *Environmental Health Perspectives* 124: A69.

Carlin, D.J., C.V. Rider, R. Woychik, and L.S. Birnbaum. 2013. Unraveling the health effects of environmental mixtures: An NIEHS Priority. *Environmental Health Perspectives* 121: A7–A8.

Carrico, C., C. Gennings, D. Wheeler, and P. Factor-Litvak. 2015. Characterization of a weighted quantile sum regression for highly correlated data in a risk analysis setting. *Journal of Agricultural, Biological, and Environmental Statistics* 20 (1): 100–120.

Fang, M., T.F. Webster, and H.M. Stapleton. 2015. Effect-directed analysis of human peroxisome proliferator-activated nuclear receptors (PPARγ1) ligands in indoor dust. *Environmental Science & Technology* 49 (16): 10065–10073.

Getzinger, G.J., M.P. O'Connor, K. Hoelzer, B.D. Drollette, O. Karatum, M.A. Deshusses, P.L. Ferguson, M. Elsner, and D.L. Plata. 2015. Natural gas residual fluids: Sources, endpoints, and organic chemical composition after Centralized Waste Treatment in Pennsylvania. *Environmental Science & Technology* 49 (14): 8347–8355.

Harrell, F.E., Jr. 2001. *Regression modeling strategies*. New York: Springer-Verlag.

Hastie, T.J., and R.J. Tibshirani. 1990. *Generalized additive models*. Boca Raton: Chapman & Hall/CRC.

Howard, G.J., and T.F. Webster. 2009. Generalized concentration addition: A method for examining mixtures containing partial agonists. *Journal of Theoretical Biology* 259: 469–477.

———. 2013. Contrasting theories of interaction in epidemiology and toxicology. *Environmental Health Perspectives* 121: 1–6.

Howard, G.J., J.J. Schlezinger, M.E. Hahn, and T.F. Webster. 2010. Generalized concentration addition predicts joint effects of aryl hydrocarbon receptor agonists with partial agonists and competitive antagonists. *Environmental Health Perspectives* 118 (5): 666–672.

Institute of Medicine (IOM). 2014. Board on Population Health and Public Health Practice, Roundtable on Environmental Health Sciences, Research, and Medicine. *Identifying and reducing environmental health risks of chemicals in our society*. Washington (DC): National Academies Press.

Johns, D.O., L.W. Stanek, K. Walker, S. Benromdhane, B. Hubbell, M. Ross, R.B. Devlin, D.L. Costa, and D.S. Greenbaum. 2012. Practical advancement of multipollutant scientific and risk assessment approaches for ambient air pollution. *Environmental Health Perspectives* 120: 1238–1242.

Korrick, S.A., M.M. Lee, P.L. Williams, O. Sergeyev, J.S. Burns, D.G. Patterson, W.E. Turner, L.L. Needham, L. Altshul, B. Revich, and R. Hauser. 2011. Dioxin exposure and age of pubertal onset among Russian boys. *Environmental Health Perspectives* 119 (9): 1339–1344.

Narotsky, M.G., J.G. Pressman, R.J. Miltner, T.F. Speth, L.K. Teuscher, G.E. Rice, S.D. Richardson, D.S. Best, A. McDonald, E.S. Hunter, and J.E. Simmons. 2012. Developmental toxicity evaluations of whole mixtures of disinfection by-products using concentrated drinking water in rats: Gestational and lactational effects of sulfate and sodium. *Birth Defects Research Part B: Developmental and Reproductive Toxicology* 95 (3): 202–212.

Narotsky, M.G., G.R. Klinefelter, J.M. Goldman, D.S. Best, A. McDonald, L.F. Strader, Juan D. Suarez, D. Juan, A.S. Murr, I. Thillainadarajah, E.S. Hunter, S.D. Richardson, T.F. Speth, Richard J. Miltner, J.G. Pressman, L.K. Teuschler, G.E. Rice, V.C. Moser, W. Robert, R.W. Luebke, and J.E. Simmons. 2013. Comprehensive assessment of a chlorinated drinking water concentrate in a rat multigenerational reproductive toxicity study. *Environmental Science and Technology* 47: 10653–10659.

NIEHS (National Institute of Environmental Health Sciences). 2015. *Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology*. July 13–14, http://www.niehs.nih.gov/about/visiting/events/pastmtg/2015/statistical/.

Rajapakse, N., E. Silva, and A. Kortenkamp. 2002. Combining xenoestrogens at levels below individual no-observed-effect concentrations dramatically enhances steroid hormone action. *Environmental Health Perspectives* 110 (9): 917–921.

Rider, C.V., and G.A. LeBlanc. 2005. An integrated addition and interaction model for assessing toxicity of chemical mixtures. *Toxicological Sciences* 87 (2): 520–528.

Rothman, K.J. 1974. Synergy and antagonism in cause-effect relationships. *American Journal of Epidemiology* 99 (6): 385–388.

———. 1986. *Modern epidemiology*. Boston: Little, Brown and Co..

Rothman, K.J., S. Greenland, and A.M. Walker. 1980. Concepts of interaction. *American Journal of Epidemiology* 112 (4): 467–470.

Rothman, K.J., S. Greenland, and T.L. Lash. 2008. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott, Williams and Williams.

Silva, E., N. Rajapakse, and A. Kortenkamp. 2002. Something from "nothing" – eight weak estrogenic chemicals combined at concentrations below NOECs produce significant mixture effects. *Environmental Science & Technology* 36: 1751–1756.

Simon, E., M. van Velzen, S.H. Brandsma, E. Lie, K. Løken, J. de Boer, J. Bytingsvik, B.M. Jenssen, J. Aars, T. Hamers, and M.H. Lamoree. 2013. Effect-directed analysis to explore the polar bear exposome: identification of thyroid hormone disrupting compounds in plasma. *Environmental Science & Technology* 47 (15): 8902–8912.

Taylor, K., J.M. Braun, C. Dilworth, C. Gennings, R. Hauser, J.J. Heindel, B.R. Joubert, C.V. Rider, T.F. Webster, and D.J. Carlin. 2016. Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology. *Environmental Health Perspectives* 124 (12): A227–A229.

USEPA. 2008. Framework for application of the toxicity equivalence methodology for polychlorinated dioxins, furans, and biphenyls in ecological risk assessment. EPA 100/R-08/004. 2008.

Watt, J., T.F. Webster, and J.J. Schlezinger. 2016. Generalized concentration addition modeling predicts mixture effects of environmental PPARγ agonists. *Toxicological Sciences* 153 (1): 18–27.

Webster, T.F. 2013. Mixtures of endocrine disruptors: How similar must mechanisms be for concentration addition to apply? *Toxicology* 313: 129–133.

Webster, T.F., and V.M. Vieira. 2015. Analyzing epidemiologic mixtures data by smoothing in exposure space. Abstract submitted to statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology studies. http://www.niehs.nih.gov/about/visiting/events/pastmtg/2015/statistical/.

Webster, T., V. Vieira, and J. Weinberg. 2006. Aschengrau A. Method for mapping population-based case-control studies using generalized additive models. *International Journal of Health Geographics* 5: 26.

Weiss, J.M., P.H. Morgan, M.W. Lutz, and T.P. Kenakin. 1996. The cubic ternary complex receptor-occupancy model. I. Model Description. *Journal of Theoretical Biology* 178: 151–167.

Chapter 11
Comparing Predicted Additivity Models to Observed Mixture Data

Chris Gennings

Abstract Dose-response relationships are generally assumed to be nonlinear. Standard multiple regression models may approximate the relationship in a narrow dose range but may not adequately approximate the relationship over a wider dose range – which may have a sigmoidal shape. Further, when the number of components in a mixture is large, the required experimental design to test for interactions becomes infeasible using factorial designs. In contrast, tests for departure from additivity may be based on comparing additivity-predicted models to those of mixtures data along fixed-ratio rays of the components. As such, tests for departure from additivity in mixtures should accommodate both nonlinear relationships and efficient experimental designs. In this chapter, we illustrate the strategy using three different basic assumptions about the underlying response surface from single chemical data.

Keywords Dose addition · Nonlinear models · Additivity · Hypothesis testing

11.1 Introduction

Previous chapters describe different types of additivity models (e.g., dose additivity, independent action) and examples of their application in toxicology (Chap. 9) and epidemiology (Chap. 10), while a later chapter deals with modeling additivity in risk assessment (Chap. 14). This chapter focuses on statistical considerations in applying concepts of dose additivity. The framework for testing hypotheses of additivity for chemical mixtures has transitioned beyond traditional multiple regression models with cross-product terms for interaction. Instead, mixture data are compared to an additivity model with statistically rigorous hypothesis tests. The motivating feature is that additivity models can be estimated with design support from only single chemical dose-response data instead of the impractical design required for estimation of a full response surface. For example, the design required to build an additivity

C. Gennings (✉)
Icahn School of Medicine at Mount Sinai, New York, NY, USA
e-mail: chris.gennings@mssm.edu

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_11

model for a mixture of J components is J dose-response curves. Assuming a control group and 4 dose groups per chemical to allow for nonlinearity, the design would include 5 J design points. With 10 components, there would be 50 design points. In contrast, consider a factorial design to support a response surface with only two levels for each component (assuming linear relationships which cannot be tested with such a design); the full factorial design includes 2^J design points. With 10 components, this would require an impractical 1024 design points. Fractional factorial designs reduce the number of design points but at the cost of assumptions about interactions that cannot be tested. As the number of components increases, the strategy of building additivity models that accommodate nonlinear dose-response shapes is appealing. Tests for interaction are based on comparing experimental mixture data or models of experimental mixture data to predictions constructed from single chemical data under an assumption of additivity.

This chapter illuminates this strategy using three different basic assumptions about the underlying response surface from single chemical data. When the dose-response curves for J components in a mixture of interest have a common maximum effect, the additivity model may be readily parameterized assuming the same parametric form for each component (e.g., Casey et al. 2004). In contrast, Rajapakse et al. (2004) consider parametric models from a set of possible models and select the "best model fit" for each component separately. The additivity surface is defined based on the dose addition definition using the selected models. This approach allows for different maximum effects per chemical. The limitation is that the prediction under additivity is constrained to be no greater than the maximum effect of the component with the lowest maximum. Finally, this limitation is not relevant when the dose-response curves are adequately represented by Hill functions with Hill parameter (slope factor) of 1 (e.g., Howard et al. 2010). Details of these strategies are described in the next section. The approaches are illustrated with mixture and corresponding single chemical data for six chemicals in an estrogen receptor-alpha reporter gene assay.

11.2 Definition of Additivity

Dose addition (used herein synonymously with concentration addition) is a widely used concept that assumes the expected combination effect of a mixture of chemicals is such that the components exert their effect without influencing each other's action (Berenbaum 1985; Casey et al. 2004; Gennings et al. 2004; Scholze et al. 2014; U.S EPA 2000). Therefore, the expected combination effect may be predicted from single chemical dose-response data. In particular, consider a mixture of J chemicals where single chemical dose-response data are available on each. Dose addition satisfies the assumption of planar contours of constant response: i.e.,

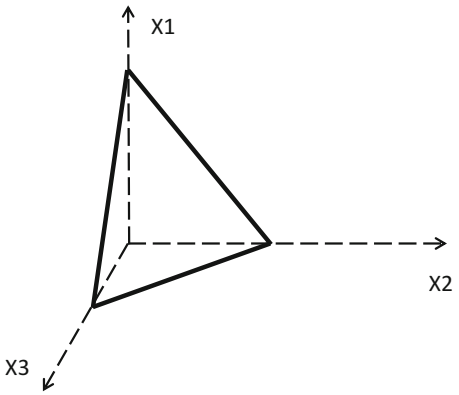

Fig. 11.1 Schematic of a planar contour of constant response on an additivity surface

$$\sum_{j=1}^{J} \frac{d_j}{\mathrm{ED}_{j,y}} = 1 \tag{11.1}$$

where d_j is the dose of the j^{th} component in the mixture and $\mathrm{ED}_{j,y}$ is the dose of the j^{th} component alone that produces the same response level (y) as the mixture. Eq. 11.1 refers to a J-dimensional plane – i.e., planar contours of constant response. A schematic of a three-dimensional planar contour is given in Fig. 11.1 for chemicals X1, X2, and X3. An important consideration is the metameter (magnitude of the observed phenomenon) of the definition. Typically, the assumption is that additivity applies on the dose scale and not on the log-dose scale; or said another way, if the assumption holds on the dose scale, it generally will not be true on another scale (e.g., log-dose scale). Another definition of additivity is independent action which is based on statistical definitions of independence (e.g., Bliss 1939; Greco et al. 1992) but is not considered herein (see Chap. 9 for a discussion of independent action).

11.3 Building Additivity Models

Several strategies have been used to build additivity models using single chemical data (see a conceptual discussion of additivity in Chap. 9). The most general is to use a "best fit" model to dose-response data from each chemical (Scholze et al. 2001). Various nonlinear regression models (e.g., logistic function, Hill function, Gompertz, exponential models; parameterizations of some are provided in Sect. 11.6) for monotonic sigmoidal (or partial sigmoidal) relationships are fit independently to the same data set, and the best fitting model is selected on the basis of a statistical goodness-of-fit test. The corresponding dose additivity model given by Eq. 11.1 is estimated for specified mixtures by defining the inverse functions for each dose-response curve (i.e., the ED values in Eq. 11.1). A limitation of this flexible approach is that the dose addition concept cannot be applied to effect levels

that exceed the maximal effect of the least efficacious compound present in the mixture (Scholze et al. 2014). Scholze et al. (2014) extend the approach by using a novel toxic unit extrapolation method, which does not have this limitation.

Less generally, the same nonlinear dose-response function (e.g., Hill function with Hill parameter equal to 1; Howard et al. 2010) is assumed for each single chemical while allowing for different maximal effects for each chemical. Under the simplifying assumption of a common slope estimate, the combined effect of any combination of full and partial agonists can be calculated using the definition of additivity in Eq. 11.1. When the assumptions of a common background effect and maximum effect for all active chemicals are appropriate, simplifying forms of predicting under additivity are available (Casey et al. 2004).

To illustrate, without loss of generality, consider the combination of two chemicals A and B. Under the best fit model strategy, assume the dose-response relationship for chemical A is best fit with a four parameter nonlinear logistic model: $\mu_0 = \alpha_A + \frac{(\gamma_A - \alpha_A)}{1+\left(\frac{d_A}{C_A}\right)^{\beta_A}}$ with inverse function $\mathrm{ED}_A = \exp\left[\frac{\log\left\{\frac{\gamma_A-\mu_0}{\mu_0-\alpha_A}\right\}+\beta_A \log C_A}{\beta_A}\right]$, where C is the inflection point (i.e., the ED_{50}) and β is the Hill slope at C. Assume the dose-response curve for chemical B is best fit with a four parameter Gompertz model: $\mu_0 = \alpha_B + (\gamma_B - \alpha_B) \exp[\exp - (\beta_{0,B} + \beta_{1,B} d_B)]$ with inverse function $\mathrm{ED}_B = \frac{-\log\left[\log\left\{\frac{\mu_0-\alpha_B}{\gamma_B-\alpha_B}\right\}\right]-\beta_{0,B}}{\beta_{1,B}}$. Under the assumption of dose addition, the combination $[d_A, d_B]$ associated with response μ_0 (i.e., the isobole) satisfies the equation $\frac{d_A}{\mathrm{ED}_A} + \frac{d_B}{\mathrm{ED}_B} = 1$, which is the equation of a line with intercept ED_A and slope $-\frac{\mathrm{ED}_A}{\mathrm{ED}_B}$. Scholze et al. (2001) estimate the combination dose with the confidence interval constructed by bootstrap sampling of the original data and re-estimation of the combination dose associated with the mean of observed dose groups. This general strategy is readily generalized to J chemicals in combination.

In comparison, Howard et al. (2010) assumed all concentration-response curves were Hill functions with Hill parameter equal to 1, i.e., $\mu_j = \frac{\alpha_j d_j}{\kappa_j + d_j}$, $j = A,B$, where κ_j is the macroscopic dissociation equilibrium constant (i.e., equivalent to the effective concentration causing 50% of maximal response) and α_j is the maximal effect level of the j^{th} ligand in the tissue or system under study. In this case, substituting the Hill function into Eq. 11.1, under dose addition, the combination of $[d_A, d_B]$ associated with response μ_0 is given by $\mu_0 = \frac{\alpha_A d_A/\kappa_A + \alpha_B d_B/\kappa_B}{1 + d_A/\kappa_A + d_B/\kappa_B}$. Howard et al. (2010) used a nonparametric Mann-Whitney test to assess the fit of the modeled response surface to experimental mixtures data. This general strategy is readily generalized to J chemicals in combination.

Finally, following the approach of Casey et al. (2004), without loss of generality that other nonlinear functions may be used, assume the dose-response curves for both chemicals are adequately represented by the nonlinear Gompertz function with common maximum effect parameter and intercept, i.e., $\mu_{A,B} = \alpha + (\gamma - \alpha) \exp[\exp - (\beta_0 + \beta_A d_A + \beta_B d_B)]$. This function is algebraically

manipulated into the form of Eq. 11.1 for contour specified by μ_0 : i.e., $\frac{d_A}{\mathrm{ED}_A} + \frac{d_B}{\mathrm{ED}_B} = 1$, where $\mathrm{ED}_j = \frac{-\log\left(\log\left\{\frac{\mu_0-\alpha}{\gamma-\alpha}\right\}\right)-\beta_0}{\beta_j}$ and $j = A,B$. Thus, the additivity model has linear contours of constant response (i.e., isoboles). Goodness-of-fit of the additivity model to the single chemical data may be assessed graphically by overlaying observed and additivity model predicted dose-response estimates on the same graph. Again, this general strategy is readily generalized to J chemicals in combination.

11.4 Hypothesis Tests Comparing Mixture Data to Predicted Model Under Additivity

Goodness-of-fit tests are used by Howard et al. (2010) to assess the fit of the modeled additivity response surface to the experimental data. For example, the Mann-Whitney test (i.e., Wilcoxon rank sum test) tests the hypothesis that the experimental data and modeled data come from the same distribution. A significant p value (e.g., $p < 0.05$) indicates that the distributions differ.

In contrast, the strategy described by Scholze et al. (2014) for predicting additivity for a mixture with fixed mixing proportions by inverting Eq. 11.1 includes statistical uncertainty by applying bootstrap samples with repeated estimation of additivity – and the total dose associated with a fixed mean response. Differences between predicted and observed effect doses are considered statistically significant when the 95% confidence belt of the prediction (from the bootstrap sampling of single chemical data) does not overlap with those of the experimentally observed mixture effects.

The assumptions of the additivity model made by Casey et al. (2004) permit a statistical test of the hypothesis of additivity, which may be a Wald-type test, likelihood ratio test, or a score test. Specifically, mixture data are assumed to be available on one or more fixed-ratio ray(s). To set notation, the mixing proportion of the j^{th} chemical in a mixture of J chemicals is a_j and $\sum_{j=1}^{J} a_j = 1$. Thus, the dose of the j^{th} chemical at total dose T is $d_j = a_j T$. From the additivity model, using a Gompertz nonlinear model for two chemicals, the predicted dose-response curve for the mixture is given by

$$\begin{aligned}
\mu_{A,B} &= \alpha + (\gamma - \alpha)\exp[\exp - (\beta_0 + \beta_A d_A + \beta_B d_B)] \\
&= \alpha + (\gamma - \alpha)\exp[\exp - (\beta_0 + \beta_A a_A T + \beta_B a_B T)] \\
&= \alpha + (\gamma - \alpha)\exp[\exp - (\beta_0 + (\beta_A a_A + \beta_B a_B) T)] \\
&= \alpha + (\gamma - \alpha)\exp[\exp - (\beta_0 + \theta_{\mathrm{add}} T)]
\end{aligned}$$

Following the approach of Casey et al. (2004), the mixture data with fixed mixing proportions are fit to a similarly parameterized model: e.g.,

$\mu_{\text{mix}} = \alpha_{\text{mix}} + (\gamma_{\text{mix}} - \alpha_{\text{mix}}) \exp[\exp - (\beta_{0,\text{ mix}} + \theta_{\text{mix}} T)]$. Then the test of additivity for the specified mixture is a test of coincidence with null and alternative hypotheses:

$$H_0 : \alpha = \alpha_{\text{mix}} \text{ and } \gamma = \gamma_{\text{mix}} \text{and } \beta_0 = \beta_{0,\text{mix}} \text{ and } \theta_{\text{add}} = \theta_{\text{mix}}$$
$$vs$$
$$H_1 : \text{any inequality}$$

An F test with 4 n–p degrees of freedom based on a Wald-type statistic can be used to test this hypothesis.

11.5 Sample Size and Power Considerations

Testing hypotheses of additivity which reject with evidence of departure from additivity should be based on study designs with adequate sample size to provide high power for detecting interaction (see Chap. 12 for further discussion of sample size and power considerations in experimental design of mixtures experiments). That is, not detecting interaction may not indicate additivity when the study design is poor (e.g., with small sample size). The location of dose/concentration groups in a study design and sample size at each group both impact the variance of slope parameters in a regression model and thereby the power for rejecting the null hypothesis of additivity. Strategies for addressing sample size and power have been described for comparison of mixture points to an additivity model (Meadows-Shropshire et al. 2005) and when comparing a model for a mixture with fixed mixing proportions compared under additivity (Casey et al. 2006).

11.6 Illustration

Data were extracted from graphs in Fig. 3 from Gennings et al. (2004) to provide an illustration of the methods described in this chapter and are provided in the appendix. The SAS code for the analysis presented herein is also in the appendix.

In short, six chemicals were selected for study by Gennings et al. (2004), including methoxychlor (MXC), o,p-DDT, beta-hexachlorocyclohexane (b-HCH), bisphenol A (BPA), octylphenol (OCT), and 2,3-bis(4-hydroxyphenyl)-propionitrile (DPN). An estrogen receptor-alpha (ER-α) gene transcription assay with MCF-7 human breast cancer cells was used to evaluate estrogenic activity. The data were assessed as units of luciferase activity normalized to the β-gal activity from individual wells. Experiments were evaluated with fold induction as the primary endpoint. A mixture was constituted with mixing proportions based on the no-observable-effect concentrations (NOECs) from preliminary concentration range-finding studies (data not shown). The resulting proportions were as follows: MXC = 0.4715;

DPN = 0.0047; DDT = 0.4715; b-HCH = 0.0471; OCT = 0.0047; and BPA = 0.0005. Details are provided in Gennings et al. (2004).

Three models were considered for illustration of the methods and parameterized as follows:

- Hill function with slope parameter 1 with background response of 1 for 100% fold induction: $\mu = 1 + \frac{\gamma x}{\mathrm{ED}_{50}+d}$.
- Four-parameter logistic model: $\mu = \alpha + \frac{(\gamma-\alpha)}{1+\left(\frac{d}{C}\right)^{\beta}}$ where α is the minimum asymptote, γ is the maximum asymptote, C is the inflection point (i.e., the ED_{50}), and β is the Hill slope at C. This model assumes symmetry around the inflection point.
- Four-parameter Gompertz model: $\mu = \alpha + (\gamma - \alpha) \exp[\exp - (\beta_0 + \beta d)]$, where α and γ are as above, β_0 is a parameter associated with the lower plateau, and β is the slope parameter.

The single chemical data with model predicted curves from the Hill function, nonlinear Gompertz, and nonlinear logistic models are presented in Fig. 11.2. There is clear evidence of varying maximum effect levels across the single chemicals. Thus the strategy of Casey et al. (2004), which assumes a common maximum effect, is not justified.

Following Howard et al. (2010), the Hill function with slope 1 was fit to each single chemical with the correction that the background response was set to 1 (i.e., the mean fold induction is 1 in the control groups): i.e., $\mu_j = 1 + \frac{\alpha_j d_j}{\kappa_j + d_j}$ and $j = 1,\ldots,6$. Under dose addition, the combination of $[d_1, d_2, d_3, d_4, d_5, d_6]$ associated with response μ_0 is given by $\mu_0 = 1 + \dfrac{\sum_{j=1}^{6} \alpha_j d_j/\kappa_j}{1 + \sum_{j=1}^{6} d_j/\kappa_j}$. A sign rank test was used to test the fit of the modeled additivity response surface to experimental mixture data with significant evidence of lack of fit ($p < 0.001$). The mixture data, predicted model, and additivity-predicted model are presented in Fig. 11.3. There is evidence that the mixture response is less than expected under additivity.

Following Scholze et al. (2001), the best fit models were used to estimate the additivity model for the mixture with fixed mixing proportions (Fig. 11.4) using minimum sum of squares error (SSE) as the model selection criterion. The selected models were the nonlinear Gompertz for β-HCH and OCT; the nonlinear logistic for MXC, DPN, BPA, and the mixture; and the Hill function with Hill slope and background parameters of 1 for DDT. The predicted curve under additivity for the specified mixture is restricted to the response region of the chemical component with the smallest maximum effect; here, β-HCH has maximum effect at 3.3 (Fig. 11.2); i.e., $t_{\mathrm{add}} = \left(\sum_{j=1}^{6} \dfrac{a_j}{\mathrm{ED}(\mu_0)_j}\right)^{-1}$ which is defined for μ_0 between 1.0 and 3.3. Scholze et al. (2014) have developed a "toxic unit extrapolation approach" to address this

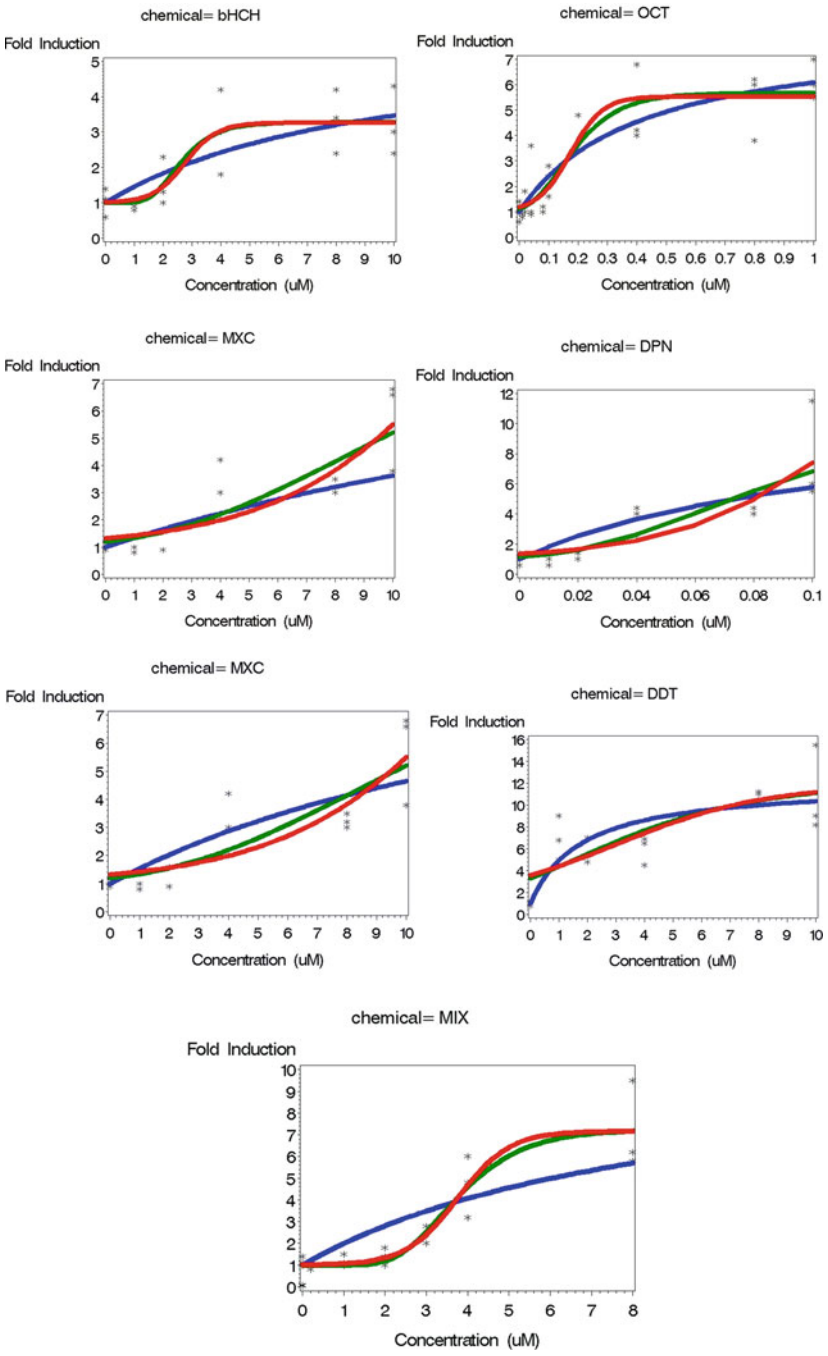

Fig. 11.2 Comparison of predicted models for single chemical data: Hill function with Hill parameter 1 (blue); nonlinear Gompertz (green); nonlinear logistic (red)

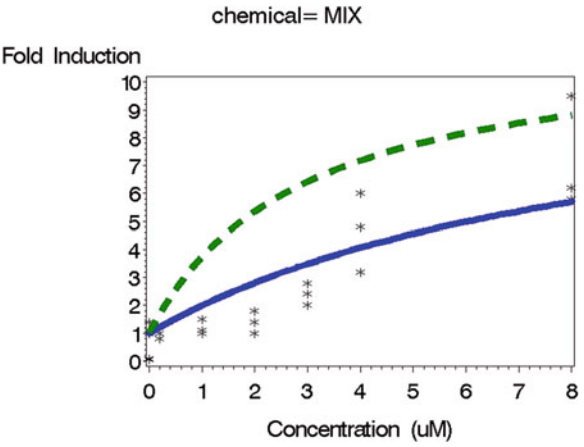

Fig. 11.3 Observed and model predicted (blue solid) mixture data (with mixing proportion as specified in the text) using a Hill function assuming a slope of 1. The additivity model (green dashed) as predicted from single chemical data is also included, based on the Howard et al. (2010) strategy

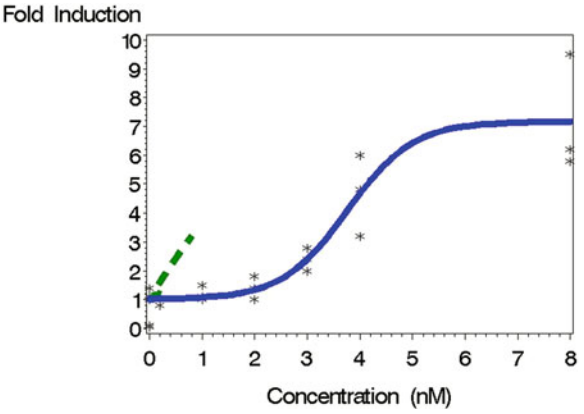

Fig. 11.4 Observed and model predicted (blue solid) mixture data (with mixing proportion as specified in the text) using best fit model (here, four parameter logistic model). The additivity model (green dashed) as predicted from single chemical data with best fit models is also included

limitation of dose addition with combinations of chemicals with differing saturating effects; however, it is beyond the scope of this chapter.

For these data, prediction of additivity was not possible for 4 of the mixture dose groups. In contrast to using a comparison of confidence bands from bootstrap samples suggested by Scholze et al. (2001), a likelihood ratio test can be conducted to test the hypothesis of additivity. The unrestricted (or full) model, based on the best fit model selection, is parameterized with each single chemical and mixture ray fit separately. Since the control group response of fold induction was set to a mean of 1.0, the models were parameterized as follows: $\mu = 1 + (\gamma_j - 1)\exp\left[\exp - (\beta_{0j} + \beta_j d)\right]$ for β-HCH ($j = 1$) and OCT ($j = 2$); $\mu = 1 + \frac{(\gamma_j - 1)}{1+\exp\left(-\left(\beta_{0j}+\beta_{1j}d\right)\right)}$ for MXC ($j = 3$), DPN ($j = 4$), BPA ($j = 5$), and the mixture ($j = 6$); and $\mu_j = 1 + \frac{\alpha_j d_j}{\kappa_j + d_j}$ for DDT ($j = 7$).

Without clear evidence of a plateau, the maximum effect parameters for MXC, DPN, and BPA were set to 10, a value somewhat beyond the observed data. Thus,

the total number of parameters estimated in the full model was 18, including an estimate for σ^2, with SSE(full) = 210.3 with $N = 152$. In comparison, the restricted model (under additivity) included only 15 parameters – those associated with the single chemical models omitting the mixture model. The prediction for the mixture is based on the dose addition model where the estimation of the restricted model is conducted with the full data set not just the single chemical data. The SSE (restricted) = 372.07. A likelihood ratio statistic is constructed as follows:

$$\begin{aligned} F^* &= \frac{(\text{SSE(rest)} - \text{SSE(full)})/\Delta df}{\text{SSE(full)}/(N - \text{df(full)})} \\ &= \frac{(372.07 - 210.31)/3}{210.31/(152 - 18)} \\ &= 34.4 \end{aligned}$$

Compared to an F distribution with (3, 134) degrees of freedom, $p < 0.001$, the hypothesis of additivity is rejected. Thus, there is evidence of departure from additivity, and the observed data are less than that predicted from additivity.

11.7 Summary

Generally, additivity models, supported from single chemical dose-response data, are statistically compared to mixture dose-response data (or models) to test the hypothesis of additivity. That is, single chemical dose-response data are used to estimate an additivity response surface (that satisfies the definition of additivity in Eq. 11.1) for any mixture of the components used in the estimation – assuming the same experimental conditions. Testing for evidence of departure from additivity using mixture dose-response data (i.e., a test of the goodness-of-fit of the additivity model) may follow standard statistical testing methods including Wald-type tests, likelihood ratio tests, and score tests. Wald-type tests may be based on comparison of model-based parameters (e.g., Casey et al. 2004) or predictions with bootstrap confidence bands (e.g., Scholze et al. 2014). In essence these tests are based on comparisons of prediction of mean responses for experimentally observed mixture data to that predicted from an additivity model. In contrast, likelihood ratio tests compare full and restricted likelihoods. Likelihood functions are joint probability distributions, which are evaluated at all data points (single chemical and mixture data) under the null hypothesis of additivity using only the parameters from the single chemical data. This restricted likelihood is compared to full likelihood using additional parameters to estimate the mixture mean response(s) (e.g., Gennings et al. 2004). The implementation of the likelihood ratio test simply requires the estimation of the full model (models for single chemical and mixture data) and the restricted model (including estimation of the model for the mixture data under additivity) with the likelihood calculated in each case and compared: i.e., −2(restricted loglikelihood

– full loglikelihood). Finally, score tests – not illustrated herein – have the advantage of being estimated only under the null hypothesis (here, of additivity) but are not generally available in many software packages.

Appendix

Extracted data from Fig. 3 in Gennings et al. (2004)

Chemical	CONC	FoldIND	Chemical	CONC	FoldIND	Chemical	CONC	FoldIND
MXC	0	0.9	b-HCH	0	0.6	BPA	0	0.6
MXC	0	1	b-HCH	0	1.1	BPA	0	1
MXC	0	1.2	b-HCH	0	1.4	BPA	0	1.4
MXC	1	0.8	b-HCH	1	0.8	BPA	0.008	1.4
MXC	1	1	b-HCH	1	0.9	BPA	0.008	1
MXC	2	0.9	b-HCH	1	1	BPA	0.008	0.08
MXC	2	1.6	b-HCH	2	1	BPA	0.01	3.2
MXC	2	1.6	b-HCH	2	1.3	BPA	0.01	2.4
MXC	4	2	b-HCH	2	2.3	BPA	0.01	2.2
MXC	4	3	b-HCH	4	1.8	BPA	0.02	3
MXC	4	4.2	b-HCH	4	3	BPA	0.02	1.4
MXC	8	3	b-HCH	4	4.2	BPA	0.02	1.4
MXC	8	3.2	b-HCH	8	2.4	BPA	0.04	1.8
MXC	8	3.5	b-HCH	8	3.4	BPA	0.04	1.2
MXC	10	3.8	b-HCH	8	4.2	BPA	0.04	1
MXC	10	6.6	b-HCH	10	2.4	BPA	0.08	1.1
MXC	10	6.8	b-HCH	10	3	BPA	0.08	1.1
DPN	0	0.6	b-HCH	10	4.3	BPA	0.08	1
DPN	0	1	OCT	0	0.6	BPA	0.1	2.5
DPN	0	1.4	OCT	0	1	BPA	0.1	1.5
DPN	0.01	0.6	OCT	0	1.4	BPA	0.1	1.5
DPN	0.01	1	OCT	0.01	0.8	BPA	0.5	2
DPN	0.01	1	OCT	0.01	0.9	BPA	0.5	2.1
DPN	0.02	1	OCT	0.01	1.2	BPA	0.5	3.2
DPN	0.02	1.4	OCT	0.02	1	BPA	1	4.4
DPN	0.02	1.4	OCT	0.02	1.2	BPA	1	8
DPN	0.04	2.6	OCT	0.02	1.8	BPA	1	10
DPN	0.04	4	OCT	0.04	0.9	MIX	0	0.06
DPN	0.04	4.4	OCT	0.04	1	MIX	0	0.09
DPN	0.08	4	OCT	0.04	3.6	MIX	0	0.09
DPN	0.08	4.4	OCT	0.08	1	MIX	0	1
DPN	0.08	4.4	OCT	0.08	1.2	MIX	0	1.1
DPN	0.1	5.5	OCT	0.08	1.2	MIX	0	1.4
DPN	0.1	6	OCT	0.1	1.6	MIX	0.2	0.8
DPN	0.1	11.5	OCT	0.1	2.4	MIX	0.2	0.8
DDT	0	0.8	OCT	0.1	2.8	MIX	0.2	1

(continued)

Chemical	CONC	FoldIND	Chemical	CONC	FoldIND	Chemical	CONC	FoldIND
DDT	0	1	OCT	0.2	3.4	MIX	1	1
DDT	0	1.2	OCT	0.2	3.4	MIX	1	1.1
DDT	1	5	OCT	0.2	4.8	MIX	1	1.5
DDT	1	6.8	OCT	0.4	4	MIX	2	1
DDT	1	9	OCT	0.4	4.2	MIX	2	1.4
DDT	2	4.8	OCT	0.4	6.8	MIX	2	1.8
DDT	2	7	OCT	0.8	3.8	MIX	3	2
DDT	2	7	OCT	0.8	6	MIX	3	2.4
DDT	4	4.5	OCT	0.8	6.2	MIX	3	2.8
DDT	4	6.5	OCT	1	5.5	MIX	4	3.2
DDT	4	6.9	OCT	1	6	MIX	4	4.8
DDT	8	11	OCT	1	7	MIX	4	6
DDT	8	11.2				MIX	8	5.8
DDT	8	11.2				MIX	8	6.2
DDT	10	8.2				MIX	8	9.5
DDT	10	9						
DDT	10	15.5						

SAS Code for Example Data

```
*****************************;
*** Gompertz function;
proc nlin data=two;
  parms g=4 b0=-.6  b1=.2;  ** for bhch, OCT, DDT, MIX;
*    parms b0=-.6  b1=.2;* g=10; ** for DPN, BPA, MXC;
    a=1;
   mu = a + (g-a)*exp(-exp(-(b0+b1*conc )));
   model foldind=mu;
   output out=predgomp p=predg;
   title 'Gompertz';
   run;

symbol1 v=star i=none;
symbol2 v=none i=join c=blue;
proc sort; by conc;
proc gplot data=predgomp;
   plot (foldind predg)*conc/overlay;
   run; quit;

*****************************;
*** logistic function;
proc nlin data=two;
  parms g=4 b0=-.6  b1=.2; a=1;  ** for bhch, OCT, DDT, MIX;
*    parms g=6 b0=-.6  b1=.2;
*    parms  b0=1 b1=.2;* g=10;   ** for DPN, BPA, MXC;
    a=1; * g=15;* a=-10;
*  b0 = -log((g-1)/(1-a));
   mu = a + 2*(g-a)/(1+exp(-(b0+b1*conc )));
   model foldind=mu;
   output out=predlogistic p=predL;
   title 'logistic';
   run;

title;
symbol1 v=star i=none;
symbol2 v=none i=join c=blue;
proc sort; by conc;
proc gplot data=predlogistic;
   plot (foldind predL)*conc/overlay;
   run; quit;

*****************************************************;
**** analysis of the mixture and tests of additivity;
*****************************************************;
**** Howard model;
** chemicals: 'bHCH', 'OCT', 'MXC', 'DPN', 'BPA','DDT';
*****************************************************;
proc nlmixed data=twob;
    parms a1=4.8 k1=9.5 a2=7 k2=.4  k3=17  k4=.11  k5=1.1
        a6=11 k6=1.8  km=9;
                    a3=10; a4=10; a5=10; am=10;
  mu = 1+ (chemical='bHCH')*a1*conc/(k1+conc) +
         (chemical='OCT' )*a2*conc/(k2+conc) +
         (chemical='MXC' )*a3*conc/(k3+conc) +
         (chemical='DPN' )*a4*conc/(k4+conc) +
         (chemical='BPA' )*a5*conc/(k5+conc) +
         (chemical='DDT' )*a6*conc/(k6+conc) +
         (chemical='MIX' )*am*conc/(km+conc)
  ;
```

```
 num = conc*(chemical='MIX')*
      (a1*mix_bhch/k1 + a2*mix_oct/k2 + a3*mix_mxc/k3 +
                    a4*mix_dpn/k4  + a5*mix_bpa/k5 + a6*mix_ddt/k6);
  den= 1+ conc*(chemical='MIX')*
      (mix_bhch/k1 + mix_oct/k2 + mix_mxc/k3 +
                    mix_dpn/k4  + mix_bpa/k5 + mix_ddt/k6);
  muadd = 1+  num/den;
  if chemical ne 'MIX' then muadd=.;
  id muadd;
  model foldind~normal(mu,sigsq);
  predict mu out=pred ;
 run;

************************************************************;
**** FSCR model Gennings and best fit model;
** chemicals: 'bHCH', 'OCT', 'MXC', 'DPN', 'BPA','DDT';
************************************************************;

*** unrestricted;
proc nlmixed data=twob ;
    parms g1=3.3 b01=-3 b1=1  g2=5 b02=-.6 b2=.1
         b03=-2 b3=.1
          b04=-2 b4=.2       b05=-2 b5=.2   a6=10    k6=3
        gm=10 b0m=-.6  bm=1;
  g4=10; g5=10; g3=10;
   a=1;
  mu = (chemical='bHCH')*(a+(g1-a)*exp(-exp(-(b01+b1*conc)))) +
      (chemical='OCT' )*(a+(g2-a)*exp(-exp(-(b02+b2*conc)))) +
      (chemical='MXC' )*(a+(g4-a)/(1+exp(-(b03+b3*conc ))))+
      (chemical='DPN' )*(a+(g4-a)/(1+exp(-(b04+b4*conc ))))+
      (chemical='BPA' )*(a+(g5-a)/(1+exp(-(b05+b5*conc ))))+
      (chemical='DDT' )*(1+a6*conc/(k6+conc)) +
                    (chemical='MIX' )*(a+(gm-a)/(1+exp(-(b0m+bm*conc ))));
  ;
  model foldind~normal(mu,sigsq);
  predict mu out=pred ;

 ED1=.; ed2=.; ed3=.; ed4=.; ed5=.; ed6=.;
 if mu>1 and chemical='MIX' and mu<g1 then do;
  ED1= (-log(-log((mu-a)/(g1-a)))-b01)/b1; end;

 if mu>1 and chemical='MIX' and mu<g2 then do;
  ED2= (-log(-log((mu-a)/(g2-a)))-b02)/b2;   end;

  if mu>1 and chemical='MIX' then do;
  if mu<g3 then ED3= (-log(-log((mu-a)/(g3-a)))-b03)/b3;
  if mu<g4 then ED4= (log((mu-a)/(g4-mu))-b04)/b4;
  if mu<g5 then ED5= (log((mu-a)/(g5-mu))-b05)/b5;
  if mu<a6 then ED6= k6*(mu-1)/(a6-mu+1);
  end;
  tadd = 1/(mix_bhch/ed1 + mix_oct/ed2 + mix_mxc/ed3 +
         mix_dpn/ed4 + mix_bpa/ed5 + mix_ddt/ed6);
  id ed1 ed2 ed3 ed4 ed5 ed6 tadd mu;

 run;

** restricted under additivity - no mixture parameters;
proc nlmixed data=twob ;
```

```
*   where foldind ne .;
    parms g1=3.3 b01=-3 b1=1.2  g2=6 b02=-1 b2=9   b03=-4 b3=.3
          b04=-4 b4=34      b05=-4 b5=3.3      a6=11 k6=1.8;
 g5=10; g4=10; g3=10;
  a=1;
 muadd = (chemical='bHCH')*(a+(g1-a)*exp(-exp(-(b01+b1*conc)))) +
     (chemical='OCT' )*(a+(g2-a)*exp(-exp(-(b02+b2*conc)))) +
     (chemical='MXC' )*(a+(g3-a)/(1+exp(-(b03+b3*conc ))))+
     (chemical='DPN' )*(a+(g4-a)/(1+exp(-(b04+b4*conc ))))+
     (chemical='BPA' )*(a+(g5-a)/(1+exp(-(b05+b5*conc ))))+
     (chemical='DDT' )*(1+a6*conc/(k6+conc))
 ;

do mu0 = 0 to 8 by .1;
 if mu0>1 and chemical='MIX' and mu0<g1 then do;
  ED1= (-log(-log((mu0-a)/(g1-a)))-b01)/b1; end;
 if mu0>1 and chemical='MIX' and mu0<g2 then do;
  ED2= (-log(-log((mu0-a)/(g2-a)))-b02)/b2;   end;

  if mu0>1 and chemical='MIX' then do;
  ED3= (-log(-log((mu0-a)/(g3-a)))-b03)/b3;
  ED4= (log((mu0-a)/(g4-mu0))-b04)/b4;
  ED5= (log((mu0-a)/(g5-mu0))-b05)/b5;
  ED6= k6*(mu0-1)/(a6-mu0+1);
  end;
  tadd_ = 1/(mix_bhch/ed1 + mix_oct/ed2 + mix_mxc/ed3 +
          mix_dpn/ed4 + mix_bpa/ed5 + mix_ddt/ed6);
 if chemical='MIX' then do;
    if (tadd_-conc)**2<0.01 then do;
        muadd=mu0;
                    tadd=tadd_;
                    end;
          end; end;
 if chemical='MIX' and (tadd-conc)**2>0.01 then muadd=.;

 model foldind~normal(muadd,sigsq);
 predict muadd out=predadd ;
 id ed1 ed2 ed3 ed4 ed5 ed6 tadd muadd;

run;
```

References

Berenbaum, M.C. 1985. The expected effect of a combination of agents: The general solution. *Journal of Theoretical Biology* 114: 413–431.

Bliss, C.I. 1939. The toxicity of poisons applied jointly. *Annals of Applied Biology* 26: 585–615.

Casey, M., C. Gennings, W.H. Carter Jr, V. Moser, and J.E. Simmons. 2004. Detecting interaction(s) and assessing the impact of components subsets in a chemical mixtures using fixed-ratio mixture ray designs. *Journal of Agricultural, Biological, and Environmental Statistics* 9 (3): 339–361.

Casey, M., C. Gennings, and W.H. Carter Jr. 2006. Power and sample size calculations for linear hypotheses associated with mixtures of many components using fixed-ratio ray designs. *Environmental and Ecological Statistics* 13 (1): 11–23.

Gennings, C., W.H. Carter Jr, E.W. Carney, D.C. Grantley, B.B. Gollapudi, and R.A. Carchman. 2004. A novel flexible approach for evaluating fixed ratio mixtures of full and partial agonists. *Toxicological Sciences* 80: 134–150.

Greco, W., H.D. Unkelbach, G. Poch, J. Suhnel, M. Kundi, and W. Moedeker. 1992. Consensus on concepts and terminology for combined action assessment: The Saariselka agreement. *Archives of Complex. Environmental Studies* 4: 65–69.

Howard, G.J., J.J. Schlezinger, M.E. Hahn, and T.F. Webster. 2010. Generalized concentration addition predicts joint effects of aryl hydrocarbon receptor agonists with partial agonists and competitive antagonists. *Environmental Health Perspectives* 118: 666–672.

Meadows-Shropshire, S.L., C. Gennings, and W.H. Carter Jr. 2005. Sample size and power determination for detecting interactions in mixtures of chemicals. *Journal of Agricultural, Biological, and Environmental Statistics* 10 (1): 104–117.

Rajapske, N., E. Silva, M. Scholze, and A. Kortenkamp. 2004. Deviation from additivity with estrogenic mixtures containing 4-nonylphenol and 4-tert-octyphenol detected in the E-SCREEN assay. *Environmental Science and Technology* 38: 6343–6352.

Scholze, M., W. Boedeker, M. Faust, T. Backhaus, R. Altenburger, and L.H. Grimme. 2001. A general best-fit method for concentration-response curves and the estimation of low-effect concentrations. *Environmental Toxicology and Chemistry* 20: 448–457.

Scholze, M., E. Silva, and A. Kortenkamp. 2014. Extending the applicability of the dose addition model to the assessment of chemical mixtures of partial agonists by using a novel toxic unit extrapolation method. *PLoS One* 9 (2): e88808.

U.S. EPA (Environmental Protection Agency). 2000. *Supplementary guidance for conducting health risk assessment of chemical mixtures*, 209. (EPA/630/R-00/002). Washington, DC: Risk Assessment Forum. http://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=4486

Chapter 12
Physiologically Based Pharmacokinetic Modeling of Chemical Mixtures

Sami Haddad

Abstract Physiologically based pharmacokinetic (PBPK) modeling is a tool that is increasingly being used for xenobiotics exposure assessment and target tissue dosimetry simulations in risk assessment and in pharmaceutical sciences. Because this tool can use chemical and physiological information/data from different sources (i.e., in vitro, in vivo, in silico), it is also being increasingly used for mixture exposures, especially for mixtures containing chemicals that toxicokinetically interact, at the physiological, physicochemical, and biochemical level. The aim of this chapter is to give an overview of what PBPK modeling is and how it can be used in the context of mixture toxicology. Known mechanisms of toxicokinetic interactions between xenobiotics are described, and mathematical representations are given when available. Existing modeling approaches that are available in the literature are presented for mixtures of various complexities. Current methods and their limitations are reported, and future directions are put forward.

Keywords Pharmacokinetics · Toxicokinetics · Interactions · Mixtures · Metabolism

12.1 Introduction: What Is PBPK Modeling?

Physiologically based pharmacokinetic (PBPK) models (note: toxicokinetics and pharmacokinetics are synonymous in the context of this chapter) are mathematical descriptions of pharmacokinetic processes (absorption, distribution, metabolism, and excretion) of xenobiotics that rely on appropriate physiological, biological, biochemical, anatomical, and physicochemical information. They allow for prediction or simulation of tissue dosimetry as a function of time and exposure scenario (dose and timing) (Krishnan and Andersen 2007). The level of detail of these models

S. Haddad (✉)
Department of Occupational & Environmental Health, School of Public Health, Public Health Research Institute of Université de Montréal, Université de Montréal, Montreal, QC, Canada
e-mail: sami.haddad@umontreal.ca

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_12

can vary extensively depending on desired dosimetry (e.g., target tissue or blood AUC, target tissue or blood C_{max}, total amount metabolized, C_{max} of metabolite in target tissue, etc.) and available information. Because of their mechanistic basis, these models are increasingly being used in toxicological risk assessment of chemicals. This is principally due to the fact that PBPK models have the advantage of allowing different types of extrapolations (e.g., animal to human, high to low dose, scenario to scenario, route to route, etc.) with much more confidence than other previously used models (e.g., non-compartmental or compartmental pharmacokinetic models) allow (Krishnan and Andersen 2007; Thompson et al. 2008; Espie et al. 2009).

PBPK models are basically composed of a series of mass-balance differential equations describing the flux of the chemical of interest in the organism (Fig. 12.1).

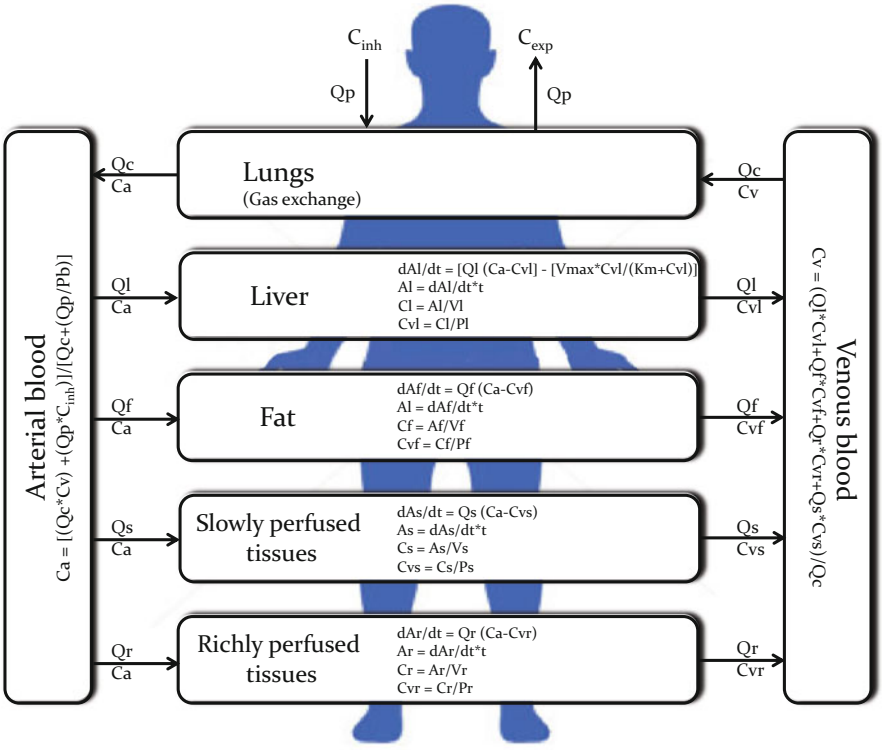

Fig. 12.1 PBPK model conceptual representation. The term d/dt refers to the derivative of the variable over time. Capital letters *A*, *C*, *Q*, *P*, and *V* refer to amount of chemical, chemical concentration, blood flow, partition coefficient, and volume. Lower case letters a, b, c, f, l, p, r, s, t, v, vf, vl, vr, and vs refer to arterial blood, blood/air, cardiac, fat, liver, pulmonary, richly perfused tissues, slowly perfused tissues, time, venous blood, venous blood leaving fat, venous blood leaving liver, venous blood leaving richly perfused tissues, and venous blood leaving slowly perfused tissues, respectively. Subscript inh and exp refer to inhaled and exhaled air. V_{max} and K_m refer to the maximal rate of metabolism and the Michaelis-Menten affinity constant, respectively

When in contact with the skin, lungs, or intestinal walls, a chemical can be absorbed and enter the blood circulation (absorption). With the arterial blood flow, the chemical can then reach the different organs of the body and accumulate, depending on its affinity with the tissue components and its capacity to cross biological membranes (endothelial wall or cellular membrane of parenchymal cells) (distribution). In some tissues, such as liver, metabolism can be an important process contributing to the elimination of the compound. This results in transforming it into another molecule (i.e., a metabolite) which can be more or less toxic than or have the same toxic potency of the parent compound (metabolism). Other processes involved in the elimination of unchanged chemicals which are often described in PBPK models are renal excretion in kidneys, exhalation in lungs, or biliary excretion in the liver (excretion).

In simple PBPK models where the chemical easily crosses tissue or cell membranes, the mass-balance differential equations describing the rate of chemical accumulation (R_acc) in a tissue compartment (subscript t) would be described by the rate of the chemical leaving the tissue with the venous blood (R_out) subtracted from the rate of the chemical entering the tissue with the arterial blood flow (R_in) as follows:

$$R\_\mathrm{acc}_\mathrm{t} = R\_\mathrm{in}_\mathrm{t} - R\_\mathrm{out}_\mathrm{t} \tag{12.1}$$

$$R\_\mathrm{in}_\mathrm{t} = Q_\mathrm{t} \times \mathrm{Ca} \tag{12.2}$$

$$R\_\mathrm{out}_\mathrm{t} = Q_\mathrm{t} \times \mathrm{Cv}_\mathrm{t} \tag{12.3}$$

where Q_t, Ca, and Cv_t are, respectively, tissue blood flow, arterial blood concentration, and venous blood concentration leaving tissue. When elimination occurs in the tissue (e.g., metabolism), an additional rate (e.g., rate of amount metabolized: R_am) must be subtracted from the R_in as follows:

$$R\_\mathrm{acc}_\mathrm{t} = (Q_\mathrm{t} \times \mathrm{Ca}) - (Q_\mathrm{t} \times \mathrm{Cv}_\mathrm{t}) - R\_\mathrm{am}_\mathrm{t} \tag{12.4}$$

The amount of chemical in the tissue can then be estimated by integrating, i.e., numerically solving, the mass-balance differential equations using algorithms available in simulation software (e.g., ACSLX, Stella, MatLab). The concentration is determined by dividing the amount of chemical in tissue by the tissue volume (Eq. 12.5), and the Cv_t is determined by the tissue/blood partitioning (P_t) (Eq. 12.6).

$$C_\mathrm{t} = \frac{A_\mathrm{t}}{V_\mathrm{t}} \tag{12.5}$$

$$\mathrm{Cv}_\mathrm{t} = \frac{C_\mathrm{t}}{P_\mathrm{t}} \tag{12.6}$$

Many toxicokinetic processes in PBPK models can be described with first order mathematical descriptions, such as R_in and R_out in Eqs. 12.2 and 12.3 above. This means that the rate of the process of interest is directly proportional to the appropriate blood or tissue concentration of the chemical (e.g., passive diffusion across membranes). In some instances, kinetic processes must be described using a

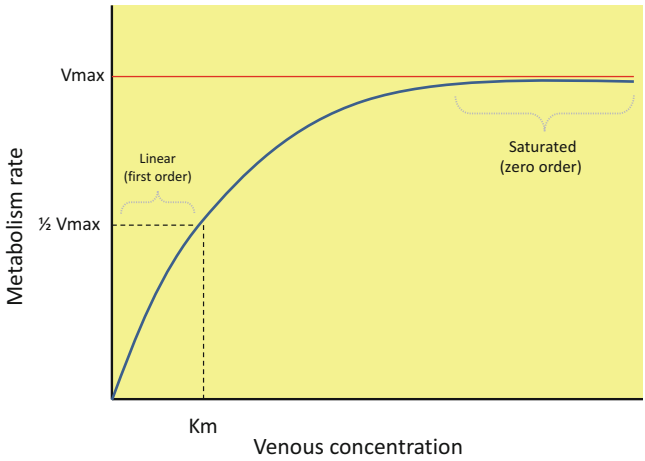

Fig. 12.2 First order process vs saturable process exemplified with metabolic rates

saturation model. This is more likely to happen at higher-dose regions. The same process may be described as a first order process when only low exposure levels are of interest but must be changed to a saturable process when simulating higher exposure levels. In such cases, at low levels of exposure, the simulated rate of the process seems to increase with dose, but at a certain dose range, it levels off to a maximal rate to attain zero order (i.e., rate becomes a constant that is invariable with dose) (Fig. 12.2). This is often the case for describing metabolic rates (R_am). For instance, at low exposure levels, the use of an intrinsic clearance constant (Cl_{int}; first order constant) is often sufficient (Eq. 12.7) for describing the metabolic rate of a chemical, whereas, at higher exposure doses, the metabolic rate becomes saturated and must be described using a maximal rate of metabolism (V_{max}) and a Michaelis-Menten affinity constant (Km) (Eq. 12.8).

$$\text{First order} : R\_\text{am} = \text{Cl}_{\text{int}} * \text{Cvl} \tag{12.7}$$

$$\text{Saturable} \quad : R\_\text{am} = \frac{\text{Vmax} * \text{Cvl}}{\text{Km} + \text{Cvl}} \tag{12.8}$$

where Cvl refers to venous blood leaving the liver.

The equations that compose the PBPK model are populated with parameters that are specific to the organism (physiological parameters), the chemical or chemicals under investigation (physicochemical parameters), and the reactions between the organism and the chemical (biochemical parameters) (Table 12.1).

When the organism is exposed to additional chemicals, the toxicokinetics may be unaffected, and therefore no further considerations in terms of PBPK modeling are necessary. But in many circumstances, co-exposure to one or more chemicals may change the relationship between external dose and internal dosimetry of the chemical of interest.

Table 12.1 PBPK model parameters

Physiological	Physicochemical	Biochemical
Cardiac output[a]	Partition coefficients	Rate constants (V_{max}, K_m) for
Alveolar ventilation rate[a]	Tissue/blood[a]	Enzymatic reactions[a]
Body weight[a]	Blood/air[a]	Active transport in:
Tissue volume[a]	Skin/air	Urinary excretion
Tissue blood flow[a]	Skin/water	Biliary excretion
Tissue blood content	Permeability coefficients	Tissue uptake
Tissue lipid and water content		Tissue efflux
Skin surface area		GI absorption
Glomerular filtration rate		Macromolecular binding constants
		B_{max}
		Kd

[a]Most frequently used PBPK model parameters

12.2 Pharmacokinetic Interactions

When the tissue or blood concentration vs time profile of one chemical (chemical *A*) is modified by co-exposure to another chemical (chemical *B*), it is a clear indication that a pharmacokinetic interaction is occurring. These interactions are the result of a pharmacokinetic mechanism being affected by the other chemical. In terms of PBPK modeling, it can either be the result of an alteration of a physiological parameter value (e.g., increased ventilation rate), a physicochemical parameter (e.g., increased lipophilicity due to complexation), or modified biochemical parameters (e.g., increased V_{max} due to enzyme induction). In the next section of the chapter, common mechanisms of pharmacokinetic interactions will be reviewed.

12.2.1 Chemicals Altering Physiology

Exposure to some xenobiotics can lead to alterations of physiological factors that are critical determinants of toxicokinetic processes. Hence, when such an event occurs, the toxicokinetics of all co-exposed chemicals are modified if the physiological parameter in question plays a role in their absorption, distribution, metabolism, or elimination. Examples of physiological parameters that are altered by exposure to chemicals are provided below.

One parameter that can be altered by the presence of chemicals is the alveolar ventilation rate (Qalv). It has been shown that salicylate poisoning and amphetamines cause increased ventilation by raising carbon dioxide. This increase in ventilation is the body's attempt to compensate for excess carbon dioxide (Crisp and Taylor 2012). Other chemicals can augment the alveolar ventilation rate by diminishing the cellular respiration, for example, carbon monoxide diminishes hemoglobin capacity in oxygen binding, hydrogen cyanide inhibits cytochrome c oxidase which plays a crucial role in the electron transport respiratory chain in the

mitochondria, DDT inhibits ATP-synthase, and dichlorovinyl cystein inhibits pyruvate dehydrogenase in the Krebs cycle (Gregus 2008). In PBPK modeling, a modification of Qalv will lead to changes in the pulmonary absorption rate (and pulmonary elimination) of volatile chemicals as it is used to determine the concentration in arterial blood as follows:

$$\mathrm{Ca} = \frac{\mathrm{Qc} \times \mathrm{Cv} + \mathrm{Qalv} \times \mathrm{Ci}}{\mathrm{Qc} + \mathrm{Qalv}/\mathrm{Pb}} \tag{12.9}$$

where Qc refers to the cardiac output, Ci to the inhaled concentration of chemical, Cv to the venous blood concentration, and Pb is the blood air partition coefficient of the chemical.

Many xenobiotics are known to affect hemodynamics (i.e., blood flow) in humans or animals. This has the result of changing Qc or Qt. Necessarily, the toxicokinetics of all co-exposed chemicals would be affected accordingly, through altered tissue distribution, altered elimination in tissue, or even pulmonary absorption. For instance, ethanol and phenobarbital increase hepatic blood flow (Ql), hence increasing clearance of all co-exposed chemicals that have a high hepatic extraction ratio (Krishnan et al. 1994). Vasodilators and vasoconstrictors will influence the distribution of co-exposed chemicals by altering tissue blood flows. For example, many drugs have been shown to alter renal hemodynamics (e.g., hypertensive agents, nonsteroidal anti-inflammatory drugs (NSAIDs), some immunosuppressants, aminoglycosides, amphotericin *B*) (Hsu and Wu 2012). This may lead to changes in renal clearance of other chemicals due to decreased glomerular filtration rates which can sometimes be irreversible.

Upon exposure to some xenobiotics, gastric emptying may be affected, and, hence, the absorption of ingested chemicals or orally administered drugs can be altered. For example, Nimmo et al. (1975) demonstrated that absorption of orally administered acetaminophen was considerably delayed when subjects were administered pethidine or diamorphine by intramuscular injection. Other drugs have been shown to delay gastric emptying and upper gastrointestinal tract motility (e.g., opioids, anticholinergics, and adrenergic receptor agonists). Several drugs are known to increase the motility of the upper gastrointestinal tract. Among them are the gastrointestinal prokinetic drugs (e.g., metoclopramide, cisapride, domperidone), which may increase rates of absorption but in some instances also decrease bioavailability because of reduced available time for total absorption (Greiff and Rowbotham 1994). Another physiological factor that can be altered to modify the rate of absorption of orally exposed chemicals is the gastric or intestinal pH (De Castro et al. 1996; Budha et al. 2012).

Skin structure or composition can be modified by exposure to a chemical. This can lead to changes in dermal absorption rates of other xenobiotics. Isobutanol has been shown to change skin composition (dehydration) and reduce the absorption of m-xylene (Riihimaki 1979). Dermal permeability of lipophilic compounds has been shown to be increased by dimethyl sulfoxide (DMSO) skin exposure (Hayes and Pearce 1953; Jacob et al. 1964; Choi et al. 1990). DMSO causes swelling of basal

cells of the stratum corneum as well as a disruption of keratin matrices in skin (Kurihara-Bergstrom et al. 1987; Qiao et al. 1996).

Although all these xenobiotic-induced physiological changes have been observed, to our knowledge, none have been described mathematically to ultimately be used in a multichemical PBPK model. Proper dose-response relationships for physiological changes would need to be characterized for them to be incorporated into a PBPK model.

12.2.2 Chemicals Altering Physicochemical Properties

There are very few examples where physicochemical properties of a chemical are modified by the presence of another xenobiotic. A documented example of this is the increased membrane permeability of lead in the presence of dithiocarbamates (Oskarsson and Lind 1985). Indeed, dithiocarbamates can form a complex with lead that is more lipophilic than lead alone, and therefore distribution to brain is increased. Organic chelators such as EDTA can also increase the lipophilicity of ionic metals and therefore alter their capacity to distribute.

Another example of alteration of physicochemical properties is co-exposure to ethanol and mercury. Ethanol is known to depress the conversion of elemental mercury to the ionic form (Kudsk 1965). Elemental mercury, being more volatile than the ionic form, is therefore more easily eliminated by exhalation.

12.2.3 Chemicals Affecting Chemical-Biological Interactions

In terms of published literature on toxicokinetic interactions, chemical-biological interactions are by far the most cited, and many examples of mathematical descriptions exist between multiple xenobiotics. This category of interactions basically results in a modification of biochemical parameters affecting metabolic rates, transport rates, or protein binding. The mechanisms that are affected therefore involve proteins implicated in a critical kinetic process of the xenobiotic of interest. The different types of interactions existing in this category can be divided into two categories: (1) mechanisms affecting the level of active proteins (concentration of enzymes, transporters, or binding proteins) and (2) mechanisms affecting the activity of a protein.

(a) Mechanism Affecting the Level of Active Proteins

The concentration of an active protein (P_a) will basically depend on its synthesis rate (R_P_{synth}), its degradation rate (R_P_{deg}), and its inactivation rate (R_P_{inact}) which is usually as follows:

$$\mathrm{RC\_}[P_a] = R\_P_{synth} - R\_P_{deg} - R\_P_{inact} \tag{12.10}$$

where RC_[P_a] refers to the rate of change in active protein concentration involved in xenobiotic metabolism, transport, or binding. The consequence of a change in [P_a] will be a proportional change in V_{max} for enzymes and active transporters as $V_{max} = K_{cat}$ [P_a] where K_{cat} is the turnover rate or of B_{max} (maximal binding capacity) for binding proteins as $B_{max} = n$ [P_a] where n refers to the number of binding sites.

Increased Protein Synthesis Many xenobiotics are known to increase (i.e., induce) the activity of proteins, which can occur through different mechanisms. One way to achieve this is through an increase in protein concentration. There are many compounds that are known to interact with and activate transcription factors (e.g., PXR, CAR, FXR, AhR, PPARα, etc.) which in turn activate the transcription and synthesis of different enzymes and other proteins (binding proteins or transporters). In terms of mathematical representation of such phenomena, the increased synthesis of CYP1A1 and 1A2 by TCDD was described by a factor representing aryl hydrocarbon receptor binding (Andersen et al. 1997; Leung et al. 1990), and Sarangapani et al. (2002) similarly described induction of CYP2B1/2 by octamethylcyclotetrasiloxane via an unknown receptor. In both cases, the R_P_{synth} is modulated as a function of the fraction of receptor occupancy (F_{RO}) typically modeled using a Hill model as follows:

$$R\_P_{synth} = R\_P_{synth_0} + \left(\left[R\_P_{synth_{max}} - R\_P_{synth_0}\right] \times F_{RO}\right) \tag{12.11}$$

$$F_{RO} = \frac{[FL]^n}{[FL]^n + Kd^n} \tag{12.12}$$

where RP_{synth_0} and $RP_{synth_{max}}$ are the basal and maximal rates of protein synthesis, Kd is the dissociation constant, FL is the free ligand concentration in cells where the receptor is present, and n is the hill coefficient which is dependent on the receptor.

Increased Protein Stability An increase in protein activity can also be achieved by stabilizing the protein, i.e., by reducing the value of R_P_{deg}. Indeed, an example of such a mechanism was described by Chien et al. (1997) where ethanol stabilizes CYP2E1, which consequently increases the overall concentration of the enzyme and therefore its activity. In this example, schematized in Fig. 12.3, the enzyme (i.e., CYP 2E1) would be found in two forms, distinguishable by rate of degradation: the form that is rapidly degraded (P_a^1) and another form which is the slowly degraded enzyme (P_a^2). The authors' hypothesis was that CYP2E1 was synthesized at a given rate into the pool of P_a^1 and it can be converted to P_a^2 according to a transfer rate constant K_{trans}. In the form P_a^1, the degradation rate of the protein is rapid when not bound to ligand and slow when bound. When P_a^1 is highly bound with ligand, the concentration of P_a increases, and therefore the turnover to the P_a^2 form increases. Equation 12.10 can therefore be modified as follows to describe the rate of change in protein concentration for the CYP2E1 in the rapid degradation form:

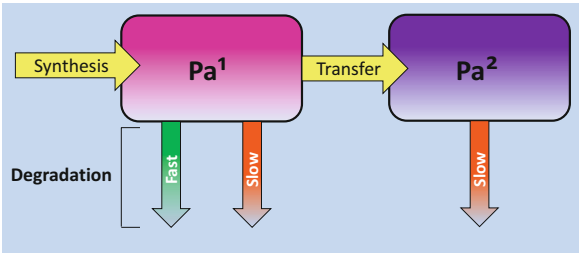

Fig. 12.3 Conceptual representation of CYP2E1 induction by ligand stabilization according to Chien et al. (1997). The CYP2E1 enzyme can occur in the rapidly degraded form ($P_a{}^1$) and the slowly degraded form ($P_a{}^2$) (Modified from Chien et al. 1997)

$$\mathrm{RC\_}\left[P_a{}^1\right] = R\_E_{\mathrm{synth}} - R\_P_{\mathrm{deg}} - R\_P_{\mathrm{trans}} \tag{12.13}$$

$$\text{where } R\_P_{\mathrm{deg}} = \left[P_a{}^1\right] \times \left(f_{\mathrm{unbound}} \times K_{\mathrm{deg}}^{\mathrm{fast}} + f_{\mathrm{bound}} \times K_{\mathrm{deg}}^{\mathrm{slow}}\right) \tag{12.14}$$

$$\text{and } R\_P_{\mathrm{trans}} = \left[P_a{}^1\right] \times K_{\mathrm{trans}} \tag{12.15}$$

The rate of change of concentration of the slowly degraded form of the enzyme is determined as follows:

$$\mathrm{RC\_}\left[P_a{}^2\right] = R\_P_{\mathrm{trans}} - R\_P_{\mathrm{deg}}{}' \tag{12.16}$$

$$\text{and } R\_P_{\mathrm{deg}}{}' = \left[P_a{}^2\right] \times K_{\mathrm{deg}}^{\mathrm{slow}} \tag{12.17}$$

where R_P_{trans} is the rate of transfer, f_{unbound} is the fraction of E_a that is not bound to ligand, and f_{bound} is the fraction of $P_a{}^1$ that is bound to the stabilizing ligand. The slow and fast degradation rate constants are $K_{\mathrm{deg}}^{\mathrm{fast}}$ and $K_{\mathrm{deg}}^{\mathrm{slow}}$, and the transfer rate constant is K_{trans}. The fractions bound and unbound can be calculated using the information of dissociation constant (Kd) for the ligand.

Protein Inactivation Proteins may be inactivated in many ways by xenobiotics. There are several published examples of enzymes being irreversibly inhibited/inactivated by xenobiotics. Generally, the inhibitor or inactivator binds irreversibly to the active site, consequently stopping all catalytic activity. As can be deduced from Eq. 12.10, an increase in R_P_{inact} will lead to a decrease in P_a levels. The same logic would apply for binding proteins or active transporters. Enzyme inactivation has been mathematically described for binary mixtures where a component inhibits the metabolic rate of the other component by this mechanism: triazolam and erythromycin (Kanamitsu et al. 2000a), 5-fluorouracil and sorovidine (Kanamitsu et al. 2000b), and trichloroethylene and its metabolite dichloroacetate (Keys et al. 2004). In the absence of the inactivator, the R_P_{inact} is nil, and levels of P_a remain stable. Upon introduction of the inactivator into the system, the R_P_{inact} becomes positive according to the following equation:

$$R\_P_{\text{inact}} = \frac{K_{\text{inact}} \times [P_a] \times f_{\text{bound}} \times \frac{[I]_t}{P_t}}{\text{Ki}_{\text{app}} + f_{\text{bound}} \times \frac{[I]_t}{P_t}} \tag{12.18}$$

where K_{inact} represents the maximum inactivation rate constant, P_t represents the tissue-to blood partition coefficient, f_{bound} is the unbound fraction in blood, and $[I]_t$ is the inactivator's concentration in tissue where P_a is located.

(b) Mechanisms Affecting the Activity of Proteins

The activity of proteins can be modified without actually changing their concentration. The most frequently published mechanism of toxicokinetic interactions is found in this category, the conventional reversible inhibitions, including competitive, noncompetitive, and uncompetitive inhibition. Other, less frequently reported types of interactions affecting activity of proteins are the allosteric interactions. Also, the depletion of cofactor reserves is another way to alter the rate of protein activity.

Competitive Inhibition When two chemicals compete for the same active site (on an enzyme or active transporter), competitive inhibition occurs. This competition may occur between two substrates for the same active site or between a substrate and another chemical that simply acts as an inhibitor. The consequence of this type of interaction is the apparent decrease in ligand affinity (i.e., increase in apparent Michaelis-Menten affinity constant: $K_{m_{app}}$) as a function of inhibitor concentration ($[I]$) and affinity (K_i) and hence a reduction in rate of activity (R_{activity}) (i.e., metabolism or transport), and V_{max} remains unchanged (Segel 1974), as follows:

$$K_{m_{app}} = K_m \times \left(1 + \frac{[I]}{K_i}\right) \tag{12.19}$$

$$R_{\text{activity}} = \frac{V_{max_{app}} \times \text{Cv}_t}{K_{m_{app}} + \text{Cv}_t} \tag{12.20}$$

There can also be competition between two or more chemicals for a binding site on a binding protein or transporter leading to binding displacement (e.g., tolbutamide and sulfonamides for plasma protein binding) (Sugita et al. 1982). The principle is the same as for metabolism or transport, and the apparent affinity of the xenobiotic for the protein is reduced (i.e., increase in apparent dissociation constant: $K_{d_{app}}$) as a function of inhibitor concentration ($[I]$) and its dissociation constant (K_{d_i}) leading to a decrease in the concentration of the chemical that is bound (C_{bound}), as follows:

$$K_{d_{app}} = K_d \times \left(1 + \frac{[I]}{K_{d_i}}\right) \tag{12.21}$$

$$C_{\text{bound}} = \frac{B_{\max} \times C_{\text{free}}}{K_{d_{\text{app}}} + C_{\text{free}}} \tag{12.22}$$

Noncompetitive Inhibition In some cases, a binding site modulating enzyme or transporter activity, which is different from the active site, may exist on the active protein. A noncompetitive inhibitor may bind this modulating binding site and affect the metabolic or transport rate of another chemical. The change in protein conformation by the inhibitor is mathematically reflected by a reduction in apparent $V_{\max}$, and K_{m} remains unaffected, as follow (Segel 1974):

$$V_{\max_{\text{app}}} = \frac{V_{\max}}{\left(1 + \frac{[I]}{K_{\text{i}}}\right)} \tag{12.23}$$

Uncompetitive Inhibition An inhibitor that binds only the protein-substrate complex (ES) is an uncompetitive inhibitor. Such an inhibitor, observed mostly with enzymes, will affect the catalytic function but not the substrate binding by causing structural distortion of the active site. Because free enzyme is temporarily reduced, $V_{\max_{\text{app}}}$ is reduced as well (as in Eq. 12.23), and the apparent affinity seems to be increased due to a shift of the reaction (Enzyme + Substrate→ES) to the right, as follows (Segel 1974):

$$K_{\text{m}_{\text{app}}} = \frac{K_{\text{m}}}{\left(1 + \frac{[I]}{K_{\text{i}}}\right)} \tag{12.24}$$

An interesting study by Barton et al. (1995) modeled the disappearance of trichloroethylene (TCE) and vinyl chloride from a closed vapor uptake chamber during concomitant rat exposure and showed how the three types of inhibition descriptions (competitive, noncompetitive, and uncompetitive) best fit the co-exposure data. This allowed elimination of noncompetitive inhibition as a mechanism of interaction between both chemicals but could not discriminate between competitive and uncompetitive inhibition with this particular exposure data set. The authors further pointed out that competitive inhibition could simulate all data sets with the same parameter values. In contrast, with uncompetitive inhibition, although multiple data sets were well simulated using the same kinetic parameters, several key mixture data sets were simulated only by varying parameter values.

Allosteric Interactions There are other examples of increased enzyme activity related to co-exposure that do not implicate a change in concentration of protein. A few enzyme kinetic studies on interacting xenobiotics have shown that some enzymes with multiple binding sites, particularly CYP 3A4, demonstrate unusual kinetics in the presence of another substrate or inhibitor. Different models were proposed for such cooperative binding (Kenworthy et al. 2001). The authors describe three different allosteric interaction models: (a) a two-site model with

competition between substrate and effector which can activate the enzyme at low concentrations but inhibit it at high concentrations; (b) a three-site model for heteroactivation where two substrates can bind cooperatively and stimulate metabolism at the activator site; and (c) a three-site model with inhibition including a substrate and an inhibitor (a more detailed description and complex equations can be found in (Kenworthy et al. 2001)). As an example of such cooperative activation and inhibition interactions, the rate of formation of 3-hydrodiazepam from diazepam increases up to nearly almost 400% in the presence of testosterone, and the formation of 6β-hydroxytestosterone from testosterone is inhibited 45% by diazepam.

Cofactor Depletion The cellular reserves of cofactors for phase 2 metabolic reactions (UDPGA, PAPS, GSH, etc.) are usually considered to be amounts well over saturation levels, making the description of reaction rate limited only by substrate concentrations. In some instances, the reserves of cofactors may be depleted well below saturation levels, rendering the reaction rate of substrate also dependent on levels of cellular concentrations of cofactors. The reaction rate, therefore, must be described as a bi-enzyme kinetic reaction where both cofactor and substrate concentration must be considered for the calculation of the reaction rate (Marangoni 2003). Of course, co-exposure to chemicals utilizing cofactors will affect the metabolic rate of other compounds using the same cofactor, independently of isoenzyme used. A description for this inhibition mechanism was used in a PBPK modeling study by (Zurlinden and Reisfeld 2015) for acetaminophen and its major metabolites (i.e., the sulfo-conjugate and the glucurono-conjugate) in humans. In this particular example, substrate and co-substrate inhibition (Forrest et al. 1982; Mutlib et al. 2006; Nagar et al. 2006) must also be considered, and the description of conjugation is as follows:

$$R_{\text{conjugation}} = \frac{V_{\max} \times C_t^S \times F_t^{cf}}{\left(\mathrm{K_m}^S + \times C_t^S + \times \frac{\left(C_t^S\right)^2}{\mathrm{Ksi}}\right)\left(K_m{}^{cf} + F_t^{cf}\right)} \tag{12.25}$$

where $F_t^{cf}$is the fraction of available cofactor in the metabolizing organ and K_{si} is the inhibition constant for the substrate inhibition. Superscripts S and cf are for substrate and cofactor, respectively. Although this particular example with acetaminophen very nicely describes the cofactor depletion phenomena, it is not in the context of co-exposures to other chemicals. But clearly, concomitant, or even subsequent, exposure of acetaminophen to another chemical metabolized by UGTs would be affected by the depletion of the UDPGA cofactor, and the rate of conjugation would have to be described accordingly.

12.3 PBPK Modeling Strategies for Mixtures

In this section, an overview of techniques or strategies is presented for pharmacokinetic modeling of mixtures and for describing or predicting the kinetics of mixture components. Having a multichemical exposure does not necessarily mean that pharmacokinetic interactions occur between all mixture components. None may actually occur, or some or all components may be affected by the presence of others. A simple way of determining which mixture components' kinetics are affected by the other mixture constituents is by comparing the pharmacokinetics (e.g., blood concentration vs time profiles) of the mixture's constituents when administered as a mixture with the kinetics of each constituent administered alone (at the same dose and exposure scenario). The components that show the same pharmacokinetic profiles in mixture and single chemical exposures are not affected by mixture constituents, unless the impact of multiple interactions caused by different chemicals cancel each other, which is rather unlikely. When developing a PBPK model for a mixture, it is important to identify the components that interact with each other and those that do not. The simplest situation is a mixture with no interaction between constituents. In this case, the PBPK model for the mixture can be developed exactly in the same way as if each chemical component were administered alone (single chemical exposures). When interactions occur (i.e., toxicokinetic profiles differ between single and mixture exposure), then chemical culprits (inhibitor, inducer, chelator, etc.) and mechanisms of interaction (see previous section) should be identified, if possible, to develop the mixture model appropriately. This may become very tedious depending on the level of complexity of the mixture, or even the level of knowledge of chemical components within the mixture.

12.3.1 Binary Mixtures

To date, many PBPK models have been published for binary mixtures for all sorts of chemicals, e.g., aromatic and chlorinated solvents (Andersen et al. 1987), petroleum products (Ali and Tardif 1999; Jang et al. 2001), drinking water contaminants (Niu et al. 2015; Tan et al. 2007; Isaacs et al. 2004), and medicinal drugs (Ishigam et al. 2001; Boom et al. 1998; Sugita et al. 1982; Russel et al. 1987, 1989), among others. When developing a model for such mixtures, it is common practice to start with single chemical PBPK models of both mixture components or to develop them if not available.

The next step would be to link both models together by the hypothesized mechanism of interaction using the appropriate mathematical description (see Chap. 9). The hypothesis can be supported by available information from in vitro experiments, known biotransformation pathways, pharmacokinetic experiments, etc. In a pharmacokinetic interaction, a mixture component may be the culprit (chemical modifying the pharmacokinetics of the other chemical) or the victim (the chemical

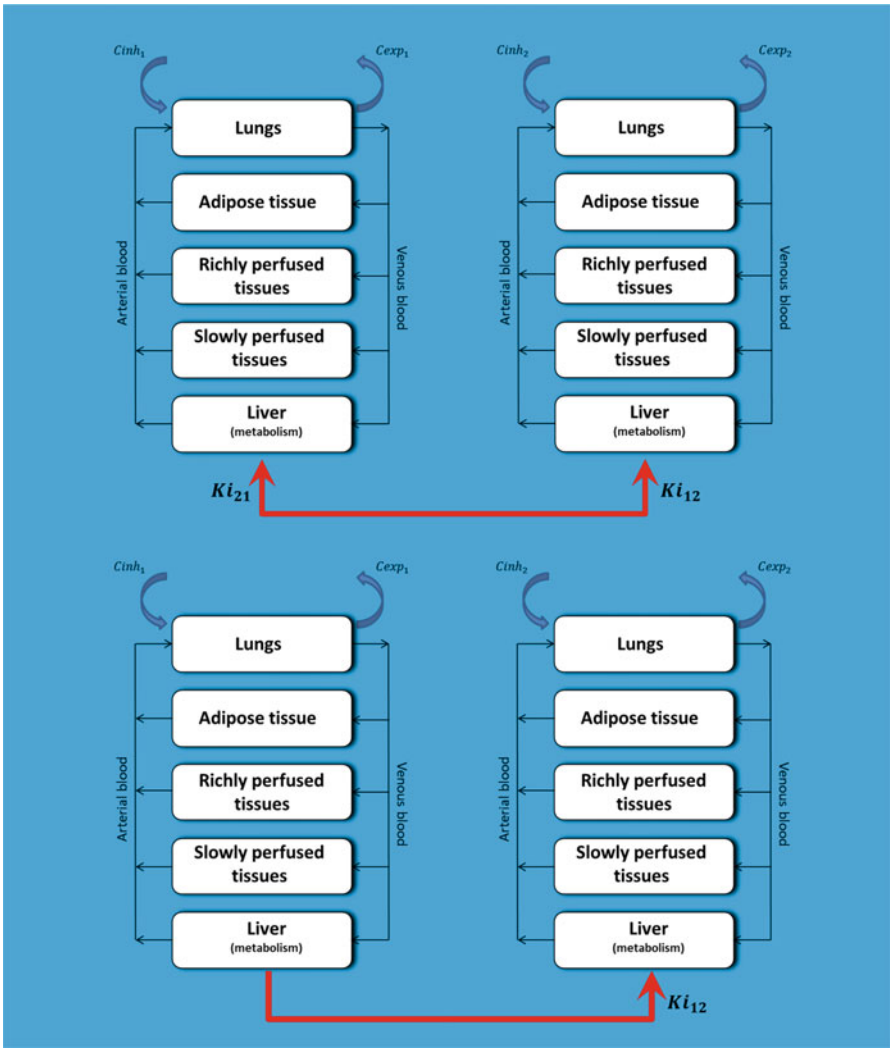

Fig. 12.4 Illustration of two examples of PBPK models for interacting pairs of chemicals. Top portion illustrates a mutual interaction, whereas bottom portion illustrates a one-way interaction where the chemical on the left is the culprit and the chemical on the right is the victim of the interaction

whose pharmacokinetics is being modified by the culprit) or both (Fig. 12.4). If chemical *A* is the culprit and chemical *B* is the victim in the interaction, then only the mathematical description of the pharmacokinetics of *B* will be modified to change absorption, metabolism, distribution and/or excretion as a function of the appropriate tissue dose of chemical *A* (e.g., liver concentration of *A* will affect hepatic metabolic rate of *B* if metabolic inhibition occurs). Alternatively, if both chemicals affect the

kinetics of the other chemical, both their mathematical descriptions will be modified according to the mechanism of interaction as a function of the culprit's appropriate tissue dose.

12.3.2 Mixtures of Greater Complexity

When pharmacokinetic interactions occur between more than two chemicals, it can become more complicated to describe the situation. An impact on the tissue dose of one chemical (chemical *A*) by the presence of another (chemical *B*) will affect all other chemicals with which chemical *A* interacts. The chemicals with modulated tissue dose will in turn affect the tissue levels of all chemicals they interact with and so on.

The situation can become even more complicated when a mixture component is biotransformed into a metabolite which interacts with its parent compound (e.g., product inhibition) or even other mixture components/metabolites. Another hurdle to tackle in mixture toxicology is that mixture components are not always fully identified. The following subsections relate to approaches that have been proposed to deal with these problems.

12.3.2.1 Extrapolating In Vivo Binary Interactions to Complex Mixtures

To resolve the problem of PBPK modeling of mixtures with more than two interacting components, Krishnan's laboratory at Université de Montréal proposed and validated an approach that allows the interlinkage of the pharmacokinetics of all interacting chemicals in a single model (Tardif et al. 1997; Haddad et al. 2001; Haddad and Krishnan 1998; Haddad et al. 1999a, 1998; Krishnan et al. 2002). This is done by linking each of the chemical models by the description of the binary interaction, forming a "web of interactions" (Fig. 12.5). This generates a web of pharmacokinetic interconnections, and all chemicals in that web are affected by a modulation of tissue dose of one of the mixture components. This also applies to metabolites that interact with the mixture components. In the latter case, the metabolite needs to be added to this chemical web (an additional PBPK model should be made for this metabolite) and linked appropriately. When a chemical in the mixture does not interact, it can be in the web but without any linkage to other chemicals.

This binary interaction-based PBPK modeling of mixtures has been validated in vivo in rats with ternary and quaternary mixtures of aromatic hydrocarbons, i.e., toluene, ethylbenzene, and m-xylene by Tardif et al. (1997) and benzene, toluene, ethylbenzene, and m-xylene by Haddad et al. (1999a) and with the addition of a fifth component, dichloromethane (Haddad et al. 2000b), as well S-8 and JP-8 jet fuel mixtures (Martin et al. 2012). Although this approach offers an accurate model to describe the mixture, it has high data requirements. Because all binary interactions

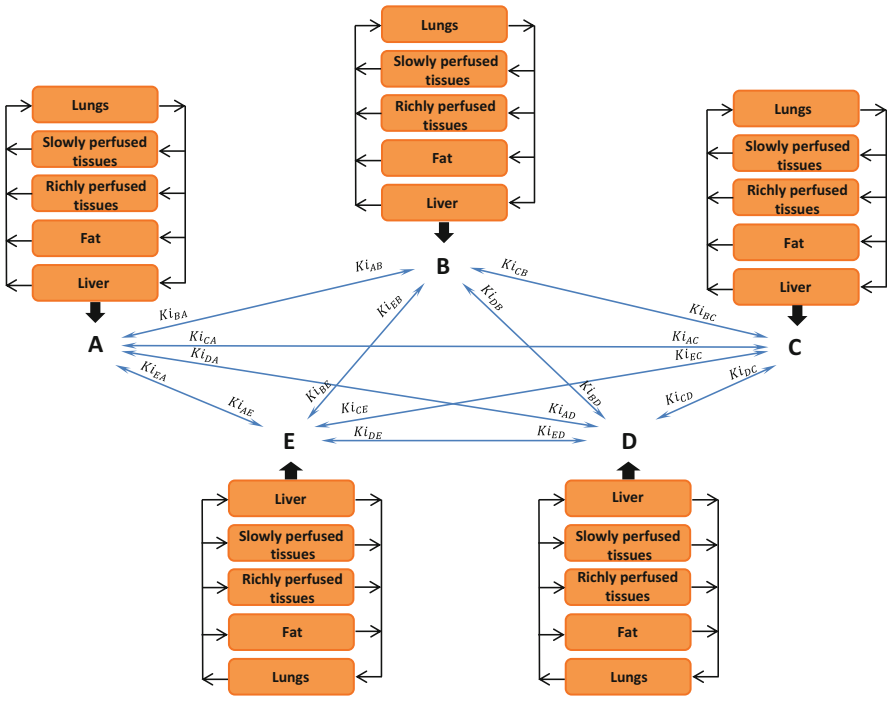

Fig. 12.5 Conceptual representation of a web of binary interactions in a quinary PBPK mixture model

must be described between all interacting components, they must all be characterized. This task may become time-consuming, animal intensive, and costly when the mixture components are numerous. The number of binary interactions (N) to characterize in a whole mixture of "n" components is determined as follows:

$$N = n\,(n-1)/2 \tag{12.26}$$

According to this equation, for a mixture of ten interacting chemicals, 45 binary interactions should be characterized to apply this interaction-based approach. To overcome this hurdle in mixture PBPK modeling, alternative methods have been proposed for mixtures with large number of components. In these mixtures, it is simply currently too costly to characterize all in vivo binary interactions, and oftentimes the identity of all components has simply not yet been determined.

12.3.2.2 QSAR Approach

Building a mixture PBPK model can prove to be difficult when data on binary interactions, chemical biotransformation, and partitioning are not fully characterized. Instead of going through full parameter characterization to develop a PBPK

model, a useful initial step can be to use quantitative structure-activity relationships (QSAR) to estimate the model parameter values. Price and Krishnan (Price and Krishnan 2011) developed QSAR algorithms for volatile organic chemicals to estimate partition coefficients, V_{max}, and K_m based on chemical structures of 53 different chemicals. Using estimated parameters, they predicted the toxicokinetics of different mixtures by assuming competitive inhibition and assuming K_i values were equal to K_m. The use of QSAR approaches can rapidly give health assessors an idea of the amplitude of interactions if competitive inhibition is the expected mechanism between mixture constituents.

12.3.2.3 Unidentified Components: Chemical Lumping

Many or most complex environmental mixtures to which humans are exposed to are not completely characterized, but some constituents may be of interest to estimate exposure. Such is the case with gasoline where a few components, such as benzene, toluene, ethylbenzene, xylene, and n-hexane, are chemicals of toxicological interest, and thousands of other hydrocarbons (isoalkanes, n-alkanes, aromatic derivatives, and smaller amounts of alkenes and alkynes) have lesser or no known toxicological significance. Simulating all constituents of such mixtures would be an enormous task and is not feasible due to resource and time constraints. In addition to difficulties in determining all parameter values for all components, validation would also prove to be nearly impossible because of limitations in chromatographic separation of all mixture components.

To circumvent such a problem, Dennison et al. (2003) devised an original strategy for gasoline mixtures. In their study, they proposed to lump most of the gasoline chemical components together by considering them as a single chemical entity and leaving the known toxicologically relevant components as separate entities, assuming competitive inhibition among components. The approach was similar to the binary mixtures approach assuming competitive inhibition between all mixture components (Haddad and Krishnan 1998). Known parameter values for partition coefficients, V_{max} and K_m, were used (i.e., for benzene, toluene, ethylbenzene, xylene, and n-hexane), and K_i values were set equal to K_m values. All other parameters related to the chemical lump were mathematically optimized to observed values, and again K_i was set equal to K_m. The characteristics of the lumped compartment changed with gasoline blend (winter blend vs summer blend). This study demonstrated the feasibility of reducing the number of model parameters in a mixture model enabling a targeted focus on toxicologically relevant mixture components.

12.3.2.4 Physiological Limits of Interactions

An alternative to the interaction-based PBPK model approach is to simply consider the physiological limits to determine the plausible range of internal exposure

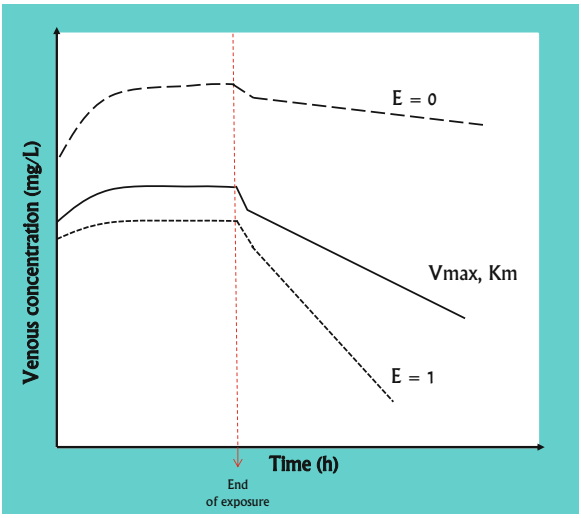

Fig. 12.6 Simulations of blood concentration vs time profiles of toluene in rats following a 4 h inhalation exposure considering average rat clearance (solid line), maximal inhibition of hepatic clearance (dashed line), and maximal induction of hepatic clearance (dotted line). Hepatic extraction ratio is represented by E (Modified from Haddad et al. 2000a)

(Haddad et al. 2000a). The logic is simple: if the sole mechanism of interaction occurs essentially at the level of hepatic metabolism, then we can determine the maximal and minimal tissue dose that a mixture component can attain if its pharmacokinetics are modulated by co-exposure. These limits are determined by fixing chemical hepatic clearance equal to hepatic blood flow (i.e., maximal impact of enzyme induction on clearance leading to a hepatic extraction ratio of 1) and to zero where the biotransformation is totally inhibited (i.e., hepatic extraction ratio = 0) (Fig. 12.6).

Although this method does not allow the risk assessor to determine with precision the concentration time profile of mixture constituents, it does allow clear estimation of the maximal value of tissue dose which would be protective/conservative in terms of health risk assessment toward potential increased internal exposure due to combined exposures. Furthermore, this method is independent of the number of mixture components and identity of mixture components. Such an approach would also be applicable to other types of interactions where physiology can be rate limiting (e.g., renal excretion, biliary excretion, extrahepatic metabolism, etc.). A limitation of this approach is for compounds having only biotransformation as a mode of elimination and only one metabolic pathway; under such circumstances, the estimated limits would yield very large concentration intervals. It works well with VOCs because they are also eliminated by exhalation.

12.3.2.5 IVIVE of Interactions

In vitro to in vivo extrapolations are acknowledged as the way toxicity testing for environmental agents should be conducted in the twenty-first century (NRC 2007). Obtaining in vivo data for interactions is not always feasible or desirable because

(i) the workload and time associated with experiments create practical limitations, (ii) the associated costs are important, and/or (iii) the number of animals required for such studies is incompatible with the call for animal reduction in research. In the pharmaceutical industry, in vitro assays for screening and preclinical research in drug metabolism and pharmacokinetics are conducted on a routine basis. Large amounts of in vitro data are collected on drug-drug metabolic interactions using cell cultures, cell suspensions, or subcellular fractions (e.g., microsomes, S9 fractions, etc.). Metabolic constants V_{max}, K_m, or intrinsic clearances are often measured as well as inhibition constants (K_i) for drug-drug interactions for different enzymes associated with drug clearance rates. Many empirical clearance models have been proposed to extrapolate in vitro clearance and metabolic interactions data to the in vivo situation, but they have had varying success rates (Wilkinson 1987; Robinson et al. 1991; Robinson 1992; Saville et al. 1992). To increase predictability of these models, several studies have proposed adjusting equations to account for in vitro non-specific binding in the incubation medium to better reflect the free concentrations of substrates and/or inhibitors at enzyme active sites, both in vitro and in vivo (Obach 1997, 1999; Mclure et al. 2000).

$$\mathrm{CL}_{\mathrm{invivo}} = \frac{Q_{\mathrm{liver}} \times \mathrm{RBP} \times \mathrm{CL}_{\mathrm{int,met}} \times \mathrm{Fu_p}/\mathrm{Fu_{inc}}}{Q_{\mathrm{liver}} \times \mathrm{RBP} + \mathrm{CL}_{\mathrm{int,met}} \times \mathrm{Fu_p}/\mathrm{Fu_{inc}}} \tag{12.27}$$

where Q_{liver}, RBP $CL_{int,met}$, Fu_p, and Fu_{inc}, respectively, refer to blood flow in the liver, blood to plasma ratio, metabolic intrinsic clearance, fraction unbound in plasma, and fraction unbound in incubation medium.

Recently, a physiologically based model for hepatic metabolic interactions has been proposed (Theil et al. 2003; Haddad et al. 2010), where the liver is described as a multicompartmental model, representing the vascular, the interstitial, and the cellular matrix (Fig. 12.7). Exchanges between these compartments consider active transport and passive diffusion. Metabolism and metabolic interactions are considered to occur inside the hepatocytes and are related to unbound concentrations in the cells. This unbound concentration is a result of different input and output processes that influence intracellular concentrations (i.e., active efflux, biliary excretion, active uptake, simple diffusion, metabolism, intracellular and extracellular protein binding, and solubility in lipids). Additionally, the chemical concentration gradient along the sinusoids of the liver lobule has also been simulated by representing the liver as seven segments linked in series. Compared to other models described above, this model fared best in predictions of drug-drug binary interactions between three cytochrome 2D1 substrates (i.e., bufuralol, bunitrolol, and debrisoquine) in an isolated perfused liver system. Accordingly, for extrapolation to work, data taken from in vitro assays must also be adjusted to eliminate bias from non-specific binding. A very recent study suggested a more complex hepatic model that incorporates hepatic lobule geometry and many of the processes described here in order to predict the magnitude of metabolic interactions (Cherkaoui-Rbati et al. 2017). These models should be compared to assess their predictive power.

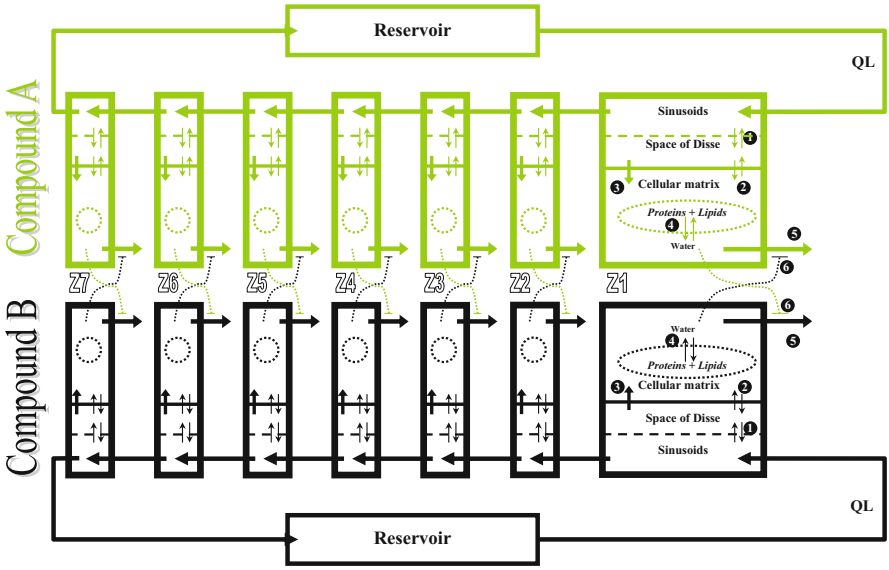

Fig. 12.7 Conceptual representation of the physiologically based pharmacokinetic model of a binary mixture (compounds *A* and *B*) in an isolated perfused liver (IPRL) system. The liver is separated into seven segments (Z1–Z7) connected in series, and each segment is further subdivided into three subcompartments (sinusoids, space of Disse, and cellular matrix). Chemicals enter the sinusoidal space of the first segment by the hepatic portal vein. From the sinusoidal space, the chemical can go to the next liver segment or distribute to other subcompartments. The exchange between the sinusoid and the space of Disse (1) is very rapid due to the presence of large fenestrae. The chemicals in the space of Disse can then enter the cellular matrix by partitioning processes (2) or by active uptake. In the cellular matrix, the chemicals freely distribute between the lipids and proteins and the water (4). The compounds are eliminated from the cellular matrix subcompartment by metabolism (5). The rate of metabolism is dependent on the substrate unbound concentration in the cell and influenced by the intracellular unbound concentration of competing substrate (6). Before leaving the liver, the chemical must pass through all liver segments where all the same processes occur. QL refers to the perfusate flow in the recirculating IPRL system (Modified from Haddad et al. 2010)

Current methods for estimating hepatic clearance have been shown to be quite ineffective in predicting in vivo clearance of compounds that are highly bound to albumin (i.e., $Fu_p < 0.05$) (Poulin et al. 2012). Recent studies have shown that to accurately predict the clearance of this category of compounds, it must be assumed that there is a mechanism facilitating the distribution of the bound drug in the organ by albumin, hence leading to an apparent unbound fraction in the organ that is greater than in blood (Poulin et al. 2012). A clearance algorithm (Eq. 12.27) that adjusted Fu_p in Eq. 12.28 was proposed and validated for IVIVE of in vitro metabolic rate from microsomes (Poulin and Haddad 2013) and hepatocytes (Poulin and Haddad 2013).

$$Fu_{p-adjusted} = \frac{PLR \times Fu_p \times \frac{Funionized_{plasma}}{Funionized_{cells}}}{1 + (PLR - 1) \times Fu_p \times \frac{Funionized_{plasma}}{Funionized_{cells}}} \quad (12.28)$$

where PLR and Funionized refer, respectively, to the plasma to liver albumin concentration ratio and the chemical's fraction that is the unionized form in the matrix.

Although nothing has been published for the impact of albumin binding on the prediction of metabolic interactions, interactions between naproxen and bisphenol A for glucuronidation were studied in vitro (Verner et al. 2010) and in isolated perfused rat liver (IPRL) (Bounakta et al. 2017; Poulin et al. 2017). In this IPRL study, liver co-exposure to both compounds in the presence and absence of albumin showed that competitive inhibition was observed. But in the presence of albumin, the clearance and interaction of these two highly albumin-bound compounds were clearly affected confirming the occurrence of an albumin-facilitated uptake mechanism, suggesting that predictions of metabolic interactions for highly albumin-bound compounds from in vitro data must be treated in a similar fashion (i.e., inhibition constants should be adjusted by $Fu_{p\text{-adjusted}}$).

12.4 PBPK Modeling and Mixture Risk Assessment

PBPK modeling can prove to be practical and useful in mixture risk assessment. Exposure assessment of mixture components and their potential for toxicokinetic interactions are among the many challenges that mixtures pose to risk assessors. As shown above, mixture PBPK modeling can allow for prediction of internal exposure, or target tissue dose, even in contexts where toxicokinetic interactions occur.

Unless toxicodynamic interactions are known to occur, the PBPK modeling approaches discussed above can be used in the context of mixture risk assessment. Haddad et al. (Haddad et al. 1999b, 2001) demonstrated an approach in which risk assessors could use PBPK modeling to estimate biological/target organ hazard indices (BHI or THI) in lieu of calculating hazard indices (HI) for mixtures using external exposures concentrations. This approach consists of summing the internal doses of mixture constituents that have similar modes of actions or the same target tissues (refer to Chap. 14). The internal doses of mixture components are normalized by the internal dose obtained during single exposure to guideline values (e.g., *T*hreshold *L*imit *V*alues, *Ref*erence *C*oncentrations). If the sum is greater than unity, exposure to the mixture is considered to pose a health risk. This allows for consideration of the pharmacokinetic interactions between mixture components in the health risk evaluation.

PBPK modeling is also useful in mixture risk assessment because it confers greater confidence in different types of extrapolations. A lot of data come from animal studies, and the kinetics of chemicals (mixtures or single) can be translated to the human situation by changing the animal parameter values to those of humans.

Tardif et al. (1997) extrapolated a ternary PBPK model for toluene, ethylene, and xylene from rat to human, and simulations successfully predicted experimental data from exposed human volunteers.

PBPK models incorporating Monte Carlo simulations can also be used to estimate exposure in the population of interest and determine the range of internal exposure levels in the population to protect sensitive populations (Niu et al. 2015; Hinderliter et al. 2011). Additionally, these models can be extrapolated to different lifestages (e.g., neonates, infants, teenagers, etc.) and sexes and to different polymorphisms to further characterize exposure in different subpopulations (Haddad et al. 2006; Verner et al. 2008, 2009; Hinderliter et al. 2011).

12.5 Conclusions, Current Needs, and Research Perspectives

Although many advances have been made in recent years to predict exposure resulting from toxicokinetic interactions and to use this information in risk assessment, hurdles still remain. In vivo characterization remains the gold standard for identifying and characterizing binary interactions but is too costly and time-consuming. Although not 100% accurate, we can now put more confidence in the prediction of in vivo metabolic interactions from in vitro data, but toxicokinetic interactions can still occur at other levels, and we cannot, therefore, solely rely on metabolic in vitro data. Effects on absorption, distribution, and excretion processes of chemicals, especially for chemicals that interact with proteins (in binding or transport processes), are frequent, and additional research emphasis should be put on IVIVE of these processes to increase capacity in high-throughput data generation and PBPK model predictions. Development of QSAR for toxicokinetic interactions is also very much needed, but the generation of such knowledge is limited by the available data on toxicokinetic interactions. Predictive environmental toxicology could certainly rely on data available from the pharmaceutical industry to generate QSAR algorithms, but more data for ADME of environmental contaminants are required. Data generated from projects such as TOXCAST (Dix et al. 2007; Judson et al. 2010) offer promise in this context.

References

Ali, N., and R. Tardif. 1999. Toxicokinetic Modeling of the Combined Exposure to Toluene and n-Hexane in Rats and Humans. *Journal of Occupational Health* 41: 95–103.

Andersen, M.E., L.S. Birnbaum, H.A. Barton, and C.R. Eklund. 1997. Regional hepatic CYP1A1 and CYP1A2 induction with 2,3,7,8-tetrachlorodibenzo-p-dioxin evaluated with a multicompartment geometric model of hepatic zonation. *Toxicology and Applied Pharmacology* 144: 145–155.

Andersen, M.E., M.L. Gargas, H.J.I. Clewell, and K.M. Severyn. 1987. Quantitative evaluation of the metabolic interactions between trichloroethylene and 1,1-dichloroethylene in vivo using gas uptake methods. *Toxicology and Applied Pharmacology* 89: 149–157.

Barton, H.A., J.R. Creech, C.S. Godin, G.M. Randall, and C.S. Seckel. 1995. Chloroethylene mixtures: Pharmacokinetic modeling and in vitro metabolism of vinyl chloride, trichloroethylene, and trans-1,2-dichloroethylene in rat. *Toxicology and Applied Pharmacology* 130: 237–247.

Boom, S.P., I. Meyer, A.C. Wouterse, and F.G. Russel. 1998. A physiologically based kidney model for the renal clearance of ranitidine and the interaction with cimetidine and probenecid in the dog. *Biopharmaceutics & Drug Disposition* 19: 199–208.

Bounakta, S., M. Bteich, M. Mantha, P. Poulin, and S. Haddad. 2017. Predictions of bisphenol A hepatic clearance in the isolated perfused rat liver (IPRL): impact of albumin binding and of co-administration with naproxen. *Xenobiotica* 3: 1–13.

Budha, N.R., A. Frymoyer, G.S. Smelick, J.Y. Jin, M.R. Yago, M.J. Dresser, S.N. Holden, L.Z. Benet, and J.A. Ware. 2012. Drug absorption interactions between oral targeted anticancer agents and PPIs: Is pH-dependent solubility the Achilles heel of targeted therapy? *Clinical Pharmacology and Therapeutics* 92: 203–213.

Cherkaoui-Rbati, Mohammed H., Stuart W. Paine, Peter Littlewood, Cyril Rauch, and Jinn-Moon Yang. 2017. A quantitative systems pharmacology approach, incorporating a novel liver model, for predicting pharmacokinetic drug-drug interactions. *PLoS One* 12 (9): e0183794.

Chien, J.Y., K.E. Thummel, and J.T. Slattery. 1997. Pharmacokinetic consequences of induction of CYP2E1 by ligand stabilization. *Drug Metabolism and Disposition* 25: 1165–1175.

Choi, H.K., G.L. Flynn, and G.L. Amidon. 1990. Transdermal delivery of bioactive peptides: The effect of n-decylmethyl sulfoxide, pH, and inhibitors on enkephalin metabolism and transport. *Pharmaceutical Research* 7: 1099–1106.

Crisp, J., and C. Taylor. 2012. *Potter & Perry's fundamentals of nursing*. 3rd ed. Chatswood: Elsevier.

De Castro, J., C. Aguirre, J.M. Rodriguez-Sasiain, E. Gomez, M.J. Garrido, and R. Calvo. 1996. The effect of changes in gastric pH induced by omeprazole on the absorption and respiratory depression of methadone. *Biopharmaceutics & Drug Disposition* 17: 551–563.

Dennison, J.E., M.E. Andersen, and R.S. Yang. 2003. Characterization of the pharmacokinetics of gasoline using PBPK modeling with a complex mixtures chemical lumping approach. *Inhalation Toxicology* 15: 961–986.

Dix, D.J., K.A. Houck, M.T. Martin, A.M. Richard, R.W. Setzer, and R.J. Kavlock. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicological Sciences* 95: 5–12.

Espie, P., D. Tytgat, M.L. Sargentini-Maier, I. Poggesi, and J.B. Watelet. 2009. Physiologically based pharmacokinetics (PBPK). *Drug Metabolism Reviews* 41: 391–407.

Forrest, J.A., J.A. Clements, and L.F. Prescott. 1982. Clinical pharmacokinetics of paracetamol. *Clinical Pharmacokinetics* 7: 93–107.

Gregus, Z. 2008. Mechanisms of toxicology. In *Casarett & Doull's toxicology. The basic science of poisons*, ed. C.D. Klaassen, 7th ed. New York: McGraw Hill.

Greiff, J.M., and D. Rowbotham. 1994. Pharmacokinetic drug interactions with gastrointestinal motility modifying agents. *Clinical Pharmacokinetics* 27: 447–461.

Haddad, S., M. Beliveau, R. Tardif, and K. Krishnan. 2001. A PBPK modeling-based approach to account for interactions in the health risk assessment of chemical mixtures. *Toxicological Sciences* 63: 125–131.

Haddad, S., G. Charest-Tardif, and K. Krishnan. 2000a. Physiologically based modeling of the maximal effect of metabolic interactions on the kinetics of components of complex chemical mixtures. *Journal of Toxicology and Environmental Health-Part A* 61: 209–223.

Haddad, S., G. Charest-Tardif, R. Tardif, and K. Krishnan. 2000b. Validation of a physiological modeling framework for simulating the toxicokinetics of chemicals in mixtures. *Toxicology and Applied Pharmacology* 167: 199–209.

Haddad, S., and K. Krishnan. 1998. Physiological modeling of toxicokinetic interactions: Implications for mixture risk assessment. *Environmental Health Perspectives* 106: 1377–1384.

Haddad, S., P. Poulin, and C. Funk. 2010. Extrapolating in vitro metabolic interactions to isolated perfused liver: Predictions of metabolic interactions between R-bufuralol, bunitrolol, and debrisoquine. *Journal of Pharmaceutical Sciences* 99: 4406–4426.

Haddad, S., G.C. Tardif, and R. Tardif. 2006. Development of physiologically based toxicokinetic models for improving the human indoor exposure assessment to water contaminants: Trichloroethylene and trihalomethanes. *Journal of Toxicology and Environmental Health. Part A* 69: 2095–2136.

Haddad, S., R. Tardif, G. Charest-Tardif, and K. Krishnan. 1999a. Physiological modeling of the toxicokinetic interactions in a quaternary mixture of aromatic hydrocarbons. *Toxicology and Applied Pharmacology* 161: 249–257.

Haddad, S., R. Tardif, and K. Krishnan. 1998. *Physiological modeling of higher order metabolic interactions in complex chemical mixtures*. Soc Computer Simulation: San Diego.

Haddad, S., R. Tardif, C. Viau, and K. Krishnan. 1999b. A modeling approach to account for toxicokinetic interactions in the calculation of biological hazard index for chemical mixtures. *Toxicology Letters* 108: 303–308.

Hayes, W.J., and G.W. Pearce. 1953. Pesticide formulation: Relation to safety in use. *J Agr Food Chem* 1: 466.

Hinderliter, P.M., P.S. Price, M.J. Bartels, C. Timchalk, and T.S. Poet. 2011. Development of a source-to-outcome model for dietary exposures to insecticide residues: An example using chlorpyrifos. *Regulatory Toxicology and Pharmacology* 61: 82–92.

Hsu, C.Y., and M.-S. Wu. 2012. Drug-Induced Alterations in Renal Hemodynamics. *Acta Nephrologica* 26: 121–133.

Isaacs, K.K., M.V. Evans, and T.R. Harris. 2004. Visualization-based analysis for a mixed-inhibition binary PBPK model: Determination of inhibition mechanism. *Journal of Pharmacokinetics and Pharmacodynamics* 31: 215–242.

Ishigam, M., M. Uchiyama, T. Kondo, H. Iwabuchi, S. Inoue, W. Takasaki, T. Ikeda, T. Komai, K. ITO, and Y. Sugiyama. 2001. Inhibition of in vitro metabolism of simvastatin by itraconazole in humans and prediction of in vivo drug-drug interactions. *Pharmaceutical Research* 18: 622–631.

Jacob, S.W., M. Bischel, and R.J. Herschler. 1964. Dimethyl Sulfoxide: Effects on the Permeability of Biologic Membranes (Preliminary Report). *Current Therapeutic Research, Clinical and Experimental* 6: 193–198.

Jang, J.Y., P.O. Droz, and S. Kim. 2001. Biological monitoring of workers exposed to ethylbenzene and co-exposed to xylene. *International Archives of Occupational and Environmental Health* 74: 31–37.

Judson, R.S., K.A. Houck, R.J. Kavlock, T.B. Knudsen, M.T. Martin, H.M. Mortensen, D.M. Reif, D.M. Rotroff, I. Shah, A.M. Richard, and D.J. Dix. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project. *Environmental Health Perspectives* 118: 485–492.

Kanamitsu, S., K. Ito, C.E. Green, C.A. Tyson, N. Shimada, and Y. Sugiyama. 2000a. Prediction of in vivo interaction between triazolam and erythromycin based on in vitro studies using human liver microsomes and recombinant human CYP3A4. *Pharmaceutical Research* 17: 419–426.

Kanamitsu, S.I., K. Ito, H. Okuda, K. Ogura, T. Watabe, K. Muro, and Y. Sugiyama. 2000b. Prediction of in vivo drug-drug interactions based on mechanism-based inhibition from in vitro data: Inhibition of 5-fluorouracil metabolism by (E)-5-(2-Bromovinyl)uracil. *Drug Metabolism and Disposition* 28: 467–474.

Kenworthy, K.E., S.E. Clarke, J. Andrews, and J.B. Houston. 2001. Multisite kinetic models for CYP3A4: Simultaneous activation and inhibition of diazepam and testosterone metabolism. *Drug Metabolism and Disposition* 29: 1644–1651.

Keys, D.A., I.R. Schultz, D.A. Mahle, and J.W. Fisher. 2004. A quantitative description of suicide inhibition of dichloroacetic acid in rats and mice. *Toxicological Sciences* 82: 381–393.

Krishnan, K., and M.E. Andersen. 2007. Physiologically-based pharmacokinetic modeling in toxicology. In *Principles and methods of toxicology*, ed. A.W. HAYES. New-York: Taylor and Francis.

Krishnan, K., H.J. Clewell 3rd, and M.E. Andersen. 1994. Physiologically based pharmacokinetic analyses of simple mixtures. *Environmental Health Perspectives* 102 (Suppl 9): 151–155.

Krishnan, K., S. Haddad, M. Beliveau, and R. Tardif. 2002. Physiological modeling and extrapolation of pharmacokinetic interactions from binary to more complex chemical mixtures. *Environmental Health Perspectives* 110: 989–994.

Kudsk, F.N. 1965. The influence of ethyl alcohol on the absorption of mercury vapour from the lungs in man. *Acta Pharmacol Toxicol (Copenh)* 23: 263–274.

Kurihara-Bergstrom, T., G.L. Flynn, and W.I. Higuchi. 1987. Physicochemical study of percutaneous absorption enhancement by dimethyl sulfoxide: Dimethyl sulfoxide mediation of vidarabine (ara-A) permeation of hairless mouse skin. *The Journal of Investigative Dermatology* 89: 274–280.

Leung, H.W., D.J. Paustenbach, F.J. Murray, and M.E. Andersen. 1990. A physiological pharmacokinetic description of the tissue distribution and enzyme-inducing properties of 2,3,7,8-tetrachlorodibenzo-p-dioxin in the rat. *Toxicology and Applied Pharmacology* 103: 399–410.

Marangoni, A.G. 2003. *Enzyme kinetics: A modern approach*. Hoboken, NJ: John Wiley & Sons.

Martin, S.A., J.L. Campbell, R.T. Tremblay, and J.W. Fisher. 2012. Development of a physiologically based pharmacokinetic model for inhalation of jet fuels in the rat. *Inhalation Toxicology* 24: 1–26.

Mclure, J.A., J.O. Miners, and D.J. Birkett. 2000. Nonspecific binding of drugs to human liver microsomes. *British Journal of Clinical Pharmacology* 49: 453–461.

Mutlib, A.E., T.C. Goosen, J.N. Bauman, J.A. Williams, S. Kulkarni, and S. Kostrubsky. 2006. Kinetics of acetaminophen glucuronidation by UDP-glucuronosyltransferases 1A1, 1A6, 1A9 and 2B15. Potential implications in acetaminophen-induced hepatotoxicity. *Chemical Research in Toxicology* 19: 701–709.

Nagar, S., S. Walther, and R.L. Blanchard. 2006. Sulfotransferase (SULT) 1A1 polymorphic variants *1, *2, and *3 are associated with altered enzymatic activity, cellular phenotype, and protein degradation. *Molecular Pharmacology* 69: 2084–2092.

Nimmo, W.S., R.C. Heading, J. Wilson, P. Tothill, and L.F. Prescott. 1975. Inhibition of gastric emptying and drug absorption by narcotic analgesics. *British Journal of Clinical Pharmacology* 2: 509–513.

Niu, Z., X. Zang, and Y. Zhang. 2015. Using physiologically based pharmacokinetic models to estimate the health risk of mixtures of trihalomethanes from reclaimed water. *Journal of Hazardous Materials* 285: 190–198.

NRC. 2007. *Toxicity testing in the 21st century: A vision and a strategy*. Washington DC: The National Academies Press.

Obach, R.S. 1997. Nonspecific binding to microsomes: Impact on scale-up of in vitro intrinsic clearance to hepatic clearance as assessed through examination of warfarin, imipramine, and propranolol. *Drug Metabolism and Disposition* 25: 1359–1369.

———. 1999. Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: An examination of in vitro half-life approach and nonspecific binding to microsomes. *Drug Metabolism and Disposition* 27: 1350–1359.

Oskarsson, A., and B. Lind. 1985. Increased lead levels in brain-after long term treatment with lead and dithiocarbamate or thirum derivatives in rats. *Acta Pharmacologica et Toxicologica* 56: 309–315.

Poulin, P., and S. Haddad. 2013. Toward a new paradigm for the efficient in vitro-in vivo extrapolation of metabolic clearance in humans from hepatocyte data. *Journal of Pharmaceutical Sciences* 102: 3239–3251.

Poulin, P., J.R. Kenny, C.E. Hop, and S. Haddad. 2012. In vitro-in vivo extrapolation of clearance: Modeling hepatic metabolic clearance of highly bound drugs and comparative assessment with existing calculation methods. *Journal of Pharmaceutical Sciences* 101: 838–851.

Poulin, Patrick, Michel Bteich, and Sami Haddad. 2017. Supplemental analysis of the prediction of hepatic clearance of binary mixtures of bisphenol A and naproxen determined in an isolated perfused rat liver model to promote the understanding of potential albumin-facilitated hepatic uptake mechanism. *Journal of Pharmaceutical Sciences* 106 (11): 3207–3214.

Price, K., and K. Krishnan. 2011. An integrated QSAR-PBPK modelling approach for predicting the inhalation toxicokinetics of mixtures of volatile organic chemicals in the rat. *SAR and QSAR in Environmental Research* 22: 107–128.

Qiao, G.L., J.D. Brooks, R.E. Baynes, N.A. Monteiro-Riviere, P.L. Williams, and J.E. Riviere. 1996. The use of mechanistically defined chemical mixtures (MDCM) to assess component effects on the percutaneous absorption and cutaneous disposition of topically exposed chemicals. I. Studies with parathion mixtures in isolated perfused porcine skin. *Toxicology and Applied Pharmacology* 141: 473–486.

Riihimaki, V. 1979. Percutaneous absorption of m-xylene from a mixture of m-xylene and isobutyl alcohol in man. *Scandinavian Journal of Work, Environment & Health* 5: 143–150.

Robinson, D.S., D.A. Cunningham, S. Dave, J. Fleming, and D.M. Mitchell. 1991. Diagnostic value of lung clearance of 99mTc DTPA compared with other non-invasive investigations in Pneumocystis carinii pneumonia in AIDS. *Thorax* 46: 722–726.

Robinson, P.J. 1992. Physiologically based liver modeling and risk assessment. *Risk Analysis* 12: 139–148.

Russel, F.G., A.C. Wouterse, and C.A. van Ginneken. 1987. Physiologically based pharmacokinetic model for the renal clearance of phenolsulfonphthalein and the interaction with probenecid and salicyluric acid in the dog. *Journal of Pharmacokinetics and Biopharmaceutics* 15: 349–368.

———. 1989. Physiologically based pharmacokinetic model for the renal clearance of iodopyracet and the interaction with probenecid in the dog. *Biopharmaceutics & Drug Disposition* 10: 137–152.

Sarangapani, R., J. Teeguarden, K.P. Plotzke, J.M. Mckim Jr., and M.E. Andersen. 2002. Dose-response modeling of cytochrome p450 induction in rats by octamethylcyclotetrasiloxane. *Toxicological Sciences* 67: 159–172.

Saville, B.A., M.R. Gray, and Y.K. Tam. 1992. Models of hepatic drug elimination. *Drug Metabolism Reviews* 24: 49–88.

Segel, I.H. 1974. *Enzyme kinetics: Behavior and analysis of rapid equilibrium and steady state enzyme systems*. New York: John Wiley.

Sugita, O., Y. Sawada, Y. Sugiyama, T. Iga, and M. Hanano. 1982. Physiologically based pharmacokinetics of drug-drug interaction: A study of tolbutamide-sulfonamide interaction in rats. *Journal of Pharmacokinetics and Biopharmaceutics* 10: 297–316.

Tan, Y.M., K.H. Liao, and H.J. Clewell 3rd. 2007. Reverse dosimetry: Interpreting trihalomethanes biomonitoring data using physiologically based pharmacokinetic modeling. *Journal of Exposure Science & Environmental Epidemiology* 17: 591–603.

Tardif, R., G. Charest-Tardif, J. Brodeur, and K. Krishnan. 1997. Physiologically based pharmacokinetic modeling of a ternary mixture of alkyl benzenes in rats and humans. *Toxicology and Applied Pharmacology* 144: 120–134.

Theil, F.P., T.W. Guentert, S. Haddad, and P. Poulin. 2003. Utility of physiologically based pharmacokinetic models to drug development and rational drug discovery candidate selection. *Toxicology Letters* 138: 29–49.

Thompson, C.M., B. Sonawane, H.A. Barton, R.S. Dewoskin, J.C. Lipscomb, P. Schlosser, W.A. Chiu, and K. Krishnan. 2008. Approaches for applications of physiologically based pharmacokinetic models in risk assessment. *Journal of Toxicology and Environmental Health. Part B, Critical Reviews* 11: 519–547.

Verner, M.A., P. Ayotte, G. Muckle, M. Charbonneau, and S. Haddad. 2009. A physiologically based pharmacokinetic model for the assessment of infant exposure to persistent organic pollutants in epidemiologic studies. *Environmental Health Perspectives* 117: 481–487.

Verner, M.A., M. Charbonneau, L. Lopez-carrillo, and S. Haddad. 2008. Physiologically based pharmacokinetic modeling of persistent organic pollutants for lifetime exposure assessment: A

new tool in breast cancer epidemiologic studies. *Environmental Health Perspectives* 116: 886–892.

Verner, M.A., T. Magher, and S. Haddad. 2010. High concentrations of commonly used drugs can inhibit the in vitro glucuronidation of bisphenol A and nonylphenol in rats. *Xenobiotica* 40: 83–92.

Wilkinson, G.R. 1987. Clearance approaches in pharmacology. *Pharmacological Reviews* 39: 1–47.

Zurlinden, T.J., and B. Reisfeld. 2015. Physiologically based modeling of the pharmacokinetics of acetaminophen and its major metabolites in humans using a Bayesian population approach. *European Journal of Drug Metabolism and Pharmacokinetics* 41: 267–280.

Chapter 13
Mixture Experimental Design

Jane Ellen Simmons, Ingvar Eide, Glenn Rice, and Paul Feder

Abstract There is a general consensus, based on a number of surveys and analytic efforts, that mixture study designs have historically been lacking. Although there has been considerable progress over the past decades, further improvement is necessary both in the development and application of experimental designs to yield data suitable for quantitative analytic methods and in the implementation of appropriate statistical analyses. This chapter reviews the state of the science with regard to the experimental and statistical quality of mixture studies. The importance of properly powering mixture experiments is emphasized; in particular when the focus is on the low-dose/low-effect region. Issues with powering defined mixture and complex mixture experiments are explored. Some designs that have proven useful in mixture experimentation are reviewed, including full and fractional factorial designs and statistical mixture designs such as the isobologram and the fixed ratio ray. General considerations are provided that will aid in the development of both experimental design and analysis strategies that address the question (s) being asked.

Disclaimer The views expressed in this chapter are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

J. E. Simmons (✉)
National Health and Environmental Effects Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, Durham, NC, USA
e-mail: Simmons.Jane@epa.gov

I. Eide
Statoil ASA, Trondheim, Norway

G. Rice
Fort Mitchell, KY, USA

P. Feder
Battelle Memorial Institute, Columbus, OH, USA

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_13

Keywords Experimental design · Power · Factorial design · Ray design · Isobologram · Chemical mixtures · Statistical mixture design

13.1 Introduction

The design, conduct, analysis, and interpretation of mixture experiments are a daunting challenge. Designing the "right" experiment is difficult as the number of mixtures to which humans are exposed is essentially infinite. Combining the number of mixtures with the myriad health endpoints and outcomes of interest reveals the true nature of the problem. This brings out two important points. The first is that mixture experiments should focus on mixtures of "concern." Identification of priority mixtures is an area of active investigation; Chaps. 5, 6, and 7 describe innovative methods for intelligent design and prioritization of mixtures. The priority setting process of deciding "which mixture(s) to test" may include consideration of the frequency of human exposure and the number of individuals exposed; the level of concern with regard to health effects on the part of decision makers, health-care providers, risk analysts, risk managers, and stakeholders; the seriousness of the health concern (e.g., pregnancy loss compared to odor threshold exceedance); and the characteristics of the affected population, such as young children or the elderly. The second point is that while there are considerable challenges facing even the most capable experimentalists and statisticians, designing quality mixture experiments, although not always easy, is an achievable goal.

13.2 State of the Science

13.2.1 Limitations of Currently Available Data

There is considerable interest in understanding the likelihood of chemical mixtures exhibiting nonadditive toxicity. Greater-than-additive toxicity can be either beneficial or disadvantageous. For example, there is significant motivation to optimize cancer treatment with "cocktails" of chemotherapeutic agents with greater-than-additive anticarcinogenic effects (Sun et al. 2015) due to the enhanced therapeutic effects. In contrast, greater-than-additive effects of environmental chemicals are generally considered undesirable (Cedergreen 2014). The ability to use the available literature to understand the potential for nonadditive interactions is hampered by issues associated with experimental design and analysis. This is a source of considerable frustration across many disciplines. Chou (2006) succinctly pointed out the dangers inherent in faulty or erroneous claims of greater-than-additive interactions in the clinical environment as such claims are frequently the basis of therapies chosen for patient treatment. Consistent with this, erroneous claims of greater-than-additive toxicity could result in lower allowable exposure limits that do not provide additional public health benefit. Such erroneous claims could be

harmful if, for example, lowered allowable exposure thresholds resulted in substitution with processes or chemicals that are perhaps less studied and either unintended consequences resulted or costs associated with reduced releases were incurred that resulted in no health benefit.

Concerns with regard to the state of the science of mixture studies, in particular with the quality of experimental design, analysis, and interpretation, are longstanding. This section reviews several pertinent examples of such analytic efforts.

13.2.2 Early Assessments of Clinical Use Applications of Pharmaceutical Agents

A particular emphasis for pharmaceutical agents is identification of chemicals that provide greater-than-additive therapeutic benefit when given in combination. Jawetz and Gunnison (1953) delivered a scathing condemnation of the quality of publications investigating the joint toxic action of antimicrobial agents. The strength of their concern is evident in the statement "while it may seem redundant to mention it, the results arrived at by any method must have quantitative validity in terms of the criteria chosen by the investigator. It is regrettable that some reports do not live up to that postulate. It is even more unfortunate that certain articles in scientific publications read like advertisements for specific commercial preparations of antimicrobial drug combinations based on most tenuous evidence."

Noting that several reviews were available that provided summaries of publications and disagreements in conclusions between them, Jawetz and Gunnison (1953) questioned the need for another "summary" review of that type and focused instead on an examination of the possible reasons for disagreement between the claims of different research groups. They looked at five points, several of which are directly relevant to the present discussion: definitions of terms and methods of study. The authors also included interpretations and significance of results, with a particular focus on correlation between laboratory findings and clinical observations. The authors concluded that the varying meanings associated with the terms "synergism" and "antagonism" were partly responsible for the confusion. Additionally, they noted an argument familiar to many current mixture investigators, i.e., the failure to consider the shape, particularly in the lower regions, of the individual-agent dose-response curves. Responses different from control become increasingly difficult to detect as the dose decreases, and differences from control are either not present or not detectable under the experimental design being employed.

Several years later, Veldstra (1956) stated in a review on the topic of "synergism" that "a great number of questions with respect to synergism cannot be answered satisfactorily at the moment" and attributed this, in part, to his perception that "a purposeful analysis has not yet been carried out." The following year, Goldin and Mantel (1957) published a review of combination therapies for neoplasia. Observing a common flaw in experimental design, they stated that the

"underlying" weakness in all the papers reviewed was "the less than comprehensive way in which they are conducted." Goldin and Mantel (1957) characterized coverage of "sufficient" portions of the dose-response curves for the individual chemicals and for the mixture as being of prime importance in understanding the joint toxic action of drugs and noted this deficiency in the studies they reviewed. They brought out this point again, with regard to testing a drug deemed active in combination with a drug deemed "inactive" and the failure to test the "inactive" drug across a sufficient range of doses. The perception or the reality of the issue persisted over the years, as McInnes and Brodie (1988) noticed that suitable study design was seldom employed in investigation of drug interactions.

13.2.3 Early Evaluations of Consumer, Industrial, and Environmental Agents

The emphasis in this section is on combinations of chemicals where the outcome of interest is adverse (e.g., decreased birth weight, increased liver necrosis, decreased cholinesterase activity in serum or brain, increased mortality). This includes naturally occurring agents (e.g., arsenic in water, silica); pharmaceutical agents; mixtures of chemicals from consumer products, agricultural practices, and industrial activities (including those associated with energy production); and other anthropogenic sources of chemicals in the air, water, or soil.

A technical support document developed by the U.S. EPA (1990) in response to the Science Advisory Board review of the 1986 Guidelines for Risk Assessment of Chemical Mixtures (U.S. EPA 1986) provided clear evidence of the historical magnitude of the problem and highlighted the need for improved design and analysis of mixture experiments. A twofold analysis of the available peer-reviewed literature on interactions between and among the chemicals present in chemical mixtures was conducted. To accomplish this, the EPA Mixture Toxicity Database (MixTox, no longer available) was used. The first analysis surveyed the statistical methods used in studies of chemical mixtures. References were deemed relevant if they included both the experimental methods and the resulting data; this was intended to remove presentation abstracts and review articles from further consideration. Of the 331 references in the MixTox database, 307 met these criteria. The authors reported that another 155 studies were included (but did not specify how these studies were selected for inclusion) for a total of 462 studies. Studies were then catalogued as to whether they were (a) mechanistic or descriptive and (b) whether the mixture studied was binary, simple, or complex. A simple mixture was defined as "a mixture containing more than two identifiable components but few enough that mixture toxicity can be adequately characterized by a combination of the component toxicities," and a complex mixture was defined as "a mixture containing so many components that any estimation of its toxicity based on its component toxicities contains too much uncertainty and error to be useful." Studies

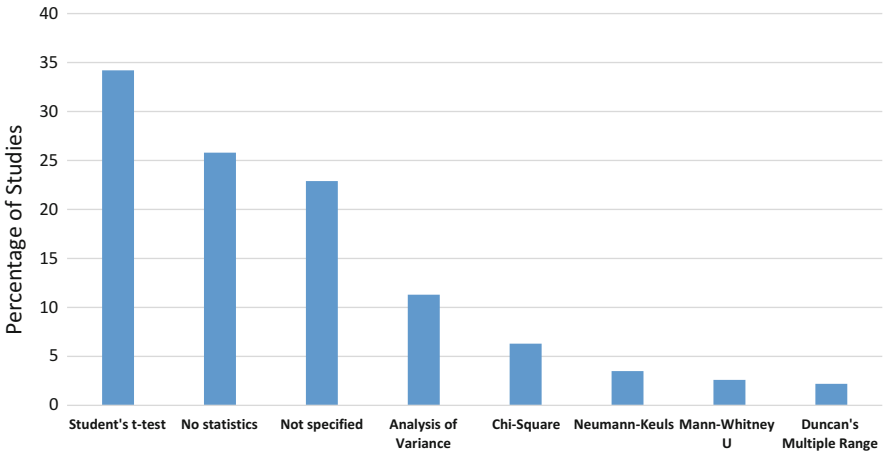

Fig. 13.1 Statistical analysis of mixture data state of the science prior to 1990

fell into more than one category if they included either both binary and simple mixtures or both mechanistic and descriptive elements. Binary mixtures were evaluated in 96.1% of studies, with simple and complex mixtures in 8.7% and 5.2% of studies, respectively. The vast majority of the studies were acute (73%), studying either mortality or the influence of one chemical administered at a nontoxic dose on the toxicity of another chemical. The statistical breakdown was particularly revealing (see Fig. 13.1) with 26% of studies including no statistical analysis and another 23% of studies reporting statistical significance levels (e.g. p values) but not specifying the statistical methods used to derive them. Among the studies that used statistics, student's t-test was the most popular choice, with 34% of researchers employing it.

The second analysis that used the MixTox Database attempted to examine elements of experimental design quality and included a detailed analysis and evaluation of the appropriateness of the statistical method used and the correctness of interpretation of the statistical results. To accomplish these goals, 32 references contained in the MixTox database were randomly selected. Of these 32, two were abstracts and were excluded from further consideration. Of the remaining 30 studies, no statistical analysis appeared to have been conducted in eight (27%), the statistical analysis was unspecified in seven (23%), and quantitative data were not presented in one. It is interesting how closely the percentages for studies with either no- or not-specified statistical analysis mirror those of the population from which the sample was taken. Considering the remaining 14 papers, four (13% of the total) were judged inadequate because baseline controls were not included, and nine (30%) were judged inadequate as the methods were deemed inappropriate. They found the experimental design, the use of statistics, and the conclusion appropriate and justified for only one (3%) of the 30 papers.

13.2.4 More Recent Assessments

Although considerable progress has been made over the past decades in development and application of experimental designs yielding data for quantitative analytic methods, problems still remain. Chou, as recently as 2006, found that there are many common errors associated with declarations of synergy and cited three recurring problems: (1) declaring synergy when A + B is greater than either A alone or B alone, (2) using effect addition in dose ranges where it is not biologically plausible to do so, and (3) analyzing toxicologically similar chemicals with a model designed for toxicologically dissimilar (i.e., independent) chemicals (Chou 2006; see Chap. 9 for limitations of the use of effect summation and discussion of toxicological similarity concepts, which guide the researcher to the use of either dose addition or independent action models). The models of Chou, based on the combination index, have been published and coded in software (CompuSyn©, www.combosyn.com/index.html) that is freely available. It is noteworthy that the articles describing this approach are extremely highly cited, with the article introducing the key concepts of the combination index (Chou and Talalay 1984) cited 4281 times (Web of Science, Web of Science Core Collection, accessed 1-10-2018).

A second noteworthy effort was undertaken by Borgert and colleagues (Borgert et al. 2001) that focused on defined mixture studies, in which dose-response information from the individual chemicals contained in the mixture are used to predict the response of the mixture under an explicit definition of additivity. This predicted response is then compared to the experimentally observed response of the mixture tested at the same mixing ratio and dose levels. The authors described a set of five criteria that can be applied to chemical mixture interaction studies to evaluate the quality of both the data and the interpretations of the data. They envisioned the criteria would be useful to mixture risk assessors in (1) identification of studies that can be used in component-based mixture risk assessments which are, by far, the most commonly used mixture risk assessments (see Chap. 14); (2) determination of studies that are less useful due to deficiencies in either design, interpretation, or both; and (3) application of a weight of evidence approach to evaluation of studies that reach opposing conclusions with regard to either the presence or absence of deviations from additivity or the direction (greater-than-additive, less-than-additive) of detected deviations. In addition to application to mixture risk assessment, the authors declared that meaningful criteria would also serve larger scientific goals and should enable (a) experimentalists to generate data that are more interpretable, (b) interpretations more consistent with the data, and (c) generation of testable hypotheses. Further, the authors stated that "the criteria should provide a basis for better understanding differences between studies and methods of analysis, and should also facilitate understanding the differences between various analysts with respect to their assumptions and conclusions. Application of the criteria in this context would also lead to better risk assessments because the quality of published studies would be strengthened overall and because

Table 13.1 Five criteria[a] of Borgert et al. (2001) for evaluating interaction studies for mixture risk assessment

1. Dose-response curves for the individual components contained in the mixture should be adequately characterized
2. An appropriate "no-interaction" hypothesis should be explicitly stated and used as the basis for assessing interactions
3. Combinations of mixture components should be assessed across a sufficient range of dose combinations to support the goals of the study
4. Formal statistical tests should be used to evaluate whether the response of a mixture dose combination is distinguishable (larger than or smaller than) from the response expected under the "no-interaction" hypothesis
5. Interactions should be assessed at relevant levels of biological organization

[a]Reworded from text of Borgert et al. (2001)

uncertainties could be better quantified according to a replicable standard of measure."

The five criteria are provided in Table 13.1. The first criterion addresses the need to characterize relevant portions of the dose-response curves of the component chemicals. The rationale for this has been clearly explained by many authors, frequently in the context of the "sham experiment" in which the same chemical is combined with itself. Berenbaum (1989) and Greco et al. (1995) provide theoretical discussions of the "sham experiment" concept, and Gennings et al. (1990) describe a combined experimental and statistical assessment of righting reflex in mice following exposure to sodium hexobarbital "combined" with sodium hexobarbital. The second criterion is the requirement that a "no-interaction" hypothesis (i.e., the null hypothesis) be explicitly stated and then used in the evaluation and interpretation of the data. Although it is obvious that a combination of chemicals cannot be declared as deviating from additivity without first establishing what would be expected in the additive state, this is still seen all too frequently. As described briefly by Borgert et al. (2001) and in more detail in Chap. 9, the null hypothesis is formed based on a judgement of whether the chemicals are toxicologically similar or dissimilar.

Criterion 3, the requirement that the mixture be assessed across a sufficient range of mixing ratios and dose levels to support the goals of the study, might appear at first glance to overlap with criterion 1. A key difference is that these two criteria have different focuses; Criterion 1 considers the individual components and Criterion 3 the mixture itself. Borgert et al. (2001) view Criterion 3 as being used to evaluate the interpretations and inferences made and not being used to judge the quality of the data itself. Under the conditions of a full factorial, fractional factorial or response-surface study that covers the entire dose range of interest, global conclusions about the nature of the interactions can be drawn. Under the much more common circumstance of experiments that provide less than full coverage of the entire dose-response range of interest, care must be taken to confine the interpretation to the conditions actually studied. For example, the results of a ray design experiment at a 1:1 mixing ratio should not be used as evidence of the effects

of the combination of the same two chemicals at an environmentally realistic ratio, where the mixing ratio may be far from 1:1. For example, Moser et al. (2012) observed different interactive properties for different mixing ratios when studying the impact of mixing ratio on the joint toxic action of seven carbamates in both pre-weanling and adult male Long-Evans rats. For one mixture, the mixing ratio was based on equitoxic doses; the mixing ratio of the second mixture was based on sales of these pesticides in California. For the three endpoints reported (brain cholinesterase inhibition, red blood cell cholinesterase inhibition, motor activity), the equitoxic mixture appeared dose additive for most endpoints at both ages and only showed marginal deviations from dose additivity for adult brain cholinesterase inhibition (greater-than-additive, one dose) and red blood cell cholinesterase inhibition (less-than-additive, middle doses) in pre-weanling rats. In contrast, for the sales-based mixture, greater-than dose-additive responses were observed in both age groups for all three endpoints.

Criterion 4 focuses on the need to conduct statistical analysis to determine whether the mixture response predicted from the individual-chemical dose-response curves deviates significantly from the experimentally observed mixture response. Borgert et al. (2001) stress that it is important that the statistical model used and the underlying assumptions do not introduce bias that favors one of the outcomes. It seems equally important that bias not be introduced that disfavors one of the outcomes. The need to assure sufficient statistical power is mentioned briefly. The final criterion, Criterion 5, is that interactions should be assessed at relevant levels of biological organization. For this criterion, it is emphasized that for complex biological systems, evaluation at multiple levels of organization may be required to characterize an interaction and understand its significance in organisms or populations.

The authors then apply the five criteria to eight studies; three are judged to satisfy the five criteria and potentially would be useful in risk assessment of the tested mixtures. Of the remaining five studies, one is deemed to fail all five criteria, with the other four failing one or more criteria. It is important to note these criteria are based on the usefulness for risk assessment of mixtures. Thus, the criteria may not be applicable to experiments conducted for purposes other than to provide data and interpretations for risk assessment. In a follow-up effort, Borgert et al. (2005) applied the five criteria to two studies on drug-dietary supplement interactions. In addition, they presented a scoring algorithm, by which a composite score for the study is calculated, with zero being the lowest score and one representing a perfect score for a study where all five criteria were fully satisfied. Both papers (Borgert et al. 2001, 2005) are informative to the risk assessor, who will likely read papers with a more informed eye, and the experimentalist, who will be able to incorporate, as appropriate, aspects of the criteria into their future investigations.

13.2.5 Evolution of the State of the Science

While focusing on a different research question, the work of a Risk Assessment Methodologies Technical Committee established by the ILSI (International Life Sciences Institute) Health and Environmental Sciences Institute provides significant insight into the experimental quality of mixture studies conducted primarily since 1990 (Boobis et al. 2011). This group initially surveyed the literature from 1990 to 2008 and then expanded to earlier studies, identifying mixture studies relating to "synergism" (i.e., the effect of the mixture was greater than additive) in the low-dose region. "Low doses" were defined as those near or at point-of-departure (POD) doses, such as no-observed-adverse-effect levels (NOAELs) or benchmark dose levels (BMDLs), that could be used in derivation of health-based guidance values. The survey had an initial emphasis on chronic exposures; however, as few studies were found, the search was broadened to include studies with shorter exposure periods where nonlethal endpoints were measured. This resulted in the inclusion of acute studies with dose levels the authors characterized as "well above chronic PODs." A number of scientific article databases and search engines were searched, such as PubMed, MEDLINE, TOXLINE, Google Scholar, and Scirus. Additionally, searches were conducted for specific chemicals whose interactions were evaluated as greater-than-additive in either the interaction profiles developed by the Agency for Toxic Substances and Disease Registry (ATSDR) for priority mixtures or the MixTox database developed by the U.S. Environmental Protection Agency. Greater-than-additive combinations of pesticides were extracted also from Carpy et al. (2000). Boobis et al. (2011) performed a number of other search activities including a gray literature investigation they termed "aggressive," contacting "key" scientists in both toxicology and risk assessment, and posting in online discussion communities.

Identified papers were compared against multiple inclusion/exclusion rules:

1. Only mammalian systems were included; others, including reptile and fish, were excluded.
2. Only epidemiology studies with validated exposures were included, whereas epidemiology studies without validated exposures were excluded.
3. Chemical stressors were preferred; noise was given as an example of an excluded stressor.
4. In vivo data were preferred over in vitro/in silico studies.
5. Studies were included where the effect of the mixture conformed to greater-than-additive interactions.
6. Included studies/work were either peer-reviewed or agency produced (e.g., ATSDR, EPA, WHO).

After these inclusion/exclusion criteria were applied, the authors asked if "quality were apparent" (see Boobis et al. (2011), Fig. 1.1); this resulted in identification of 90 papers. These 90 papers were then reviewed and studies excluded if the findings were not novel or if it was deemed that the claim of a greater-than-additive

interaction was not supported by the information provided in the paper. Forty-three papers remained that met the inclusion criteria and were judged to provide novel data on greater-than-additive interactions in mammals. Boobis et al. (2011) then went on to critically examine papers where the magnitude of the greater-than-additive interaction was quantified.

13.2.6 Conclusions Regarding the State of the Science

Early studies of environmental mixtures were overwhelmingly focused on binary combinations; for example, more than 90% of MixTox mixtures were binary. However, the field has begun to include higher-order mixtures ranging from defined mixtures of 5 (Moser et al. 2005; Howdeshell et al., 2015), 6 (Hass et al. 2017), 7 (Moser et al. 2012; Rider et al. 2008), 10 (Rider et al. 2010), and 18 chemicals (Crofton et al. 2005) in in vivo studies and 12 (Hadrup et al. 2013), 17 (Ermler et al. 2011), 21 (Scholze et al. 2014), and 30 (Orton et al. 2014) chemicals in in vitro studies. In these studies, dose/concentration-response relationships determined for the component chemicals were used to develop the estimated effect of the mixture under "additivity"; this predicted mixture effect was then compared to the experimentally observed effect of the mixture. The early investigations of both pharmaceutical and environmental mixtures provided clear evidence that many studies were flawed. Common problems included lack of clearly articulated null and alternative hypotheses, deficient experimental designs, either no statistical analysis or application of inappropriate statistical analyses, and drawing conclusions unsupported by the data. These earlier quality reviews provide insight into the long-standing nature of the problem and indicate that these concerns are not limited to one type of mixture (e.g., only pharmaceutical mixtures or only industrial or environmental mixtures). It is also important to note that it is not uncommon for even highly qualified statisticians to disagree on the appropriateness of a particular mixture design and statistical analysis. It is very encouraging that of the 90 studies that met inclusion/rejection criteria that had no bearing on the quality of the work, almost half (43 out of 90) passed the quality review of Boobis et al. (2011). The number of quality studies is almost certainly higher as many of the exclusion criteria did not bear on the quality of the work, and papers that met the inclusion criteria also had to pass a review of the novelty of the findings. Thus, comparing the results of Boobis et al. (2011) to the findings of the EPA review of studies in the MixTox database, substantial progress has been made with regard to the quality of mixture studies. It is noteworthy that while mixture researchers have been grappling with issues of study quality for a very long time, the quality and reproducibility of scientific investigations in general have been a topic of considerable interest in the scientific community. George et al. (2015) note that the increasing concern about research reproducibility is not limited to one field of study. Reproducibility issues have been raised within diverse scientific disciplines, including "omics" (Ioannidis

and Khoury 2011), computational science (Peng 2011), field biology (Ryan 2011), and preclinical studies for cancer therapeutics (Begley and Ellis 2012).

13.3 Power

13.3.1 Introduction to Power

Within the context of statistical analysis, the power of a particular analysis or hypothesis test is defined as the probability of making the correct decision if the alternative hypothesis, rather than the null hypothesis, is true. Thus, the power of a hypothesis test is the probability of rejecting the null hypothesis, H_0, when the alternative hypothesis, H_A, is the hypothesis that is "true." Two types of errors may occur: Type I and Type II. Type I error is sometimes called a false positive, as it occurs when the null hypothesis of "no statistically detectable effect" is "true," but the null hypothesis is rejected. Type II error, sometimes called a false negative, occurs when the null hypothesis of "no statistically detectable" effect is false, but the null hypothesis is not rejected. Power and the level of significance impact the risks of Type I and Type II errors, with the risk of Type I and II errors being inversely related. With each hypothesis test, both Type I and Type II errors are possible. The probability of a Type I error is α, the level of significance set by the investigator. Typical α levels of 0.05 and 0.01 indicate a 5% and a 1% chance, respectively, that the null hypothesis is incorrectly rejected, i.e., Type I error. The probability of a Type II error is β, with power defined as $1\text{-}\beta$. Traditionally, power of 0.80 (also referred to as 80% power) is considered adequate. For mixture experiments, the typical situation appears to be underpowered studies, but it is noteworthy that "overpowering" is also possible. An undersized/underpowered study will not yield meaningful or useful results and wastes resources. An oversized/overpowered study is more likely to yield statistical significance for effect size changes that are not biologically or toxicologically meaningful and also wastes resources as more are used than are necessary.

Although typical levels of Type 1 error (0.05, 0.01) and Type 2 error (0.20) (1 – power 0.80) are often taken as convention, they are not always most appropriate. Clinical acceptability of error levels depends on the consequences of the errors. For example, for a screening test for a disease such as tuberculosis, power of 0.80 corresponds to a false-negative rate of 0.20, which would likely be deemed unacceptable. To increase the likelihood of "true" cases being detected, screening tests often tolerate increased false-positive rates to gain reductions in the false-negative rates. For tuberculosis screens, if the initial test is positive, the person is assumed to be more likely to have the disease, and a more sensitive diagnostic test is carried out. However if the initial test is negative, if there is any possibility that the person was exposed, the test is repeated several weeks later to reduce the possibility of a false-negative result. As another example, mammograms have sensitivity (power) of approximately 0.70 to 0.80. To increase the sensitivity, it is recommended that

the test be repeated every year. This leads to increased false-positive rates. According to Cancer.gov "...the test also comes with a fairly high false positive rate: half of women who get annual mammograms each year for 10 years in the United States will experience a false positive result...." However this is considered to be an acceptable trade-off for reducing the chances of missing a true positive result.

It is interesting that the opposite is true for drug screens and for other tests that have legal consequences. An individual is considered innocent until proven guilty. Operationally this means that a drug screen needs to have a very low false-positive rate even at the expense of increasing the false-negative rate. For example, Peace et al. (2000) in Table III report the overall specificity level in donor specimens for a "Rapid Drug Screen" of 96.1% (corresponding to a false-positive rate of 0.039). By contrast the overall sensitivity level is 81.8% (corresponding to a false-negative rate of 0.182).

Similar considerations would apply to pollution sites to identify pollution hotspots in need of remediation. A trade-off arises between an increased false-positive rate which could increase cost and reduce the area that could be screened and an increased false-negative rate which could result in toxic materials remaining unidentified, with adverse health consequences.

To assess the power of a hypothesis test, the response(s) being considered, the response variability, the nature of the model to be fitted, the null and alternative hypotheses, and the statistical strategy to be implemented need to be specified quantitatively. Clarity in specification of both the null and alternative hypotheses is essential. A typical null hypothesis might be "deviation from dose additivity, as defined according to _____ (here the researcher inserts the reference for their definition of dose additivity) is not statistically detectable," with the corresponding alternative hypothesis being "deviation from dose additivity is statistically detectable." Although these may be the most commonly encountered null and alternative hypotheses, other hypotheses are credible and likely to be encountered. Most important is clear specification of the null and alternative hypotheses. Under the null hypothesis of additivity, dose-response relationships are fitted to each component. The expected response for the joint doses and the variability in this estimate are determined from the individual component responses under the additivity hypothesis being evaluated. This is compared to the experimentally observed response and its associated variability, based on either an unsmoothed response or on a smoothed model-based response. The deviations that can be detected with specified power are determined.

When carrying out such comparisons to assess additivity, it is essential to distinguish between statistical significance and toxicological importance. (Note that the "type" of additivity is not specified here as the concepts apply equally to additivity models based on toxicological similarity and those based on independent action). Statistical significance is a mathematical concept that specifies differences in responses that can be detected with specified power based on the amount of data available, its variability, the test design used, and the statistical test carried out. In contrast, toxicological importance is a biological concept. It pertains to the extent of

difference in response that is important from a health or toxicological standpoint (e.g., a decrease in body weight that indicates declining health or a change in hormone levels that are known to be associated with "downstream" health effects). If there are insufficient amounts of test data, if the data are too variable, or if data collection efforts are inappropriately allocated, the test may be "underpowered." In this case, toxicologically meaningful deviations from additivity may not be statistically significant, because of low power. If power calculations suggest that such is the case, (i.e., the study is underpowered to detect statistically significant deviations from additivity), then even though the hypothesis of additivity is not rejected, it cannot be concluded that important deviations from additivity do not exist. It can only be stipulated that the test is not powerful enough to detect toxicologically important additivity deviations, if present. The opposite situation occurs if the test is "overpowered." In that case, toxicologically "trivial" deviations from additivity may be highly statistically significant because the large amounts of data allow statistically significant deviations from additivity to be detected that are not toxicologically meaningful. If power calculations suggest that such is the case, then even though the hypothesis of additivity is rejected, it cannot be concluded that toxicologically important deviations exist. What can be said with certainty is that the test is more than powerful enough to detect small effects; these small deviations from additivity should be acknowledged along with the interpretation of the study team with regard to their toxicological meaning or lack thereof. In either case, both notions, of statistical significance and toxicological importance, need to be considered together. The objective when designing a study is to choose sample sizes, large enough to be able to detect toxicologically meaningful deviations with high statistical power but not so large as to detect small, toxicologically unimportant deviations. It is important to report estimates of observed effect size, in addition to statistical significance/nonsignificance. This helps indicate whether the study is underpowered, overpowered, or properly powered to meet toxicological objectives.

13.3.2 Importance of Power to Study of Low-Dose/Low-Effect Region

Researchers studying mixtures (whether defined or complex) are increasingly encouraged to examine the low-effect/low-dose region where having sufficient statistical power to detect effects, when present, becomes increasingly important. In 2002, both the Expert Working Group, formed in response to a call by the Society of Toxicology and the Society for Environmental Toxicology and Chemistry (Teuschler et al. 2002), and the EPA's Four Lab Steering Committee (Simmons et al 2002) emphasized the importance of the low-dose/low-effect region. An organizing principle used by EPA's Four Lab team for their experimental design was that "mixtures studies should be conducted, to the extent feasible, in the low-dose region of health effects dose-response curves" (Simmons et al. 2002).

The Expert Working Group noted that "toxicology experiments on whole mixtures or mixture components should include doses at or below the no observed (adverse) effect levels (NOELs/NOAELs) for individual mixture components," further making the strong statement that "experimental paradigms characterizing only the interactions of chemicals at high doses relative to actual environmental exposures will not provide the necessary data to support scientifically informed health policy decisions" (Teuschler et al. 2002). A summary of the outcomes of the 2005 Contemporary Concepts in Toxicology workshop, "Charting the Future: Building the Scientific Foundation for Mixtures Risk Assessment," identified the need to rethink risk assessments for chemical mixtures, in particular how to add the respective risks of low concentrations/dose levels of chemicals in a mixture, and called for identification of the data requirements for improved assessments (Mason et al. 2007). The workshop participants agreed that the development of useful experimental data sets was both important and technically challenging. Risk assessors and experimental scientists relayed their frustrations with the use of high-dose studies to estimate the risks posed by exposure to actual environmental concentrations and mixtures. As a result, the workshop encouraged development of experimental data sets to improve low-dose risk evaluations of mixtures (Mason et al. 2007). Before (Simmons, 1995) and since then, agencies and advisory groups (e.g., IGHRC 2008; Kortenkamp et al. 2009) have either noted the difficulty in assessing the risks resulting from low-level exposure to mixtures from high-dose experimental data or have made calls for efforts in the low-dose/low-effect region.

The importance of considering power is increased when studies are not conducted according to accepted guideline protocols. Guideline studies typically include considerations of power (e.g., see Crosier 1984; U.S. EPA 1998; National Toxicology Program standard study protocols found at https://ntp.niehs.nih.gov/testing/types/cartox/protocols/perinatal/index.html; and Organization for Economic Cooperation and Development (OECD) Guidelines for the Testing of Chemicals found at http://www.oecd-ilibrary.org/content/package/chem_guide_pkg-en). The sample size per treatment group and group placement/dosage selection with regard to the regions of the dose-response curve may not be appropriate for studies of chemical mixtures, in particular when assessing the low-effect/low-dose region. As important as power considerations are for mixture studies, there are few examples of power being considered explicitly as an element of study design.

Statistical power can be increased through consideration of the number of experimental units (e.g., rodents for an in vivo study, number of wells for an in vitro study), the number of dose groups, selection of dose levels, assignment of experimental units to dose groups, sources of variability, and the use of blocking designs to reduce variability of comparisons (Cochran and Cox 1992). An additional important element is the degree of departure from additivity (i.e., the effect size) specified in the alternative hypothesis, where the null hypotheses is consistent with the type of additivity specified by the investigator. The smaller the difference the investigator desires to detect between the experimentally observed mixture effect and the predicted mixture effect under an assumption of additivity, the greater the number of experimental units per treatment group will be required;

i.e., holding power constant, the sample size/group to detect a 10% difference between the predicted and observed mixture effect will be 16 times larger than the sample size/group required to detect a 40% difference.

13.3.3 Powering Defined Mixture Experiments

Given the importance of properly powering experiments to avoid a false negative (failing to detect a nonadditive interaction due to insufficient statistical power), relatively little research has been conducted in this arena. Several collaborative efforts between statisticians and toxicologists are reported here.

An example of how considerations of power can influence the experimental design for defined mixture studies is provided by a collaborative effort between Virginia Commonwealth University statisticians and U.S. Environmental Protection Agency toxicologists (Coffey et al. 2005). The purpose of the investigation was assessment of the impact on power of both mixture dosage and allocation of experimental units within the dose groups. The endpoint was inhibition of cholinesterase activity in erythrocytes of adult, male Long-Evans rats administered an acute oral dose of either one of two pesticides or the binary pesticide mixture. A particular advantage was the availability of both individual chemical and binary mixture data for the effects of chlorpyrifos alone, carbaryl alone, and a 2:1 mixture of chlorpyrifos-carbaryl on cholinesterase activity in red blood cells (Gordon et al. 2006). It is important to note that inferences about dose additivity and associated power were restricted to combinations along the 2:1 ratio ray. Although individual chemical data may be available, mixture data in advance of the experiment being designed are rare but, as can be seen in this case study, allow for extremely valuable insights into the usefulness of various designs. Similar to the majority of in vivo studies that are constrained due to technical and resource limitations, the design under construction was constrained by a maximum of 50 animals in the study. Because nonadditive interactions may result in greater-than or less-than expected toxicity, experimental designs were optimized for these opposing situations and for "mixed interactions," in which greater-than-additive interactions occur in one mixture dose region, and less-than-additive interactions occur in another mixture dose region. A D-optimal design strategy, based on a specified parametric dose-response model, was selected in which the design is optimized to decrease the variability associated with the model parameters. (Note: D-optimal designs are computer-aided designs, that result in generated experimental designs with minimized variance of the model parameter estimates under model assumptions (NIST/SEMATECH, accessed 1-10-18). The definition of dose additivity provided by Berenbaum (1985, 1989) (see Chap. 9), based on Loewe and Muischnek (1926), was adopted by the authors.

For detection of a 33% change in the dosage resulting in a 20% effect level (ED_{20}) based on either greater-than-dose-additive or less-than-dose-additive interactions, the authors found that equal spacing of dose groups combined with equal

Table 13.2 D-optimal experimental designs for detection of departure from dose additivity

Study design	Mixture dosage (mg/kg)	N/group	Power[a]	Variance[b]
	Detection of greater-than-additive toxicity			
A	0, 5, 10, 15, 20	10, 10, 10, 10, 10	0.08	72,720
B	0, 3, 6, 9, 12	10, 10, 10, 10, 10	0.39	1.50
C	0, 0.74, 2.1, 14.0, 17.9	8, 13, 15, 7, 7	0.64	0.00054
D	0, 0.74, 2.46, 16.4	7, 11, 26, 6	0.74	0.00044
	Detection of less-than-additive toxicity			
A	0, 5, 10, 15, 20	10, 10, 10, 10, 10	0.05	0.32
B	0, 3, 6, 9, 12	10, 10, 10, 10, 10	0.08	0.0028
E	0, 1.57, 3.8, 4.2, 16.9	7, 11, 12, 13, 7	0.26	0.00032
F	0, 1.57, 4.2, 18.3	7, 13, 23, 7	0.27	0.00030
	Detection of dose-dependent mixed nonadditivity[c]			
A	0, 5, 10, 15, 20	10, 10, 10, 10, 10	0.10	773.7
G	0, 4, 8, 12, 16	10, 10, 10, 10, 10	0.14	5.02
H	0, 1.75, 3.2, 11.6, 13.4	9, 11, 15, 7, 8	0.78	0.00068
I	0, 1.75, 3.2, 12.0	9, 15, 19, 7	0.84	0.00056

Adapted from Coffey et al. (2005)

[a]Powered for the effect size of a 33% change in the experimentally observed mixture ED_{20} relative to the predicted mixture ED_{20} estimated from the individual chemical dose-response curves under an assumption of dose additivity

[b]The variance of the model parameters, in this case the slope, and intercept parameters of the statistical additivity model

[c]The dose-dependent mixed nonadditivity model was for the case where less-than-additive interactions are observed at lower doses and greater-than-additive interactions are observed at higher doses

allocation of animals to dose groups resulted in low power. Table 13.2 provides examples of the designs considered by the authors. Comparing Designs A and B with equally spaced doses and equal allocation of animals to groups, it can be seen that neither resulted in acceptable power. For less-than-additive effects, power was quite low for both Designs A and B; for greater-than-additive effects, more power was achieved for Design B (0.39) than for Design A (0.08). For the less-than-additive design options, inequality in spacing of both mixture doses and in allocation of animals to groups (Designs E and F) achieved small increases in power; however, power never exceeded 0.27. In contrast, for the greater-than-additive design options, both designs shown in Table 13.2 with unequal dose spacing and unequal allocation of animals to groups (Designs C and D) achieved higher power, with the four dose group alternative (Design D) achieving almost 0.80 power. Unequal dose spacing and unequal allocation of animals to dose groups yielded 0.78 power for Design H and 0.84 power for Design I, although the equally spaced/allocation designs (A and G) had poor power. It seems unlikely that any reasonable alteration in experimental design would result in the less-than additive design achieving 0.80 power. Although not discussed by the authors (Coffey et al. 2005), several potentially feasible design adjustments may have resulted in achieving 0.80 power for the greater-than-additive design. If an extra five animals (10%

increase in total experimental sample size) would increase power to 0.80, the experimentalists may have been able to accommodate the increased sample size. If the rate limiting factor were the number of animals that could be handled on the day of tissue collection, perhaps a blocked design, in which a second cohort of animals increased the overall experimental sample size, may result in the desired power. Or, perhaps the investigators could consider whether the goals of the study would be compromised if the effect size were increased. If one or more of these alternatives were acceptable to the experimentalists and the goals of the project, the study team could determine: how many animals would need to be added to the study, whether blocking would be required to accommodate the extra animals, and the effect size increase necessary to achieve 0.80 power (e.g., compared to a 33% change, would a 40% change improve power?).

An example of sample size and power calculations for conducting tests of additivity on parameter estimates is found in Casey et al. (2006). The experimental design was a fixed ratio ray of a mixture of five organophosphate (OP) pesticides at a mixing ratio of 0.040 acephate, 0.002 diazinon, 0.031 chlorpyrifos, 0.825 dimethoate, and 0.102 malathion. Dose-response curves for the individual chemicals and the five OP mixtures were developed in adult, male Long-Evans rats. In addition to the control group, six mixture dose groups were included in the experiment design, to ensure coverage of the dose-response space, with total mixture dose levels ranging from 10 to 450 mg/kg (for a total of seven experimental groups). Rats were dosed acutely by oral gavage; assessment of motor activity, measured as the total activity count for the duration of the session, began 15 min after dosing and lasted 20 min. Employing a threshold dose additivity model based on the Berenbaum definition of dose addition (Berenbaum 1985, 1989) and parameter estimates derived from the individual chemical dose-response curves, an additivity model for the five OP mixtures was constructed. Sample size and power calculations were made for two different types of changes in the mixture dose-response function that might occur as the result of nonadditive interactions: change in threshold and change in slope. For illustration and convenience in simulation, the two situations were dealt with as isolated events – i.e., it was assumed that the change in threshold occurred without a change in slope and that the change in slope occurred without a change in threshold. In both cases, the effect size remained the same, a mean total motor activity of 125 activity counts. First, the authors considered the situation where a change in the intercept of the mixture dose-response curve resulted in a 25% shift to the left in the mean total motor activity, at an effect level of 125 activity counts. The shift to the left indicates that the effect (in this case 125 activity counts) is achieved at a lower dosage of the mixture (see Fig. 13.2, panel A). Under this scenario, equal allocation of 18 animals/group resulted in 0.77 power to detect a 25% shift to the left in mean total activity, where the null hypothesis is no change in threshold, and the alternative is a threshold change resulting in a 25% activity shift. Unequal allocation, as shown in Table 13.3, resulted in 80% power.

The second situation considered was a change in slope. In this scenario (illustrated in panel B of Fig. 13.2), the investigators determined the change in slope

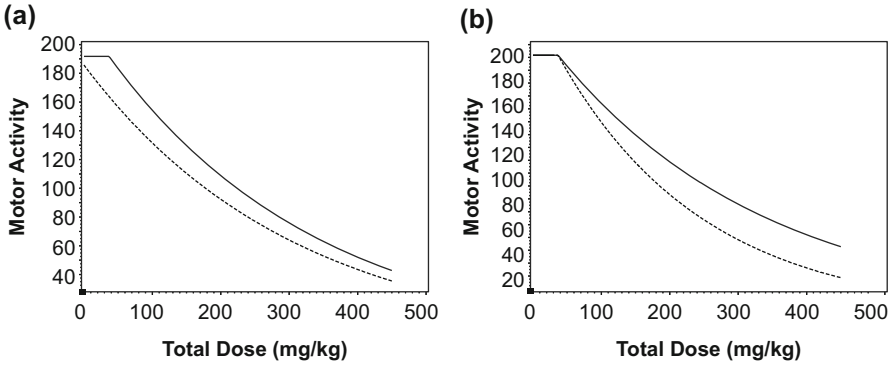

Fig. 13.2 Conducting tests of additivity on model parameter estimates (from Casey et al. 2006)

Table 13.3 Powering for changes in model parameter estimates

	Shift in threshold resulting in a 25% shift to the left at 125 activity counts	Shift in slope resulting in a 25% shift to the left at 125 activity counts
Total mixture dose	N/group	N/group
0.0	30	1
10	31	1
55	14	11
100	19	3
200	20	39
300	13	44
450	1	51
Total N[a]	128	150
Estimated power	0.800	0.799
Estimated power, equal allocation	0.770	0.648

From Casey et al. (2006)

[a]The total *N* is the number of animals required to achieve ~80% power under unequal allocation, as for this example, unequal allocation of the same total number of animals resulted in more power

required for a 25% shift to the left at a motor activity count of 125, where the null hypothesis is no change in slope, and the alternative hypothesis is a change in slope sufficient to result in a 25% shift to the left in mean total activity. With unequal allocations, weighted toward the upper end of the dose range, power of 79.9% was achieved (see Table 13.3). Under approximately equal allocation of the same number of animals, power was estimated to be 64.8%. This study assumed that the mixture dose levels were fixed, could not be altered, and still meet the objectives of the study. Such a scenario might occur where the study is intended to match exposure scenarios that are known and of interest with regard to additivity;

examples would be consumer use of a commercial product containing a blend of pesticides present in invariant ratios or patients infected with antibiotic-resistant bacteria being dosed with multiple antimicrobial agents at set dose levels at fixed time intervals. The results illustrate that the allocation of subjects to treatment groups differs based on the type of change expected in the experimentally derived mixture dose-response curve and that reasonable power can be achieved even when the actual dose levels are fixed. Without preliminary experimental data or information that enables the study team to develop expectation(s) of the type of change expected in the mixture dose-response function, the experimentalist may wish to examine whether increasing the total sample size, sufficient to increase power to 80% under both scenarios (change in slope, change in threshold), is feasible. If so, running the experiment in multiple blocks, without confounding, could be explored to accommodate the extra animals.

13.3.4 Powering Complex Mixtures

Consideration of power appears to be even more rarely mentioned in complex mixture toxicology studies than in defined mixture studies. Complex mixture experiments share the constraints outlined above for defined mixture studies and are further constrained by limitations of sample volume. Studies that strive to use environmentally realistic complex mixtures have further practical constraints on the sample concentration factors that can be achieved as well as the sample volume that can be produced. Dingus et al. (2011) described methodology for calculation of statistical power for nonindependent observations for a multi-generational rat reproductive/developmental bioassay being conducted as part of the EPA's Four Lab study (Simmons et al. 2002, 2008; Pressman et al. 2010; Narotsky et al. 2013).

It was recognized (Simmons et al. 2002, 2008) that the sample sizes typically associated with traditional experimental designs would likely not be adequate to detect effects resulting from exposure to environmental exposure levels of the complex mixtures formed during oxidant disinfection of water. As a multi-generational rat bioassay was being conducted, the extent that water could be concentrated while conserving the volatile and nonvolatile organic chemicals initially present in the water and retaining palatability, so that the animals would not voluntarily restrict their consumption of the water concentrates (Simmons et al. 1994), were important design considerations. Although a dose-response study was desirable, this objective was not possible due to the limited amount of water concentrate available. The study team decided it was better to have one well-powered treatment group compared to the control than to have multiple underpowered treatment groups, where it would not be possible to discern if failure to detect was due to "there being no effect to detect" or due to the lack of statistical power resulting in a false negative (Type II error).

Pup weight at birth and prenatal loss were the primary endpoints for the complex mixture multi-generational bioassay (Narotsky et al. 2013). Data for these

endpoints were available from a previous study (Narotsky et al. 2008) in which pregnant Sprague-Dawley rats were exposed during gestation days 6–16 to chlorinated water concentrated ~130-fold by reverse osmosis membrane techniques with volatile DBPs lost during the concentration process spiked back into the concentrates. Data from Narotsky et al. (2008) were used to establish effect sizes for pup weight and prenatal loss. Power calculations for pup weight were based on a two-sided test. In contrast, as it was deemed biologically implausible for exposure to water concentrates to decrease the level of prenatal loss in the exposed group, a one-sided hypothesis was constructed and powered. As statistical tests were to be performed independently for these two endpoints, the Type I error rate was set at 0.05 for each of the two tests. For both endpoints, the experimental unit was the litter.

For pup weight, each pup within a litter represented a repeated measurement. Weights of pups from different litters were assumed to be independent, i.e., uncorrelated, while a constant correlation was assumed between pups from the same litter. Expected control group and treatment group means, variances and intra-litter correlations, were estimated from the earlier study (Narotsky et al. 2008). Although not biologically realistic, a simplifying approximation was an equal number of repeated measures per experimental unit, so it was assumed that an equal number of live pups would be produced in each litter. Several designs were considered, examining power and technical constraints with different approaches that varied the number of study blocks, the number of dams in the parental generation, and the number of females per litter selected for breeding in the next generation. Greater than 0.80 power would be achieved for pup weight for each design, except when the unlikely assumption was made that the intra-litter correlation would be zero (Dingus et al. 2011).

Although any of the considered designs would meet the desired objectives with regard to pup weight, none of the designs appeared to provide sufficient power for prenatal loss. A two-block design, with unequal allocation of first-generation dams to control (40 dams) and treatment (60 dams) groups yielded the highest power estimates for prenatal loss, providing 0.57 power when assuming 13 implants per dam. Despite the continued concern with regard to inadequate power for prenatal loss, the decision was made to proceed with the design that afforded the best power estimate for this endpoint. Interestingly, when analyzing the prenatal data from the experiment, it was noticed that the original power calculations (Dingus et al. 2011) used untransformed prenatal data, rather than arcsine-square root transformed data, the transformation that was used in the statistical analysis of the collected data (Narotsky et al. 2013). A retrospective analysis was conducted to detect the same prespecified effect size (Thomas 1997), with arcsine square-root transformed data. The result showed that the study design had more than 99% power to detect prenatal loss.

13.3.5 Conclusions Regarding Power

Power is frequently neglected in toxicological investigations of mixtures. This may be a result of many toxicologists having experience with guideline studies, where frequently the number of dose groups and the number of animals per dose group are specified, as is the top dose level. Consideration of power is "built into" these types of study designs. Having specific designs used coherently by multiple laboratories enables comparisons between chemicals that are not confounded by differences in study design and analysis schemes. However, as discussed above, these types of "guideline" experimental studies are generally not feasible for complex mixtures and often are less than desirable for defined mixture studies. For both defined mixture and complex mixture toxicological experiments, the increasing call to consider the low-effect/low-dose region leads to experimental designs tailored to specific mixtures and specific hypotheses and the need to power them properly. Achieving adequate power in mixture studies often requires consideration and use of non-equal group sizes. For experiments examining an assumption of additivity, designing the experiment to achieve sufficient power requires not only specification of the form of additivity (concentration addition or independent action) but whether the investigator is interested in greater-than-additive or less-than-additive interactions. In addition to clear specification of the null and alternative hypotheses, the importance of preliminary or pilot data cannot be overemphasized.

13.4 Types of Experimental Designs

13.4.1 Defined and Complex Mixtures

Defined mixtures are often constructed by experimentalists under controlled conditions to be used for various types of testing, such as toxicological, medical, physical, or other chemical applications. Mixtures of pharmaceuticals as administered to patients are an example of a defined mixture, with very tightly controlled constituent proportions. In contrast, the constituent proportions of mixtures of pharmaceuticals present in wastewaters or surface waters are much more variable. Since the constituents of experimentally constructed defined mixtures are well controlled, studies can be designed to determine the relative contributions of individual constituents (main effects) and specified combinations of constituents (interactions) to the overall mixture effect. This is accomplished by varying the mixture constituent proportions as specified by the design matrix and comparing the effects of different mixtures. Sometimes complex mixtures are treated as if they were defined mixtures by focusing on a specified subset of components and characterizing this subset mixture. Effects due to individual constituents or combinations of constituents can be determined from these subset defined mixtures if the variation in component proportions among the data set of mixtures is sufficiently

large and if the components selected for analysis account for nearly all the effect under investigation. Nearly all the designed experiments involving mixtures reported in the literature pertain to defined mixtures.

Examples of complex mixtures are effluents from wastewater treatment plants, industrial effluents, seepage from hazardous waste disposal sites, emissions from the burning of wood or coal, and emissions from gasoline-powered and diesel-powered vehicles. These mixtures may exhibit apparently random variation across neighboring water treatment or industrial plants or across time within plants (e.g. wastewater treatment, industrial facilities) or across different sites (e.g. hazardous waste sites) or across time within the same location (e.g. wastewater treatment, industrial facilities, hazardous waste sites), due to burning of different varieties of fuel or from different types of engines and running conditions (gasoline-powered and diesel-powered vehicles). Designing tests and carrying out inferences with complex mixtures is more difficult than with defined mixtures.

13.4.2 Designs for Defined Mixtures

The composition of mixtures made in the laboratory may be based on statistical experimental design (e.g., Groten et al. 1991; Johnsen et al. 1994; Eide 1996; Eide and Zahlsen 1996). Factorial designs (Box et al. 2005) imply complete independence between the variables (orthogonality). Factorial designs may be expanded to response surface designs to describe nonlinearity. Fractionated factorial designs and complete and incomplete block designs (Box et al. 2005; Cochran and Cox 1992) are alternatives with many variables. Other classes of designs are ray designs (Stork et al. 2007) and statistical mixture designs (Cornell 2002), the latter if the variables are proportions of a blend. Statistical experimental design is a structured and systematic way of combining a number of variables with as few combinations as needed for meeting statistical power objectives and simultaneously maximizing the variation in mixture composition. In mixture toxicology, the purpose is to make a number of mixtures with variation in both composition and concentration of predefined chemical compounds (variables).

13.4.2.1 Factorial Designs

Among factorial designs, the two-level factorial design is the simplest and the most commonly used, implying complete independence between the variables (orthogonality). Figure 13.3 illustrates a two-level full factorial design with three chemicals (termed a 2^3 design, with the first number indicating the number of dose levels and the second number the number of chemicals in the mixture). Each chemical (variable) is varied between a low and a high value (−1,1) giving eight different combinations (mixtures). In addition, the center point is often added to the design with all three variables at an intermediate (average) level. The center point

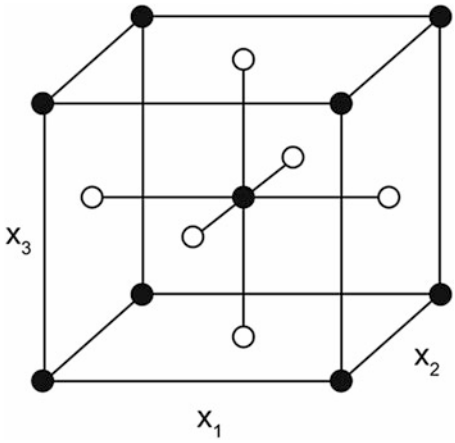

Fig. 13.3 Illustration of factorial design with three variables at two levels plus center point (black dots) for linear models and additional combinations (open circles) for nonlinear models

implies that all variables are at three levels, providing the ability to identify departures from linear dose trends. The center point may also be valuable for measuring repeatability in the tests or analyses to estimate pure error. This can be achieved by making and testing, for example, three center point mixtures independently. However, to describe nonlinearity individually for each chemical, more observations are required, e.g., by selecting combinations of variables such that all chemicals are tested at five levels instead of three (response surface design). This involves samples representing the center of each surface as shown in the central composite face design in Fig. 13.3. Alternatively, the full factorial design may be expanded with "star points" extending with a factor outward from the center of each surface. To have some desirable mathematical properties, it is recommended that this factor is taken as [number of factorial runs]$^{1/4}$. Based on a full factorial 2^3 design with 8 runs, the factor becomes $[8]^{1/4} = 1.68$. The six star points would therefore be at a distance ± 1.68 along the axis for each chemical. Statistical experimental design is described in more detail by Box et al. (2005).

The design matrix corresponding to Fig. 13.3 can be drawn up manually. In addition, there are options for statistical design in most software for multivariate data analysis. The use of software is particularly useful with many variables and with designs with more than two levels and also when there are constraints in the possible combinations. However, for practical reasons, the application of complete statistical experimental designs is only possible with a limited number of chemicals (variables). For example, a full factorial design at two levels requires a significant increase in the number of dose combinations as the number of chemicals increases:

3 chemicals: 2^3 design = 8 mixtures (+ center point + replicates)
5 chemicals: 2^5 design = 32 mixtures (+ center point + replicates)
10 chemicals: 2^{10} design = 1024 mixtures (+ center point + replicates)

An alternative to the full factorial design is the fractional factorial design, which is a structured way of selecting a fraction of the mixtures from a full factorial design

and testing only those with respect to toxicity. The disadvantage is some confounding, as described by Box et al. (2005).

Cochran and Cox (1992, Chapter 6) discuss and present tables of factorial designs run in blocks of manageable size and the associated confounding structure of effects with blocks. The blocks are designed so that main effects and low-order interactions (e.g., two-factor interactions) can be estimated independently of block effects. For example, for a 2^3 design, Cochran and Cox present a plan that includes two blocks of four test runs each (see Plan 6.1, Cochran and Cox 1992). The blocks are organized so that only the three-factor interaction is confounded with block effects and all the main effects, and two-factor interactions are determined independent of block effects. As a more complicated example, a 2^6 design (six chemicals, each at two-dose levels) has 64 combinations. This may be too many dose groups to test at one time. Cochran and Cox present a plan to test the 64 dose combinations in four blocks of 16 dose groups per block (see Plan 6.3, Cochran and Cox 1992). The blocks are constructed so that only four-factor interactions are confounded with block effects. The main effects and two-factor and three-factor interactions are determined independent of block effects.

An advantage with factorial design of defined mixtures is that the mixture composition can be controlled, and the different variables can be combined as desired, to estimate the desired effects with predicted precision. There are always more observations than test conditions (i.e. dose combinations), and regression analysis may provide empirical models with linear, interaction (cross), and square terms (Box et al. 2005) describing cause-effect relationships and not just correlations.

A full factorial design (2^3) was used in the evaluation of mixtures of three polychlorinated organic compounds followed by in vitro toxicity testing (Søfteland et al. 2009). The design matrix (**X**) corresponds directly to the eight corners and the center point in Fig. 13.3. An example with a central composite face design (i.e., response surface design) corresponding to all points in Fig. 13.3 was used by Figueiredo et al. (2015). The data from both of these studies were interpreted by partial least squares regression (see Chap. 15 for a discussion of PLS).

A fractional factorial design was used to combine five fractions of produced water in various proportions prior to toxicity testing (Johnsen et al. 1994). Fractional factorial design was also used in a study in which organic extracts of diesel exhaust particles were spiked with four individual PAHs prior to mutagenicity testing (Bostrøm et al. 1998). The purpose was to evaluate possible interactions between the individual PAHs and the organic extract. These two approaches described above make it possible to treat the complex mixtures as simple mixtures. However, neither of these approaches gives information about all the individual compounds in the complex mixture.

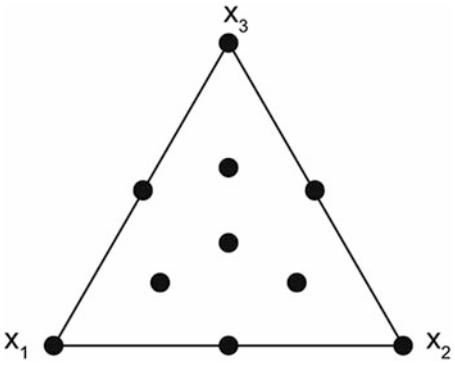

Fig. 13.4 Illustration of mixture design with three variables

13.4.2.2 Statistical Mixture Design

Statistical mixture design (Crosier 1984; Cornell 2002), rather than factorial design, may be preferred or required when the variables are proportions of a blend. A statistical mixture design for three chemicals is illustrated by the triangle (simplex) shown in Fig. 13.4. In such a mixture design, there is not complete independence between the chemicals. For example, in a three chemical mixture, at given percentages of x_1 and x_2, x_3 is fixed to give a total of 100%. Statistical mixture design is particularly useful when blending liquid chemicals or solutions. Statistical mixture design has been used in combination with bioassay-directed fractionation (Østby et al. 1997). After fractionation of organic extracts of diesel exhaust particles, three fractions were recombined in different proportions based on the design and then tested with respect to mutagenicity to identify interactions between the fractions. A similar design was used by Eide and Zahlsen (1996) in inhalation experiments with three hydrocarbons. Both the **X** and **Y** matrices are shown in these articles.

13.4.2.3 Isobolograms

An experimental approach that directly applies to the dose additivity definition of binary mixtures is the isobolographic analysis which actually represents the simplest form of mixture design (Plackett and Hewlett 1952; Könemann and Peters 1996). For a binary mixture, an isobologram is a two-dimensional graph with the doses of agents A and B along coordinate axes and with one or several lines, isoboles that connect different dose combinations producing the same magnitude of effect (Fig. 13.5). Dose additivity is characterized by straight lines in an isobologram with linear dose scales touching the dose axes at the single-agent doses D_A and D_B. Deviations from this straight line directly indicate greater-than-additivity or less-than-additivity as shown in Fig. 13.5. Examples of the isobole approach are presented by Bernhoft et al. (2004). Combined effects of selected *Penicillium* mycotoxins on in vitro proliferation of porcine lymphocytes were

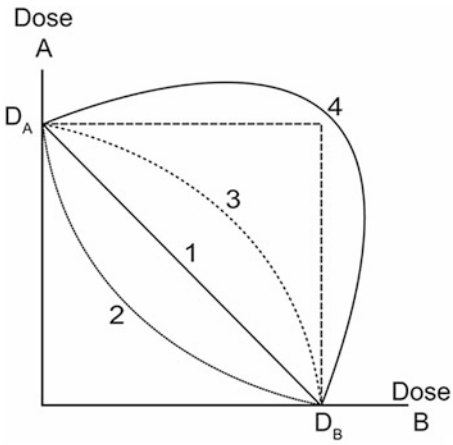

Fig. 13.5 Isobologram. Line 1, dose additivity; line 2, greater-than additivity; lines 3 and 4, less-than additivity

studied in binary mixtures. Dose-response curves for each mycotoxin and mycotoxin combinations were generated. The combined effects of toxin pairs based on IC_{20} were illustrated in isobolograms.

Although isobolograms usually are made for binary mixtures, multidimensional isobolograms may in principle be used implying statistical mixture designs with equipotent doses or concentrations (Eide and Johnsen 1998).

13.4.2.4 Ray Designs

An alternative, simple design strategy is based on using "ray designs" (e.g., Gennings et al. 2004; Moser et al. 2005; Hertzberg et al. 2013). In this design, dose-response relations are determined for individual mixture components as well as for mixtures of the components tested with graded total (sum across components) doses tested along a fixed *ray*. A ray is a one-dimensional surface in high-dimensional space along which the component doses are in fixed proportion, but the total dose varies. For example, if there are three components, then along the ray, the relative component doses are fixed (e.g., 0.3, 0.5, 0.2), with the absolute component doses varying. Ray mixture designs require many fewer test points than factorial designs, but their inferences are more limited. If the individual components have the same mode of action and can be thought of as dilutions or concentrations of one another with respect to toxic effects, then the responses along the ray can be predicted from the individual component responses as the effects associated with a specified linear combination of the component doses. A limitation of ray designs is that if departures from dose additivity are detected, there is no direct indication which component or combinations of components are responsible. In their paper, Hertzberg et al. (2013) provide a method, termed the expected component contribution score, that calculates the percentage of the mixture response under dose addition that is expected from each component. This involves

additional exploratory analysis steps in which components are examined one at a time.

13.4.3 Summary

Factorial designs, response surface designs, statistical mixture designs, isobolograms, and ray designs are each useful for studying the joint toxicity of mixtures of chemicals, but they have different advantages. Factorial designs lead to more general inferences. Response surface designs are good for identifying joint mixture conditions resulting in maximum (minimum) response or in moving toward the direction of the extreme response. Ray designs are much simpler to carry out since they involve mainly single-component tests, but their utility is limited to the detection of dose additivity or departures from it and do not provide an obvious path forward for the next stage of statistical analysis in the absence of dose additivity. Statistical mixture designs may be preferred when the variables are proportions of a blend. For detailed explanations of these various designs, the reader is referred to Box et al. (2005), Cochran and Cox (1992), Crosier (1984), and Cornell (2002).

13.5 What is the Right Experiment?

Although it might be easy to be discouraged after reading the chapter to this point, it is important to keep several points in mind. First, the recent widespread concerns with reproducibility of studies and concerns with design, power, and reproducibility extend far beyond mixtures. With the increased emphasis in many scientific disciplines for studies that are well designed, properly powered, appropriately conducted, analyzed, and interpreted, all fields of science, including mixtures, will benefit.

It is also important to note that appropriate experimental designs and associated statistical methods for mixture analyses are now more widely implemented. The right experiment is often one that is best if designed by a multidisciplinary team, with the experimentalist, the statistician, and the consumer of the results being the essential trio, with other expertise included as appropriate. Examples of common "consumers of the results" include risk assessment and risk management practitioners, as well as regulatory agencies and decision makers, pharmaceutical manufacturers, medical professionals, and the chemical industry.

The right experiment is one that is tailored to address the question being asked. Careful and clear articulation of both the effects to be examined (e.g., mutagenicity in Salmonella strain TA100 in the presence of metabolic activation or change in pup weight on postnatal day 6) and the null and alternative hypotheses are crucial to conducting the right experiment. The optimal design will vary based on the purpose of the experiment; in turn, the purpose of the experiment drives the elements of the

null and alternative hypotheses. The experimental design should optimize the accuracy and precision of the values being estimated; the optimal design depends on the values being estimated and the "true" nature of the underlying dose-response relationship, i.e., the shape of the underlying dose-response curve(s) and variability. The total number of experimental units is important as is the number of dose groups, the selection of doses, and the allocation of experimental units within dose groups. Developing the analysis plan is important as the most elegant design will not be successful if the analysis is not appropriate; conversely a superior analysis strategy will flounder if the data collected are not suitable for that type of analysis.

Consideration of power in advance of conducting the experiment is vital. Neglect of power considerations appears to be one of the most common mistakes in mixture experimental design. The importance of advance power analysis increases as investigators move into the low-effect/low-dose region of dose-response curves. As shown in this chapter, different designs are likely to be required dependent on the complexity of the dose-response trend, the complexity of the interactions, and whether the focus is on detection of a greater-than-additive, less-than-additive, or a dose-dependent interaction. The utility of a pilot experiment cannot be overemphasized as it provides highly useful data to use in planning the more definitive experiment, including determination of the sources and degree of variability, the selection of an appropriate design (e.g., isobole vs ray vs response surface vs factorial), power calculations, dose selection decisions, the allocation of experimental units per dose group, and optimization of experimental design.

In sum, the right experiment:

- Addresses an important question
- Focuses on mixtures of concern
- Uses relevant levels of biological organization for evaluation of the mixture
- Includes relevant mixing ratios for defined mixture investigations and environmentally realistic mixtures for complex mixture investigations
- Is planned by cross-disciplinary teams with essential expertise in mixtures toxicology, experimental design, and statistical analysis and draws freely on other expertise based on the purpose of the experiment and the use of the resulting data/information

Finally, the experimental design and the analysis should always match the question being asked.

Acknowledgments The authors thank Drs. Cynthia Rider and Mike Narotsky for thoughtful review of the chapter.

References

Begley, C.G., and L.M. Ellis. 2012. Drug development: Raise standards for preclinical cancer research. *Nature* 483: 531–533.

Berenbaum, M.C. 1985. The expected effect of a combination of agents – the general-solution. *Journal of Theoretical Biology* 114: 413–431.

———. 1989. What is synergy? *Pharmacological Reviews* 41: 93–141.

Bernhoft, A., M. Keblys, E. Morrison, H.J. Larsen, and A. Flåøyen. 2004. Combined effects of selected Penicillium Mycotoxins on in vitro proliferation of porcine lymphocytes. *Mycopathologia* 158: 441–450.

Boobis, A., B. Budinsky, S. Collie, K. Crofton, M. Embry, S. Fleter, R. Hertzberg, D. Kopp, G. Mihlan, M. Mumtaz, P. Price, K. Solomon, L. Teuschler, R. Yang, and R. Zaleski. 2011. Critical analysis of literature on low-dose synergy for use in screening chemical mixtures for risk assessment. *Critical Reviews in Toxicology* 41: 369–338.

Borgert, C.J., B. Price, C. Wells, and G.S. Simon. 2001. Evaluating interaction studies for mixture risk assessment. *Human and Ecological Risk Assessment* 7: 259–306.

Borgert, C.J., S.A. Borgert, and K.C. Findley. 2005. Synergism, antagonism, or additivity of dietary supplements: Application of theory to case studies. *Thrombosis Research* 117: 123–132.

Bostrøm, E., S. Engen, and I. Eide. 1998. Mutagenicity testing of organic extracts of diesel exhaust particles after spiking with PAHs. *Archives of Toxicology* 72: 645–649.

Box, G.E.P., W.G. Hunter, and J.S. Hunter. 2005. *Statistics for experimenters: Design, innovation, and discovery*. 2nd ed. New York: Wiley. 664 p.

Carpy, S.A., W. Kobel, and J. Doe. 2000. Health risk of low-dose pesticides mixtures: A review of the 1985–1998 literature on combination toxicology and health risk assessment. *Journal of Toxicology and Environmental Health B* 3: 1–25.

Casey, M., C. Gennings, W.H. Carter Jr., V.C. Moser, and J.E. Simmons. 2006. Power and sample size calculations for linear hypotheses associated with mixtures of many components using fixed-ratio ray designs. *Environmental and Ecological Statistics* 13: 11–23.

Cedergreen, N. 2014. Quantifying synergy: A systematic review of mixture toxicity studies within environmental toxicology. *PLOSOne* 9 (5): e96580.

Chou, T.C. 2006. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacological Reviews* 58: 621–682. Correction: *Pharmacological Reviews*, 59: 124.

Chou, T.C., and P. Talalay. 1984. Quantitative analysis of dose-effect relationships: The combined effects of multiple drugs or enzyme inhibitors. *Advances in Enzyme Regulation* 22: 27–55.

Cochran, W.G., and G.M. Cox. 1992. *Factorial designs. In experimental designs*. 2nd ed. - New York: Wiley.

Coffey, T., C. Gennings, J.E. Simmons, and D.W. Herr. 2005. D-optimal experimental designs to test for departure from additivity in a fixed-ratio mixture ray. *Toxicological Sciences* 88: 467–476.

Cornell, J.A. 2002. *Experiments with mixtures: Designs, models, and the analysis of mixture data*. 3rd ed. New York: Wiley. 680 p.

Crofton, K.M., E.S. Craft, J.M. Hedge, C. Gennings, J.E. Simmons, R.A. Carchman, W.H. Carter Jr., and M.J. DeVito. 2005. Thyroid hormone disrupting chemicals: Evidence for dose-dependent additivity and synergism. *Environmental Health Perspectives* 113: 1549–1954.

Crosier, R.B. 1984. Mixture experiments: Geometry and pseudocomponents. *Technometrics* 26: 209–216.

Dingus, C.A., L.K. Teuschler, G.E. Rice, J.E. Simmons, and M.G. Narotsky. 2011. Power calculations for a multigenerational reproductive toxicity rodent bioassay of complex mixtures of drinking water disinfection by-products. *International Journal of Environmental Research and Public Health* 8: 4082–4101.

Eide, I. 1996. Strategies for toxicological evaluation of mixtures. *Food and Chemical Toxicology* 34: 1147–1149.

Eide, I., and K. Zahlsen. 1996. Inhalation experiments with mixtures of hydrocarbons: Experimental design, statistics and interpretation of kinetics and possible interactions. *Archives of Toxicology* 70: 397–404.

Eide, I., and H.G. Johnsen. 1998. Mixture design and multivariate analysis in mixture research. *Environmental Health Perspectives* 106 (Suppl. 6): 1373–1376.

Ermler, S., M. Scholze, and A. Kortenkamp. 2011. The suitability of concentration addition for predicting the effects of multi-component mixtures of up to 17 anti-androgens with varied structural features in an in vitro AR antagonist assay. *Toxicology and Applied Pharmacology* 257: 189–197.

Figueiredo, M.A.O., I. Eide, M. Reynier, A.B. Villas-Bôas, F.T.S. Tâmega, C.G. Ferreira, I. Nilssen, R. Coutinho, and S. Johnsen. 2015. The effect of sediment mimicking drill cuttings on deep water rhodoliths in a flow-through system: Experimental work and modeling. *Marine Pollution Bulletin* 95: 81–88.

George, B.J., J.R. Sobus, L.P. Phelps, B. Rashleigh, J.E. Simmons, R.N. Hines, and the Community of Practice for Statistics Guidance Documents Working Groups. 2015. Raising the Bar for Reproducible Science at the US Environmental Protection Agency Office of Research and Development. *Toxicological Sciences* 145: 16–22.

Gennings, C., W.H. Carter Jr., E.D. Campbell, J.G. Staniswalis, T.J. Martin, B.R. Martin, and K.L. White Jr. 1990. Isobolographic characterization of drug interactions incorporating biological variability. *Journal of Pharmacology and Experimental Therapeutics* 252: 208–217.

Gennings, C., W.H. Carter Jr., M. Casey, V. Moser, R. Carchman, and J.E. Simmons. 2004. Analysis of functional effects of a mixture of five pesticides using a ray design. *Environmental Pharmacology and Toxicology* 18: 115–125.

Goldin, A., and N. Mantel. 1957. The employment of combinations of drugs in the chemotherapy of neoplasia: A review. *Cancer Research* 17: 635–654.

Gordon, C.J., D.W. Herr, C. Gennings, J. Graff, M. McMurray, L. Stork, T. Coffey, A. Hamm, and C. Mack. 2006. Thermoregulatory response to an organophosphate and carbamate insecticide mixture: Testing the assumption of dose-additivity. *Toxicology* 217: 1–13.

Greco, W.R., G. Bravo, and J.C. Parson. 1995. The search for synergy: A critical review from a response surface perspective. *Pharmacological Reviews* 47: 331–385.

Groten, J.P., E.J. Sinkeldam, T. Muys, J.B. Luten, and P.J. van Bladeren. 1991. Interaction of dietary Ca, P, Mg, Mn, Cu, Fe, Zn, and Se with the accumulation and oral toxicity of cadmium in rats. *Food and Chemical Toxicology* 29: 249–258.

Hadrup, N., C. Taxvig, M. Pedersen, C. Nellemann, U. Hass, and A.M. Vinggaard. 2013. Concentration addition, independent action and generalized concentration addition models for mixture effect prediction of sex hormone synthesis in vitro. *PLoS One* 8 (8): e70490. https://doi.org/10.1371/journal.pone.0070490.

Hass, U., S. Christiansen, M. Axelstad, M. Scholze, and J. Boberg. 2017. Combined exposure to low doses of pesticides causes decreased birth weight in rats. *Reproductive Toxicology* 72: 97–105.

Hertzberg, R., Y. Pan, R. Li, L.T. Haber, R. Lyles, D.W. Herr, V. Moser, and J.E. Simmons. 2013. A four step approach to evaluate mixtures for consistency with dose addition. *Toxicology* 313: 134–144.

Howdeshell, K.L., C.V. Rider, V.S. Wilson, J.R. Furr, C.R. Lambright, and L.E. Gray Jr. 2015. Dose addition models based on biologically relevant reductions in fetal testosterone accurately predict postnatal reproductive tract alterations by a phthalate mixture in rats. *Toxicological Sciences* 148: 488–502.

IGHRC (The Interdepartmental Group on Health Risks from Chemicals). 2008. Chemical mixtures: A framework for assessing risks to human health. Institute of Environment and Health Cranfield, UK.

Ioannidis, J.P., and M.J. Khoury. 2011. Improving validation practices in "Omics" research. *Science* 334: 1230–1232.

Jawetz, E., and J.B. Gunnison. 1953. Antibiotic synergism and antagonism: An assessment of the problem. *Pharmacological Reviews* 5: 175–192.

Johnsen, S., A.T. Smith, J. Brendehaug, H. Riksheim, and A.L. Gjøse. 1994. Identification of Sources of Acute Toxicity in Produced Water. In: *Proceedings SPE International Conference on Health, Safety and Environment*. Jakarta, Indonesia. SPE 27138.

Könemann, W.H., and M.N. Pieters. 1996. Confusion of concepts in mixture toxicology. *Food and Chemical Toxicology* 34: 1025–1031.

Kortenkamp, A., T. Backhaus, and M. Faust. 2009. State of the art Report on Mixture Toxicity. http://ec.europa.eu/environment/chemicals/effects/pdf/report_mixture_toxicity.pdf.

Loewe, S., and H. Muischnek. 1926. On combination effects (D. Doty Trans.). *Archiv für Experimentalle Pathologie und Pharmakologie* 114: 313–326.

Mason, A.M., C.J. Borgert, J.S. Bus, M. Mumtaz, J.E. Simmons, and I.G. Sipes. 2007. Improving the scientific foundation for mixtures joint toxicity and risk assessment: Contributions from the SOT mixtures project. *Toxicology and Applied Pharmacology* 223: 99–103.

McInnes, G.T., and M.J. Brodie. 1988. Drug interactions that matter: A critical reappraisal. *Drugs* 36: 83–110.

Moser, V.C., M. Casey, A. Hamm, W.H. Carter Jr., J.E. Simmons, and C. Gennings. 2005. Neurotoxicological and statistical analyses of a mixture of five organophosphorus pesticides using a ray design. *Toxicological Sciences* 86: 101–115.

Moser, V.C., S. Padilla, J.E. Simmons, L. Haber, and R.C. Hertzberg. 2012. Impact of chemical proportions on the acute neurotoxicity of a mixture of seven carbamates in preweanling and adult rats. *Toxicological Sciences* 129: 126–134.

Narotsky, M.G., D.S. Best, E.H. Rogers, A. McDonald, Y.M. Sey, and J.E. Simmons. 2008. Integrated disinfection by-products mixtures research: Assessment of developmental toxicity in Sprague-Dawley rats exposed to concentrates of water disinfected by chlorination and ozonation/postchlorination. *Journal of Toxicology and Environmental Health A* 71 (17): 1216–1221.

Narotsky, M.G., G.R. Klinefelter, J.M. Goldman, D.S. Best, A. McDonald, L.F. Strader, J.D. Suarez, A.S. Murr, I. Thillainadarajah, E.S. Hunter, S.D. Richardson, T.F. Speth, R.J. Miltner, J.G. Pressman, L.K. Teuschler, G.E. Rice, V.C. Moser, R.W. Luebke, and J.E. Simmons. 2013. Comprehensive assessment of a chlorinated drinking water concentrate in a rat multigenerational reproductive toxicity study. *Environmental Science and Technology* 47: 10653–10659.

NIST/SEMATECH e-Handbook of Statistical Methods. n.d. http://www.itl.nist.gov/div898/handbook/. Accessed 1-10-18.

Orton, F., S. Ermler, S. Kugathas, E. Rosivatz, M. Scholze, and A. Kortenkamp. 2014. Mixture effects at very low doses with combinations of anti-androgenic pesticides, antioxidants, industrial pollutant and chemicals used in personal care products. *Toxicology and Applied Pharmacology* 278: 201–208.

Østby, L., S. Engen, A. Melbye, and I. Eide. 1997. Mutagenicity testing of organic extracts of diesel exhaust particles after fractionation and recombination. *Archives of Toxicology* 71: 314–319.

Peace, M.R., L.D. Tarnai, and A. Poklis. 2000. Performance evaluation of four on-site drug testing devices for detection of drugs of abuse in urine. *Journal of Analytical Toxicology* 24: 589–594.

Peng, R.D. 2011. Reproducible research in computational science. *Science* 334: 1226–1227.

Placket, R.L., and P.S. Hewlett. 1952. Quantal responses to mixture of poisons. *Journal of the Royal Statistical Society, series* B14: 143–163.

Pressman, J.G., S.D. Richardson, T.F. Speth, R.J. Miltner, M.G. Narotsky, E.S. Hunter III, G.E. Rice, L.K. Teuschler, A. McDonald, S. Parvez, S.W. Krasner, H.S. Weinberg, A.B. McKague, C.J. Parrett, N. Bodin, R. Chinn, C.F. Lee, and J.E. Simmons. 2010. Concentration, chlorination, and chemical analysis of drinking water disinfection byproduct mixtures

for health effects research: The four lab study. *Environmental Science and Technology* 44: 7184–7192.

Rider, C.V., J. Furr, V.S. Wilson, and L.E. Gray Jr. 2008. A mixture of seven antiandrogens induces reproductive malformations in rats. *International Journal of Andrology* 31: 249–262.

Rider, C.V., J.R. Furr, V.S. Wilson, and L.E. Gray Jr. 2010. Cumulative effects of in utero administration of mixtures of reproductive toxicants that disrupt common target tissues via diverse mechanisms of toxicity. *International Journal of Andrology* 33: 443–462.

Ryan, M.J. 2011. Replication in field biology: The case of the frog eating bat. *Science* 334: 1229–1230.

Scholze, M., E. Silva, and A. Kortenkamp. 2014. Extending the applicability of the dose addition model to the assessment of chemical mixtures of partial agonists by using a novel toxic unit extrapolation method. *PLoS One* 9 (2): e88808. https://doi.org/10.1371/journal.pone.0088808.

Simmons, J.E. 1995. Chemical mixtures: Challenge for toxicology and risk assessment. *Toxicology* 105: 111–119.

Simmons, J.E., S.D. Richardson, T. Speth, R.J. Miltner, G. Rice, K. Schenck, and L.K. Teuschler. 2002. Development of an research strategy for integrated technology-based toxicology studies on drinking water disinfection byproducts. *Environmental Health Perspectives* 110 (Supplement 6): 1013–1024.

Simmons, J.E., S.D. Richardson, L.K. Teuschler, R.J. Miltner, T.F. Speth, K.M. Schenck, E.S. Hunter, and G. Rice. 2008. Research issues underlying the four-lab study: Integrated disinfection by-products mixtures research. *Journal of Toxicology and Environmental Health A*. 71: 1125–1132.

Simmons, J.E., R.S.H. Yang, D.J. Svendsgaard, M.B. Thompson, J.C. Seely, and A. McDonald. 1994. Toxicology studies of a chemical mixture of 25 groundwater contaminants: Hepatic and renal assessment, response to carbon tetrachloride challenge and influence of treatment-induced water restriction. *Journal of Toxicology and Environmental Health* 43: 305–325.

Søfteland, L., I. Eide, and P.A. Olsvik. 2009. Factorial design applied for multiple endpoint toxicity evaluation in Atlantic Salmon (*Salmo Salar* L.) hepatocytes. *Toxicology In Vitro* 23: 1455–1464.

Stork, L.G., C. Gennings, W.H. Carter Jr., R.E. Johnson, P.M. Darcy, J.E. Simmons, E.D. Wagner, and M.J. Plewa. 2007. Testing for additivity in chemical mixtures using a fixed-ratio design and statistical equivalence testing methods. *Journal of Agricultural, Biological and Environmental Statistics* 12: 514–533.

Sun, Y., Z. Sheng, C. Ma, K. Tang, R. Zhu, Z. Wu, R. Shen, J. Feng, D. Wu, D. Huang, D. Huang, J. Fei, Q. Liu, and Z. Cao. 2015. *Nature Communications* 6: 8481–8490.

Teuschler, L., J. Klaunig, E. Carney, J. Chambers, R. Conolly, C. Gennings, J. Giesy, R.H. Hertzberg, C. Klaassen, R. Kodell, D. Paustenbach, and R. Yang. 2002. Support of science-based decisions concerning the evaluation of the toxicology of mixtures: A new beginning. *Regulatory Toxicology and Pharmacology* 36: 34–39.

Thomas, L. 1997. Retrospective power analysis. *Conservation. Biology*. 11: 276−280.

U.S. Environmental Protection Agency (EPA). 1998. Health effects test guidelines OPPTS 870.3800. Reproduction and fertility effects. EPA 712-C-98-208.

———. 1986. Guidelines for the health risk assessment of chemical mixtures. (EPA/630/R-98/002). *Federal Register* 51 (185): 34014–34025.

———.1990. Technical support document on health risk assessment of chemical mixtures. EPA/600/8–90/064.

Veldstra, H. 1956. Synergism and potentiation with special reference to the combination of structural analogues. *Pharmacological Reviews* 8: 339–387.

Part IV
Chemical Mixtures Risk Assessment

Chapter 14
Component-Based Risk Assessment Approaches with Additivity and Interactions

Richard C. Hertzberg and M. Moiz Mumtaz

Abstract Mixture risk assessments are most often based on toxicology information for the component chemicals, primarily because dose-response information on the mixture of concern is inadequate. The most widely used component methods in mixture risk assessment are based on dose addition or response addition. Two methods described here are derived from dose addition: the hazard index and the relative potency factor method. The hazard index informs safety decisions for specific environmental exposures, while the relative potency factor formula actually estimates the health risk for specific environmental exposures. Use of dose addition is supported when chemicals are toxicologically similar (see Chap. 9). The hazard index is motivated by dose addition and is applied to chemicals having a common toxicological target organ, which is fairly weak evidence of toxicological similarity. A weight of evidence system for binary mixtures allows the hazard index to be modified to incorporate toxicological interactions. Relative potency factors are applied in a dose addition formula and are developed for chemicals that have good evidence of toxicological similarity, e.g., chemicals that share a common adverse outcome pathway (AOP) and are assumed not to elicit toxicological interactions. Two other methods described here are related to response addition, both requiring the assumption of toxicological independence (see Chap. 9). When response is a risk (probability), response addition follows the statistical formula for independent events. When response is a measured effect, the addition is called effect summation, where the component incremental effects are added. This chapter reviews the history and concepts related to component-based methods for health risk assessment of chemical mixtures, illustrates commonly used methods and some modifications, and discusses strengths, limitations, and likely future development.

R. C. Hertzberg (✉)
Emory University, Atlanta, GA, USA
e-mail: hertzberg@comcast.net

M. M. Mumtaz
Agency for Toxic Substances Disease Registry, Atlanta, GA, USA

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_14

Keywords Dose addition · Hazard index · Relative potency factors · Toxicity equivalence factors · Binary weight of evidence · Interaction-based hazard index · Toxicological similarity · Toxicological independence · Response addition · Effect summation · Integrated addition

List of Abbreviations

AOP	Toxicological adverse outcome pathway
ATSDR	U.S. Agency for Toxic Substances and Disease Registry
BMD	Benchmark dose
BMR	Benchmark response
DBP	Drinking water disinfection by-product
DVFA	Danish Veterinary and Food Administration
ECJRC	European Commission Joint Research Centre
ED_x	Effective dose causing x response, usually $x\%$
EFSA	European Food Safety Authority
U.S. EPA	U.S. Environmental Protection Agency
FQPA	U.S. Food Quality Protection Act
ICED	Index chemical equivalent dose
IRIS	U.S. EPA's Integrated Risk Information System
MOA	Toxicological mode of action
NOAEL	No-observed-adverse-effect level
PBPK	Physiologically based pharmacokinetic
QSAR	Quantitative structure-activity relationship
RfD	Reference dose
RIVM	Dutch National Institute for Public Health and the Environment
SF	Carcinogenicity slope factor (U.S. EPA)
UF	Uncertainty factor
WHO	World Health Organization

14.1 Background

The main purpose of a chemical mixture risk assessment is to provide scientific support to inform a risk management decision concerning multichemical exposure. With regulatory agencies that protect public health, that decision often relates to determinations of safety, and if the exposure is deemed unsafe, it leads to consideration of priorities and consequent action. Because a regulatory risk management decision can be mandated by the courts to occur by a certain date, a risk assessment strategy is needed that can proceed even when the scientific information is incomplete; that usually results in the establishment of several default methods to fill those information gaps. Nonregulatory risk management can have similar safety goals and

time constraints but might also be able to incorporate into the mixture assessment more extensive information on the full dose-response curve, to use in comparative evaluations such as cost-benefit prioritization. Particularly for occupational settings, such assessments might have specific information on the exposures and characteristics of the human population of concern.

Risk assessment is then not the same as quantitative toxicology, and approaches suitable for research on joint toxicity, such as those in other sections in this book on dose-response analysis from mathematical and toxicological perspectives (Chaps. 8, 9, 10, 11, 12, and 13), are often not suitable for regulatory or safety decision-making. Mixture risk assessment methods in particular involve several compromises and approximations, where the desirable characteristics are that they are feasible, roughly correct, and able to be improved. For mixture exposures, the common lack of dose-response information on the whole mixture itself has led to assessments based primarily on information for the component chemicals, with occasional enhancement by information on pairwise toxicological interactions (ATSDR 2004a; Meek et al. 2011; SCHER et al. 2012; U.S. EPA 2000b). Chemical mixture exposures can easily become very complex with multiple types of uncertainty. Consequently, many safety and regulatory approaches include simplifications, often resulting in dramatic reduction in the dimensionality (number of variables) but usually with a corresponding increase in assumptions.

This chapter reflects the practice of health risk assessment for chemical mixtures, and so is limited to methods contained in official regulatory risk guidance or methods used in public health risk assessments. Regulatory guidance often emphasizes consistency and transparency of the risk assessment approach. Consequently, such guidance usually includes proscriptive lists of assessment steps and criteria for evaluating the approaches (U.S. EPA 2000a, 2001b, 2002a). Agency reports that focus more on concepts are often called frameworks (Meek et al. 2011; U.S. EPA 2003c). The governmental guidance summarized in this chapter reflects that of the U.S. EPA, U.S. ATSDR, WHO, EFSA, DVFA, RIVM, and some European Union commissions (see Kortenkamp et al. 2009 and Kienzler et al. 2014 for summaries of approaches for different agencies and various EU regulations). The focus of this chapter is on component-based risk assessment approaches for chemical mixtures, along with their attendant uncertainties. While the application is primarily to public health, most principles and approaches also work well with ecological populations. Several examples are included that are motivated by actual environmental mixture risk assessments.

14.1.1 Scope

Most of the methods developed for mixture risk assessment involve dose-response formulas, and consequently this section focuses on dose-response assessment. The other steps in risk assessment, namely, hazard identification and exposure assessment, are only briefly discussed (Sect. 14.1.2). For mixture dose-response

assessment, the translation from experimental results to human risk estimation is similar to that process used for individual chemicals, where main concerns involve cross-species extrapolation (or scaling) of equitoxic dose and duration and relevance of animal results to human health. The dominant extrapolation and assumption with mixture risk estimation are that the observed experimental toxicity results, whether whole-mixture response or evidence for toxicological interaction, also apply to humans. The quantitative integration of animal and human dose-response data on mixtures into a mixture risk assessment is rarely performed and can be considered still in the early stages of development. Some concepts and examples do exist of comparing and combining toxicological and epidemiological studies (see Chap. 10).

14.1.2 Exposure Assessment and Hazard Characterization of Mixtures

The bulk of the exposure assessments and toxicological hazard characterizations for environmental chemicals have been performed for individual chemicals, not mixtures (Egeghy et al. 2012; Meek et al. 2011). Some work has considered exposure assessment of chemical combinations by focusing on those most commonly encountered, e.g., those included in the ATSDR HazDat database for hazardous waste sites in the USA (Fay and Mumtaz 1996). Other work has targeted combination exposures in specific media, e.g., U.S. EPA characterizations of disinfection by-product concentrations in finished drinking water systems (Miltner et al. 2008) and exposure research on ambient air quality including the U.S. EPA Cumulative Exposure Project (Fox et al. 2004) and the Cumulative Communities Research Program (Zartarian and Schultz 2010). Another targeted combination exposure is to specific chemical groups, e.g., multimedia exposure to organophosphate pesticides for specific geographic regions and populations (U.S. EPA 2002b). Other work has focused on specific subpopulations, e.g., the Detroit area (Fann et al. 2011; Morello-Frosch and Shenassa 2006; Wesson et al. 2010), or has developed exposure grouping by common media and exposure time (i.e., concurrent or nearly so) in order to assume that members of a specific population are all exposed to the mixture of chemicals in that exposure group (MacDonell et al. 2017 (in press); Rice et al. 2008; U.S. EPA 2007a). Research on exposure measurement and modeling of multiple chemicals is evolving, such as investigations into potential application of ecosystem approaches to include economics and geolocation into estimates of likely coexposures (Tornero-Velez et al. 2012). Other concepts and approaches point the way toward improved exposure assessments for mixture risk (see Chaps. 2, 3, and 4).

The hazard characterization step includes descriptions of the nature of observed effects, which for human data with mixtures has mostly been performed for whole mixtures, often commercial (intentional) combinations or compositionally consistent by-products (e.g., diesel engine exhaust emissions in urban air). The exception is the hazard descriptions from binary mixture research on toxicological interactions,

exemplified by the many references summarized in the ATSDR interaction profiles (ATSDR 2004a). A main limitation of many interaction studies is the focus on mechanistic understanding with less information on interaction magnitude or potential exposure levels and often involves a limited dose range. To address such limitations with binary interactions, U.S. EPA and ATSDR have developed structured evaluations of the weight of evidence for pairwise interactions that includes both scientific quality and the relevance to human health (see Sect. 14.4).

14.1.3 Focus on Dose-Response Assessment

Dose-response assessment is a focus of this chapter, reflecting the dominance of that assessment step in the research literature and in the regulatory guidance dealing with mixtures of chemicals. Similarly, the methods in this chapter focus on those using only information on the mixture component chemicals, reflecting the relatively wide availability of dose-response information on individual chemicals. Dose-response assessment for complete mixtures can be performed. For high-priority mixtures with widespread exposure, such as diesel engine emissions and aroclors (commercial mixtures of polychlorinated biphenyls), especially complex mixtures of hundreds of components, the evaluations of the whole mixtures often treat the mixture as if it were an individual chemical substance, with only minor changes from individual chemical dose-response methods (see Chap. 15). For example, one can determine a mixture's virtually safe dose for chronic oral exposures by following the approach for the individual chemical oral reference dose (RfD) that is in the U.S. EPA IRIS database, e.g., dividing a lower bound on the benchmark dose (e.g., ED_{10}) by a product of uncertainty factors, as detailed in Chap. 15, and U.S. EPA (2000b).

The two most common component-based biological concepts for mixture dose-response assessment are toxicological similarity and toxicological independence (Teuschler et al. 2002). The preferred initial step in selecting the mixture prediction model is to determine which biological concept is most appropriate for the mixture component chemicals. The next step is to decide on the particular quantitative approach that reflects the chosen concept and corresponds to the available data. In contrast to research approaches to mixture dose-response modeling, in risk assessment the calculation method selected is also strongly influenced by the risk assessment goal. For example, when the goal is prioritization of waste sites for remediation, the level of toxicological information related to similarity can be quite general, such as same target organ. With stronger goals and more detailed information, more precise mixture risk formulas can be used. For toxicological similarity, the most commonly used assessment approach is dose[1] addition (see Chaps. 9, 10, and 11), with the more common implementations using the hazard

[1] "Dose" is used in this section to represent exposure in general. Thus, discussions of dose addition can usually be applied to concentration addition; any exceptions will be identified.

index (HI), often a screening method to decide safety, and relative potency factors (RPFs), often used to estimate the expected toxic effects from the mixture. For toxicological independence, the most common risk assessment approach is called response addition (more often called independent action in the toxicology literature; see Chap. 9), with implementation using formulas based on probabilities of response or the fraction of population responding. Those common approaches and some variants, including modification to reflect evidence on interactions, are described and illustrated in this section.

The HI, RPF, and response addition approaches are particularly attractive for regulatory risk assessment because they are easy to apply and because official sets of risk-based values for individual chemicals have been published, promoting use and consistency across the mixture assessments, at least for those chemicals of higher concern that have been formally evaluated. For example, U.S. EPA's IRIS database contains risk-reference values on over 500 chemicals including the oral reference doses (RfDs) and inhalation reference concentrations (RfCs) that can be used in the HI formula and cancer unit risk values that can be used in response addition formulas (U.S. EPA 2017). The U.S. EPA has published RPF values for a few groups of pesticides (U.S. EPA 2002b, 2006, 2007d), and both the U.S. EPA and the World Health Organization (WHO) have published RPF values, called toxicity equivalence factors (TEFs; see Sect. 14.2.3.2) for the dioxin-like chemicals (U.S. EPA 2010; van den Berg et al. 2006, 2013).

14.2 Risk Assessment Methods for Toxicologically Similar Chemicals

14.2.1 Background on Use of Toxicological Similarity in Risk Assessment

The first extensive regulatory guidance on risk assessment of chemical mixtures was published by U.S. EPA in 1986 (U.S. EPA 1986b). That report noted that several agencies and organizations had been addressing risk or safety assessments of mixtures, citing the American Conference of Governmental Industrial Hygienists, the U.S. Occupational Safety and Health Administration, the World Health Organization, and the U.S. National Research Council. Those that recommended a specific calculation approach adopted some form of a dose-additive formula, primarily for application to mixtures of toxicologically similar chemicals. For example, ACGIH (ACGIH 1983) applies their concentration-additive formula to combinations of volatile organic compounds. The U.S. EPA's guidance for chemicals to be addressed under the 1996 Food Quality Protection Act (FQPA) focuses on establishing a common toxicological mechanism or mode of action (MOA) as the basis for including chemicals, mainly pesticides, in a similarity group (U.S. EPA 1999). The U.S. EPA's detailed supplementary guidance on mixture risk assessment

expands that consideration to include similarity of other processes affecting chemical toxicity, such as pharmacokinetics. Instead of specific criteria for deciding toxicological similarity, the guidance notes the implicit assumptions when similarity methods such as dose addition are used (U.S. EPA 2000b p. 68):

> In the simplest terms, two chemicals are dose additive if chemical B is functionally a clone of chemical A. In this ideal case, the chemicals are assumed to behave similarly in terms of the primary physiologic processes (uptake, metabolism, distribution, elimination) as well as the toxicological processes.

While several research articles have evaluated the biological evidence supporting dose addition for mixtures of varying degrees of component similarity, no more extensive criteria have yet been published by U.S. regulatory agencies (Lambert and Lipscomb 2007). In determining the dose-additive prediction model for a mixture, some methods have been previously published that bypass the component RPFs and estimate the model directly from the component dose-response information (Gennings 1995; Hertzberg et al. 2013). The methods presented in this chapter are constrained to those described in official risk assessment guidance and related publications.

14.2.2 Hazard Index

The original U.S. EPA guidance for using the HI was quite limited; however the application to specific groups of chemicals clearly conveyed an assumption of toxicological similarity, e.g., solvents that caused neurotoxicity. The HI has been used most often in the risk assessment of hazardous waste sites to implement the National Oil and Hazardous Substances Pollution Contingency Plan (NCP) of 1988 and the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) of 1980, the latter also called the Superfund Act.

The main concept behind the HI is toxicological similarity across the mixture component chemicals. The formula for the HI is motivated by dose addition, where the toxicity of the mixture exposure can be represented by the toxicity of the sum of the component doses, each scaled by its relative toxic potency. For the HI, however, the relative potency of a chemical is replaced by the inverse of an estimate of a risk-based reference level, an acceptable level, or exposure of minimal risk, e.g., the U.S. EPA's RfD and RfC values and ATSDR's similarly determined minimal risk levels (MRLs). In the following, the U.S. EPA version of the HI formula is used to illustrate the approach. For hazardous waste sites, each component exposure is scaled by its RfD (oral exposure) or RfC (inhalation exposure). For example, with oral exposures to J chemicals, the HI formula is:

$$\mathrm{HI} = \sum_{j=1}^{J} {}^{E_j}/_{\mathrm{RfD}_j} \tag{14.1}$$

where both the exposure level, E, and RfD are in the same units so that the HI is dimensionless. Thus the same formula can be used to evaluate soil contamination levels where the scaling uses values such as the U.S. EPA's soil screening level instead of the RfD (U.S. EPA 2016). The dimensionless requirement also means that chemicals can be included even if their exposures use different units, such as oral intake of mg/kg per day for cadmium (Cd) and blood levels of μg/dL for lead (Pb). One short-hand presentation of this formula uses the hazard quotient (HQ) for the ratio of exposure to reference level. For oral exposure, Eq. 14.1 then becomes:

$$\mathrm{HI} = \sum_{j=1}^{J} \mathrm{HQ}_j \tag{14.2}$$

where HQ = E/RfD for each component. With individual chemicals, for both the U.S. EPA and ATSDR reference levels, the interpretation is that individual chemical exposures at or below those levels pose minimal or no health risk (ATSDR 2016; U.S. EPA 2017). The mixture exposure has a similar interpretation: acceptable mixture exposure when HI equals or is less than unity.

The HI is mostly used as a decision index (U.S. EPA 1989b) to determine which sites or scenarios require further action, i.e., focusing on situations where HI > 1. In those instances, the mixture exposure is usually judged to have the potential to pose unacceptable risk, often leading to further investigation or to controls that can lower the exposures until HI = 1. For an initial screening assessment by U.S. EPA of hazardous waste ("Superfund") sites, all chemical exposures are included in the HI. If HI > 1, then the next step is to calculate separate target organ specific HI values, i.e., one for each effect or target organ (U.S. EPA 1989b). That means a liver HI would only include chemicals that affect the liver. That refinement is more aligned with the assumption that the included chemicals are roughly toxicologically similar.

14.2.2.1 Characteristics of the Hazard Index

The principal strength of the HI approach is its ease of use. As long as HI <1, usually no further action is deemed necessary. When HI >1, the information used in calculating the HI should be carefully examined, both for relevance and uncertainties. Because the HI is conceptually derived from dose addition (Svendsgaard and Hertzberg 1994; U.S. EPA 2000b), its use is better justified if the toxicological similarity assumptions are roughly met. The formula in Eq. 14.1 looks similar to that of dose addition, where the exposure is multiplied by a measure of toxic potency, in this case the inverse of the RfD. Because of the common lack of information on MOA and pharmacokinetics, the evidence for toxicological similarity for the HI is

usually relaxed to that of similarity of target organs, which allows application to a wider variety of scenarios. This less rigorous interpretation is similar to the recommendations of the NRC where chemicals to be included can share an AOP without requiring the same exact sequence of toxicological steps (NRC 2008). When animal studies are the basis of the toxicity assessment, however, additional information on MOA or on other factors that could affect tissue exposure (e.g., deposition pattern in the nose) should be reviewed to ensure that toxicological similarity is still roughly supported. If that desirable characteristic is not present and the HI is applied anyway, important errors and difficulties can arise. For example, for a discussion of the HI when interactions exist and when one chemical is toxicologically inert, see Svendsgaard and Hertzberg (1994).

There are many indices used in environmental assessment, usually to condense complex information into a simpler form both for decision-making and for communication (Kienhuis et al. 2015; Ott 1978). The decision focus for the HI is on values near unity, where the dose-response complexity for multiple chemicals is reduced to a single number that indicates acceptable vs. unacceptable exposure, e.g., a site poses no significant risk of toxic effects if HI $\leq$1. When HI $>$1, conclusions are less clear and often inconsistent. For some serious effects, mitigating action might be indicated if HI $>$5, while for other less serious effects, action might be deferred.

Contaminated Soil Example Consider the case of exposure to soil contaminated by heavy metals (Table 14.1). Assume that soil reference values for lead (Pb) and cadmium (Cd) are 400 and 70 ppm, respectively. The motivation for the HI from dose addition implies that for a given mixture, a specific value for HI should indicate the same level of concern, regardless of the individual chemical exposure levels. The different exposures in ppm to Pb and Cd of (1000, 35) at location A and (600, 105) at location B, respectively, would both give HI = 3 and so should be interpreted the same, even though the individual levels of Pb and Cd differ between the two sites.

When mixtures of different components are involved, their HI values should not be compared. Thus if HI = 7 for site 1 that has four pesticides in its groundwater and HI = 3 for site 2 that has seven solvents in its groundwater, site A is not necessarily of greater concern than site B. A similar uncertainty applies when considering different endpoints, even for the same mixtures. HI values are strongly dependent on the data used in deriving the reference values, requiring judgment in each case.

This decision-oriented focus on doses where HI = 1 is visually similar to the definition of dose addition by Berenbaum (1985):

$$\sum_{j=1}^{J} \frac{d_j}{\mathrm{ED}_{x,j}} = 1 \tag{14.3}$$

where d_j is the dose of chemical j in the mixture with a specific joint effect (x) and $\mathrm{ED}_{x,j}$ is the dose of that chemical alone that produces the same effect as the mixture. Under Berenbaum's definition, chemicals are said to be dose additive when Eq. 14.3 is true (see Chaps. 9 and 11). For both the HI and Berenbaum formulas, the

Table 14.1 Hazard index example of two sites with different exposure levels but the same HI

Chemical	Soil ref level ppm[a]	Site A ppm	HQ A	Site B ppm	HQ B
Cd	70	35	0.5	105	1.5
Pb	400	1000	2.5	600	1.5
HI			3.0		3.0

[a]Assumed RfD values for this example; always obtain current regulatory risk values for specific environmental risk assessment calculations

evaluation is at a single-dose combination. The key difference is that the HI value provides an indication of the *potential for health risk* using only component information, while the Berenbaum formula is a *test for consistency with dose addition* using information on both the components and the mixture. While the HI can be easily interpreted for different doses of the same mixture, the Berenbaum formula only is applied to a single mixture and single set of doses.

14.2.2.2 Uncertainties with the Hazard Index and Improvements

The commonly used HI method of Eq. 14.1 is useful and simple, but it does have important uncertainties and potential biases. For some of these, there are refinements and modifications. The main issues discussed here include:

- No statistical test for HI accuracy
- Bias toward overestimation because of reference value in the denominator
- Focus on a single exposure pathway
- Inclusion of chemicals with very low exposure levels
- Lack of consideration of toxicological interactions

14.2.2.2.1 Issue: No Statistical Test for Accuracy. Improvement – Use ED*x* Instead of RfD

Numerically, the HI is only interpreted well for HI = 1 and is only a decision index for higher or lower values. That is, if evidence is good for toxicological similarity among the component chemicals and HI <1, then a conclusion that the mixture exposure is safe seems reasonable. Because the HI is an index of concern, not an estimate of probabilistic risk, the statistical evaluations related to its scientific support usually focus on empirical support for its basis concept of dose addition. There is a way that HI can be calculated directly from dose additivity. Consider the case where both the doses and responses are in the same organism, say the rat, and the response level of interest is the ED_{10} for some specified effect. If the ED_{10} for the rat is used in the denominator in Eq. 14.1 instead of the RfD, then when the mixture component doses give HI = 1 and dose addition is assumed, the mixture should be at its ED_{10} value for the rat (Svendsgaard and Hertzberg 1994). This is easily shown by

starting with the Berenbaum definition of dose addition for two chemicals, slightly rewritten from Eq. 14.3 as:

$$1 = \frac{d_1}{\mathrm{ED}_{10,1}} + \frac{d_2}{\mathrm{ED}_{10,2}} \tag{14.4}$$

This is conceptually the HI formula for the rat with HI = 1. In this case, we set each reference value to the rat ED_{10} value for that chemical. When dose addition is valid, then the mixture combination dose (d_1, d_2) produces the same effect as the denominators, so the test is whether that combination dose produces a 10% response. Let the total mixture (d_1 + d_2) dose be represented by d_{mix}. For a fixed mixture composition (fixed fractions or ratios), each component dose can be replaced by the product of the total mixture dose and the mixture fraction ($q_i = d_i/d_{\mathrm{mix}}$). When dose addition is valid, Eq. 14.4 is true, and the total mixture dose is the mixture ED_{10}; thus $\mathrm{ED}_{10,\mathrm{mix}} = d_1 + d_2$. Replacing each dose by its fraction multiplied by the mixture dose (d_{mix}) and then dividing Eq. 14.4 by d_{mix} give:

$$\frac{1}{d_{\mathrm{mix}}} = \frac{\left(\frac{q_1 \cdot d_{\mathrm{mix}}}{\mathrm{ED}_{10,1}} + \frac{q_2 \cdot d_{\mathrm{mix}}}{\mathrm{ED}_{10,2}}\right)}{d_{\mathrm{mix}}} = \frac{q_1}{\mathrm{ED}_{10,1}} + \frac{q_2}{\mathrm{ED}_{10,2}} \tag{14.5}$$

and inverting both sides gives an estimate of the mixture ED_{10}. This is well known as the harmonic mean formula, historically used for estimating common measures of joint toxicity such as the mixture LD_{50} (Finney 1971; Smyth Jr. et al. 1969). Thus we have another test for dose additivity: one compares the observed ED_{10} for the mixture with the ED_{10} predicted from Eq. 14.5. That would provide quantitative support for dose addition for that particular mixture of concern. Note that Eq. 14.5 applies only to a mixture of a specified composition (the fractions q_1, q_2 are fixed). A different mixture (different component fractions) of the same total mixture dose would not be expected to elicit a 10% response. With these results in mind, if the reference value (e.g., RfD) in the HI formula could be given an ED_x interpretation, then a corresponding quantitative interpretation for HI = 1 is therefore possible by invoking dose addition. For higher exposures, where HI >1, a quantitative interpretation is more difficult.

14.2.2.2.2 Issue: Bias Toward Overestimation. Improvement – TTD

In general applications to human exposures, when HI = 1 or less, the mixture exposure is assumed to carry no health concern. Such an interpretation is deemed plausible based on the view that the HI is a rough application of dose addition, which relies on an interpretation of the reference values (e.g., RfDs) as being roughly equitoxic doses. That latter interpretation is generally incorrect and at best highly uncertain (Hertzberg and Teuschler 2002). More specifically, even when HI = 1, no precise quantitative interpretation can be given, partly because most reference levels are derived from animal studies and partly because the reference levels

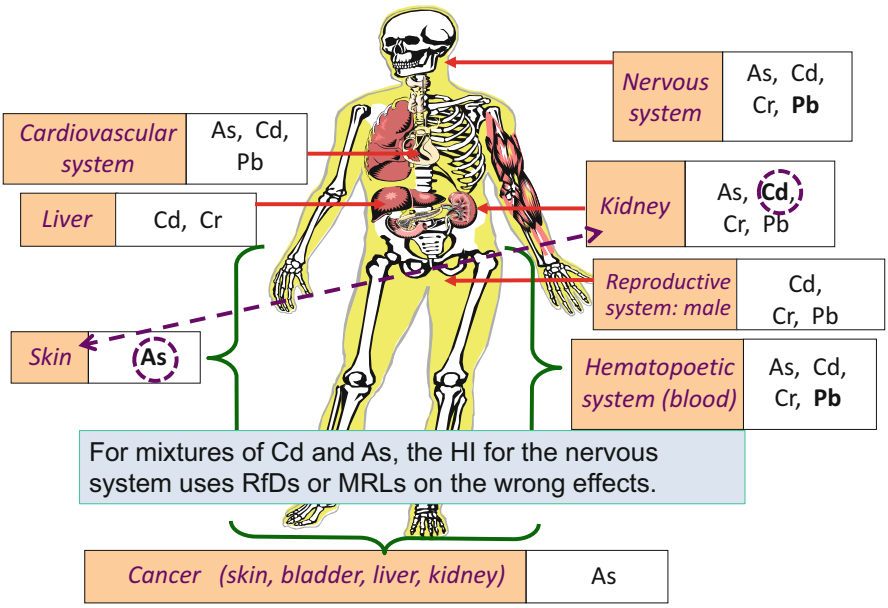

Fig. 14.1 Multiple target organs result in uncertainties with the hazard index when using the reference values for the critical effect. Chemical is boldface when aligned with its critical effect, e.g., Cd for kidney and As for skin (Adapted from MacDonell 2014)

(denominators in Eq. 14.1) are usually not derived from probabilistically determined values such as an ED_{20}. When RfD values are used instead of ED_x values, there can be increased uncertainty. Because the RfD is based on the most sensitive toxic effect, there can be a bias toward overestimation of the mixture health concern beyond that usually attributed to RfDs for individual chemical assessments. The first refinement, calculation of a target organ-specific HI, differs only by including just the chemicals affecting that target organ; even so, one or more reference values (e.g., RfDs) could be for a different and more sensitive target organ, again resulting in an overestimate of the HI. In the example depicted in Fig. 14.1, a combined exposure to arsenic (As) and cadmium (Cd) evaluated for joint neurotoxicity would give a HI using reference values for other effects (see dashed circles), namely, dermal and renal toxicity, respectively. In the second refinement, ATSDR has improved the target organ HI method by using an HI based on *target organ toxicity doses* (TTDs), which ATSDR defines as a risk-reference value derived in a similar fashion to ATSDR MRLs or U.S. EPA RfDs but only using data on the target organ of concern (Mumtaz et al. 1997). Because the RfD is defined by the U.S. EPA as representing the critical effect (first toxic effect seen with increasing dose), it will be equal to or smaller than the corresponding TTD for the effect being assessed. ATSDR has gradually been developing TTD replacement values for other effects as needed, e.g., the health assessment for the Conrail Rail Yard (ATSDR 2005a). Other examples are in some of the ATSDR interaction profiles (ATSDR 2017). Although

TTD values are only available for a few chemicals, the substitution of the RfD with the TTD can make a significant difference: Mumtaz et al. (1997) show some examples where the chemical HQ is reduced by a factor of 10–100 by replacing the RfD by the TTD. In its 2012 report to Congress, U.S. EPA announced its plans to include toxicity values for multiple effects associated with the chemical being evaluated (U.S. EPA 2012). Those planned toxicity values seem similar in concept to the ATSDR TTD values.

14.2.2.2.3 Issue: Focus on a Single Exposure Route. Solution – Multiroute HI

Many multichemical exposures involve several exposure routes (e.g., oral and inhalation), also described as exposures by multiple pathways or media (e.g., water, air). The HI formula can be slightly modified to address such exposures. The resulting multiroute HI is calculated by simple summing of HIs across the exposure routes (U.S. EPA 1989b, 2007a). Specific guidance exists from U.S. EPA for multiroute assessments for individual pesticides (U.S. EPA 2001a) and for Superfund sites (U.S. EPA 2001b), including the full characterization of exposure pathways from source to population. The multiroute HI (MHI) is simply the sum of HQs across all chemicals and exposure routes.[2] For K exposure routes and J chemicals, the formula is:

$$\mathrm{MHI} = \sum_{k=1}^{K} \left(\sum_{j=1}^{J} \mathrm{HQ}_{jk} \right) \tag{14.6}$$

Similar to the single-route HI in application, a screening assessment MHI can include all chemicals, and a second level MHI can be restricted to a common target organ.

The MHI can be calculated in two equivalent ways with different intermediate steps that can assist priority setting by the regulator or decision-maker. When the interest is on which media to control first, such as ambient air, Eq. 14.6 is useful. The first calculation (inner sum) in Eq. 14.6 yields the HI value for each exposure route (e.g., inhalation), easily showing which route poses the highest concern and thus which media should be controlled first. If the interest is on which chemicals to control first, then the order of summation is reversed. First, the multiroute HQ is calculated for each chemical (for $j = 1, ..., J$):

$$\mathrm{HQ}_j = \sum_{k=1}^{K} \mathrm{HQ}_{jk} \tag{14.7}$$

[2]The MHI has also been called a cumulative HI (U.S. EPA 2007b).

Chemicals: Contaminated Media	Exposure Level and Route	Noncancer Toxicity Reference Value[a]	Tap Water *Oral*	HQ Fish *Oral*	Air *Inhalation*	Multiroute HQ or HI
Methyl mercury: Fish	3.5×10^{-5} *Oral mg/kg-d*	1×10^{-4} *mg/kg-d*	-	0.35	-	0.35
Aroclor1254: Fish	4.4×10^{-6} *Oral mg/kg-d*	2×10^{-5} *mg/kg-d*	-	0.22	-	0.22
BDCM[b]: Tap water	1.0×10^{-2} *Oral mg/kg-d*	2×10^{-2} *mg/kg-d*	0.50	-	-	0.94
Air	8.7×10^{-3} *Inh mg/m³*	2×10^{-2} *mg/kg-d*	-	-	0.44	
DCA: Tap water	1.2×10^{-2} *Oral mg/kg-d*	4×10^{-3} *mg/kg-d*	3.00	-	-	3.00
Air	1.0×10^{-2} *Inh mg/m³*	Not available	-	-	-	
			4.10		0.44	4.5
			Route-specific HIs			Multiroute HI

Fig. 14.2 Example of multiroute hazard index for exposure to four chemicals by two routes. Missing (−) entries indicate the value is not applicable (e.g., oral column and inhalation row). Dermal absorption, such as during bathing, is not considered in this example. Final MHI rounded to two significant figures. BDCM, bromodichloromethane; DCA, dichloroacetic acid; inh, inhalation route. [a] These chronic reference values are based on those available from U.S. EPA. For regulatory applications, always obtain the latest official reference values. [b] BDCM has only a subchronic reference concentration, which might overestimate the chronic value and thus underestimate the HQ for inhalation

easily showing which chemical has the multiroute exposure of most concern. Those multiroute HQs are then summed across all *J* chemicals to give the MHI for all chemicals by all routes. The example in Fig. 14.2 is devised to represent residential exposure to a contaminated river, which is the source of fish and drinking water. Because of showering and other indoor volatilization sources, the exposure involves inhalation as well as oral intake and thus has four chemicals and two exposure routes. The oral route has two sources: fish and tap water. The combined oral HI, 4.1, then reflects the four chemicals via the two sources. The sum of the multiroute HQs (right column) equals the sum of the two route-specific HIs (bottom row), giving the multiroute HI of 4.5.

14.2.2.2.4 Issue: Inclusion of Chemicals with Very Low Exposure Levels. Solution – Screen Out Chemicals Likely to Have Insignificant Exposures

Encountered mixtures, such as at hazardous waste sites or with urban air, often contain many chemicals, with some at or near detection limit exposure levels where

those limits are well below apparent toxicity thresholds. Those small exposures may contribute little to the overall mixture risk. One further simplification in the approach for Superfund waste sites is to remove a chemical from the list of chemicals of concern (CoC's) if its HQ <0.1 or if a lifetime component cancer risk is less than 10^{-6} (U.S. EPA 1993). By using the HQ and not merely the exposure level, the toxic potency of the chemical is automatically considered. Clearly, complex mixtures of 100 chemicals could be of concern if all chemicals had HQ = 0.09 (so HI = 9). This approach to screen out chemicals should then be limited to assessments involving defined mixtures of only a few chemicals.

14.2.2.2.5 Issue: Lack of Consideration of Toxicological Interactions. Solution – Weight of Evidence Judgment

As a rough application of the dose addition concept, the HI includes chemicals that are toxicologically similar and assumes no toxicological interactions. Yet some chemicals have strong evidence of interaction, including greater than dose-additive joint response (see example of lead (Pb) and cadmium (Cd) in Table 14.9), so ignoring such interactions can be another uncertainty and possible bias. As discussed in Chap. 13, most interaction studies only involve two chemicals (Mumtaz and Hertzberg 1993; Svendsgaard and Hertzberg 1994). The U.S. EPA and ATSDR have thus developed categorical weight of evidence (BINWOE) schemes for binary interactions and modifications of the HI to incorporate those BINWOE designations (ATSDR 2004a; U.S. EPA 2000b). Details are shown in Sect. 14.4.

14.2.3 Relative Potency Factor and Toxicity Equivalence Factor

In this section, dose addition is defined by the dilution concept (Bliss 1939) where the components behave as toxicological clones of each other, though with differing toxic potencies (see Chap. 9 for extended discussion of this and other versions of dose addition). Thus exposure to one chemical can be converted into an equivalent exposure of another chemical by simple potency-weighted scaling. Evidence for toxicological similarity is required for chemicals to be considered for such toxicity scaling (U.S. EPA 1999, 2000b, 2002a). The relative potency factor (RPF) approach is the more general of the two discussed here. As used by the U.S. EPA and the European Commission, usually a single RPF value is set for each chemical and assumed to be constant over the dose range of interest (SCHER et al. 2012; U.S. EPA 2000b). If information so indicates, different RPF values can be developed for different scenarios: exposure routes (e.g., oral vs. inhalation), durations (e.g., acute vs. lifetime), effects or target organs (e.g., hepatic vs. renal effects), or dose ranges. That potential scenario-specific option is one way U.S. EPA differentiates the RPF definition from that of the TEF. The toxicity equivalence factor (TEF)

approach requires more extensive evidence of essentially complete toxicological similarity because each TEF is assumed to apply to any exposure setting and to any effect (U.S. EPA 2000b, 2010). The TEF is then considered to be a special case of the RPF. The U.S. EPA has only used TEFs with dioxins and dioxin-like chemicals, where a key assumption is that the aryl hydrocarbon receptor mediates most if not all of their biologic and toxic effects.

One consequence of the dilution definition is that all components of the mixture should have geometrically congruent dose-response curves, which means that once each component dose (on the *x*-axis of a dose-response plot) has been scaled for toxic potency, the component response curves will be identical in shape (Hertzberg, et al., 2013). This leads to another consequence: that the component dose-response models are the same, except for each one's potency-based dose coefficient. For example, with a mixture of two toxicologically similar chemicals, if the model for chemical 1 is a simple exponential function:

$$y_1 = \gamma + \exp(\alpha + \beta_1 d_1) \tag{14.8}$$

then the model for chemical 2 is also a simple exponential function:

$$y_2 = \gamma + \exp(\alpha + \beta_2 d_2) \tag{14.9}$$

where only the dose coefficient, β, has changed. A related consequence is that the prediction model for the mixture response is the same dose-response model as for the components but with the dose term replaced by a linear combination of the component doses. This prediction model is the dose addition formula. For this exponential example, the dose-additive mixture prediction model is:

$$y_{\text{mix}} = \gamma + \exp(\alpha + \beta_1 d_1 + \beta_2 d_2) \tag{14.10}$$

Once dose addition is assumed, the dose coefficients from the individual chemical models in Eqs. 14.8 and 14.9 are plugged into the prediction model (Eq. 14.10) to calculate the estimated mixture response. The above properties are key to the applications of relative potency factors discussed in the next section.

14.2.3.1 Relative Potency Factors

The most often used direct application of the dose addition model in health risk assessment is the relative potency factor (RPF) approach. In contrast to the HI approach, which is used as a decision index, the RPF approach actually estimates the mixture risk or response and is a direct application of dose addition as defined by the concept of simple similar action and discussed previously in Eqs. 14.8, 14.9, and 14.10. Once toxicological similarity has been decided, then dose addition needs the constants representing the proportional differences in toxic potency. While one could estimate the RPFs as the ratio of the mixture model coefficients (e.g., the βs in Eq. 14.10 if developed from mixture data), mixture data are usually lacking, so the

most common component-based application uses the ratio of individual chemical toxicity-specific values of the mixture component chemicals. If the application is to fairly low doses, then each chemical's toxicity-specific value might be set at, say, the ED_{10}. Then the RPF is the ratio of ED_{10} values:

$$RPF_{21} = \frac{ED_{10,1}}{ED_{10,2}} \tag{14.11}$$

That ratio of equitoxic doses is the potency of chemical 2 compared with chemical 1. To see how this calculation works, if chemical 2 is more toxic than chemical 1, then RPF_{21} should be >1; in Eq. 14.11 its ED_{10} will be smaller than that of chemical 1, and so indeed $RPF_{21} > 1$.

In an RPF-based assessment as presently applied (U.S. EPA 2003a), actual exposure levels are potency-weighted by the set of RPFs, often official values set by regulatory agencies. Those RPFs are assumed to be constant across all response levels. Specifically, the exposure levels of the component chemicals are converted into equivalent exposure levels of another chemical, usually called the index chemical (IC), using the RPFs. The index chemical is usually selected to be the chemical in the group or mixture that has the most robust dose-response data and is most representative of the toxicological profile of the other chemicals in the group or mixture being assessed. Because the more toxic chemicals are often better funded for research than weakly toxic chemicals, in many cases the index chemical is also the most toxic component (e.g., benzo(a)pyrene for polycyclic aromatic hydrocarbons). Each RPF is the ratio of a chemical's toxic potency to that of the index chemical, so in Eq. 14.11 chemical 1 is the IC. Several toxicological and statistical methods have been described for estimating RPFs (Altenburger et al. 2005; Budinsky et al. 2006; Chen et al. 2003; Dinse and Umbach 2011, 2012; Wolansky et al. 2006). The following example uses the ED_{10} as the toxicity value.

Carbamate Example The U.S. EPA has published RPFs for carbamate pesticides along with the oral ED_{10} values for rat brain cholinesterase inhibition used in their calculation (U.S. EPA 2007d). In this example, the RPF is effect-specific, applying here to the primary effect of concern, brain cholinesterase inhibition. Values for the index chemical, oxamyl, and for two other carbamates, carbaryl and propoxur, are in Table 14.2 along with the resulting RPF values. For carbaryl, its oral ED_{10} is larger than that of oxamyl so carbaryl is less toxic, and its RPF is correspondingly less than 1.0. An oral dose of 5 mg/kg of carbaryl is then roughly equitoxic to an oxamyl oral dose of 0.75 (= 5*0.15) mg/kg.

This relationship between the similar chemicals allows the toxic response for a poorly studied chemical to be predicted from that of a well-studied chemical merely by scaling the dose. In a strict interpretation of the dilution concept, not only is the mean response predicted from the index chemical but also the response variance, i.e., the data spread for the poorly studied chemical should be the same as the data spread observed for the well-studied index chemical and depend only on the response level, not the chemical in the exposure (Hertzberg et al. 2013). More common is for just the

Table 14.2 Example oral RPF values for three carbamates with oxamyl as the index chemical for brain cholinesterase inhibition

Chemical	ED10 (mg/kg)	RPF
Carbaryl	1.58	0.15
Propoxur	2.09	0.11
Oxamyl	0.24	1.00

Adapted from U.S. EPA (2007d)

mean effect to be so estimated. For example, an ED_{05} for the dioxin 1,2,3,7,8,9-HxCDD could be estimated from the ED_{05} for the dioxin index chemical, 2,3,7,8-TCDD, by multiplying the latter dose by 10, which is the inverse of the HxCDD's TEF of 0.1 (van den Berg et al. 2006). The RPFs (and TEFs) are not always developed from ED_x values or other derivatives of in vivo dose-response data, so much of the uncertainty in such extrapolations could be in the RPF value itself.

The most common use of the RPFs is in estimating a mixture response. The exposure levels of all the mixture components are scaled (multiplied) by their RPF values and then summed to give the mixture exposure level as the equivalent index chemical's exposure level. For oral exposure as a daily dose, the index chemical equivalent dose (ICED) is then the weighted sum of the component doses (d_i):

$$\text{ICED}_{\text{mix}} = \sum_{i=1}^{n} d_i{}^{*}\text{RPF}_i \tag{14.12}$$

Note that the dose for the index chemical does not change in Eq. 14.12 because its RPF by definition is 1. For an exposure (mg/kg) of 2.0, 3.5, and 2.2 to carbaryl, propoxur, and oxamyl (the three pesticides in Table 14.2), the equivalent dose of oxamyl is calculated, using Eq. 14.12, as:

$$\text{ICED}_{\text{oxamyl}} = 2.0{}^{*}0.15 + 3.5{}^{*}0.11 + 2.2 = 2.9\ (\text{mg/kg})$$

This concept of summing the equivalent doses is shown in Fig. 14.3 for a two-chemical example, with the ICED on the *x*-axis. The incremental exposure from each component increases the ICED but still follows the same dose-response curve of the index chemical, which is chemical 1 in that figure.

14.2.3.2 Toxicity Equivalence Factors

The toxicity equivalence factor (TEF, also called toxic equivalency factor) is mathematically similar to the RPF (ratio of equitoxic doses), and the associated mixture methods are mathematically equivalent, including that the TEF is assumed to be constant across the dose range of interest. Thus far, the TEF method has been used to assess mixtures of chlorinated and brominated dioxins and dioxin-like compounds

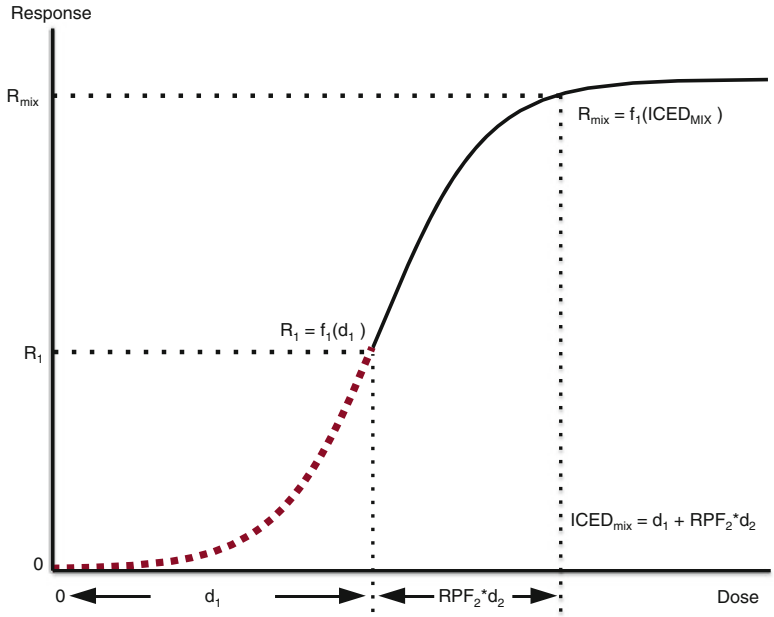

Fig. 14.3 Dose addition using RPFs to determine the mixture's response from its index chemical equivalent dose (ICED). U.S. EPA's dilution concept means both chemicals follow the same dose-response model, once the non-index chemical doses are scaled by their RPFs. Curve is for chemical 1. *Red-dashed* portion indicates response from chemical 1 alone. The dose of chemical 2 has been scaled to its equivalent dose of chemical 1

(DLCs), and values have been developed for humans and for wildlife (Alexeeff et al. 2007; U.S. EPA 2001c; van den Berg et al. 1998, 2013). Because the TEF approach assumes essentially complete similarity, where each component acts as a true dilution or concentration of any other component, that allows one TEF to be calculated for each chemical and applied to any endpoint, any response level, and any exposure setting (duration, route). Such an interpretation requires a high degree of evidence for toxicological similarity. The rationale for the TEFs to be applicable to multiple endpoints is that the endpoints are all assumed to result from a common mechanism of action or causal pathway, that is, the DLC binding to the aryl hydrocarbon receptor or AhR. For the dioxins and dioxin-like chemicals, U.S. EPA (2010) on p. 14 explains the evidence supporting use of TEFs:

> The EPA recommends these TEFs be used for all cancer and noncancer effects that appear to be mediated through AhR binding by the DLCs. EPA recognizes that this issue will require further evaluation as additional toxicity data become available. Eventually, endpoint specific TEFs or separate TEFs for systemic toxicity and carcinogenicity endpoints may need to be developed.

This extent of understanding and consistency of the mechanism of action rarely exists, and so TEFs are expected to apply to only a few chemical groups

Table 14.3 Comparison of TEF and RPF characteristics

Toxicity equivalence factor	Relative potency factor
Specific type of RPF	*Generalized case*
All health endpoints	May be limited (e.g., reproductive toxicity)
All routes	May be limited (e.g., oral)
All time frames of exposure	May be limited (e.g., acute)
Encompasses all doses	May be limited to specific dose range
Implies more abundant data are available	May be based on lower quality/fewer data
Implies greater certainty that mechanism of action is common to all effects	Assumes similar MOA or AOP May be more accurate because application can be constrained to the available data
One TEF set for all scenarios	Can generate different RPF sets for various scenarios (e.g., exposure route, duration)

Adapted from U.S. EPA (2000b)

(Table 14.3). This strict interpretation of the TEF is mainly endorsed by U.S. EPA and is a useful short-hand notation for a decision that the chemicals in the group are highly similar in their toxicological processes. When mixture component chemicals show toxicological differences in some aspects, such as different potencies by endpoint or by exposure route, the RPF designation should be used to clearly denote this lack of complete toxicological similarity.

14.2.3.3 Uncertainties with RPFs and TEFs

The RPF and TEF approaches can have significant uncertainties, both toxicologically and quantitatively. Evidence for toxicological interaction, i.e., departure from dose additivity, is not reflected in the numerical values for the RPF or TEF. While some of the uncertainty is reflected by the custom of only assigning order of magnitude values (e.g., 0.1, 10) or one or two significant digits, it is common, especially with official regulatory values, for the numbers to be incorporated into an index chemical equivalent dose (ICED) that is presented alone, with no quantitative uncertainty discussion or representation as a range (e.g., confidence interval). While TEF values are described as based on strong understanding of a common toxicological pathway across the components, information on some components could be much less than desired, e.g., from in vitro studies, different exposure routes, or quantitative structure-activity relationship (QSAR) extrapolation. Many such examples are in the original U.S. EPA dioxin report (U.S. EPA 1989a). Even if the same endpoint measure is used for RPF estimation across all components, the value depends on the specific response level (e.g., ED_x) chosen. When ED_x values are used for the endpoint measure, the ED_{50} is usually considered more statistically robust, while those for lower responses, e.g., ED_{10} or ED_{05}, are usually considered more relevant to exposures used in environmental assessments. The RPF values can also vary with the dose metric used, e.g., administered dose rate, cumulative dose over a time frame (area under the curve), or concentration in target tissue. Only a few

researchers have investigated the potential for alternative values and the sensitivity of the final assessment (Budinsky et al. 2006; Parvez et al. 2013; U.S. EPA 2007b).

Numerical uncertainties can arise from inconsistent data (high variance) or from dependence on the specific statistical method used. Some consideration should always be given to statistical approaches deemed most applicable to the data being analyzed. Regulatory consistency would improve, however, if official statistical methods were listed, if not required, such as those in the U.S. EPA benchmark dose procedure as implemented in their BMDS project (U.S. EPA 2016). Some uncertainties derive from policy decisions, such as numbers being outdated because of official values remaining unchanged pending formal revision and review processes.

Both the RPF and TEF approaches use an index chemical. Once component doses are scaled to their ICED values, the predicted mixture response is determined from the dose-response model for the index chemical. Because of variations in the underlying data across the component chemicals, a change of choice of index chemical could lead to changes in the predicted mixture risk even though conceptually they should be the same (Chen et al. 2003; Chen et al. 2001). The sensitivity to choice of index chemical is also affected by any dissimilarity of the component dose-response curves, e.g., when isoboles are curvilinear (Bosgra et al. 2009).

14.2.3.4 U.S. EPA Margin of Exposure

The Office of Pesticide Programs (OPP) in U.S. EPA has developed conceptual and risk estimation guidance for addressing mixtures of pesticides under the Food Quality Protection Act (FQPA) of 1996. This section discusses some of that guidance as it pertains to the interpretation and application of dose addition. While developed for pesticides, the following concepts and approaches (common mechanism group and margin of exposure) can also be applied to mixtures of other toxicologically similar chemicals.

Common Mechanism Group Toxicological guidance was published for determining which chemicals should be included in a common mechanism group, with the first application to the organophosphates (U.S. EPA 1999, 2002a). That guidance is fairly detailed and carefully written including clear definitions of key terms and should be applicable to any set of chemicals under consideration for grouping by toxicological similarity. Only a very brief summary of the steps is given here. The guidance stresses that "mechanism," the term explicitly used in the FQPA text, is not always well defined and so interprets that term as similar to the term "mode of action" (MOA) that is used elsewhere. To be consistent with the U.S. EPA risk guidance for pesticide mixtures, this section uses the term "mechanism," even though MOA might be more appropriate.

Step 1. Identify a candidate set of substances that might cause a common toxic effect by a common mechanism of toxicity.

Step 2. Definitively identify those substances from Step 1 that cause a common effect.
Step 3. Determine the toxic mechanism(s) by which each substance causes a common toxic effect.
Steps 4 and 5. Comparison of mechanisms of toxicity (step 4) and refined grouping of substances (step 5).

As practiced by U.S. EPA, this grouping process also includes literature searches and, when deemed necessary, external reviews.

Margin of Exposure Calculation for Pesticide Mixtures The margin of exposure (MOE) construct is a hybrid dose-response approach that uses RPFs in the calculation but results in a decision index similar to the HI. The MOE for a single pesticide is the ratio between a toxicity benchmark or point of departure (POD) and a human exposure level. The POD is typically a no-observed-adverse-effect level (NOAEL), lowest-observed-adverse-effect level, or a lower bound on the benchmark dose (BMDL) from an animal bioassay. In general, as the MOE decreases below a specified value called the "target MOE," the concern for toxicity increases. That "specified value" depends on the POD and the way uncertainty and extrapolation factors are used. For example, if the POD is a NOAEL from a chronic rodent assay, then the target MOE is typically 100, similar to the two tenfold uncertainty factors of intraspecies and interspecies that are associated with scaling from a rodent assay to humans (see Chap. 15 for discussion of the basic uncertainty factors used by U.S. EPA). For this example, an exposure with MOE $>$ 100 would then be considered of no cause for concern. The target MOE for an individual chemical is then chemical specific. MOEs higher than the target value are considered more health protective.

For mixtures of pesticides that are in a common mechanism group, a similar POD-to-human exposure ratio is used but is based on the relative potency factor (RPF) approach. First, the component exposures (E) are converted into their equivalent exposure levels for the index chemical ($E \times$ RPF). For a single exposure route, the J-chemical mixture MOE is calculated as the ratio of the index chemical POD and the human index chemical equivalent exposure for that route (U.S. EPA 2002a). In Eq. 14.13, the denominator sum is the mixture exposure expressed as the equivalent index chemical exposure:

$$\mathrm{MOE} = \frac{\mathrm{POD}_{\mathrm{Index}}}{\sum_{j=1}^{J} E_j \times \mathrm{RPF}_j} \tag{14.13}$$

For multiroute exposures, the route-specific MOE values are combined into a total MOE. For example, with combination exposures by oral, dermal, and inhalation routes, the total MOE is:

$$\mathrm{MOE}_{\mathrm{Total}} = \frac{1}{\frac{1}{\mathrm{MOE}_{\mathrm{Oral}}} + \frac{1}{\mathrm{MOE}_{\mathrm{Dermal}}} + \frac{1}{\mathrm{MOE}_{\mathrm{Inhalation}}}} \tag{14.14}$$

With the MOE in Eq. 14.13, the POD represents the toxicology study value for the index chemical. The specific uncertainty factors for the individual chemicals in the mixture are compared (e.g., interspecies compared to interspecies, intraspecies compared to intraspecies). Pesticide assessments by U.S. EPA under the FQPA also can include an additional uncertainty factor, one for child sensitivity, called here the FQPA factor. When the values of a specific uncertainty factor vary across the assessment group, the uncertainty factor value specific to each chemical is incorporated into the "adjusted RPF" value for that chemical. For example, consider a mixture where most components have chronic data but chemical 2 only has subchronic data. Then the RPF for chemical 2 using subchronic data for chemical 2 but chronic data for the index chemical would be multiplied by a duration uncertainty factor; the resulting ratio approximates a chronic RPF for that chemical, making the adjusted RPF conceptually equivalent to the RPF for chemical 1. When the value of an uncertainty factor within a group is the same for each chemical, it is used in setting the target MOE (U.S. EPA 2007d).

Carbamate Example Four carbamates were evaluated for dermal exposure and had differing uncertainty factors for interspecies and child sensitivity (see Table 14.4). The only uncertainty factor in common is the intraspecies factor (10), sometimes called the extrapolation from healthy to sensitive people, and is the only factor not accounted for in the adjusted RPFs; thus the target MOE for that mixture is 10.

Strengths and Weaknesses of the MOE Approach The similarity to the HI is the interpretation, where the mixture MOE is compared with a decision value and so provides an indication of risk to inform decisions. Thus, even though RPFs are used in the calculation, the MOE is not an explicit estimate of either probabilistic risk or of the measured effect from the mixture, as could be obtained with the standard RPF approach described in Sect. 14.2.3.1. On the other hand, the mixture MOE formula as used by the U.S. EPA/OPP for pesticide combinations is restricted to common mechanism chemicals, and so is more akin to the TTD-based HI, albeit with more extensive requirements of evidence about toxicological similarity (U.S. EPA 1999,

Table 14.4 Adjusted dermal relative potency factors for four carbamates, for children and adults based on interspecies and Food Quality Protection Act (FQPA) specific factors

Chemical[a]	Dermal RPF	Interspecies factor	Adjusted RPF adults	FQPA factor children only	Adjusted RPF children
Carbaryl	0.71	10	7.1	1.8	13
Methiocarb	0.09	10	0.9	10	9
Oxamyl	1.00	3	3	3.48	10
Propoxur	0.03	10	0.3	10	3

Adapted from U.S. EPA (2007d)
[a]Oxamyl is the index chemical for this group

2002a, 2007d). Because the MOE approach described here uses RPFs, it more closely reflects the dose addition concept than does the HI approach. The MOE approach as used by U.S. EPA/OPP is more flexible than the usual RPF approach because, if dose-response data are not adequate for the index chemical formula (i.e., for modeling and estimating RPFs), the mixture MOE can be calculated using NOAEL or LOAEL values for the individual pesticides. One weakness is in risk communication. Unlike the HI, which is always compared with 1.0, the MOE has no fixed comparison value because the target MOE varies with the data. As with any dose-additive approach, evidence of toxicological interaction would increase the uncertainty in the mixture MOE assessment and should be included in the uncertainty discussion.

14.3 Risk Assessment Methods for Toxicologically Independent Chemicals

Toxicological independence is perhaps the simplest concept of joint action: chemicals in a joint exposure have no influence on each other's toxicity. Of course, all effects interact when they become severe, such as liver failure resulting in damage and dysfunction of virtually all other organs. For most chemical risk assessments, the potential for affecting multiple organs and tissues is acknowledged and sometimes reflected in the risk assessment. In this section, the focus is on chemicals that have some aspect of toxicity in common but where the sequences of steps from exposure to effect are considered independent across the chemicals. Usually the commonality is expressed by a general category of toxicity (e.g., liver pathology) and the independence evidenced by differing types of cellular damage or toxic mechanisms. For example, if several chemicals cause cancer but in different target organs, the overall cancer assessment might assume that at low doses, the chemicals' toxic actions are independent. To help clarify the difference between toxicological research and risk assessment, this section focuses on the formulas and terminologies used in risk assessment. Mixture risk assessment formulas that reflect the concept of independent toxicological processes are called "response addition" for probabilities and "effect summation" for measured endpoints; those terms (defined below) may not have the same use and definitions in other publications.

14.3.1 Background on Use of Toxicological Independence in Risk Assessment

Risk assessment formulas based on toxicological independence have been used in safety and regulatory practice for many years. In fact, the first mixture risk estimates for cancer published by the U.S. EPA used simple addition of the individual

chemical risks (U.S. EPA 1986b). The easiest implementation is where the predicted mixture toxicity is merely the collection of the component toxic effects, often assuming only one effect per chemical. The next simplest approach is the summation of individual responses, regardless of the response measure, and is often called effect summation. The next and most common approach is response addition where response is a probabilistic risk (or fraction of population affected). The three formulas for these approaches are, for a simple mixture of two chemicals:

$$y_{\text{mix}}(d_1, d_2) = (y_1, y_2) \tag{14.15}$$

$$y_{\text{mix}}(d_1, d_2) = y_1(d_1) + y_2(d_2) \tag{14.16}$$

$$y_{\text{mix}}(d_1, d_2) = p_1(d_1) + p_2(d_2) - p_1(d_1) \cdot p_2(d_2) \tag{14.17}$$

where y_i is the response (however measured) caused by the ith chemical. Note that p_i is used in Eq. 14.15 to clarify if the response is a probability or population fraction (thus bounded by [0,1]). As an example of the collection approach in Eq. 14.13, if one chemical causes cancer (risk p_1) and the other reproductive effects (risk p_2), then the joint risk could be described by the pair, p_1 (cancer) and p_2 (reproductive effects), with possible concern for different population groups. In all three formulas, the prediction is usually considered plausible for small incremental effects and for chemicals with distinct critical effects (thus no other effects seen until a much higher dose). The collection approach of Eq. 14.13 is trivial to implement and was previously used by U.S. EPA and other agencies for simple mixtures of a carcinogen with a noncarcinogen; the result was an estimate of cancer risk (from the first chemical) and an HQ for the noncancer effect (from the second chemical). With more emphasis now on multiple effects, it is rarely used and is not further discussed here. The assumptions and uncertainties for response addition and effect summation are discussed in more detail in the corresponding subsections that follow.

14.3.2 Response as Probability

Risk assessments using probability as the response measure are relatively rare when based on mammalian experiments because the measured dose-response information is usually not expressed using probability. The most well-known exception in toxicology literature is where mortality is the endpoint, especially that of the target species of various pesticides (Bliss 1939). For human risks, the most common probabilistic endpoints have been mortality and cancer, but the concept can be applied to any toxic endpoint where the response data include probabilities or population fractions. The term "risk" is used here as a synonym for response probability.

When exposures are low and toxicological independence can be assumed, the mixture toxicity risk is the statistical combination of the component toxicity risks. The risk from exposure to the mixture then follows the probabilistic formula for independent events (see Chap. 9 for a more complete discussion, including several

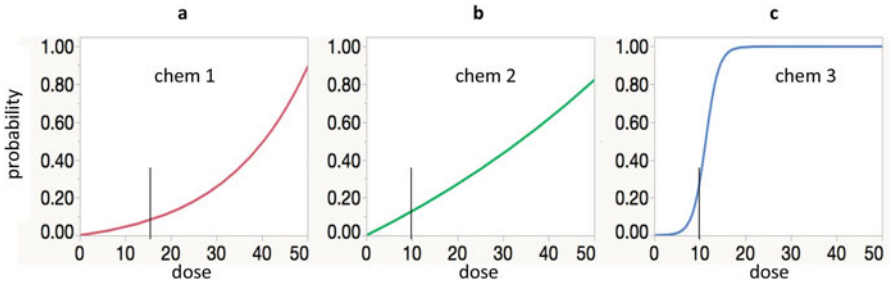

Fig. 14.4 Response addition for probabilities showing that with independence-based formulas, curve shape does not matter. Doses and probabilities at vertical lines are (**a**) $d1 = 15, p1 = 0.08$; (**b**) $d2 = 10$, $p2 = 0.13$; and (**c**) $d3 = 10$, $p3 = 0.29$. Mixture estimated response uses Eq. 14.20: $p_{\text{mix}} = 1-(1-0.08)(1-0.13)(1-0.29) = 1-(0.92*0.87*0.71) = 0.43$

variations on the concept). The general formula uses the assumption that the individual chemical survival events are independent. For a mixture of two chemicals, if *w(d)* is the survival probability from dose *d*, then:

$$w_{\text{mix}}(d_1, d_2) = w_1(d_1) \cdot w_2(d_2) \tag{14.18}$$

The risk of toxic effects, *p*, is then $1-w$:

$$\begin{aligned} p_{\text{mix}} &= 1 - w_{\text{mix}} \\ &= 1 - (1 - p_1) \cdot (1 - p_2) \\ &= p_1 + p_2 - (p_1 p_2) \end{aligned} \tag{14.19}$$

where each p_i is the risk (or dose-response) function for chemical *i* evaluated at dose d_i. Equation 14.19 is the same as those shown in Chap. 9 (Sect. 9.3 on independent action). The general response addition formula for *J* chemicals is:

$$p_{\text{mix}} = 1 - \prod_{j=1}^{J} (1 - p_j) \tag{14.20}$$

In contrast to the dilution definition of dose addition (Sect. 14.2.3), for response addition, there is no constraint on the shape of the component dose-response curves. For example, a three-chemical mixture could have component dose-response curves of quite different shapes (Fig. 14.4) because the only values entering the formula are the risks at the component doses, whether a component curve has a threshold dose; is nonlinear, convex, or concave; and has no bearing on the mixture risk calculation. The expected response for the mixture is calculated with Eq. 14.20 using only the three response values indicated: where each vertical line crosses the curve (0.08, 0.13, 0.29).

Cancer risk assessments by the U.S. EPA during the agency's early years involved multiplying the exposure level by a cancer potency value (U.S. EPA 1986a). Those potency values are still determined for many chemicals and are

included in the U.S. EPA's IRIS database of risk values as plausible upper-bound estimates of the increased cancer risk from lifetime exposure to one exposure unit of the chemical. Those estimates include the oral slope factor (SF) for oral intake of 1 mg/kg/day and unit risk value (UR) for inhalation exposure to a concentration of 1 μg/m^3 or ingestion exposure to a water concentration of 1 μg/L. For one chemical, U.S. EPA can then have three cancer potency values depending on the exposure route and units used. For example, application of Eq. 14.17 to a combination exposure of two cancer-causing chemicals, one by oral intake with dose d_1 (thus p_1=d_1*SF$_1$) and one by inhalation at concentration c_2 (thus $p_2 = c_2$*UR$_2$), is then calculated as:

$$p_{\text{mix}}(d_1, c_2) = (\text{SF}_1 \cdot d_1) + (\text{UR}_2 \cdot c_2) - (\text{SF}_1 \cdot d_1)(\text{UR}_2 \cdot c_2) \tag{14.21}$$

Note that the units of each potency estimate are the inverse of the corresponding exposure units, so that each term in parentheses in Eq. 14.21 is dimensionless, allowing their combination into a dimensionless risk value for the mixture.

14.3.3 Effect Summation

Effect summation is infrequently used for mixture response estimation. It has been strongly criticized on theoretical grounds but for reasons not related to the usual setting for environmental health risk assessment (Berenbaum 1989; Muska and Weber 1977; Plackett and Hewlett 1948). All criticisms located thus far focus on the implausible estimates for high-component response levels (see Chap. 9, Sect. 9.5). For example, with measured responses, such as changes in relative liver weight, the simple sum could exceed physiological limits. For response such as the population fraction showing adverse effects, the simple sum could exceed 100%. Effect summation can be useful if it is restricted in application in a similar way as is done for cancer risk (e.g., risk less than 1% per chemical) and noncancer effects (e.g., less serious effects, well below doses causing frank or lethal effects), that is, to the dose range where component effects are small. With such a restriction, effect summation can be plausible both numerically and in terms of the assumption of toxicological independence. Because of inadequate investigation into extrapolation issues, such as cross-species concordance of the percent incremental change for some endpoints, the U.S. EPA guidance recommends restricting effect summation to the toxic effects and dose ranges in the cited studies. The risk estimate for the mixture then is determined using the estimated risk for the study mixture and its scenario (U.S. EPA 2000b). For example, if a study on rats used effect summation for cholinesterase inhibition by three pesticides, with raw effect measures of 3%, 5%, and 1%, then the estimated response for the mixture is 9% reduction in cholinesterase activity from the control activity. That estimate would then be translated into a human mixture risk estimate using standard approaches for rat to human extrapolation and any other quantitative adjustments applied to animal studies.

14.3.4 Uncertainties with Independence Methods

The key biological uncertainty in application of the risk formulas in this section is their underlying assumption of toxicological independence. Rarely is that assumption statistically tested and rarely is there adequate AOP information on all of the mixture chemicals to support such independence as a biological concept. At present, no guidance from regulatory agencies has been found for determining whether multiple chemicals are toxicologically independent, a complicating factor in mixture risk assessment (Lambert and Lipscomb 2007; McCarty and Borgert 2006). Consequently the choice of an independence formula is usually based on judgment, such as being selected because of the lack of sufficient evidence for toxicological similarity or other contrary information.

Regulatory applications of response addition raise a concern similar to that of the HI: each chemical's risk estimate is conservatively calculated, so combining many of them will exacerbate that bias into a gross overestimate of the mixture risk. Response addition is easier to address because its dose-response data are numerical, e.g., the fraction of a dose group with adverse effects. For example, individual chemical cancer risk estimates by U.S. EPA have often been described as upper bounds where the true risk can be lower (U.S. EPA 2005). Statistical research has shown that the extent of the overestimate is not large. For mixtures of up to 20 chemicals, one theoretical statistical study showed that if each potency estimate were a statistical upper 95% confidence limit, the true 95% upper bound potency estimate for the mixture was likely to be 1.6-fold or less below the sum of the upper-bound potency estimates (Cogliano 1997). While that suggests a simple correction factor could be used, the U.S. EPA's human cancer potency value is not a statistical limit. Instead, several conservative steps are involved to ensure that the true risk is unlikely to be higher than the estimate. The statistical results, however, do provide some indication of the degree of conservatism in the mixture risk estimate. More of such results should be obtained, involving a wider variety of carcinogens, before quantitative modifications to the mixture risk estimation can be recommended.

14.4 Interactions Using Qualitative and Quantitative Approaches

14.4.1 Weight of Evidence Approach for Binary Toxicological Interactions

The HI is an exposure-based decision approach. While the HI is loosely based on dose additivity (see Sect. 14.2.2), when applied to site assessments, it is used in conjunction with biomedical judgment, community-specific health outcome data, and community health concerns to assess the degree of public health hazard. The HI approach is most useful when augmented by information on toxicological

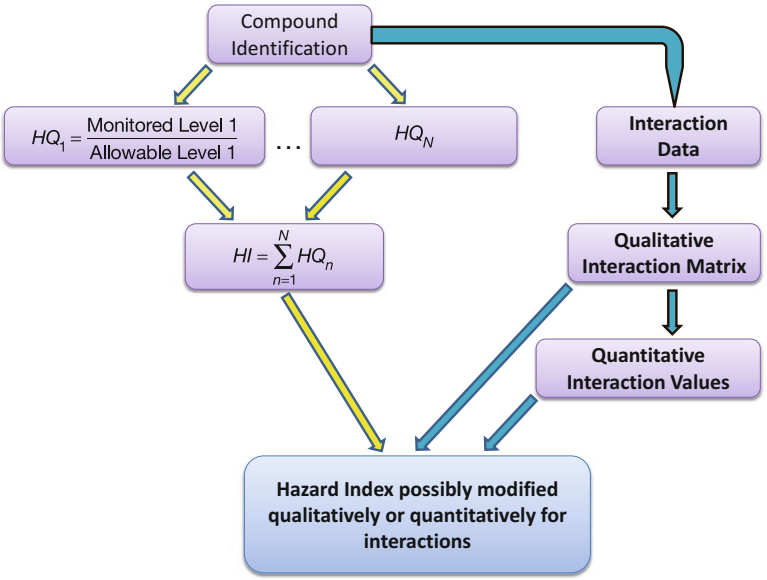

Fig. 14.5 Steps in developing a hazard index and modifying it by incorporating information on toxicological interactions. The qualitative interaction matrix (ATSDR) and quantitative interaction values (ATSDR and U.S. EPA) are described in Sect. 14.4.1. The U.S. EPA's interaction-based HI, where each HQ is adjusted by the interaction values, is described in Sect. 14.4.2

interactions. Although quantitative estimates of interaction are rare, a feasible approach is a weight of evidence evaluation of the potential for interactions (departures from dose addition) among the components in the mixture. Several methods have been proposed to incorporate interaction data into chemical mixture risk assessment that use dose additivity. The goal has been to express the available interaction and experimental data in a sufficiently simple way for use in the risk assessment process but also to be both scientifically plausible and procedurally practical (Arcos et al. 1995; Mumtaz and Durkin 1992). The idea is to initially use the information qualitatively and then if possible quantitatively (Fig. 14.5). Both of these methods, qualitative and quantitative, are based on the use of binary (pairwise) interaction data, the most common type of interaction data found in scientific literature.

A binary weight of evidence (BINWOE) scheme for use of interaction data from studies of systemic toxicity has been described in guidelines and used for chemical mixtures risk assessment (ATSDR 2004a; Mumtaz and Durkin 1992; Pohl et al. 2009; U.S. EPA 2000b). It is a subjective/judgmental method, motivated by the International Agency for Research on Cancer (IARC) classification scheme for carcinogens (International Agency for Research on Cancer (IARC) 1982), that can be used to integrate the interaction data into the overall risk assessment of the mixture. The process starts with a thorough literature search and review of all relevant information on all possible binary combinations of chemicals in the mixture

Table 14.5 U.S. EPA's BINWOE classification

Category	Description	Score for greater than D-A	Score for less than D-A
I	The interaction has been shown to be relevant to human health effects, and the direction of the interaction is unequivocal	1.0	−1.0
II	The direction of the interaction has been demonstrated in vivo in an appropriate animal model, and the relevance to potential human health effects is likely	0.75	−0.5
III	An interaction in a particular direction is plausible, but the evidence supporting the interaction and its relevance to human health effects is weak	0.50	0.0
IV	The assumption of additivity has been demonstrated or must be accepted because of insufficient interaction data	0.0	0.0

Adapted from U.S. EPA (2000b)

of interest. This review considers, in addition to available binary interaction data, information on the toxicity and pharmacokinetics of the individual compounds as well as interaction data on related compounds. All pertinent information is collapsed into a qualitative weight of evidence determination using defined criteria.

The U.S. EPA BINWOE classification is simple (Table 14.5), mainly involving judgment of the extent of extrapolation required. Scores are assigned based on the quality and direction of the interaction. The associated scores are biased toward health protection in that less-than-additive interaction has a lower absolute score than greater-than-additive interaction (unless evidence is very strong, where both have absolute values = 1). The scores modify the hazard quotients in the additive HI to reflect the likely pairwise interactions (see details in Sect. 14.4.2), e.g., strong evidence of greater-than-additive interactions and no evidence for less-than-additive interactions would increase the HI value.

For each pair of chemicals in the mixture of concern, two BINWOEs can be derived: one for the effect of the first chemical on the toxicity of the second chemical and another for the effect of the second chemical on the toxicity of the first chemical. This is because the interactions of a pair are not always bidirectional and usually consist of varying quality and degrees of detail. For example, a substantial amount of information could be known that PCBs will enhance the toxicity of carbon tetrachloride and the mechanisms of this enhancement are well understood. Much less, however, could be known about the effect of carbon tetrachloride on the toxicity of PCBs. For some binary chemical combinations, upon the availability of data, multiple BINWOEs can then be developed, including for multiple target organs or multiple specific effects.

The BINWOE classification of ATSDR is quite detailed (Table 14.6) and will be used for the remainder of this section. As illustrated in Fig. 14.6, each BINWOE

Table 14.6 ATSDR's BINWOE classification

Direction of interaction
= additivity, > greater than additivity, < less than additivity
Mechanistic understanding
I. Direct and unambiguous mechanistic data: the mechanism(s) by which the interactions could occur has been well characterized and leads to an unambiguous interpretation of the direction of the interaction
II. Mechanistic data on related compounds: the mechanism(s) by which the interactions could occur has not been well characterized for the chemicals of concern, but structure-activity relationships, either quantitative or informal, can be used to infer the likely mechanisms(s) and the direction of the interaction
III. Inadequate or ambiguous mechanistic data: the mechanism(s) by which the interactions could occur has not been well characterized, or information on the mechanism(s) does not clearly indicate the direction of the interaction
Toxicological significance
A. The toxicological significance of the interaction has been directly demonstrated
B. The toxicological significance of the interaction can be inferred or has been demonstrated for related chemicals
C. The toxicological significance of the interaction is unclear
Modifiers
1. Anticipated exposure duration and sequence
2. Different exposure or sequence
a. In vivo data
b. In vitro data
i. Anticipated route of exposure
ii. Different route of exposure

Adapted from Mumtaz and Durkin (1992) and ATSDR (2004a)

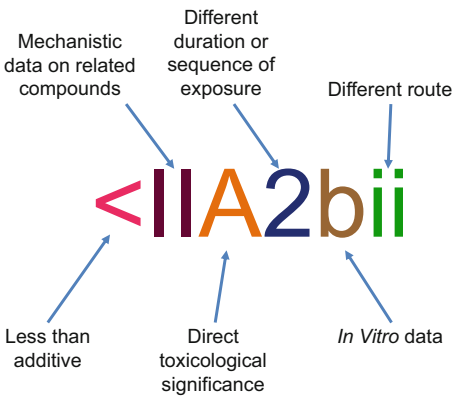

Fig. 14.6 Example interaction weight of evidence categorization used by ATSDR (Adapted from Mumtaz et al. 1994)

classification consists of a symbol indicating the direction of the interaction followed by an alphanumeric expression (Mumtaz et al. 1994). That expression can contain up to five components, but what was found to be more useful in actual applications is the direction and two major categories (mechanistic understanding, toxicological significance) along with target organ (Mumtaz et al. 2007; Pohl et al. 2009).

The BINWOE determination is an endpoint-specific classification that indicates the expected direction of an interaction (greater than dose-additive, less than dose-additive, dose-additive, or indeterminate) and categorizes the data qualitatively by using an alphanumeric scheme that considers mechanistic understanding, toxicological significance, and relevance of the exposure duration, sequence, bioassay (in vitro versus in vivo), and route of exposure (ATSDR 2004a). The first two components are major factors that capture the mechanistic information that supports the assessment and its toxicological significance. The first component expresses the level of understanding of the potential interaction, while the second determines the health impact of the interaction. Examples of these two BINWOE components are given below in the subsection, "Mixtures Evaluated."

The rating based on mechanistic understanding could be I, II, or III reflecting the quality of the available mechanistic information supporting the observation or assumption of a toxicological interaction and the extent to which the information supports its direction. A rating of I is given for the greatest confidence, i.e., the mechanism by which the interaction that occurs is well characterized and leads to a clear direction. A rating of II is given for the next level of confidence, e.g., the mechanism of interaction is not well understood, but structure-activity relationships (SAR) may be used to infer the mechanism. This rating is mainly intended to encourage the use of qualitative, quantitative, or informal SAR relationships. A rating of III is given for weakly understood interactions that have supporting mechanistic information that is poor.

The ratings for toxicological significance (A, B, or C) are parallel to the above ratings (I, II, or III) and are used to express confidence that the chemicals will interact in a way that will have significant impact on health. The highest rating (A) is given when the chemical interaction has been observed directly and is linked to a toxicologically significant endpoint. The B rating is given to those interactions that can be inferred. Finally, a rating of C is given to unclear interactions. The last three components of the BINWOE are modifiers that express the closeness of the available data to the conditions of the specific risk assessment in terms of the duration, sequence, routes of exposure, as well as the experimental models.

Once all of the qualitative BINWOE determinations have been made for each pair of compounds in the mixture, these are arrayed in a qualitative BINWOE matrix (Fig. 14.7). This matrix shows the 12 binary evaluations for four metals, lead, manganese, zinc, and copper, with the chemicals in each binary classification along the two axes (ATSDR 2004c). The diagonal line running from the upper left-hand corner to the lower right-hand corner corresponds to chemical identities, which are, by definition, dose additive and are left blank in the interaction matrix. The column headings indicate the chemicals that are affected by the compounds listed in the row headings. For example, the classification "<IIA" for the effect of copper on the toxicity of zinc is given in row four (copper) of column three (zinc). Similarly, the classification for the effect of lead on the toxicity of copper is "=IIIC," given in row one (lead) of column four (copper). Pairwise interactions are not always symmetric, e.g., the circled cells in Fig. 14.7 for Mn + Pb.

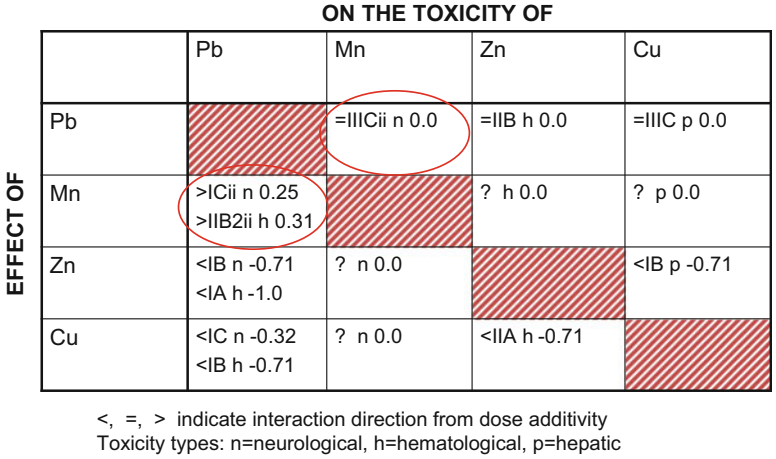

Fig. 14.7 Example results for four metals of their interaction weight of evidence categories and scores. Other types of toxicity were included in the evidence search; these three types were the most important. Note multiple toxicity types and lack of symmetry of interaction in circled cells for Mn + Pb. A zero value can show evidence of dose additivity (=) or lack of information (?) (Adapted from ATSDR 2004c)

The qualitative BINWOE matrix may then be used as a tool for qualifying the risk assessment for a particular site. Certainly, if all the individual BINWOE determinations indicated a high degree of confidence in one type of interaction – e.g., all the BINWOEs were ">I.A.1.a.i" – the risk assessor would have a strong indication that interactions that increased the toxicity were likely to be important for this mixture. Often, however, the qualitative BINWOE matrix will show a variety of potential interactions and degrees of confidence in the interactions, including several combinations with inadequate or missing data.

Mixtures Evaluated ATSDR performs health assessment of waste sites across the USA (ATSDR 2005b). Of the 1706 hazardous waste sites assessed, 743 have been found to have one or more completely characterized exposure pathways (Fay and Mumtaz 2004; Pohl et al. 2009). About 588 of these sites (79%) have at least two chemicals in a completed exposure pathway, and 475 sites (64%) have at least three. The most frequently found binary, tertiary, and quaternary combinations of chemicals have become the subject of interaction profiles (Table 14.7), documents that capture all pertinent information and assess the weight of evidence for interactions for these mixtures (ATSDR 2017; Mumtaz and Durkin 1992). Using the weight of evidence scheme, as of 2009, a total of 380 BINWOE evaluations were determined for target organ toxicities (Pohl et al. 2009). Of these, 156 (41%) indicated possible dose-additivity of effects ("="), 76 indicated synergism (greater than dose-additive effects ">"), and 57 indicated antagonism (less than dose-additive effects "<"). But a substantial number, 91, lacked the minimum information needed for making any assessments and, hence, were undetermined, thus highlighting data

Table 14.7 Mixtures evaluated in ATSDR interaction profiles using the BINWOE approach

Year	Chemicals in mixture
2004	Chlorinated dibenzo-p-dioxins (CDDs), hexachlorobenzene, dichlorodiphenyl dichloroethane (p,p0-DDE), methyl mercury, and polychlorinated biphenyls (PCBs)
2004	1,1,1-Trichloroethane, 1,1-dichloroethane, trichloroethylene, and tetrachloroethylene
2004	Arsenic, cadmium, chromium, and lead
2004	Copper, lead, manganese, and zinc
2004	Cesium, cobalt, PCBs, strontium, and trichloroethylene
2004	Arsenic, hydrazines, jet fuels, strontium, and trichloroethylene
2004	Cyanide, fluoride, nitrate, and uranium
2006	Atrazine, deethylatrazine, diazinon, nitrate, and simazine
2006	Chlorpyrifos, lead, mercury, and methylmercury
2007	Carbon monoxide, formaldehyde, methylene chloride, nitrogen dioxide, and tetrachloroethylene
2007	Chloroform, 1,1-dichloroethylene, trichloroethylene, and vinyl chloride

Adapted from http://www.atsdr.cdc.gov/interactionprofiles/index.asp

gaps. Of the 76 that indicated possible greater than dose-additive effects, only 16 pairs were ranked in Group I for mechanistic understanding, i.e., there was a good confidence in the data to determine the mechanism of interaction. These pairs include arsenic:benzo(a)pyrene (and vice versa); arsenic:chloroform (and vice versa); arsenic:polychlorinated biphenyls (PCBs); benzene:chloroform; benzo(a) pyrene:benzene; benzo(a)pyrene:chloroform (and vice versa); cadmium:chloroform (and vice versa); manganese:lead; PCBs:benzene; PCBs:benzo(a)pyrene; and PCBs: vinyl chloride (and vice versa). Of the 57 determinations that indicated possible less than dose-additive effects, only 12 pairs are in Group I for mechanistic understanding. These pairs include aspirin:butyl hydroxyanisole; copper:lead; nitrate:cyanide; zinc:cadmium; zinc:chloroform; zinc:lead; zinc:mercury; chloroform:trichloroethylene; 1, 1-dichloroethylene:trichloroethylene (and vice versa); and trichloroethylene: vinyl chloride (and vice versa). From all of the BINWOEs, only 46 were categorized as Group I for mechanistic understanding and 35 as Group A for toxicological significance.

As an example, the explanation of the hematotoxic effect of zinc (Zn) on lead (Pb) shown in Fig. 14.7 can be given as follows (ATSDR 2004c). A code of "<IA" determination indicates that mechanistic understanding and toxicological significance are well established for the interaction between these two chemicals and the evidence is substantiated from the published literature. The direction of interaction between zinc and lead is predicted to be less than additive based on a study in orally exposed children indicating a protective effect of zinc on the hematopoietic effects of lead and several intermediate-duration oral studies in rats that show protection by supplemental zinc against a number of hematological effects of lead related to heme synthesis, particularly at higher lead doses. The evidence for zinc inhibition of lead hematotoxicity is clear and toxicologically significant and is supported by clear mechanistic understanding that excess zinc protects and reactivates lead-inhibited

Table 14.8 ATSDR's BINWOE narrative for four organics common in rural well water

	On toxicity of			
Effect of	Atrazine	Simazine	Diazinon	Nitrate
Atrazine		Additive	Greater than additive	Greater than additive
		Reproductive effects	Neurological effects	Cancer effects
		High confidence	Medium confidence	Low confidence
Simazine	Additive		[Lack of information]	Greater than additive
	Reproductive effects			Cancer effects
	High confidence			Low confidence
Diazinon	Greater than additive	Greater than additive		[Lack of information]
	Neurological effects	Neurological Effects		
	Medium confidence	Medium confidence		
Nitrate	Greater than additive	Greater than additive	[Lack of information]	
	Cancer effects	Cancer effects		
	Low confidence	Low confidence		

Adapted from ATSDR (2006)

ALAD (delta-aminolevulinic acid dehydratase), decreases the absorption and tissue distribution of lead, and may induce proteins that sequester lead and donate zinc to ALAD and for other tissue needs. For risk communication with the public, this can be simply summarized as follows: oral administration of Zn could mitigate blood effects induced by Pb, and the confidence is high in this assessment.

The BINWOE process is different for atrazine and diazinon (Table 14.8). No studies were found that studied atrazine and diazinon interactions, but data were available for interaction between atrazine and chlorpyrifos, another organophosphate. It is known that all organophosphates cause toxicity by the same mechanism of action. Based on this structure-activity relationship understanding, the weight of evidence of atrazine and diazinon interaction was determined to be "$>$IIB," which is weaker in comparison to the zinc and lead interaction data ("$<$IA") where direct test results exist of the pairwise interaction. Explanation of the effect of atrazine on diazinon can be given as follows (ATSDR 2006). A code of "$>$IIB" indicates a greater than dose-additive effect for this interaction. Diazinon is a phosphorothioate organophosphorus insecticide that is metabolically activated through oxidative desulfuration to diazoxon by cytochrome P450. Diazoxon binds to acetylcholinesterase, inhibiting its ability to hydrolyze the neurotransmitter acetylcholine. The resulting accumulation of acetylcholine at the nerve endings causes continual neurological stimulation. This mechanism of action applies to both invertebrates and

mammals. Atrazine induced the metabolic activation of a similar phosphorothioate organophosphorus insecticide, chlorpyrifos, and potentiated its acute neurotoxicity to midges. Based on the similarity in structure and mechanism of action of diazinon and chlorpyrifos, a similar mechanism (induction of metabolic activation) can be inferred for atrazine's potentiation of the acute neurotoxicity of diazinon to midges in the same study. Because the mechanism of interaction is inferred from a similar chemical, a rating of II is chosen for mechanistic understanding, and because it is based on a similar structure and it is inferred, the rating of B is given for significance. Because of the many technical issues in inference and extrapolation of the interaction evidence, for the public, the ">IIB" might be more effectively communicated in words, such as the following: greater than dose-additive toxicity is possible when these two chemicals are present together, but the confidence is only medium in this assessment.

Experience shows that this kind of binary interaction determination can be successfully used in health assessments at hazardous waste sites, and the methodology works particularly well for chemicals with similar mechanism of action as demonstrated by laboratory confirmation of predictions for some nephrotoxicants (Mumtaz et al. 1998). Major data gaps do exist in our understanding of joint toxicity of many component mixtures. Hundreds of chemicals are introduced into the marketplace on a weekly basis, and the number of structure-searchable chemicals is in the millions (Demchuk et al. 2008). Testing all chemicals and their combinations is not economically feasible, so relying on expert judgment and computational methodologies (e.g., PBPK, QSAR modeling) will be increasingly used in future risk assessment. To facilitate this process, toxicological interaction data need to be assembled into a database that can be used to characterize the consistency of interactions and to define classes of compounds that display similar toxicological interactions (Durkin et al. 1995).

At ATSDR, the BINWOE determinations are used to qualitatively adjust the HI. For example, if the analysis indicates several binary combinations will have more than dose-additive joint toxic action, the HI is further analyzed. Conversely, if the analyses indicate several binary combinations will have less than dose-additive joint toxic action, the HI is considered adequate for the hazard presented by the exposure scenario. Such results are used for potential public health actions, including surveillance, health studies, community education, exposure investigations, and research.

14.4.2 Interaction-Based Hazard Index

The advantages of the HI include its simplicity and its practicality, mainly its easily met information requirements. Its main disadvantage is its constrained application to toxicologically similar chemicals that show no evidence of toxicological interaction. Most environmental exposures involve several types of chemicals, many with well-known demonstrated interactions, such as the greater than dose-additive joint

toxicity of Pb and Cd (ATSDR 2004b). In this chapter, we use the U.S. EPA definition of toxicological interaction as a deviation from dose additivity (Hertzberg et al. 2013). With the goal of broader applicability and more realistic mixture risk estimates, U.S. EPA developed a modification to the HI that incorporates interactions. The interaction-based HI adjusts each HQ in the HI formula to account for data on toxicological interactions among all pairs of chemicals in the mixture (Hertzberg et al. 1999; Hertzberg and Teuschler 2002; U.S. EPA 2000b, 2007a). For n chemicals, this modified HI is:

$$\mathrm{HI}_{\mathrm{INT}} = \sum_{j=1}^{J} \mathrm{HQ}_j \left(\sum_{i \neq j}^{J} f_{ji} M_{ji}^{B_{ji} g_{ji}} \right) \tag{14.22}$$

The factors in this formula are described in detail in those references. In summary, for a binary mixture of chemicals j and i:

M_{ji} = Pairwise interaction magnitude, the ratio of the expected to observed isotoxic dose (e.g., ED10), where "expected" refers to the dose-additive prediction. If that ratio ≤ 1, then the inverse should be used so that $M \geq 1$. U.S. EPA's default assumption is that, for a specified target organ, the binary interaction magnitudes are symmetric, so $M_{ji} = M_{ij}$.

B_{ji} = The binary weight of evidence score (see Sect. 14.4.1 for details).

f_{ji} = The index of toxic hazard (per its HQ) of the ith chemical relative to the total hazard from all chemicals potentially interacting with the jth chemical. Using $\mathrm{HI}_{\mathrm{add}}$ to clarify that it represents the usual additive HI:

$$f_{ji} = \frac{\mathrm{HQ}_i}{\mathrm{HI}_{\mathrm{add}} - \mathrm{HQ}_j} \tag{14.23}$$

g_{ji} = The degree to which chemicals j and i are present in equitoxic amounts, as indicated by their HQ values. Using an assumption that interaction magnitude is higher when chemicals are present at equitoxic levels, g is the ratio of the geometric mean to the arithmetic mean of the two HQ values:

$$g_{ji} = \frac{\sqrt{\mathrm{HQ}_j \cdot \mathrm{HQ}_i}}{\left(\mathrm{HQ}_j + \mathrm{HQ}_i\right)/2} \tag{14.24}$$

The exponent B_{ji} is the BINWOE score (described in the previous subsection) for interactions involving those two chemicals. Interaction magnitude (M) is included, defined as the ratio of equitoxic doses. Specifically, M is the larger of the ratio of the observed mixture response-specific dose (e.g., ED_{10}) to the isotoxic dose predicted by dose addition and its inverse ratio, so that $M \geq 1$. When pairwise interaction magnitude is not available, the default value of 5 is used by U.S. EPA. That value is not empirically based but is close to the very rough estimate for low-dose synergy of 4 found in a study by the International Life Sciences Institute (Boobis et al. 2011).

Table 14.9 Example results for the HI and interaction-based HI

Chemical	High exposure (ppm)	SSL (ppm)	HQ or HI	
			Additive	Interaction-based[a]
Pb	3600	400	9.0	25.3
Cd	800	70	11.4	32.3
Zn	32,200	23,000	1.4	1.4
Mixture			21.8	59.0

Adapted from ATSDR (2004b, c)

[a]The interaction-based HI in this example is calculated using the U.S. EPA formula (Eq. 14.22) but the ATSDR BINWOE scheme with results from two interaction profiles

An example of how the HI changes with the inclusion of interaction evidence is based on exposure levels estimated for the Superfund site in Palmerton, Pennsylvania (USA), involving a huge zinc waste pile. The main chemicals were lead, cadmium, and zinc. The levels summarized (Table 14.9) are close to the upper range of reported levels for soil contamination (U.S. EPA 2007c) and are compared with the U.S. EPA's soil screening levels (SSLs) used with Superfund sites (U.S. EPA 1996). The binary interaction evidence is fairly strong for Zn inhibiting Pb, is strong for Pb and Cd enhancing each other's toxicity, is weak for dose additivity of Pb affecting Zn, and is unknown for Zn with Cd. Because of the high levels of both Pb and Cd, their greater than additive interaction dominated the antagonistic interaction influence of Zn, resulting in an interaction-based HI almost threefold higher than the additive HI without interactions (Table 14.9).

The U.S. EPA's interaction-based HI is a fairly simple modification to an already simple mixture risk approach. In addition to U.S. EPA applications, it has been suggested as a default method (Sarigiannis and Hansen 2012) appropriate for the first level assessment in tiered approaches such as those of WHO/IPCS (Meek et al. 2011).

14.4.3 Uncertainties

A main uncertainty when incorporating interactions is the lack of interaction information for many chemical pairs of concern, as indicated by the "?" notation in the ATSDR BINWOE matrices (e.g., Fig. 14.7). Another uncertainty occurs when the BINWOE evaluations are applied to exposure levels much lower than those tested or are based on extensive extrapolations, e.g., if the studies were in vitro assays or for a different exposure duration and route than the assessment scenario. An uncertainty specific to the U.S. EPA interaction-based HI formula relates to the dependence of interaction magnitude on the degree of equitoxicity of the component ratios in the mixture (the g function in Eq. 14.24). That function is not an empirical deduction but an assumption based on judgment by toxicologists. For example, a study on phthalates was intentionally designed to involve a mixture of near equitoxic

components: "We employed a mixture dose with equipotent contribution from each individual phthalate, as opposed to using an environmentally-relevant ratio, so that we would be more likely to observe unexpected interactions of the phthalate mixture on postnatal reproductive endpoints" (Howdeshell et al. 2015). The uncertainty in that assumption could be reduced somewhat by more binary mixture studies with varying component ratios that represent different types of chemicals and endpoints.

Both the ATSDR and U.S. EPA BINWOE schemes focus on credibility of the interaction evidence and relevance to human health. What is missing in both those schemes is any consideration of toxicological importance of the *magnitude* of the observed interaction. While Table 14.9 shows results when interactions are included in the formula, it lacks any indication of whether the *change* in mixture toxicity from that indicated by dose addition is toxicologically important. A similar weakness is prevalent in the many articles on detecting and quantifying interactions using statistical models: statistical significance is reported but not biological significance. Reporting biological significance of interactions in some standard fashion is not simple. For example, consider an approach that is similar to how the individual chemical benchmark response is often calculated: use the response of some incremental change, say 1 standard deviation (SD), away from the control mean (Crump 2002). If a greater than dose-additive interaction was measured by a mean response at least 1 SD away from the response estimated under dose addition, one could perhaps define that interaction as toxicologically important. The difficulty with such a definition is that "interaction" is a relative term, e.g., a departure from dose addition. The extent of departure that is toxicologically meaningful most likely depends on the response expected under dose addition. Consequently, unlike the benchmark response for individual chemicals, there would not be a corresponding standard criterion for a toxicologically meaningful interaction, suggesting instead a case-by-case evaluation. On rare occasion, researchers have evaluated the observed mixture response relative to the expected variability for that toxic response. For example, Crofton et al. (2005) decided that the observed, statistically significant greater than dose-additive response in a mixture of polyhalogenated aromatic hydrocarbons was not very important toxicologically: "... the magnitude of underestimation of the experimental data by the additivity model is not large. ... even in the high mixture-dose region, the effects of this mixture are predicted by additivity with a fair degree of accuracy."

Another key uncertainty in using interaction BINWOE results is the lack of standardization of methods for determining and describing interactions. Many reported interactions are only qualitatively determined and often use different phrases, e.g., "supra-additive" or "greater than dose additive," based on the audience and the authors' scientific backgrounds (see Chap. 9). Even quantitative determinations can vary because of differing definitions of the expected "noninteractive" response. Many statistical methods exist for evaluating departure from dose addition, including refinements to address different toxicological scenarios, such as mixtures with partial agonists and mixture data from a fixed ratio ray design (see Chap. 11). In contrast, comparatively few methods have been developed for departure from response addition. Further, conclusions based on apparent evidence of no interaction (e.g., statistical consistency with dose addition) could be weak because of poor

statistical power to detect such an interaction. As discussed in Chap. 13, statistical power is rarely reported (Meadows-Shropshire et al. 2005; Stork et al. 2007).

14.5 Future Directions for Mixture Risk Assessment

The component-based approaches to mixture assessment that are based on or derived from dose addition are highly appealing. There are obvious tiers of assessment (HI-based decisions to RPF-based estimates of response), and for the chemicals identified as sufficiently toxicologically similar, the required information is usually available, often with official endorsement by governmental agencies. Several research articles and government reports have identified important uncertainties with these approaches and in some cases have suggested ways for improvement. Most of the arguments have identified issues with data analysis and modeling, but recently they have focused on toxicological understanding and experimental design. This section includes a specific enhanced approach to address combinations of both similar and independent chemicals, followed by general discussion of new directions that take advantage of better toxicological understanding of the joint toxicity pathway and of more sophisticated quantitative concepts and tools.

14.5.1 Combining Concepts: Integrated Addition

Several publications have presented dose addition and response addition as the two alternative concepts for component-based mixture risk assessment, placing importance on toxicological MOA for deciding between the two and sometimes comparing the corresponding results from those two formulas. Many if not most real exposures to environmental chemicals include both toxicologically similar chemicals and independently acting chemicals. Such a scenario arose with the evaluation of the feasibility of a health risk assessment of mixtures of disinfection by-product (DBP) chemicals in drinking water, which led to the development of a hybrid additivity approach that incorporated both dose addition and response addition for dichotomous endpoints (U.S. EPA 2000a, 2003b). That approach is here termed "integrated addition" to comport with similar published methods applied to other chemical combinations (Altenburger et al. 2005; Kim et al. 2014; Mwense et al. 2004; Olmstead and LeBlanc 2005; Rider et al. 2009; Teuschler et al. 2004). Integrated addition is yet to be officially defined in any agency or regulatory guidance documents, but it is mentioned as a theoretically sound option for mixture risk assessment in U.S. EPA methodology documents published from 2000 to 2007 (Teuschler et al. 2004; U.S. EPA 2000a, 2003b, 2007a). Of the examples of integrated addition presented in those U.S. EPA and the journal publications cited above, the DBP example is shown here because it includes clearly identified steps and requires only standard risk-based component information, similar to what is

described for the dose addition and response addition approaches discussed in Sects. 14.2 and 14.3.

The integrated addition approach begins with partitioning the mixture chemicals into nonoverlapping subgroups of toxicologically similar chemicals. For the example here, each subgroup is assumed to share a common MOA distinctly different from the other subgroups. The assumption is that while all chemicals in the assessment share a common health endpoint, the subgroups differ in the AOP leading to that common endpoint. A risk estimate is then calculated for each subgroup using methods based on similarity: the RPF method is used in this example. The subgroups are considered to be toxicologically independent of each other so that response addition is used to combine risks from all subgroups. The specific steps for applying integrated addition that we suggest are adapted from those of Teuschler et al. (2004). "Dose" is used here to generally represent exposure: it can be external exposure (e.g., mg/kg/day in drinking water) or internal dose (e.g., target tissue concentration). While only oral exposure is used in this example, those original U.S. EPA publications address multiroute exposures along with more enhanced characterizations of dose (Teuschler et al. 2004). An index chemical for each subgroup (for the RPF method) is used here to simplify the separation of the dose addition and response addition steps:

1. Partition the chemicals into subgroups by common toxicological MOA.
2. Develop exposure estimates for every chemical.
3. Develop dose-response models for every chemical.
4. For each subgroup, convert the doses to toxicologically equivalent doses for the same animal species (e.g., rats), and, if not already so, convert the response measure into a probabilistic risk measure.
5. Identify an index chemical for each subgroup and calculate the ICED for that subgroup mixture and then the subgroup risk estimate for the chosen animal species.
6. Convert each subgroup risk estimate into a human risk estimate, and using response addition, sum up the subgroup human risks to estimate the total mixture risk; develop a full risk characterization, including an analysis of uncertainty.

For a mixture of J chemicals divided into S similarity subgroups, the final step uses the response addition formula:

$$p_{\text{mix}}\left(\vec{d}\right) = 1 - \prod_{s=1}^{S}\left(1 - p_s(\text{ICED}_s)\right) \tag{14.25}$$

The dose $\vec{d}$ is the vector of all J component doses. For the s^{th} similarity subgroup, $p_s(*)$ is the dose-response model for the index chemical, and ICED_s is the index chemical equivalent dose for that subgroup mixture (see Eq. 14.12).

The example is adapted from Teuschler et al. (2004), which uses a mixture of six carcinogenic DBP chemicals divided into two similarity subgroups: genotoxic vs. nongenotoxic MOA. The RPFs are estimated from the ratio of cancer slope factors (SFs). One chemical ($CHCl_4$) has an RfD instead of a cancer slope factor.

Table 14.10 Example of integrated addition approach for cancer risk of five DBPs

DBP	RPF (SF*i*/SF1)	Total absorbed dose for 70 kg male	Component ICED	Subgroup ICED	Subgroup risk[a]
Genotoxic subgroup					
BDCM	1.00	1.20E−03	1.20E−03		
DBCM	1.35	7.84E−04	1.06E−03		
CHBr3	0.13	4.29E−04	5.46E−05		
				2.32E−03	1.32E−05
Nongenotoxic subgroup					
DCA	1.00	4.49E−04	4.49E−04		
TCA	0.84	4.77E−04	4.01E−04		
				8.50E−04	1.19E−06
Total mixture expected cancer risk					1.44E−05

Adapted from Teuschler et al. (2004)

[a]Subgroup risk = ICED*(index chemical SF). Genotoxic subgroup index chemical, BDCM: maximum likelihood estimate (MLE) of cancer SF = 5.7E−3. Nongenotoxic subgroup index chemical, DCA: MLE of cancer SF = 1.4E−3

Because its dose was below its RfD, it was assumed to contribute no cancer risk to the mixture, and so it is not included in our example (Table 14.10). The published example is much more complex and includes details on the nuances considered in the estimates of slope and absorbed dose (Teuschler et al. 2004).

Application of integrated addition to mixture risk assessment requires three conditions:

1. A common adverse effect can be caused by multiple toxicological processes (i.e., mechanisms, MOAs, AOPs).
2. The chemicals can be partitioned into several nonoverlapping subgroups, each of which composed of toxicologically similar chemicals.
3. The subgroups are toxicologically independent of each other.

Those conditions usually would only be assumed to exist; while supporting evidence could range from poor to good, most likely it would be minimal. The appeal to risk assessment is that integrated addition allows a reasoned, structured judgment to be made about safety for those mixtures previously not addressed, i.e., those that involve combinations of similar and dissimilar chemicals. To become practical, decision criteria are needed for when those conditions can be assumed. While considerable guidance exists related to toxicological similarity, little exists at present for toxicological independence, particularly when applied to multiple subgroups of chemicals.

14.5.2 Discussion and Conclusions

Improvements to mixture risk assessment, particularly regulatory assessment, are expected to be slow but possibly dramatic. Three interconnected processes are involved: toxicological understanding, data analysis and modeling, and regulatory

operations. Dramatic changes in the science are possible if regulatory policy changes. The first policy shift toward mixture risk occurred with passage of CERCLA in 1980, which forced such consideration at Superfund sites. The second policy shift came with FQPA in 1996, where similarity of toxicological processes and pathways was encoded in pesticide risk guidelines. These three interconnected processes are now discussed in terms of mixture risk improvements.

14.5.2.1 Improved Toxicological Study and Understanding

The use of the HI approach with adjustment for nonlinear deviation from dose-additivity gives a reasonable approximation of the toxic potency of a chemical mixture and is the most practical approach currently available. Our experience at ATSDR and U.S. EPA shows that this approach can be successfully used in health assessments at hazardous waste sites and the methodology works particularly well for mixture components with fairly similar toxicological action. Major data gaps do exist in our understanding of joint toxicity, particularly of more complex mixtures (more than ten components). However, testing even a representative sample of all environmental chemical mixtures is resource limiting and is not feasible. The U.S. National Academies have recommended a new systems biology approach, using experimental high-throughput screening studies and alternative testing protocols of human cell lines, tissues, or in vitro systems to gain insights into happenings at molecular and cellular levels (NRC 2007). That approach, when pragmatically used, could allow testing of biologically effective dose of the mixture components in the target tissues. All the data thus generated need to be captured in strategically designed, curated, and searchable databases to allow exploration of appropriate questions and assumptions leading to new testable hypotheses about joint toxicity mechanisms (Kienzler et al. 2014). A prudent way forward in the near term would be the increased use of presently available computational methodologies (e.g., PBPK, QSAR modeling) in mixture risk assessment to increase the number of mixtures addressable by component methods.

14.5.2.2 Improved Modeling and Data Analysis

Many of the quantitative approaches for component methods were developed years ago. Some risk-based values are primarily judgment, and most are deterministically set. The newer reference values are more likely based on dose-response models and statistical descriptions. For the HI approach, federal agencies such as the U.S. EPA, ATSDR, and NIOSH have transitioned to reference values based on the benchmark dose, usually an ED_{10} from modeling the full-dose range, as a replacement for the no-observed-adverse-effect level (NOAEL), which was merely a single data point (U.S. EPA 2014). RPF values have also been improved by several statistical estimation methods (U.S. EPA 2003a). Judgments about similarity of dose-response curves in support of dose additivity have been recast as statistical comparisons using

response estimates for general and specific dose-response models, including mixtures with partial agonists (Altenburger et al. 2000; Hadrup et al. 2013; Hertzberg et al. 2013; Howard and Webster 2009; Kienhuis et al. 2015).

The RPF and TEF approaches remain among the preferred implementations of dose addition and in many cases are the mixture risk approaches with the strongest toxicological support. Many of the uncertainties discussed in this chapter are not routinely considered in mixture risk assessments. As more detailed information, including data and concepts of dosimetry and toxicodynamic pathways, becomes more widely available, the RPF and TEF approaches should evolve. Arguments and examples have been published supporting a variable RPF, i.e., one that is not a constant but a function of the dose and/or response level (Dinse and Umbach 2011, 2012). When data are extensive (though rarely available with in vivo mammalian studies), the variable RPF approach can be implemented totally empirically (Altenburger et al. 2000). In general, a preferable and likely more feasible approach would be a biologically based response-dependent RPF model.

14.5.2.3 The Role of Governmental Regulatory Approaches and Numbers

A different issue arises from the official nature of governmental databases, in this case, the collection of risk-based reference values (e.g., RfD, MRL), relative potency values (e.g., RPF, TEF), and interaction evidence (e.g., BINWOE classifications and scores). The availability of such official resources promotes their use, and their official stature improves consistency across scenarios and over time. That consistency, however, is also a weakness, because governmental updates to reflect new information are notoriously slow and underfunded. In addition, adoption of new methods for experiment and analysis is often slow because of the corresponding revision of so many existing regulatory numbers, reports, and conclusions. Recommended changes in risk assessment methods and formulas are unlikely to occur soon, in spite of theoretical merits. For example, changes to the U.S. EPA version of dose addition suggested by two National Academies reports are not feasible because they require direct testing of each mixture or full dose-response curves for every component on every endpoint, neither of which is economically feasible. At the least, the conclusions about predicted mixture response in governmental assessments should include when possible the numerical impact of variations in those underlying values (e.g., RfDs in a HI calculation), especially if comparisons included values based on new research results with those using the existing, official regulatory values.

Those changes to the conclusions are likely to be minor. What could dramatically change mixture risk assessment is a shift in regulation toward more realistic exposures and concomitant risks, especially to specific population groups. For population-centric assessments that include chemicals and nonchemical exposures

or factors, the U.S. EPA calls that activity "cumulative" risk assessment[3] and emphasizes the role of population characteristics, such as nutrition and psychological stress, in altering sensitivity to chemical toxicity (U.S. EPA 2003c). Once regulatory policy changes to require the so-called twenty-first century approaches (NASEM 2016), including in silico modeling and collection of high-throughput screening data that quantitatively describe fairly complete AOPs involving multiple chemicals, then corresponding research will expand, and risk methods will adapt to incorporate the new information.

References

ACGIH. 1983. *TLVs: theshold limit values for chemical substances and physical agents in the work environment with intended changes for 1983–1984*, Cincinnati.

Alexeeff, S.E., A.A. Litonjua, H. Suh, D. Sparrow, P.S. Vokonas, and J. Schwartz. 2007. Ozone exposure and lung function: Effect modified by obesity and airways hyperresponsiveness in the VA normative aging study. *Chest* 132 (6): 1890–1897.

Altenburger, R., T. Backhaus, W. Boedeker, M. Faust, M. Scholze, and L.H. Grimme. 2000. Predictability of the toxicity of multiple chemical mixtures to *Vibrio fischeri:* Mixtures composed of similarly acting chemicals. *Environmental Toxicology & Chemistry* 19 (9): 2341–2347.

Altenburger, R., H. Schmitt, and G. Schuurmann. 2005. Algal toxicity of nitrobenzenes: Combined effect analysis as a pharmacological probe for similar modes of interaction. *Environmental Toxicology and Chemistry* 24 (2): 324.

Arcos, J.C., M.F. Argus, and Y. Woo. 1995. *Chemical induction of cancer: Modulation and combination effects. An inventory of the many factors which influence carcinogenesis*. Boston: Birkhäuser.

ATSDR. 2004a. *Guidance manual for the assessment of joint toxic action of chemical mixtures, agency for toxic substances and disease Registry*. Atlanta: U.S. Department of Health and Human Services, Public Health Service.

———. 2004b. *Interaction profile for: Arsenic, cadmium, chromium, and lead, agency for toxic substances and disease registry*. Atlanta: U.S. Department of Health and Human Services, Public Health Service.

———. 2004c. *Interaction profile for: Lead, manganese, zinc, and copper, agency for toxic substances and disease registry*. Atlanta: U.S. Department of Health and Human Services, Public Health Service.

———. 2005a. *Public health assessment for Conrail Rail Yard, Elkhart, Elkhart County, Indiana, EPA Facility ID: IND000715490 agency for toxic substances and disease registry,* U.S. Department of Health and Human Services, Atlanta: Public Health Service. https://www.atsdr.cdc.gov/HAC/pha/ConrailRailYd/ConrailRailYardPHA081105.pdf.

———. 2005b. *Public health assessment guidance manual (2005 update)*. Atlanta: Agency for Toxic Substances and Disease Registry, U.S. Department of Health and Human Services, Public Health Service.

[3]Note that previous "community based" assessments by U.S. EPA also focused on the population, but could involve exposure to only one chemical, and often include recommendations for community action. A cumulative risk "population" in contrast could be the entire U.S. population.

——— (2006) Interaction profile for: Atrazine, deethylatrazine, diazinon, nitrate, and simazine, agency for toxic substances and disease registry, U.S. Department of Health and Human Services, Public Health Service., Atlanta.

———. 2016. *Minimal risk levels.*Atlanta: U.S. Department of Health and Human Services. http://www.atsdr.cdc.gov/mrls/index.asp.

———. 2017. *Interaction profiles for toxic substances*. U.S. Department of Health and Human Services. https://www.atsdr.cdc.gov/interactionprofiles/index.asp. Accessed October 2017.

Berenbaum, M.C. 1985. The expected effect of a combination of agents: The general solution. *Journal of Theoretical Biology* 114 (3): 413–431.

———. 1989. What is synergy? *Pharmacological Reviews* 41 (2): 93–141.

Bliss, C.I. 1939. The toxicity of poisons applied jointly. *Annals of Applied Biology* 26 (3): 585–615.

Boobis, A., R. Budinsky, S. Collie, K. Crofton, M. Embry, S. Felter, R. Hertzberg, D. Kopp, G. Mihlan, M. Mumtaz, P. Price, K. Solomon, L. Teuschler, R. Yang, and R. Zaleski. 2011. Critical analysis of literature on low-dose synergy for use in screening chemical mixtures for risk assessment. *Critical Reviews in Toxicology* 41 (5): 369–383.

Bosgra, S., J.C.H. van Eijkeren, and W. Slob. 2009. Dose addition and the isobole method as approaches for predicting the cumulative effect of non-interacting chemicals: A critical evaluation. *Critical Reviews in Toxicology* 39 (5): 418–426.

Budinsky, R.A., D. Paustenbach, D. Fontaine, B. Landenberger, and T.B. Starr. 2006. Recommended relative potency factors for 2,3,4,7,8-pentachlorodibenzofuran: The impact of different dose metrics. *Toxicological Sciences* 91 (1): 275–285.

Chen, J.J., Y.-J. Chen, G. Rice, L. Teuschler, K. Hamernik, A. Protzel, and R. Kodell. 2001. Using dose addition to estimate cumulative risks from exposures to multiple chemicals. *Regulatory Toxicology and Pharmacology* 34: 35–41.

Chen, J., Y.-J. Chen, L. Teuschler, G.E. Rice, K. Hamernik, A. Protzel, and R.L. Kodell. 2003. Cumulative risk assessment for quantitative response data. *Environmetrics* 14 (4): 339–353.

Cogliano, V.J. 1997. Plausible upper bounds: Are their sums plausible? *Risk Analysis* 17 (1): 77–84.

Crofton, K.M., E.S. Craft, J.M. Hedge, C. Gennings, J.E. Simmons, R.A. Carchman, W.H. Carter Jr., and M.J. DeVito. 2005. Thyroid-hormone disrupting chemicals: Evidence for dose-dependent additivity or synergism. *Environmental Health Perspectives* 113 (11): 1549.

Crump, K.S. 2002. Critical issues in benchmark calculations from continuous data. *Critical Reviews in Toxicology* 32 (3): 133–153.

Demchuk, E., P. Ruiz, J.D. Wilson, F. Scinicariello, H.R. Pohl, M. Fay, M.M. Mumtaz, H. Hansen, and C.T. De Rosa. 2008. Computational toxicology methods in public health practice. *Toxicology Mechanisms and Methods* 18 (2–3): 119–135.

Dinse, G.E., and D.M. Umbach. 2011. Characterizing non-constant relative potency. *Regulatory Toxicology and Pharmacology* 60 (3): 342–353.

———. 2012. Parameterizing dose-response models to estimate relative potency functions directly. *Toxicological Sciences* 129 (2): 447–455.

Durkin, P., R. Hertzberg, W. Stiteler, and M. Mumtaz. 1995. The identification and testing of interaction patterns. *Toxicology Letters* 79 (1–3): 251–264.

Egeghy, P.P., R. Judson, S. Gangwal, S. Mosher, D. Smith, J. Vail, and E.A. Cohen Hubal. 2012. The exposure data landscape for manufactured chemicals. *Science of the Total Environment* 414: 159–166.

Fann, N., H.A. Roman, C.M. Fulcher, M.A. Gentile, B.J. Hubbell, K. Wesson, and J.I. Levy. 2011. Maximizing health benefits and minimizing inequality: Incorporating local-scale data in the design and evaluation of air quality policies. *Risk Analysis* 31 (6): 908–922.

Fay, R.M., and M.M. Mumtaz. 1996. Development of a priority list of chemical mixtures occurring at 1188 hazardous waste sites, using the hazdat database. *Food and Chemical Toxicology* 34 (11–12): 1163–1165.

Fay, M., and M. Mumtaz. 2004. Frequency of exposure to contaminant mixtures at hazardous waste sites. *Toxicology* 78 (108): 522.

Finney, D.J. 1971. *Probit analysis*. 2nd ed. Cambridge: Cambridge University Press.

Fox, M.A., N.L. Tran, J.D. Groopman, and T.A. Burke. 2004. Toxicological resources for cumulative risk: An example with hazardous air pollutants. *Regulatory Toxicology and Pharmacology* 40 (3): 305–311.

Gennings, C. 1995. An efficient experimental design for detecting departure from additivity in mixtures of many chemicals. *Toxicology* 105 (2–3): 189–197.

Hadrup, N., C. Taxvig, M. Pedersen, C. Nellemann, U. Hass, and A.M. Vinggaard. 2013. Concentration addition, independent action and generalized concentration addition models for mixture effect prediction of sex hormone synthesis in vitro. *PloS One* 8 (8): e70490.

Hertzberg, R.C., and L.K. Teuschler. 2002. Evaluating quantitative formulas for dose-response assessment of chemical mixtures. *Environmental Health Perspectives* 110 (Supplement 6): 965–970.

Hertzberg, R.C., G.E. Rice, and L.K. Teuschler. 1999. Methods for health risk assessment of combustion mixtures. In *Hazardous waste incineration: Evaluating the human health and environmental risks*, ed. S. Roberts, C. Teaf, and J. Bean, 105–148. Boca Raton: CRC Press.

Hertzberg, R.C., Y. Pan, R. Li, L.T. Haber, R.H. Lyles, D.W. Herr, V.C. Moser, and J.E. Simmons. 2013. A four-step approach to evaluate mixtures for consistency with dose addition. *Toxicology* 313 (2–3): 134–144.

Howard, G.J., and T.F. Webster. 2009. Generalized concentration addition: A method for examining mixtures containing partial agonists. *Journal of Theoretical Biology* 259 (3): 469–477.

Howdeshell, K.L., C.V. Rider, V.S. Wilson, J. Furr, C.R. Lambright, and L.E. Gray. 2015. Dose addition models based on biologically-relevant reductions in fetal testosterone accurately predict postnatal reproductive tract alterations by a phthalate mixture in rats. *Toxicological Sciences* 148: 488–502.

International Agency for Research on Cancer (IARC). 1982. *IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans, Supplement 4*, Lyon.

Kienhuis, A.S., W. Slob, E.R. Gremmer, J.P. Vermeulen, and J. Ezendam. 2015. A dose-response modeling approach shows that effects from mixture exposure to the skin sensitizers isoeugenol and cinnamal are in line with dose addition and not with synergism. *Toxicological Sciences* 147 (1): 68–74.

Kienzler, A., E. Berggren, J. Bessems, S. Bopp, S. van der Linden, and A.P. Worth. 2014. *Assessment of mixtures – Review of regulatory requirements and guidance*. Luxembourg: European Commission.

Kim, J., S. Kim, and G.E. Schaumann. 2014. Development of a partial least squares-based integrated addition model for predicting mixture toxicity. *Human and Ecological Risk Assessment* 20 (1): 174.

Kortenkamp, A., T. Backhaus, and M. Faust. 2009. *State of the art report on mixture toxicity*. London: University of London.

Lambert, J.C., and J.C. Lipscomb. 2007. Mode of action as a determining factor in additivity models for chemical mixture risk assessment. *Regulatory Toxicology and Pharmacology* 49 (3): 183–194.

MacDonell, M. 2014. Grouping stressors by exposure factors for cumulative risk assessment. In workshop: Cumulative risk assessment: addressing combined environmental stressors. Society for Risk Analysis annual meeting, Denver, CO, Dec. 7, 2014.

MacDonell, M.M., R.C. Hertzberg, G.E. Rice, J.M. Wright, and L.K. Teuschler. 2017. Characterizing risk for cumulative risk assessments. *Risk Analysis*. https://doi.org/10.1111/risa.12933.

McCarty, L.S., and C.J. Borgert. 2006. Review of the toxicity of chemical mixtures: Theory, policy, and regulatory practice. *Regulatory Toxicology and Pharmacology* 45 (2): 119–143.

Meadows-Shropshire, S.L., C. Gennings, and W.H. Carter. 2005. Sample size and power determination for detecting interactions in mixtures of chemicals. *Journal of Agricultural, Biological, and Environmental Statistics* 10 (1): 104.

Meek, B., A. Boobis, K.M. Crofton, G. Heinemeyer, M. Van Raaij, and C. Vickers. 2011. Risk assessment of combined exposure to multiple chemicals: A WHO/IPCS framework. *Regulatory Toxicology and Pharmacology* 60 (Supplement (2)): S1–S14.

Miltner, R.J., T.F. Speth, S.D. Richardson, S.W. Krasner, H.S. Weinberg, and J.E. Simmons. 2008. Integrated disinfection by-products mixtures research: Disinfection of drinking waters by chlorination and ozonation/postchlorination treatment scenarios. *Journal of Toxicology and Environmental Health, Part A* 71 (17): 1133–1148.

Morello-Frosch, R., and E.D. Shenassa. 2006. The environmental "riskscape" and social inequality: Implications for explaining maternal and child health disparities. *Environmental Health Perspectives* 114 (8): 1150–1153.

Mumtaz, M., and P. Durkin. 1992. A weight of evidence scheme for assessing interactions in chemical mixtures. *Toxicology and Industrial Health* 8: 377–406.

Mumtaz, M., and R.C. Hertzberg. 1993. The status of interaction data in risk assessment of chemical mixtures. In *Hazard assessment of chemicals*, ed. J. Saxena, vol. 8, 47–79. Washington: Hemisphere.

Mumtaz, M.M., C. DeRosa, and P.R. Durkin. 1994. Approaches and challenges in risk assessments of chemical mixtures. In *Toxicology of chemical mixtures: Case studies, mechanisms, and novel approaches*, ed. R.S.H. Yang, 565–597. San Diego: Academic Press.

Mumtaz, M., K. Poirer, and J.T. Colman. 1997. Risk assessment for chemical mixtures: Fine-tuning the hazard index approach. *Journal of Clean Technology Environmental Toxicology and Occupational Medicine* 6: 189–204.

Mumtaz, M.M., C.T. De Rosa, J. Groten, V.J. Feron, H. Hansen, and P.R. Durkin. 1998. Estimation of toxicity of chemical mixtures through modeling of chemical interactions. *Environmental Health Perspectives* 106 (Suppl 6): 1353–1360.

Mumtaz, M.M., P. Ruiz, and C.T. De Rosa. 2007. Toxicity assessment of unintentional exposure to multiple chemicals. *Toxicology and Applied Pharmacology* 223 (2): 104–113.

Muska, C.F., and L.J. Weber. 1977. An approach for studying the effects of mixtures of environmental toxicants on whole organism performances. In *Recent advances in fish toxicology*, ed. R.A. Tubb, 71–87. Corvallis: Corvallis Environmental Research Laboratory, U.S. EPA.

Mwense, M., X.Z. Wang, F.V. Buontempo, N. Horan, A. Young, and D. Osborn. 2004. Prediction of noninteractive mixture toxicity of organic compounds based on a fuzzy set method. *Journal of Chemical Information and Computer Sciences* 44 (5): 1763–1773.

NASEM. 2016. *Using 21st century science to improve risk-related evaluations. 978-0-309-45348-6*. Washington: National Academies of Sciences, Engineering, and Medicine, The National Academies Press. https://www.nap.edu/catalog/24635/using-21st-century-science-to-improve-risk-related-evaluations.

NRC. 2007. *Toxicity testing in the 21st century: A vision and a strategy. 9780309151733*. Washington: National Academies Press. http://www.nap.edu/openbook.php?record_id=11970.

———. 2008. *Phthalates and cumulative risk assessment: The tasks ahead*. Washington: National Academies Press.

Olmstead, A.W., and G.A. LeBlanc. 2005. Toxicity assessment of environmentally relevant pollutant mixtures using a heuristic model. *Integrated Environmental Assessment and Management* 1 (2): 114–122.

Ott, W.R. 1978. *Environmental indices: Theory and practice*. Ann Arbor: Ann Arbor Science.

Parvez, S., A.M. Evans, M. Lorber, B.S. Hawkins, J.C. Swartout, L.K. Teuschler, and G.E. Rice. 2013. A sensitivity analysis using alternative toxic equivalency factors to estimate U.S. dietary exposures to dioxin-like compounds. *Regulatory Toxicology and Pharmacology* 67 (2): 278–284.

Plackett, R.L., and P.S. Hewlett. 1948. Statistical aspects of the independent joint action of poisons, particularly insecticides. I. The toxicity of a mixture of poisons. *Annals of Applied Biology* 35 (3): 347–358.

Pohl, H.R., M.M. Mumtaz, F. Scinicariello, and H. Hansen. 2009. Binary weight-of-evidence evaluations of chemical interactions–15 years of experience. *Regulatory Toxicology and Pharmacology* 54 (3): 264–271.

Rice, G., M. MacDonell, R.C. Hertzberg, L. Teuschler, K. Picel, J. Butler, Y.-S. Chang, and H. Hartmann. 2008. An approach for assessing human exposures to chemical mixtures in the environment. *Toxicology and Applied Pharmacology* 233 (1): 126–136.

Rider, C.V., V.S. Wilson, K.L. Howdeshell, A.K. Hotchkiss, J.R. Furr, C.R. Lambright, and L.E. Gray. 2009. Cumulative effects of in utero administration of mixtures of "Antiandrogens" on male rat reproductive development. *Toxicologic Pathology* 37 (1): 100–113.

Sarigiannis, D.A., and U. Hansen. 2012. Considering the cumulative risk of mixtures of chemicals – A challenge for policy makers. *Environmental Health: A Global Access Science Source* 11 (Suppl 1): S18.

SCHER, SCCS, SCENIHR. 2012. *Opinion on the toxicity and assessment of chemical mixtures*. Brussels: European Union.

Smyth Jr, H.F., C.S. Weil, J.S. West, and C.P. Carpenter. 1969. An exploration of joint toxic action: Twenty-seven industrial chemicals intubated in rats in all possible pairs. *Toxicology and Applied Pharmacology* 14 (2): 340–347.

Stork, L.G., C. Gennings, W.H. Carter, R.E. Johnson, D.P. Mays, J.E. Simmons, E.D. Wagner, and M.J. Plewa. 2007. Testing for additivity in chemical mixtures using a fixed-ratio ray design and statistical equivalence testing methods. *Journal of Agricultural, Biological, and Environmental Statistics* 12 (4): 514–533.

Svendsgaard, D., and R.C. Hertzberg. 1994. Statistical methods for the toxicological evaluation of the additivity assumption as used in the EPA chemical mixture risk assessment guidelines. In *Toxicology of chemical mixtures: Case studies, mechanisms, and novel approaches*, ed. R.S.H. Yang, 599–642. San Diego: Academic Press.

Teuschler, L., J. Klaunig, E. Carney, J. Chambers, R. Conolly, C. Gennings, J. Giesy, R. Hertzberg, C. Klaassen, R. Kodell, D. Paustenbach, and R. Yang. 2002. Support of science-based decisions concerning the evaluation of the toxicology of mixtures: A new beginning. *Regulatory Toxicology and Pharmacology* 36 (1): 34–39.

Teuschler, L.K., G.E. Rice, C.R. Wilkes, J.C. Lipscomb, and F.W. Power. 2004. A feasibility study of cumulative risk assessment methods for drinking water disinfection by-product mixtures. *Journal of Toxicology and Environmental Health, Part A* 67 (8–10): 755–777.

Tornero-Velez, R., P.P. Egeghy, and E.A. Cohen Hubal. 2012. Biogeographical analysis of chemical co-occurrence data to identify priorities for mixtures research. *Risk Analysis* 32 (2): 224–236.

U.S. EPA. 1986a. *Guidelines for carcinogen risk assessment. EPA/630/R-00/004*. Washington: Risk assessment Forum.

———. 1986b. *Guidelines for the health risk assessment of chemical mixtures. EPA/630/R-98/002*. Washington: Risk Assessment Forum.

———. 1989a. *Interim procedures for estimating risks associated with exposures to mixtures of chlorinated dibenzo-p-dioxins and -dibenzofurans (CDDs and CDFs) and 1989 update. EPA// 625/3-89/016*. Washington: Risk Assessment Forum.

———. 1989b. *Risk assessment guidance for superfund. vol. 1. Human health evaluation manual (Part A). EPA/540/1-89/002*. Washington: Office of Solid Waste and Emergency Response.

———. 1993. Technical guidance manual, risk assessment. In *Selecting exposure routes and contaminants of concern by risk-based screening*. *U.S. EPA/903/r-93-001*, Region III Philadelphia: Hazardous Waste Management Division, Office of Superfund Programs. https://www.epa.gov/risk/selecting-exposure-routes-and-contaminants-concern-risk-based-screening.

———. 1996. *Supplemental guidance for developing soil screening levels for superfund sites. OSWER 9355.4–24*. Washington: Office of Emergency and Remedial Response.

———. 1999. *Guidance for identifying pesticide chemicals and other substances that have a common mechanism of toxicity*. Washington: Office of Pesticide Programs.

———. 2000a. *Conducting a risk assessment of mixtures of disinfection by-products (DBPs) for drinking water treatment systems. EPA/600/R-03/040*. Cincinnati: Office of Research and Development, National Center for Environmental Assessment.

———. 2000b. *Supplementary guidance for conducting health risk assessment of chemical mixtures. EPA/630/R-00/002*. Washington: Risk Assessment Forum. http://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=4486.
———. 2001a. *General principles for performing aggregate exposure and risk assessments*. Washington: Office of Pesticide Programs.
———. 2001b. Risk assessment guidance for superfund. In *Vol. I. Human health evaluation manual (Part D), standardized planning, reporting, and review of superfund risk assessments*. *9285.7-47*, Washington: Office of Solid Waste and Emergency Response.
———. 2001c. *Workshop report on the application of 2,3,7,8-TCDD toxicity equivalence factors to fish and wildlife. EPA/630/R-01/002*. Washington: Risk Assessment Forum.
———. 2002a. *Guidance on cumulative risk assessment of pesticide chemicals that have a common mechanism of toxicity*. Washington: Office of Pesticide Programs.
———. 2002b. *Revised organophosphorous pesticide cumulative risk assessment*. Washington: Office of Pesticide Programs.
———. 2003a. *Developing relative potency factors for pesticide mixtures: Biostatistical analyses of joint dose-response. EPA/600/R-03/052*. Cincinnati: Office of Research and Development, National Center for Environmental Assessment.
———. 2003b. *The feasibility of performing cumulative risk assessments for mixtures of disinfection by-products in drinking water. EPA/600/R-03/051*. Cincinnati: Office of Research and Development, National Center for Environmental Assessment.
———. 2003c. *Framework for cumulative risk assessment. EPA/600/P-02/001F*. Washington: Risk Assessment Forum.
———. 2005. *Guidelines for carcinogen risk assessment. EPA/630/P-03/001F*. Washington: Risk Assessment Forum.
———. 2006. *Organophosphorus cumulative risk assessment – 2006 update. EPA-HQ-OPP-2006-0618-0002*. Washington: Office of Pesticide Programs.
———. 2007a. *Concepts, methods, and data sources for cumulative health risk assessment of multiple chemicals, exposures and effects: a resource document. EPA/600/R-06/013F*. Cincinnati: Office of Research and Development, National Center for Environmental Assessment.
———. 2007b. *Considerations for developing a dosimetry-based cumulative risk assessment approach for mixtures of environmental contaminants. EPA/600/R-07/064*. Cincinnati: Office of Research and Development, National Center for Environmental Assessment.
———. 2007c. *Palmerton zinc pile: compost/biosolids application to revegetate defoliated areas*. Washington: Abandoned Mine Lands Program.
———. 2007d. *Revised n-methyl carbamate cumulative risk assessment*. Washington: Office of Pesticide Programs.
———. 2010. *Recommended Toxicity Equivalence Factors (TEFs) for human health risk assessments of 2,3,7,8- tetrachlorodibenzo-p-dioxin and dioxin-like compounds. EPA/100/R-10/005*. Washington: Office of Research and Development, Risk Assessment Forum.
———. 2012. *EPA's integrated risk information system program: Progress report and report to congress*. Washington: Office of Research and Development. https://www.epa.gov/sites/production/files/2015-06/documents/iris_report_to_congress_2012.pdf.
———. 2016. Benchmark Dose Software (BMDS) User Manual, National Center for Environmental Assessment, Research Triangle Park.
———. 2016. *Regional screening levels for chemical contaminants at superfund sites*. Office of Land and Emergency Management. https://www.epa.gov/risk/regional-screening-levels-rsls. Accessed October 2017.
———. 2017. Integrated Risk Information System (IRIS). Office of Research and Development. Available via U.S. EPA Office of Research and Development, National Center for Environmental Assessment. https://www.epa.gov/iris/. Accessed October 2017.
van den Berg, M., L. Birnbaum, T. Bosveld, B. Brunstrom, et al. 1998. Toxic Equivalency Factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. *Environmental Health Perspectives* 106 (12): 775–792.

van den Berg, M., L.S. Birnbaum, M. Denison, M. De Vito, W. Farland, M. Feeley, H. Fiedler, H. Hakansson, A. Hanberg, L. Haws, M. Rose, S. Safe, D. Schrenk, C. Tohyama, A. Tritscher, J. Tuomisto, M. Tysklind, N. Walker, and R.E. Peterson. 2006. The 2005 world health organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. *Toxicological Sciences* 93 (2): 223–241.

van den Berg, M., M.S. Denison, L.S. Birnbaum, M.J. DeVito, H. Fiedler, J. Falandysz, M. Rose, D. Schrenk, S. Safe, C. Tohyama, A. Tritscher, M. Tysklind, and R.E. Peterson. 2013. Polybrominated dibenzo-p-dioxins, dibenzofurans, and biphenyls: Inclusion in the toxicity equivalency factor concept for dioxin-like compounds. *Toxicological Sciences* 133 (2): 197–208.

Wesson, K., N. Fann, M. Morris, T. Fox, and B. Hubbell. 2010. A multi-pollutant, risk-based approach to air quality management: Case study for Detroit. *Atmospheric Pollution Research* 1 (4): 296–304.

Wolansky, M.J., C. Gennings, and K.M. Crofton. 2006. Relative potencies for acute effects of pyrethroids on motor function in rats. *Toxicological Sciences* 89 (1): 271–277.

Zartarian, V.G., and B.D. Schultz. 2010. The EPA's human exposure research program for assessing cumulative risk in communities. *Journal of Exposure Science and Environmental Epidemiology* 20 (4): 351–358.

Chapter 15
Assessing Human Health Risks Using Information on Whole Mixtures

Glenn E. Rice, Ingvar Eide, Paul I. Feder, and Chris Gennings

Abstract This chapter discusses whole mixture approaches to assessing the risks of potentially hazardous chemical mixtures in the environment within the context of the risk assessment paradigm. Here, "whole mixtures" represent the combination of chemicals in the exposure being assessed. For risk assessment purposes, the environmental mixtures considered as a whole mixture can range from complex mixtures, consisting of perhaps hundreds of component chemicals, to less complex whole mixtures, such as all of the members (i.e., components) of a defined class of compounds. Whole mixture approaches are preferred to component approaches in mixture risk assessments. Because of the variability of whole mixtures encountered in the environment and the paucity of health effect studies, including dose-response studies, conducted on whole mixtures, if toxicity data are not available for an environmental mixture, the risk assessment could be based on surrogate toxicity information obtained from testing a sufficiently similar mixture. Biostatistical approaches for evaluating whether mixtures are sufficiently similar are included here as potential approaches that may, with further evaluation, prove useful in regulatory risk assessment contexts. The chapter concludes with a discussion of future directions for whole mixture risk assessment research.

Disclaimer Mention of trade names or commercial products does not constitute endorsement or recommendation of use.

G. E. Rice (✉)
Fort Mitchell, KY, USA
e-mail: GRice2@fuse.net

I. Eide
Statoil ASA, Trondheim, Norway
e-mail: IEide1950@gmail.com

P. I. Feder
Battelle Memorial Institute, Columbus, OH, USA
e-mail: pfeder121@gmail.com

C. Gennings
Icahn School of Medicine at Mount Sinai, New York, NY, USA
e-mail: Chris.Gennings@MSSM.edu

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_15

Keywords Whole mixtures · Risk assessment · Sufficient similarity · Chemical mixtures

List of Abbreviations

ARC	Aromatic ring class
BMD	Benchmark doses
BMR	Benchmark response
CDC	Centers for Disease Control and Prevention
CDF	Cumulative distribution function
DBP	Disinfectant by-product
DDE	Dichlorodiphenyldichloroethylene
ED	Effective dose
EPA	U.S. Environmental Protection Agency
ESI-MS	Electrospray ionization-mass spectrometry
FTIR	Fourier-transform infrared spectroscopy
GC-MS	Gas chromatography-mass spectrometry
HAA	Haloacetic acids
HAA6	Six haloacetic acids
HBPS	High-boiling point petroleum substances
HTP	High-throughput platforms
NHANES	National Health and Nutrition Examination Survey
NHL	Non-Hodgkin lymphoma
NOAEL	No-observed-adverse-effect-level
OP	Organophosphates
OR	Odds ratio
PAC	Polycyclic aromatic compounds
PAH	Polycyclic aromatic hydrocarbons
PBPK	Physiologically based pharmacokinetic
PC	Principal components
PCA	Principal component analysis
PCB	Polychlorinated biphenyls
PLS	Partial least squares
RfD	Reference doses
TI	Tolerance interval
THM	Trihalomethanes
TOC	Total organic carbon
TOX	Total organic halogen
TTHM	Total trihalomethanes
UF	Uncertainty factor
VOC	Volatile organic compounds
WQS	Weighted quantile sum

15.1 Introduction

In assessments of potentially hazardous chemical mixtures in the environment, analyses of human health risks can be conducted on the components of the mixture or the "whole mixture," depending on the available exposure information and health effect (e.g., dose-response) data. This chapter explores risk assessment approaches and describes important considerations when analyzing the risks using whole mixture data.

Risk assessments are conducted to help inform decisions faced by risk managers, sometimes referred to as decision-makers. Ideally, risk assessments are conducted in a decision-relevant context, conveying to a risk manager both what is known and not known about the risks associated with the exposures in the populations of interest. In addition to evaluating health risks when forming their decisions, risk managers often consider other factors including cost, feasibility, and social acceptance.

Human health risk assessments for environmental chemicals generally are comprised of the following series of interdependent steps (U.S. EPA 1998, 2000; NRC 1983):

- Problem formulation
- Hazard identification
- Exposure assessment
- Dose-response assessment
- Risk characterization

For environmental mixtures, risk assessments based on whole mixtures are preferred to those using component approaches (U.S. EPA 2000). Relative to risk assessments based on component data, risk analysts typically are more confident in assessments based, in part, on whole mixture toxicological or epidemiological studies, because such whole mixture studies account for interactions among mixture components and for the toxicity of unidentified compounds in the mixture. U.S. EPA (2000) also offered that whole mixture approaches should be used when component approaches are unlikely to estimate a health risk accurately. If the whole mixture is evaluated in an epidemiology study, its relevance to human health risk assessment is potentially increased, because the evidence of toxicity is in a human population, and the exposure route(s), exposure levels, exposure patterns, and observed dose-response could be pertinent to other human populations. A major limitation of using whole mixture health effect data is that their applicability may be limited to other mixtures judged "similar"; further, the health effects data may not be useful if the mixtures are judged to be not similar. If components of a mixture are known, component-based toxicology testing can be performed, but information on the mode of action and interactions among chemicals are either needed or assumed in order to assess the type of combined action applicable (e.g., independent action or similar action) and to estimate risk.

The meaning of the term "whole mixtures" is important when conducting a risk assessment on an environmental chemical mixture. "Whole mixtures" typically represents the combination of chemicals in the exposure being assessed. The composition of the mixture including the component chemicals and their proportions might be fully known, partially known, or unknown. The mixture can still be uniquely specified for the latter two cases when the process is known that caused the mixture exposure.

Often, the term "complex mixture" is used to describe a combination of exposures to chemical substances that involves significant uncertainty when addressed by component-based approaches. That uncertainty can arise from the large number of components or the large fraction of the mixture that is of unknown composition. Because of the uncertainty, the risks posed by exposures to such mixtures are often assessed as a single entity and described by the exposure scenario or source process and not by the component chemicals. Complex mixtures can consist of perhaps hundreds of component chemical substances and can occur in different phases (i.e., gas, liquid, solid). There are other definitions of complex mixtures; for example, some contend that mixtures are complex when ten or more chemicals are included (Feron et al. 1998; Feron and Groten 2002).

15.1.1 Polycyclic Aromatic Hydrocarbons: A Motivating Example of a Complex Mixture

As they occur in environmental media, the class of chemicals identified as the polycyclic aromatic hydrocarbons (PAHs) provide examples of some attributes of complex whole mixtures and illustrate some of the challenges associated with conducting risk assessments on such mixtures (Dybing et al. 2013). PAHs occur naturally in coal, crude oil, and gasoline. They also are produced when substances such as coal, gas, wood, garbage, and tobacco are burned. Sources that release PAH mixtures to the environment include wild fires, industrial processes, domestic heating, and motor vehicle emissions. PAH mixtures also can be formed during the cooking of food.

When they occur in the environment, PAH mixtures are not a single entity exhibiting the same components at fixed component ratios, rather, the composition of PAH mixtures differs over time and place; their composition also depends on the source and the medium in which they are encountered. PAH mixtures in the environment can consist of hundreds of components; some components of these PAH mixtures may not be identified chemically (i.e., the PAH mixture includes an unidentified fraction). PAHs can occur in different phases; for example, PAHs in the atmosphere can be measured bound to particulates (solid phase) and in the gas phase. PAH mixtures can occur in multiple media; in addition to being present in the air, PAH mixtures can contaminate soils, sediments, and water bodies, as well as aquatic and terrestrial food webs.

Considering this variability in occurrence, the composition of PAH mixtures to which individuals are exposed can vary based on the different conditions in which the mixture is produced or released to the environment, over time, or by location; the composition of PAH mixtures also can vary depending on the environmental medium (e.g., air, water, soil, food web) through which exposure occurs. Exposures to PAH mixtures can occur through multiple routes of exposure (i.e., oral, inhalation, and transdermal absorption). These differences pose challenges when considering the analysis of risks posed by PAH mixtures, including when deciding which PAH mixtures to evaluate in toxicological studies (Flowers et al. 2002) and evaluating PAH exposures in epidemiology studies.

15.1.2 Whole Mixture Risk Assessment: Considering Environmental Fate

Generally, chemical mixtures can be released to the environment from many different human activities. These include the following:

- Emissions from fixed sources (industrial emissions, municipal incinerators) and mobile sources (cars, airplanes)
- Releases and products from engineering processes (drinking water disinfectant by-products)
- Sources related to lifestyle (food, smoking, sprays, fuels)
- Occupational processes (workplace dusts)
- Product applications (pesticides and herbicides)

In addition to human activities, natural events in the environment can also release chemical mixtures (e.g., forest fires, decaying plant, or animal tissues).

In addition to releases into specific environmental media, some mixtures can occur in specific environmental media due to the properties of the chemicals. For example, urban air and indoor air typically contain mixtures of chemicals. Such mixtures might include volatile and semi-volatile chemicals, particulates released from human activities (see above list), and natural sources. Similarly, terrestrial and aquatic food webs might be contaminated with chemicals that partition to and are preferentially retained in adipose tissues (e.g., polychlorinated biphenyls [PCBs] and dioxin-like compounds) or other tissues. The composition and human health risks associated with these mixtures that co-occur in environmental media could differ markedly from the mixtures as released (e.g., Lorber et al. 1994; Fries 1995; Cogliano 1998).

Mixtures of chemicals also can co-occur in human tissues as a consequence of previous and current exposures and toxicokinetics. The levels of such chemicals or residues from these chemicals can be estimated through toxicokinetic studies or measured in biomonitoring studies; the chemicals present in these tissues could be considered a whole mixture.

Finally, the compositions (including the constituents and their proportions) of many mixtures are altered in the environment following their release (e.g., weathering of pesticides, photodegradation of volatile organic compounds [VOCs] in the atmosphere, partitioning of some components of a mixture into other media). The health risks associated with exposures to these altered mixtures may differ markedly from the risks posed by the original mixture (ATSDR 1996). For example, technical-grade toxaphene, an organochlorine insecticide heavily used in the United States until the 1980s, illustrates a complex mixture that changes markedly following release in the environment. Although originally applied to soils, technical-grade toxaphene contains over 670 chemicals, some of which have degradation products that can be transported into water bodies where they accumulate in fish that people eat (Simon and Manning 2006).

15.1.3 Chapter Overview

The remaining sections of this chapter discuss issues of concern when evaluating the risks using data on whole mixtures, focusing on the exposure assessment and dose-response assessment steps as they are conducted as part of a whole mixture risk assessment. (See Chaps. 2, 3, and 4 in this book for additional discussions of exposure assessment and dose-response assessment for chemical mixtures.) Section 15.2 provides a brief overview of problem formulation and hazard identification from a whole mixture perspective. Section 15.3 addresses exposure assessment issues for whole mixtures, including both qualitative and quantitative approaches. Section 15.4 addresses the selection of mixture studies for developing dose-response assessments. Section 15.5 discusses dose-response assessment and presents examples of reference doses (RfDs) and cancer slope factors that have been developed by regulatory bodies for whole mixtures. Section 15.6 proposes approaches for evaluating sufficient similarity. Using this concept, when suitable dose-response information regarding the mixture of concern (e.g., a mixture that people are exposed to when it occurs in the environment) is not available, surrogate data might be used from a sufficiently similar mixture or a group of similar mixtures. Consequently, determining whether the mixture of concern is sufficiently similar to a tested mixture or a group of tested mixtures is central to using whole mixture methods. This method is employed because, often, dose-response data are developed for very few environmental mixtures, often due to resource constraints. Finally, Sect. 15.7 addresses future research directions.

The statistical methods described throughout are described initially as the discussion warrants instead of within a statistical methods section to aid in the organization of the chapter focusing on risk assessment methods of whole mixtures.

15.2 Problem Formulation and Hazard Identification

Although this chapter on whole mixtures focuses on exposure assessment and dose-response assessment, to provide a complete discussion of such human health risk assessments, this section briefly considers conducting the first two steps of the risk assessment paradigm, problem formulation, and hazard identification. The goals of problem formulation include developing an initial understanding of key relationships between the types of health outcomes associated with the chemical mixtures of interest and potential human exposures. This entails analyzing the problem, defining the risk assessment objectives, and developing plans for data collection and analysis and risk characterization. The goal of hazard identification is to determine whether exposure to a mixture can cause an increase in the incidence of an adverse health outcome and whether that outcome likely occurs in humans (U.S. EPA 2017).

Collection and analysis of samples can be among the early steps in hazard identification. Such activities generally require preliminary identification of the chemical mixtures potentially of concern, because the appropriate collection practices and laboratory analysis methods often depend on the chemical properties of the mixture. Following chemical identification of the mixture in environmental media, risk assessors typically gather information from epidemiological and toxicological sources to help discern whether the mixture or an individual chemical potentially poses a human health risk and the nature of the health hazard (e.g., cancer, hepatotoxicity). Exposure data, such as measures of mixture concentrations in exposure media, can then be collected. The available exposure and toxicity information could indicate whether sufficient whole mixture or component health data are available to conduct a risk assessment. Generally, if sufficient toxicity and exposure information are available on the whole mixture, then such approaches are typically utilized. For a quantitative risk assessment, sufficient toxicity information could include data to develop a dose-response function or derive a health reference value. Sufficient exposure information could include measures of the chemical mixture concentration in environmental media and exposure information that characterizes human contact rates with the media or other exposure measures such as measures of the mixture in human tissues. If such quantitative exposure and dose-response data are not available, then a component-based assessment or a qualitative assessment could be conducted.

15.3 Exposure Assessment: Considering Whole Mixtures

U.S. EPA (1992) describes exposure assessment as "the determination or estimation (qualitative or quantitative) of the magnitude, frequency, duration, and route of exposure." People are generally exposed to chemical mixtures in environmental media through the oral, inhalation, and ingestion routes. These exposures

continually change over the course of a day and over longer time periods as human activities and chemical concentrations in environmental media change. Even when such exposures occur to the "same mixture," the chemicals comprising the mixture and the relative proportions of the chemicals in these mixtures can differ depending on an individual's exposure patterns and changing chemical occurrence patterns. (For additional discussion of exposure assessment issues, see Chaps. 2, 3, and 4 of this book.)

15.3.1 Using Drinking Water Disinfection By-Products to Illustrate Exposure Assessment Issues for Whole Mixtures

The following discussion on human drinking water disinfectant by-product (DBP) exposure assessments highlights the many complexities encountered in characterizing exposures to mixtures. It is followed by four example exposure assessment approaches.

Most U.S. public drinking water is disinfected using oxidants (e.g., chlorine compounds). While oxidants are deleterious to potentially infectious microorganisms in water, reactions between oxidizing agents and organic and other materials in source waters result in the formation of complex mixtures of chemical DBPs in potable drinking waters (Rook 1974). More than 600 DBPs have been identified in drinking waters (Richardson et al. 2007). The trihalomethanes (THM) and haloacetic acids (HAA) are typically the most abundant classes (by mass) present in these complex mixtures. Haloacetonitriles and haloketones, among other classes, are also routinely detected. Despite significant resources targeting identification of DBPs, much (often 50% or more) of the total organic halogen (TOX) formed during disinfection remains unidentified chemically (Richardson et al. 2007). For example, in a drinking water DBP study, Pressman et al. (2010) chemically identified the halogenated DBPs associated with approximately 60% of the TOX (by mass) that was present in the mixture; approximately 40% of the halogenated organic mass was not identified.

Chemical disinfection of drinking water appears to be a risk-risk tradeoff. While the reduction in pathogenic microorganisms has significantly decreased the number of diseases attributed to exposures to pathogen-contaminated drinking water, multiple epidemiological and toxicological studies suggest that there may be countervailing health effects in humans. Epidemiological studies examining the relationship between exposure to DBPs and human health risks report several different adverse pregnancy outcomes, including increased risk of the child being small for gestational age and increased risk of still birth, as well as increased risk of bladder and colon cancers, among other effects, although the underlying biological mechanisms that cause these effects are incompletely understood (e.g., see Cantor et al. 1987; Colman et al. 2011; Richardson et al. 2007; Villanueva et al. 2007;

Nieuwenhuijsen et al. 2009; Rahman et al. 2010; Wright et al. 2004). Thus, assessing exposures to DBPs is an important public health concern.

The concentrations of DBPs produced during drinking water treatment, as well as the individual classes of DBPs produced, vary with the disinfectant used, differences among source waters, as well as changes in source water composition (seasonal changes in levels of organic matter), and season/temperature fluctuations, among other factors. Within a drinking water distribution system, additional variation occurs due to degradation of some DBPs and DBP classes and enrichment of others, as both time and temperature can affect those DBP levels. DBP exposure assessments are further complicated by the varying water-use behaviors of individuals. DBP exposures occur through the following three exposure routes: water ingestion, inhalation, and dermal absorption (Weisel and Jo 1996; Lynberg et al. 2001; Wilkes 1998). Individual water use-behaviors, as well as those by other individuals in a household, can change daily, seasonally, and over the course of an individual's life. These behaviors affect the intensity, duration, and frequency of the DBP exposures as well as the primary routes of exposure and the composition of the DBP mixture (e.g., mixes of DBPs encountered in inhalation exposures likely differ substantially from those mixes of DBPs encountered in drinking water.) Because the composition of some mixtures is variable in the environment and whole mixture exposures also can be variable due to human activity patterns, it is critical to ensure that the dose-response data are consistent with the dose-response data for whole mixture risk assessment.

15.3.2 Qualitative and Semiquantitative Approaches

Subjects of environmental epidemiological studies are generally exposed to whole mixtures encountered in their environments. Sometimes these whole mixtures can be specifically measured and exposures estimated; other approaches evaluate whole mixture exposures by inference. Garshick et al. (1988) provides an example of an inferred exposure to a complex mixture based on occupational activities. The study objective was to investigate the association between lung cancer and regular occupational exposures to diesel exhaust, a complex mixture that includes both a vapor and particulate phase,[1] from diesel locomotives in white U.S. male railroad workers aged 40–64 years in 1959. Interest in this objective arose from epidemiological and rodent bioassay data in which exposures to relatively high concentrations of diesel exhaust were associated with lung cancer. The outcome of interest was fatal lung cancer listed as either the primary or a secondary cause of death on a death certificate from those who worked at least 10 years for U.S. railroads. Death

[1]Vapor phase constituents of diesel exhaust typically include the following: PAHs, hydrocarbons, aldehydes, nitrogen and sulfur oxides, and carbon monoxide; particulate phase constituents include PAHs, elemental carbon, sulfates, and hydrocarbons.

certificates obtained from the Railroad Retirement Board were searched through 1980 for all subjects who had died. Death certificates were obtained for 88% of all 19,396 deceased subjects. Lung cancers on death certificates were identified by ICD8 codes. Exposure to diesel exhaust from railroad locomotives occurred after 1945 when diesel locomotives were introduced. By 1959, 95% of all locomotives in the United States were diesels. Job groups with significant diesel exhaust exposures (i.e., train crews and locomotive repair shop workers) and without significant diesel exhaust exposures (i.e., clerks, ticketing assistants, station attendants, and signal maintainers) were identified through industrial hygiene evaluation. Worker job codes helped distinguish between the "exposed" and "unexposed" groups of workers. This study did not quantify the actual workplace exposures to diesel exhaust.

Other types of studies utilize surrogates to estimate exposure to the whole mixture using semiquantitative approaches, for example, approaches that categorize exposures based on total number of cigarettes that the subject smokes per day. Studies quantifying some components of whole mixture exposures are described in the next section.

15.3.3 Quantitative Approaches

In addition to qualitative and semiquantitative exposure assessment approaches, Paustenbach (2000) identified the following three general approaches for quantifying human exposures: exposure scenario, direct measurement, and biomonitoring. Some important considerations for quantitative approaches are listed in the following questions:

- Are all potential sources of the mixture well characterized?
- Are the environmental media in which the mixture occurs identified and the concentrations of the mixture in these media characterized?
- Are all potential routes of exposure and rates of contact with the relevant media well characterized?
- How well is the composition of mixtures characterized in the relevant environmental media at the point of human contact? Are all components measured? If not, how well characterized are the mixture components that are not identified chemically?
- Does the exposure assessment approach accurately characterize interindividual differences in the proportions of chemicals in the mixture that can be differentially encountered?

This section describes three approaches used to quantify exposures to mixtures. In each case, the study authors quantified exposures to a subset of the "whole mixture." The mixture subsets were typically selected as they are thought to include the most toxic components of the whole or they encompass much of the mass or a significant portion of the mass of the whole mixture.

15.3.3.1 Exposure Scenario

We draw on DBP exposure assessments to provide examples of exposure scenario development and direct measurement. Using predictive mathematical models, Teuschler et al. (2004) estimated lifetime daily exposures to four trihalomethanes, components of DBP mixtures. They estimated human contact rates for these DBPs through an exposure model (Wilkes 1998) and internal doses of these DBPs using a physiologically based pharmacokinetic (PBPK) model (U.S. EPA 2002). For many DBP mixtures, THMs are the identified class with the highest mass. The exposure modeling was based on the following information:

1. Human activity pattern studies that examined human contact time with drinking water as well as their indoor environments (e.g., time spent in the bathroom)
2. DBP physical-chemical properties that influence the fate of these individual components of the mixture in the indoor environment
3. Building characteristics that influence indoor air concentrations of volatile DBPs

The development of internal dose estimates for the DBPs allowed for the integration of exposures to the DBPs across oral, dermal, and inhalation exposure routes. The authors estimated internal doses for an adult female of reproductive age, as there are concerns about reproductive and developmental effects from DBP exposures. This scenario-based study has a number of limitations, including the small number of DBPs simulated relative to the number of DBPs comprising these mixtures in the environment. Sarigiannis and Karakitsios (this book) provide additional examples for other mixtures.

15.3.3.2 Direct Measurement

Exposure to the THMs or other specific classes of DBPs often serves as a surrogate for whole DBP mixture exposures in direct measurement studies. For example, DBP exposure measures in epidemiological studies often consist of a combination of THM measurements from water samples collected during distribution system monitoring of THM levels and the subjects' residential addresses (e.g., Bove et al. 2002). Rivera-Nunez et al. (2012) suggest that blood and tap water THM concentrations are correlated. This indicates that, although many studies using these approaches cannot account for spatial and seasonal variability in the concentration of these four DBPs within a distribution system and interindividual variability in tap water usage, these concentrations are reasonably correlated with human exposures, at least for this class of DBPs. Some studies refine exposure estimates by questioning subjects to obtain information on tap water usage and detailed residence information (e.g., Savitz et al. 1995; Waller et al. 1998).

15.3.3.3 Biomonitoring

For some chemical mixtures, previous exposures can be estimated based on the levels of chemicals (or chemical metabolites) present in body tissues or being eliminated from the body (e.g., via exhaled breath, feces, or urine). While individuals with chemicals present in their tissues have clearly been exposed to these mixtures, additional information, including contact rates with contaminated substances and toxicokinetic data, is needed to estimate previous intake rates associated with biomonitoring data and potential sources of exposures to these chemicals.

The ongoing National Health and Nutrition Examination Survey (NHANES) conducted by the Centers for Disease Control and Prevention (CDC) is among the most extensive collection efforts of biomonitoring data. The survey of noninstitutionalized U.S. civilians can be evaluated to reflect exposures in the U.S. population. In the most recent studies, over 250 chemical biomarkers were measured. The biomarkers evaluated include multiple chemical groups, including the PCBs, PAHs, VOCs, dioxins and furans, fungicides, herbicides, insecticides (e.g., organophosphates [OPs]), metals, phthalates, and DBPs. The measured biomarkers include only a subset of the chemicals comprising most complex environmental mixtures. Calafat (this book) provides additional information regarding the use of biomonitoring data to assess exposures.

15.4 Selecting Mixtures

This section describes strategies for selecting mixtures with different but overlapping composition, which subsequently may be tested with respect to toxicity in order to provide dose-response data for use in risk assessment. A prerequisite is that the chemical composition is known. Principal component analysis (PCA, Sect. 15.5.3) can be carried out on the compositional data to evaluate similarities and dissimilarities among the different samples, visualized in score plots. PCA provides the possibility to select a limited number of samples that are different in both composition and concentration for more extensive examination, such as toxicity testing. The selection can be carried out according to Kennard and Stone (1969) selecting a subset of samples which provide uniform coverage over the data set and includes samples on the boundary of the data set. The chemical characterization may be either detailed identification or quantification of the mixtures as comprehensively as possible or chemical "fingerprinting" (providing spectra or chromatograms) as described in Sect. 15.5.3. The goal is to carry out toxicity testing on as few mixtures as possible, to correlate measured toxicological responses to mixture composition, and later use compositional data of new mixtures to predict their toxicity and dose-response relationship, provided the new mixtures are sufficiently similar or within calibration domain of the regression model.

A structured approach for selecting field samples with sufficient variation in composition and concentration based on PCA is to some extent analogous to statistical experimental design, although there may be intercorrelations between *x*-variables. Statistical experimental design may be useful also in a whole mixture perspective, for example, to create mixtures mimicking mixtures of concern (e.g., Groten et al. 1991; Eide and Zahlsen 1996; Eide and Johnsen 1998; Softeland et al. 2009) and to combine fractions (Ostby et al. 1997) or to spike mixtures with individual compounds (Pressman et al. 2010; Bostrom et al. 1998). An important advantage with statistical experimental design is that the mixture composition can be controlled, and the different variables can be combined as desired. These approaches are particularly important in order to identify major contributors to toxicity or biological impact.

What mixtures to test? Risk assessments of complex mixtures are preferably based on toxicity and exposure data on the whole mixture. In such situations where the toxicologically important chemical constituents of the mixtures cannot be fully characterized, an alternative approach is to characterize more homogeneous subsets of the mixture based on physical characteristics, such as molecular weight, density, solubility, and boiling point, and estimate the toxicity of the entire mixture by aggregating the toxicities of the subsets. This has been used to develop a unified approach to evaluate the toxicities of a class of complex mixtures resulting from crude oil refining, the so-called high boiling point petroleum substances (HBPS), which have a boiling point greater than 650°F and that contain a wide variety of polycyclic aromatic compounds (PAC). "A Supplement: Assessing the Mammalian Toxicity of High-boiling Point Petroleum Substances" (Gray et al. 2013) presents a unified approach to evaluating toxicity of HBPS. The approach discussed in the supplement involves testing representative mixtures within the class (reference set) with various types of in vivo and in vitro toxicity tests and predicting the toxicity of untested members of the class based on their aromatic hydrocarbon content, represented as their "aromatic ring class" (ARC) profile. Aromatic ring classes fall into seven classes with successively higher boiling point in each group. These classes can be characterized analytically. The ARC profile is the percent by weight within each class. The toxicities observed in the reference set of mixtures are used to predict the toxicities of new mixtures as long as the ARC profiles fall within the range of the ARC profiles of the reference substances (i.e., are interpolations within the range of reference set ARC profiles used to develop the prediction models). The proximity of the ARC profile of the new mixture to the ARC profiles of the reference set mixtures constitutes an operational definition of "sufficient similarity" for HBPS containing PAC. The predictive models combine the ARC profiles of the mixtures with dose, biological characteristics (e.g., gender, weight) of the test animals, and test conditions to predict toxicological dose-response trends or mutagenic potential. Models have been developed to predict subchronic toxicological endpoints such as changes in selected organ weights and hematology parameters, reproductive and developmental endpoints such as fetal body weight and percent resorptions per litter, as well as mutagenic effects based on the Ames assay.

Nicolich et al. (2013) developed a class of models to predict repeated-dose, reproductive, and developmental toxicity responses for HBPS with no toxicity test information based on the toxicity test information of HBPS in the reference set. The models include terms that reflect characteristics of the animal (sex, body weight, control group response, test chemical dose, and the ARC profile and its interactions with dose). Nicolich et al. (2013) discuss their model validation procedure, which included, among other things, getting predictions based on training sets and validation sets.

Gray et al. (2013) discuss the application of this class of models to predict the relationship between subchronic endpoints and ARC profile of the HBPS. They demonstrate by example that the models result in very good agreement between toxicity predicted by their models and observed toxicity for multiple toxicity endpoints.

15.5 Dose-Response Assessment of Complex Chemical Mixtures

For whole mixture risk assessments, in the dose-response assessment step, the analyst investigates and quantifies the relationship between the magnitude of a mixture dose and a biological response. Response can be expressed as the measured or observed incidence or difference in level of response in a population of study subjects (U.S. EPA 2012). Dose is typically measured as the quantity of a mixture administered over time or, in epidemiological studies, by the concentration of a mixture in an environmental medium and the population's contact rate with that medium (e.g., in studies of contaminated drinking water, the dose would be computed as the product of the volume of drinking water consumed over time and the concentration of the mixture). The mixture dose is typically treated as a single entity. Typically, as the total dose increases, the measured response also increases. At low doses, there may be no response. Both the dose at which response begins to appear and the rate at which it increases (given increasing doses) can vary across different mixtures, populations, and exposure routes.

An important consideration for whole mixture dose-response studies is that many whole mixtures are only partially characterized with respect to constituent chemical mass per unit mass or volume of mixture, for example, mass of drinking water DBPs per liter of drinking water (see Sect. 2.1). In such cases, dose can be specified based on those chemical components that have been characterized chemically. This is just a subset of the total composition of the mixture but, if carefully selected, can account for most of the mixture toxicity. These known constituents can be controlled for in laboratory toxicity studies by appropriately spiking or diluting mixture components. If only observational data can be obtained (e.g., from an epidemiological study), then the amount of known constituents exposed to or ingested can be determined and observed but not controlled. In either the

controlled laboratory test situation or the "uncontrolled" epidemiological study, there can be an uncharacterized portion of the mixture composition that contributes to the toxicity, but which cannot be observed or controlled.

U.S. EPA (2000) published a procedure to estimate toxicity values based on dose-response assessments of whole mixtures. When considering the dose-response data, the procedure "treats" the whole mixture as a single entity. The procedure has the following five steps:

1. *Collect and evaluate available epidemiological and toxicological data on the mixture*. Generally, epidemiological data are preferred over toxicological data for human health risk assessments—a limitation of such data is that the mixture composition associated with the exposure may differ across individuals, potentially increasing the variability of the response.
2. *Evaluate the stability of the mixture*. Because variation in the composition of the mixture can affect its toxicity, examine variability in components and their relative proportions both over time within an environmental medium and across environmental media. The examination of variability in mixture composition across media may need to consider multiple media; for example, the composition of mixtures of dioxin-like congeners varies among air, plant tissues, and beef tissues as a consequence of uptake and retention of dioxin-like congeners from the air into plants and, subsequently, plant tissues to beef tissues.

Another important aspect of mixture stability is to analyze the stability of the tested mixture, as samples are prepared and administered to test subjects over time within a study. The composition of a mixture can vary due to (1) natural variations of the mixture in the environment where the mixture is collected, (2) concentration procedures, (3) preparation of the mixture for administration in a toxicological study, and (4) on-going chemical reactions of components (e.g., degradation) during storage prior to or use in a toxicological study.

3. *Decide which mixture to test* (if examining an assortment of comparable mixtures). See discussion in Sect. 15.4.
4. *Conduct dose-response assessment on the tested whole mixture using single chemical procedures* (e.g., slope factors (see Sect. 15.5.1) and RfDs (see Sect. 15.5.2).
5. *Characterize uncertainties* (e.g., relevance of health effects to environmental exposures and stability of the mixture composition [proportions and chemical concentrations] and dose over time).

15.5.1 Deriving Cancer Slope Factors for Whole Mixtures

The cancer risks associated with exposures to whole mixtures can be evaluated in epidemiological and toxicological studies. Data from these studies can be used directly to estimate cancer risks. In this section, we describe an example that

Table 15.1 Summary of technical-grade toxaphene dose-response study performed by Litton Bionetics (1978) in B6C3F1 male mice and further analyzed by U.S. EPA (1991)

Administered dose		Human equivalent dose (mg/kg-day)	Tumor incidence
(ppm)	(mg/kg-day)		
0	0.0	0	10/53
7	0.91	0.051	10/54
20	2.6	0.144	12/53
50	6.5	0.361	18/51

Source: Litton Bionetics (1978) and U.S. EPA (1991)

illustrates the use of whole mixture toxicology data to derive an oral slope factor. The data were obtained from a study of toxaphene. As discussed previously (see Sect. 15.1), technical-grade toxaphene, which was first introduced in 1947, contains over 670 chemicals. In the United States, toxaphene was banned for most uses in 1982 and was banned for all uses in 1990 (U.S. EPA 1999).

The slope factor was derived from a long-term mouse carcinogenicity bioassay (U.S. EPA 1991; Litton Bionetics 1978). Toxaphene was administered in test animals' diets for 18 months at doses of 0, 7, 20, and 50 ppm; the test animals were then observed for 6 months posttreatment. An increased incidence of hepatocellular carcinomas and adenomas was observed in the male mice (see Table 15.1). Using the linearized multistage model, the U.S. Environmental Protection Agency (EPA) estimated the oral cancer slope factor to be 1.1 per (mg/kg-day). Table 15.1 presents the original study data and the modeled human exposure estimates developed by EPA.

As toxaphene applications have been banned in the United States for decades and few people in the United States are now likely exposed to technical toxaphene, U.S. human health concerns have focused on exposures to weathered toxaphene. Because some components of technical-grade toxaphene may volatilize to air and others have limited solubility in water, the composition of the mixture will differ depending on whether the toxaphene is encountered in soil at a contaminated site, the air around the site, or in nearby lake sediments (ATSDR 1996). Further, some components of technical-grade toxaphene have been measured in shellfish and fish (ATSDR 1996). In fact, after analyzing the toxicokinetics of toxaphene for a dose-response analysis, Simon and Manning (2006) proposed a component approach for estimating cancer risk from eating fish contaminated with toxaphene congeners.

15.5.2 Deriving and Applying Reference Doses for Mixtures of PCBs

PCBs occur as mixtures in the environment, although they are no longer manufactured in the United States. PCB mixtures are comprised of different ratios of the 209 individual PCB congeners. Chemical properties among congeners vary

widely; some of these properties are due to the different number of chlorines on the different PCB congeners. After release into the environment, PCB mixture compositions change over time through the following processes:

- Partitioning
- Chemical transformation
- Preferential bioaccumulation

When assessing risks from environmental PCBs, consideration of how environmental processes alter mixture composition is critical because such changes could alter the toxicity of the mixture.

Based on several reports describing reproductive studies in monkeys, U.S. EPA (1996) published RfDs for several commercial PCB mixtures known as Aroclors. Each Aroclor has a different level of chlorination. In one set of studies, Aroclor 1016 was administered to adult female monkeys for 22 months, beginning 7 months prior to breeding and continuing until offspring were weaned at age 4 months. Relative to control animals, there was a significant decrease in the birth weights of the offspring of the high-dose group that was estimated to have received 0.028 mg/kg-day of Aroclor 1016. EPA identified the low-dose group that received 0.007 mg/kg-day Aroclor 1016 as a no-observed-adverse-effect-level (NOAEL) for the mixture. The EPA reference dose was based on the following formula:

$$\mathrm{RfD_m} = \frac{\mathrm{NOAEL_m}}{\mathrm{UF_m}} \tag{15.1}$$

where

$\mathrm{RfD_m}$ = reference dose for the mixture
$\mathrm{NOAEL_m}$ = no-observed-adverse-effect level for the mixture;
$\mathrm{UF_m}$ = uncertainty factors for the mixture

U.S. EPA (1996) applied four uncertainty factors. An uncertainty factor of 3.16 was applied to account for extrapolating from experimental rhesus monkeys to humans (uncertainty factors are based on a log scale and the actual value is $10^{0.5}$). A second uncertainty factor of 3.16 was applied to account for sensitive populations. A third uncertainty factor of 3.16 was applied to account for a subchronic-to-chronic exposure duration, and a fourth uncertainty factor of 3.16 was applied to account for the lack of a two-generation reproductive study and of reproductive studies in adult males. The total uncertainty factor for the RfD was 100; this is the product of these four individual factors ($\mathrm{UF_m} = 100 \approx 3.16 \times 3.16 \times 3.16 \times 3.16$).

U.S. EPA (1996) described the confidence in this RfD as medium because the congener pattern of mixtures of PCBs in the environment does not match those observed in Aroclor 1016. For environmental applications where it is known that Aroclor 1016 is the only form of PCB contamination, EPA recommended that this RfD could be used with high confidence; however, all other applications only merited medium confidence.

Cogliano (1998) addressed the issue of characterizing the risks posed by environmental PCB mixtures that exhibited differences in composition due to transformation in the environment. He considered the cancer slope factors for Aroclors 1260, 1254, 1242, and 1016 that have been derived by the EPA from different rat cancer bioassays. He recognized that the cancer potency of these mixtures differed markedly. Considering the composition of environmental PCBs encountered through various environmental media, he categorized three levels of PCB mixtures, each with a markedly different estimated cancer potency. Cogliano suggested that the highest slope factor among the tested Aroclors be applied to the high-risk, persistent environmental PCB mixtures that resulted from the following: food web exposures, sediment or soil ingestion, or dust or aerosol inhalation. He assigned one of the lower slope factors among the tested Aroclors to the second level that included low-risk, low-persistent environmental PCB mixtures that would result from ingestion of water-soluble congeners and inhalation of volatile congeners. For the third level that included the lowest risk and least persistent mixtures where the congeners with more than four chlorines comprise less than 0.5% of total PCBs, he assigned the lowest factors among the tested Aroclors for this group. In this way, Cogliano addressed the environmental processes that alter the cancer potential of PCB mixtures.

15.5.3 Multivariate Data Analysis

Multivariate data analysis may be useful in whole mixture risk assessment, first of all to evaluate similarities and differences between mixtures, to relate biological endpoints to chemical composition, and to predict dose-response relationships. Generally, multivariate data analysis can be used to obtain the structured information inherent in large data sets with many variables. In mixture toxicology and risk assessment, there will be an **X**-matrix describing the chemical composition of a number of samples and the **Y**-matrix the biological effects or toxicity of the same samples. This implies that in such examples, an *x*-variable corresponds to a chemical compound, and a *y*-variable corresponds to a biological parameter or toxicological response.

Frequently used multivariate data analysis techniques for obtaining the structured information in **X**- and **Y**-matrices include the following:

- PCA for exploratory data analysis (Jackson 1991), for example, in order to evaluate whether mixtures are sufficiently similar and also to select a subset of samples for further investigation
- PLS regression (Wold et al. 1983; Martens and Næs 1992) to correlate mixture composition (**X**-matrix) to measured properties such as toxicity (**Y**-matrix) and to predict dose-response relationships for new mixtures within calibration domain

PLS finds the relationship between the matrix **Y** (response variables) and the matrix **X** (predictor variables) by simultaneous projections of both the **X-** and **Y-**spaces to a plane or hyperplane. This is analogous to PCA; however, PCA is performed on one data matrix (**X** or **Y**), and PLS evaluates both (**X** and **Y**) simultaneously to develop a predictive model (e.g., predict **Y** from **X**) and to evaluate relationships between specific *x*- and *y*-variables (e.g., which chemicals covary with toxicity).

Statistically designed mixtures enable the identification of cause-effect relationships (i.e., the relative contribution from the different compounds in the mixture). Environmental samples, on the other hand, are not statistically designed, which usually implies intercorrelations between individual measured compounds and influential background factors that may not be measured or even identified. PLS regression will therefore identify compounds that may be associated (i.e., correlated) with toxicity, but not necessarily having cause-effect relationships with toxicity.

A limitation in the ability to use multivariate data analysis, for example, to evaluate similarity of whole mixtures is that, with complex mixtures, a detailed and complete characterization may be a challenge. Consequently, sometimes only a limited number of a priori selected compounds are identified and quantified. Chemical "fingerprints" (spectra, chromatograms) may be used as alternatives to the detailed identification and quantification of individual compounds for a first screening of similarities between mixtures. This screening identifies the important lines or peaks in the spectra or chromatograms for subsequent detailed identification and quantification.

Multivariate data analysis was used to predict the composition-response relationship between particle and semi-volatile organic chemical constituents in gasoline and diesel vehicle exhaust samples and toxicity as measured by inflammation and tissue damage in rat lungs and mutagenicity in bacteria (McDonald et al. 2004). Exhaust samples were collected from "normal" and "high-emitting" gasoline and diesel light-duty vehicles, and a large number of chemical compounds were identified and quantified. PCA was used to group the different endpoints, showing that the lung toxicity data and bacterial mutagenicity responded to different chemical components. The PLS regression revealed the chemical constituents covarying most strongly with toxicity and produced models predicting the relative toxicity of the samples with good accuracy. The specific nitro-PAHs important for mutagenicity were the same chemicals that have been implicated from decades of bioassay-directed fractionation. These chemicals were not related to lung toxicity, which instead was associated with organic carbon and select organic compounds that are present in lubricating oil. The results demonstrate the utility of the PCA/PLS approach for evaluating composition-response relationships in complex mixture exposures and for providing a starting point for confirming causality and determining the mechanisms of the lung effects.

In another study (Eide et al. 2002), pattern recognition and multivariate regression were used in assessing complex mixtures by correlating chemical fingerprints to the mutagenicity of the mixtures. The mixtures were 20 organic extracts of

exhaust particles, each containing 102–170 individual compounds such as PAHs, nitro-PAHs, oxygenated PAHs, and saturated hydrocarbons. The mixtures were characterized by full-scan gas chromatography-mass spectrometry (GC-MS). The data were resolved into peaks and spectra for individual compounds by an automated curve resolution procedure. Resolved chromatograms were integrated, resulting in a predictor matrix that was used as input for PCA to evaluate similarities between mixtures (i.e., classification). Furthermore, PLS regression was used to correlate the GC-MS data to mutagenicity, as measured in the Ames Salmonella assay (i.e., calibration). The regression model can be used to predict mutagenicity from GC-MS chromatograms of other organic extracts provided they are sufficiently similar or at least within calibration domain.

Synthetic mixtures of three C9 n-paraffinic, naphthenic, and aromatic hydrocarbons (n-nonane, trimethylcyclohexane, and trimethylbenzene, respectively) were studied in rats after inhalation for 12 h (Eide and Zahlsen 1996). The hydrocarbons were mixed according to principles for statistical experimental design to support an empirical model with linear, interaction, and quadratic terms (Taylor polynomial). Immediately after exposure, concentrations of hydrocarbons were measured by headspace gas chromatography in the blood, brain, liver, kidneys, and perirenal fat. The best PLS regression models were obtained after removing all interaction terms, suggesting that there were no interactions between the hydrocarbons with respect to absorption and distribution. Uptakes of paraffins, and particularly aromatics, were best described by quadratic models, whereas the uptake of the naphthenic hydrocarbons was nearly linear. All models exhibited good fits with high correlation coefficients (r^2) and prediction properties (Q^2), the latter after cross validation (Wold, 1978). The PLS models were used to create curves for tissue and blood concentrations versus exposure. The approach may be useful in risk assessment of combinations of these hydrocarbons provided the relationship between tissue or blood concentration and health impact is known.

15.5.4 Estimating Mixture Effects by Weighting Components

Another strategy for identifying the relative contributions from different compounds in a mixture is weighted quantile sum (WQS) regression, which is a weighted sum of quantiles of the components in the mixture (Carrico et al. 2015; Czarnota et al. 2015). For example, if quartiles are used, the concentrations of the *j*th component in the lowest quartile are scored $q_j = 0$, in the second quartile $q_j = 1$, in the third quartile $q_j = 2$, and the highest quartile $q_j = 3$. The sum of quantiles is used to represent exposure to multiple chemicals by binning exposure levels relative to the sample of concentrations for each chemical so that extreme values are bounded and the range of concentrations across chemicals is standardized. The approach is motivated by the fact that standard regression methods are challenged with complex correlated variables—commonly the case in environmental mixtures due to exposure patterns, human behavior, and metabolic effects. In short, WQS

regression determines an empirically weighted index of chemical concentrations with weights determined by the association with a specified health outcome. Thereby, generally, components least associated with the outcome are downweighted and those highly associated are upweighted. The constraints imposed by the model (namely, that the weights sum to one and that there is a single regression coefficient associated with the index either in a positive or negative direction) improve the ill-conditioning due to the complex correlation pattern among the variables. The advantage of reducing the dimensionality to a unidimensional index is that the test for the significance of the mixture effect is a single degree-of-freedom test, which has increased power. The approach focuses inference in a single direction—for example, increasing risk. When interest is in both an increasing and decreasing direction, two separate analyses may be conducted constraining the regression coefficients in each direction.

To illustrate the strategy, we use a study of the suspected risk factors for non-Hodgkin lymphoma (NHL). Czarnota et al. (2015) used WQS regression to model the association of a mixture of 27 correlated environmental chemicals measured in household dust, with the risk of NHL in a case-control study conducted in four geographic locations in the United States (Detroit, MI, several agricultural communities in Iowa, Los Angeles, CA and Seattle, WA). The estimated WQS index was a sum of weighted quartiles for 5 PCBs, 7 PAHs, and 15 pesticides. Chemical weights and a mixture effect were estimated for all study sites combined and for each site separately. For comparison, risk estimates for individual chemicals were computed using logistic regression. Models were adjusted for age, gender, race, education, and study site. The chemical exposures showed a complex correlation pattern, with pairwise correlations ranging from slightly negative to near-perfect correlation. There was a high degree of intragroup correlation among the PAHs (range 0.87–0.96) and PCBs (range 0.69–0.91), whereas the pesticides generally exhibited weaker correlation (interquartile range 0.06–0.26). An increase in the WQS index was significantly associated with increased risk of NHL in the full study population ($p = 0.006$). Specifically, a one-quartile increase in the index was associated with a multiplicative increase of 1.30 (95% confidence interval: 1.08, 1.56) in NHL risk. The WQS index placed non-negligible weight on several chemicals that displayed elevated and potentially meaningful odds ratios (ORs) in traditional single chemical analysis but were not found to be significant, likely due to a lack of power. The only single chemical significantly associated with increased risk of NHL was PCB 180 ($p < 0.01$), which received 32% of the weight in the WQS index. Three other PCBs received 7% of the weight and had multiplicative increases in NHL risk (4th versus 1st quartiles) of at least 1.2 but with nonsignificant p values ($p = 0.07, 0.12, 0.27$). Exposures in the highest quartile for α-chlordane (OR = 1.40; $p = 0.06$), γ-chlordane (OR = 1.35; $p = 0.09$), propoxur (OR = 1.27; $p = 0.18$), and DDE (OR = 1.26; $p = 0.19$) were associated with increased risk of NHL in the mixture, with 37% of the weight. Since all chemicals with non-negligible weights factor positively into the WQS index, the additional chemicals contributing to the mixture effect, not found significantly associated in single chemical analyses, were considered to be important.

15.6 Developing Approaches to Evaluate Sufficient Similarity

Complex mixtures typically include those mixtures whose constituents are so numerous that they cannot be fully characterized. This is in contrast to pure chemicals or simple mixtures whose content can be characterized exactly. One way to regard complex mixtures is that the portion that can be well specified can be modeled as fixed components and the portion that cannot be specified can be modeled as random components. Thus, a specified complex mixture is not a single entity, but would exhibit random variation around a central value. Sources of random mixture variation could be due to variation over time, laboratory, treatment facility, processing procedures, or ambient conditions. This would affect the fixed components to a limited extent and the random components that would be expressed as random variation in outcomes. This implies that health outcomes are associated with complex mixture regions rather than single entities. Characterizing those regions entails testing families of mixtures exhibiting natural variation in outcomes. By limiting the extent of variation within the family, one can obtain toxicities or health outcomes within specified tolerances around the central value. These are termed "similar regions." One would then determine whether a new mixture falls within this "similar region," preferably based on toxicity or health outcomes but alternatively based on closeness of the known portion of the chemical constituents if toxicity or health information is not available. Thus the analysis of complex mixtures is completely intertwined with sufficient similarity and tolerance regions.

Testing multiple doses of complex mixtures in test species and evaluating the responses are resource intensive. In addition to routine resources required to conduct a single chemical animal bioassay, preparing or purifying the mixture, characterizing the chemical composition, and analyzing the stability of such complex mixtures generally require further resources (e.g., Pressman et al. 2010; Pressman et al. 2012). As a consequence, limited numbers of complex mixtures can be tested in this manner.

Considering these limitations in resources and the variability of whole mixtures in the environment, few dose-response studies are conducted on whole mixtures. U.S. EPA (2000) offered that, if toxicity data are not available for a mixture of concern, the risk assessment could be based on surrogate toxicity information obtained from testing a *sufficiently similar* mixture. This condition assumes that the toxicological consequences of exposure to the two mixtures are nearly identical. Mixtures judged to be sufficiently similar would exhibit relatively few differences in toxicological effects between the mixtures or their components. A mixture would likely be sufficiently similar to another when its components are not very different, and the components are roughly in the same proportions. Further, similar mixtures would likely have few differences in environmental fate, bioavailability, and pharmacokinetics. U.S. EPA (2000) did not propose a specific method by which similarity could be judged. While in some cases expert toxicological or

epidemiological judgments regarding similarity among mixtures have been made, this section reviews several biostatistical studies that have developed approaches that examine similarity among mixtures. These approaches offer the advantage that they are repeatable, and the criteria and approach used to evaluate similarity among mixtures are clearly articulated.

15.6.1 Evaluating Sufficient Similarity Using Multivariate Data Analysis

The example described in Sect. 15.5.3 on the statistical relationship between particle and semi-volatile organic chemical constituents in exhaust samples and toxicity and mutagenicity (McDonald et al. 2004) can also be used to illustrate evaluation of similarities among mixtures depending on whether similarity is based on chemical composition or, for example, toxicological responses. The three score plots in Fig. 15.1 illustrate that the seven exhaust samples group differently if they are based on the chemical composition (184 compounds), lung toxicity (11 parameters), or mutagenicity (4 parameters). The corresponding loading plot shown in Fig. 15.2 (from McDonald et al. 2004) demonstrates that the mutagenicity and lung toxicity parameters group differently because they respond to different compounds in the samples. This was also verified by PLS regression.

15.6.2 Evaluating Sufficient Similarity Using Multivariate Distance Functions (Mahalanobis D^2)

Feder et al. (2009) illustrate the application of multivariate statistics to evaluate sufficient similarity for DBP mixtures. Multivariate graphical and analytical methods are applied to assess the degree of similarity of input water supplies and output water supplies from five treatment plants using data from Schenck et al. (2009). The input water is drawn from ground water (one treatment plant) and from surface water (four treatment plants). For each treatment plant, the output water sampled is either "finished water" that has just left the treatment plants or "distribution water" at various locations along the distribution pipelines, removed from the treatment plants.

To characterize similarity among treatment plant/water source process combinations, a "reference set" needs to be determined. The treatment plants/water sources within the reference set are relatively homogeneous, are characterized with respect to toxicological effects and/or chemical composition, and are considered to have been sampled from a reference distribution. For treatment plants/water sources not in the reference set, statistical methods are used to determine whether their chemical/toxicological characteristics can be treated as having been drawn

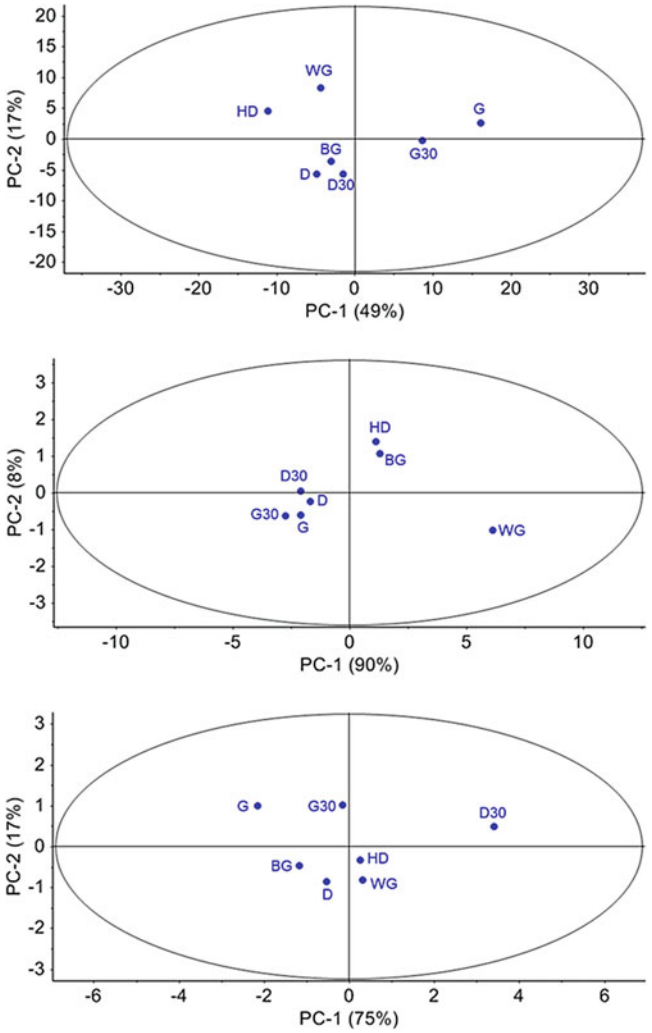

Fig. 15.1 Score plots obtained after PCA of data on 184 chemical compounds (top), 11 lung toxicity (middle), and 4 mutagenicity responses (bottom) of gasoline and diesel vehicle exhaust samples (Recalculated from McDonald et al. (2004)). The vehicles: gasoline (G), diesel (D), white (WG) or black (BG) smoke gasoline, high-emitting diesel (HD), gasoline and diesel operating at 30 °F (G30, D30)

from the same reference distribution as the reference set. If so, they can be considered sufficiently similar to the reference set, and risk assessment for these water sources can be based on risks associated with the reference set. Note that the similarity of water sources with respect to toxicological effects is of principal interest from a public health perspective. Similarity of chemical composition among water sources is of interest to the extent that similarity of chemical

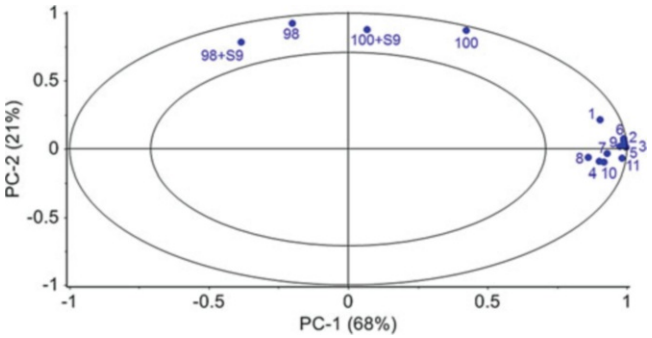

Fig. 15.2 Correlation loading plot showing the groupings among the 11 lung toxicity measurements and the 4 mutagenicity responses (TA98 ± S9 and TA100 ± S9). Separation of mutagenicity and lung toxicity groups suggested that they responded to different chemical components (Recalculated from McDonald et al. (2004))

characteristics leads to the similarity of toxicological characteristics. Since detailed toxicological characteristics of sources outside the reference set are usually not known, the similarity of chemical characteristics within the reference set is the best available and is adopted as a surrogate criterion.

The treatment plants/water sources for the Schenck et al. (2009) data are represented by six broad chemical characteristics and mutagenic activity, a toxicological characteristic:

- Total organic carbon (TOC)
- Total organic halogens (TOX)
- Mutagenic activity (revertants/L equivalent)
- Total trihalomethanes (TTHM)
- Six haloacetic acids (HAA6)
- Percent brominated TTHM
- Percent brominated HAA6

Let **X** denote the chemical and mutagenic characteristics for a treatment plant/water source within the reference set. Denote the reference set as $X_1, \ldots, X_M$, and let $\bar{X}, S$ denote the mean vector and covariance matrix of the X's, for the mth treatment plant/water source outside the reference set. Let X_m denote the measured characteristics, and let ${D_m}^2$ denote the Mahalanobis D^2 statistic (Morrison 1976):

$$D_m^2 \equiv \left(X_m - \bar{X}\right)' S^{-1} \left(X_m - \bar{X}\right) \tag{15.2}$$

The Mahalanobis D^2 is the square of the distance of each observation to the center of the distribution, accounting for the variances and the covariances in the data. An observation X_m is classified as potentially dissimilar if ${D_m}^2$ is statistically significant.

The ordered ${D_m}^2$ values are plotted in a chi-square probability plot. Treating the overall mean vector and covariance matrix as approximately known, as M gets

large, the distances are distributed approximately as chi square with degrees of freedom equal to the dimension of the vector, **p**. Under normality assumptions, if the X_m's are drawn from the reference distribution, the plot should resemble a straight line with slope 1. If some of the X's are not drawn from the reference distribution, their distances would be expected to lie above the curve through the majority of the chi-square values.

In the analysis of the Schenck et al. (2009) data, the X_m's were taken from the reference set of values as was used to calculate $\bar{X}, S$. To approximate the situation where the X_m's were taken from independent distributions, a robust version of the Mahalanobis D^2 statistic involves removing one observation at a time when calculating the mean vector and the covariance matrix. For each observation $\mathbf{X_m}$, let $\mathbf{X}_{(-m)}$ and $\mathbf{S}_{(-m)}$ represent the mean vector and the covariance matrix, respectively, of the M–1 observations that remain after omitting X_m. Repeating this for each X_m, the robust version of the Mahalanobis D^2 statistic is calculated for $m = 1, \ldots, M$, where one observation is removed for each calculation. Denote the robust version of the above distance function ${D^2}_m$ as ${RD^2}_{(-m)}$:

$$\mathrm{RD}^2_{(-m)} = \left(\mathbf{X}_m - \bar{\mathrm{X}}_{(-m)}\right)' \left[\mathrm{S}_{(-m)}\right]^{-1} \left(\mathrm{X}_m - \bar{\mathrm{X}}_{(-m)}\right) \tag{15.3}$$

An observation is classified as potentially dissimilar if ${\mathrm{RD}^2}_{(-m)}$ lies above the curve through the majority of the values. Figure 15.3 displays the robust Mahalanobis D^2 distances versus percentiles from the chi-square distribution with p degrees of freedom for the Schenck et al. data ($p = 7$). The reference line through

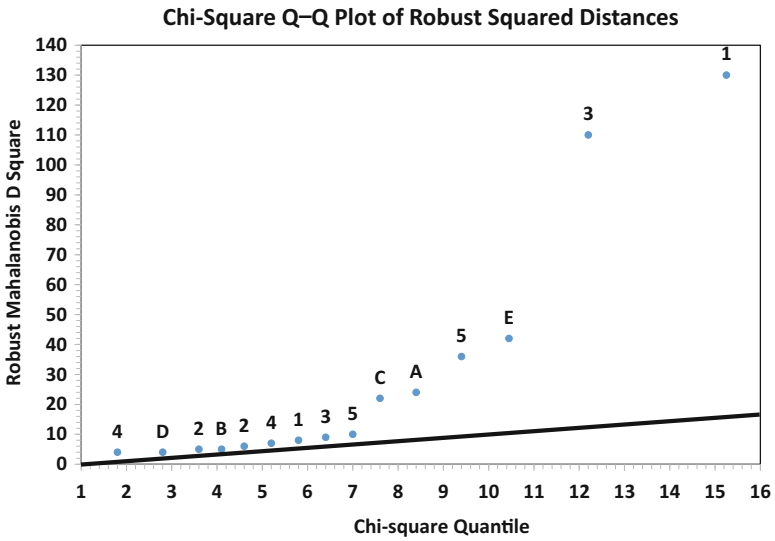

Fig. 15.3 Robust chi-square plot of Mahalanobis square distance. Symbols A–E represent finished water, and symbols 1–5 represent distribution water. Reference line corresponds to the chi-square distribution CDF with $p = 7$ degrees of freedom

(0,0) with slope 1 corresponds to the cumulative distribution function (CDF) of the chi-square distribution with $p = 7$ degrees of freedom. Two distribution water samples, from sources 1 and 3, deviate substantially from the chi-square line. These treatment plant/water source combinations would not be sufficiently similar to the reference set to carry out risk assessments based on the reference set toxicity determinations.

15.6.3 Evaluating Sufficient Similarity Using Principal Component Analysis

As discussed previously, PCA is a data analytic procedure often used for exploratory purposes that represents relatively high-dimensional response vectors in lower-dimensional space, such that most of the variation in high-dimensional data is reflected in the lower-dimensional representation. Feder et al. (2009) applied PCA to the Schenck et al. (2009) data. The analysis was carried out on the studentized residuals (i.e., scaling variance to 1 [unit variance]) after adjusting for treatment plant and water type (finished water or distribution water). The first and second principal components (PC) are those orthogonal directions that have the greatest variation among the observations. Successive PCs reflect less and less variation among the observations.

Representing the data in the space of the PCs with the largest variances reflects major relationships in the original responses, in particular if the responses cluster into subsets rather than coming from a single homogeneous distribution. Representing the data in the space of the PCs with the smallest variances, in which the reference set is most highly bunched, facilitates the detection of outlying processes corresponding to dissimilar mixtures. Donnell et al. (1994) and Jolliffe (2004, Chap. 10) discuss applications of the PCs with the smallest variances. Jolliffe states "...by examining...the last few PCs we may be able to detect observations that violate the correlation structure...imposed by bulk of the data, but that are not necessarily aberrant with respect to individual variables...."

For each treatment plant and output water sample type, the normalized mean responses expressed in PC coordinates were added back to the studentized residual scores. Figure 15.4 displays the recentered normalized values in the space of the first and second PCs. These PCs explain 60.7% and 21.3% of the total variance, respectively. Figure 15.5 displays the recentered normalized values in the space of the sixth and seventh PCs. These PCs explain 1.0% and 0.6% of the total variance, respectively. In each plot, the letter plotting symbols A, B, ..., E correspond to the five finished water samples from the five water treatment plants. The number plotting symbols 1, 1, 2, 2,...5, 5 correspond to the ten distribution water samples from the five water treatment plants.

The first and second PCs are the dimensions in which the individual data points exhibit the greatest variation. Together they explain 82.0% of the total variation in

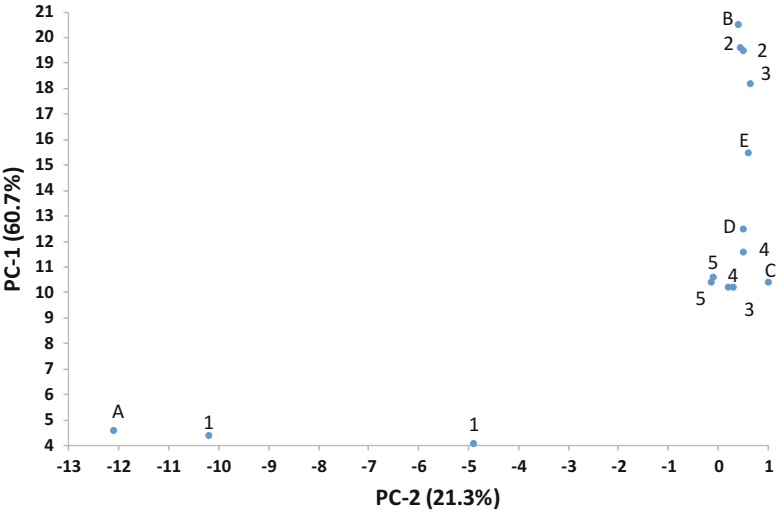

Fig. 15.4 Schenck et al. (2009) data. Score plot. First principal component versus second principal component. Symbols A, . . ., E represent finished water samples. Symbols 1, . . ., 5 represent distribution water samples

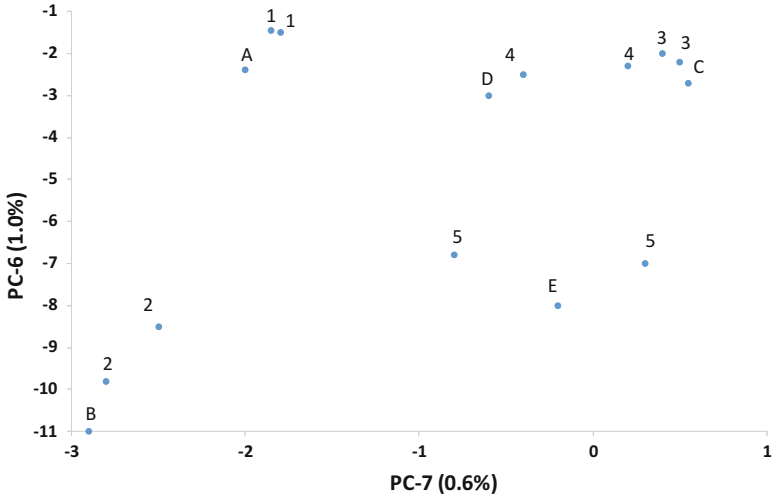

Fig. 15.5 Schenck et al. (2009) data. Score plot. Sixth principal component versus seventh principal component. Symbols A, . . ., E represent finished water samples. Symbols 1, . . ., 5 represent distribution water samples

the data. Figure 15.4 displays the structure of the primary variability in the data. The sixth and seventh PCs are the dimensions in which the individual data points exhibit the least variation (i.e., they are most tightly clustered). Together they explain just

1.6% of the total variability in the data. It is not the objective of studying these PCs to explain variability in the data. Rather, because the data are so tightly clustered, in these dimensions, individual data points that do not conform to the correlation structure of the remainder of the data points (i.e., multidimensional outliers) and finer structure in the variability of the data are magnified. Figure 15.5 illustrates more subtle contributions to the variability in the data.

Figure 15.4 shows that the characteristics of both the finished water and the distribution water from treatment plant 1 (with a groundwater source) differ considerably from those from treatment plants 2–5 (with surface water sources). Figure 15.5 shows that the output water characteristics from the five treatment plants are separated from one another. For each plant, the finished water samples and the distribution water samples cluster together. Within each cluster, the sixth PC scores for the finished water samples fall below those for the distribution water. The scores for the sixth PC are positive for total trihalomethanes and six haloacetic acids and are negative for TOC and mutagenic activity. This suggests that the finished water has higher TOC and mutagenic activity and/or lower TTHM and HAA6 than the distribution water.

Figures 15.6 and 15.7 display the loadings for the first and second PCs and for the sixth and seventh PCs, respectively. In Fig. 15.6, the coefficients of PC-1 and PC-2 for A through E are clustered and are distinct from the coefficients for percent

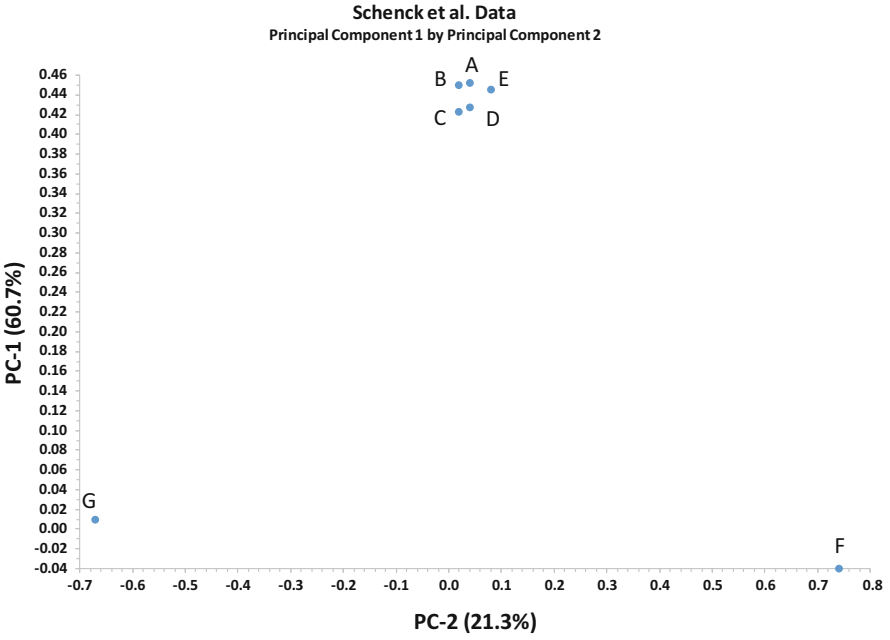

Fig. 15.6 Schenck et al. (2009) data. Loading plot. First principal component versus second principal component. Symbols A, . . ., G represent A, TOX; B, TOC; C, MUTG; D, TTHM; E, HAA6; F, Pct Br TTHM; G, Pct Br HAA6

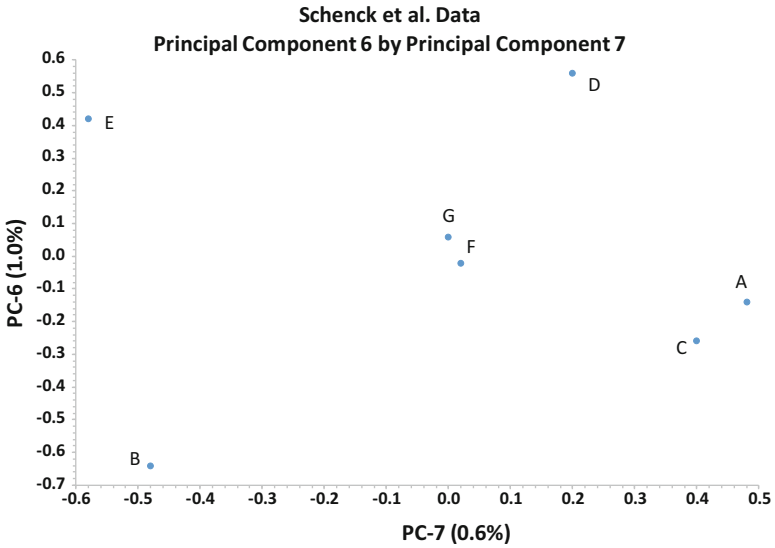

Fig. 15.7 Schenck et al. (2009) data. Loading plot. Sixth principal component versus seventh principal component. Symbols A, . . ., G represent A, TOX; B, TOC; C, MUTG; D, TTHM; E, HAA6; F, Pct Br TTHM; G, Pct Br HAA6

bromination F and G. In Fig. 15.7, the coefficients of PC-6 and PC-7 for percent bromination F and G are close to 0. The coefficients for the chemical groups and mutagenicity are scattered and contrast with one another. PC-1 is essentially an equi-weighted sum of the four chemical groups and mutagenicity. Percent bromination has essentially no contribution. PC-2 is essentially a contrast between percent bromination of the TTHM components and the percent bromination of the HAA6 components. The other components have little contribution. PC-6 is a contrast between the sum of TTHM and HHA6 versus TOC. The other components have little contribution. PC-7 is a contrast between TOX, mutagenicity, and to a lesser extent TTHM versus TOC and HAA6. The percent bromination components have little contribution. Percent bromination enters primarily into PC-2, which separates the single ground water source from the surface water sources.

15.6.4 Evaluating Sufficient Similarity Using an Equivalence Testing Approach

Marshall et al. (2013) developed a test for sufficiently similar mixtures using equivalence testing methods. The approach is in contrast to that discussed in Sect. 15.6.2 where a distance measure was used among observed mixtures. Here, exposure/epidemiology data are used to identify environmentally relevant

mixtures, and representative reference mixtures are experimentally evaluated using toxicity dose-response data. The idea is to determine the distance between benchmark doses (BMDs) (in terms of total dose) of different mixtures: a reference mixture and candidate mixtures, where dose-response data are available on the reference mixture. The analysis strategy is developed for the data-rich case where dose-response data is also assumed available for the candidate mixtures. Under simplifying assumptions, the strategy can be used when only mixing proportions are available on candidate mixtures. When the distance between BMDs is within a predetermined similarity region, the mixtures are claimed to be sufficiently similar. In essence, the similarity is based on the dose-response effect of the specified benchmark response (BMR). Statistical details are provided below.

Let c be the total number of chemicals in the candidate and reference mixtures. Consider a BMR is set and associated BMDs are estimated for both mixtures. Let $\boldsymbol{\theta_r}$ be the BMD for the reference mixture and $\boldsymbol{\theta_i}$ be the BMD for the ith candidate mixture. The test for similarity between two mixtures is based on a weighted distance between the BMDs for the two mixtures. Define $\mathbf{W}$ as a diagonal matrix with respective weights of assumed relative potency of the mixture components, w_j, subject to the constraint, $\sum_{i=1}^{C} w_i = C$. The weighted distance (d_w) is constructed in the following manner:

$$\begin{aligned} d_w &= \sqrt{(\boldsymbol{\theta}_r - \boldsymbol{\theta}_i)' \mathbf{W} (\boldsymbol{\theta}_r - \boldsymbol{\theta}_i)} \\ &= \sqrt{\sum_{j=1}^{c} w_j \left(\theta_{jr} - \theta_{ji}\right)^2} \end{aligned} \tag{15.4}$$

Using relative potencies to calculate the weighted distance is preferred to the unweighted distance (where $w_j = 1$ for all $j = 1, \ldots, c$), which may be used when relative potencies are not available. Relative potency values are obtained from external sources; for example, Marshall et al. (2013) described single chemical toxicity data where potency factors were determined.

Two BMDs are considered to be sufficiently similar when $d_w \leq \Delta$, where Δ is a positive number, set a priori. Using equivalence testing methodology, we test the following hypothesis:

$$\begin{aligned} H_0 &: d_w > \Delta \\ H_1 &: d_w \leq \Delta \end{aligned}$$

This can be tested using the principle of confidence interval inclusion (Berger and Hsu 1996); that is, an alpha-level test rejects H_0 when the upper limit of the one-sided confidence interval on d_w does not exceed Δ. For example, assuming $\widehat{d}_w$ has a bell-shaped distribution,

$$\widehat{d}_w + t_{1-\alpha;N-p} \sqrt{\mathrm{Var}\left(\widehat{d}_w\right)} \leq \Delta \tag{15.5}$$

where

$\widehat{d}_w$ is estimated from available dose-response data

$t_{1-\alpha;N-p}$ is the critical value from the t distribution, associated with Δ (e.g., 5%) in the upper tail, with $N - p$ degrees of freedom (i.e., total sample size minus number of model parameters in the dose-response data for the mixtures)

$\mathrm{Var}\left(\widehat{d}_w\right)$ is the variance of the estimate for distance, estimated from the dose-response data

In the usual data-poor case, dose-response data are only available on a reference mixture. To set notation, define T_{r} as the BMD in terms of total dose (i.e., the summed doses of each mixture component in $\boldsymbol{\theta}_r$). For the reference mixture with proportions of each chemical given in the c-dimensional vector of proportions $\mathbf{a}_r$, ($\mathbf{a}_{rj} \geq 0$, $\sum_j \mathbf{a}_{rj} = 1$), the BMD is the proportion $\mathbf{a}_r$ times the total dose, T_{r}, for each chemical; that is, using vector notation, $\boldsymbol{\theta}_r = T_r \mathbf{a}_r$.

In this data-poor case, we assume we know the mixing ratio weights, $\mathbf{a}_i$, for the ith mixture, but as there are no studies available to estimate the BMD (in terms of total dose) for this mixture, we cannot estimate T_i or $\boldsymbol{\theta}_i$. Thus, we cannot estimate d_w (from Eq. 15.3) without simplifying assumptions.

Without loss of generality, consider a two-chemical mixture with mixing proportions of, say, 0.3 and 0.7 and total dose for the BMD of, say, 10 units (i.e., 3 units of chemical 1 and 7 units of chemical 2). It is reasonable to expect that when the mixing proportions are, for example, 0.31 and 0.69, the total dose BMD should be close to 10 units. Thus, it is reasonable to assume that mixtures with similar mixing proportions have similar BMDs, that is, $T_i \approx T_r$. Thus, in the data-poor case, we propose to estimate the BMD for the ith mixture by adjusting the total dose of the mixture to that of the reference mixture; that is,

$$\boldsymbol{\theta}_i^{\mathrm{adj}} = T_r \mathbf{a}_i \tag{15.6}$$

where

T_r is the total dose of BMD for the reference mixture.

In this scenario, the distance measure is

$$d_w = \sqrt{\sum_{i=1}^{c} w_j \left(\theta_{jr} - \theta_{ji}^{\mathrm{adj}}\right)^2} = T_r \sqrt{\sum_{i=1}^{c} w_j \left(a_{jr} - a_{ji}\right)^2} \tag{15.7}$$

which can be estimated as

$$\widehat{d}_w = \widehat{T}_r \sqrt{\sum_{i=1}^{c} w_j \left(a_{jr} - a_{ji}\right)^2} \tag{15.8}$$

The distance estimate increases as the difference between corresponding mixing proportions increases. The variance of distance estimate is

$$\mathrm{Var}\left(\widehat{d}_w\right) = \mathrm{Var}\left(\widehat{T}_r\right)\left\{\sum_{i=1}^{c} w_j\left(a_{jr} - a_{ji}\right)^2\right\} \tag{15.9}$$

Thus, the variance is the product of a term that accounts for biological variability and the sum of weighted squared differences between the mixing proportions. The estimate for $\mathrm{Var}\left(\widehat{T}_r\right)$ is found using the delta method in the analysis of the mixture dose-response data for the reference mixture.

To illustrate the method, Marshall et al. (2013) use pesticide residue data from the Child Care Center Study, a nationally representative survey of pesticide concentrations from floor wipes in rooms where children spend most of their time in day care. Fifteen pesticides were measured from 168 child care centers; residue concentrations from roughly 25% of the centers were all below the limit of detection of the assay. The centers with the top 10% of total concentrations were used to identify a mixing proportion used in a toxicity study of neurodevelopment using motor activity as the endpoint. The BMD and lower confidence limit were estimated from this study. The objective was to determine the proportion of candidate mixtures from the 126 centers with detectable concentrations that were sufficiently similar to the BMD from the toxicity study. The authors outlined their strategy for determining the similarity region for the motor activity assay, which, in short, was based on the change in total dose with a 20 percentage point change in response from the benchmark reference. Potency weights were used to calculate the weighted distance (Eq. 15.8) between observed mixing proportions and the reference mixture with 95% confidence intervals on each (from Eq. 15.5). On the weighted total dose scale, the BMD of the reference mixture was 4.2 mg/kg; the critical value for the similarity region was determined to be the difference in the estimated BMD (from an effective dose [ED] for the reference mixture at which 20% of the subjects respond [ED_{20}] BMR) and the ED_{40} (6.3 mg/kg)—a 20 percentage point shift in response. Thus, the critical value was 2.1 in units of the weighted total dose. There were 114 centers with the upper confidence limit on the distance between BMDs (in the data-poor case) less than the similarity boundary of 2.1. In this study, the authors concluded that 90% of the child care centers with detectable pesticide concentrations were sufficiently similar to the reference mixture.

15.7 Future Directions for Assessing Risks Posed by Whole Mixtures

Evaluating the risks associated with exposures to whole mixtures can be accomplished using information from the fields of epidemiology, toxicology, cell biology, and biochemistry that inform the scientific understanding of responses to chemical mixture-mediated insults at subcellular, cellular, and tissue levels. The types of toxicology data can include in vitro and in vivo data, as well alternative information platforms.

Human exposure-response data from epidemiology studies play an increasingly important role in the risk assessment of whole mixtures through the use of biomonitoring data and increasingly sophisticated exposure and toxicokinetic models that relate chemical concentrations measured in internal matrices (e.g., blood, urine, hair, toe/fingernails, teeth) to concentration measures collected in environmental media (e.g., air, water, foods, and dust). While biomonitoring measures reflect "actual" human exposures, additional information and, typically, mathematical models are required to account for complexities, such as the sources of the exposures and the actual magnitude of mixture intake, when studying the potential impact on human health. Matrices, such as hair, nails, urine, and teeth, provide opportunities for measuring retrospective exposures for evaluation of health effects. For example, concentrations measured from the growth rings in teeth (e.g., Andra et al. 2015) allow for estimates of prenatal and postnatal early-life exposures to be linked with health effects later in life (e.g., autism spectrum disorder).

15.7.1 Using Toxicity Data on Whole Mixtures from Alternative Information Platforms

In the near future, risk assessors will have opportunities to consider toxicity data on whole mixtures based on high-throughput platforms (HTP), in addition to "traditional" sources of toxicity information. Prior to testing a whole mixture in such platforms, the mixture may need to be isolated and concentrated. The whole mixture toxicity information from these platforms would likely be more useful in risk assessments, if the tested mixture is characterized chemically, to the extent possible; for example, see discussions of the chemical characterization of a DBP mixture analyzed in a toxicological study by the EPA (Simmons et al. 2002, 2008; Pressman et al. 2010; Speth et al. 2008). Alternatively, the chemical mixture could be treated with metabolizing enzymes prior to treatment. Dose-response assessments and sufficient similarity could be examined using data generated through such high-throughput platforms, which potentially include toxicogenomics, proteomics, metabolomics, chemoinformatics, bioinformatics, and cell-based bioactivity screening assays, although, at this time, information from these sources is more useful qualitatively rather than quantitatively in risk assessments. To utilize HTP data for dose-response assessment in humans, the relationship between the test outcome (e.g., receptor-binding) and human disease would need to be addressed (e.g., what is the toxicity pathway from the endpoint measured in the HTP assay to the health outcome of concern?). When evaluating similarity among mixtures, how informative the HTP data are likely will depend on the overall understanding of the relationship between the HTP endpoint being tested and the human disease.

Chemical characterization is essential in exposure assessment of whole mixtures and also in the evaluation of similarities between mixtures. With complex mixtures, a

detailed and complete characterization may be a challenge. Chemical "fingerprints" (spectra, chromatograms) may be used as alternatives to the detailed identification and quantification of individual compounds for a first screening of similarities between mixtures. Generally, modern analytical instruments and particularly hyphenated techniques (e.g., GC-MS, ESI-MS) generate huge amounts of data. These data can provide great opportunities for insights into toxicity and toxicity pathways through the actions of chemical mixtures on specific targets and potentially through the propagation of such interactions through cellular-level, tissue-level, or organ-level responses. When considering data generated from such techniques, data handling and analysis become increasingly important. The study described in Sect. 15.5.3 (Eide et al. 2002) on correlating chromatograms to the mutagenicity of a number of mixtures is one example of using chemical "fingerprints" instead of identifying and quantifying numerous compounds. Only important peaks may subsequently be subject to detailed characterization. In addition to chromatograms, spectra obtained by, for example, electrospray ionization-mass spectrometry (ESI-MS) or Fourier-transform infrared spectroscopy (FTIR) can be used to evaluate similarities between mixtures. This kind of fingerprinting and subsequent multivariate data analysis has been used on biofuels and fossil fuels (Eide and Zahlsen 2007; Eide and Neverdal 2014); however, these studies did not include toxicological parameters.

Additionally, automated online multivariate analysis of data from multiple sensors may be useful, for example, in long-term exposure assessment. The time series provide basis for risk assessment and in addition the possibility for early detection of changes, not necessarily a significant change in one parameter, but minor changes in several parameters simultaneously. One example, although related to the marine environment, illustrates the possibilities. At the LoVe Ocean Observatory (http://love.statoil.com), a total of 91 parameters are measured frequently, and data are submitted online for automated interpretation using PCA over intervals spanning a few minutes. In the future perspective, this same concept will be used to monitor the impact of discharges and emissions and may also include analysis of spectra and chromatograms. The software used are the Unscrambler X and Process Pulse from Camo Software, Oslo, Norway.

15.7.2 Biomonitoring Data

The U.S. Army Center for Environmental Health Research has developed and implemented an aquatic biomonitoring system to monitor drinking water quality from water treatment plants and effluent water discharges from industrial plants. The system uses blue gills as sentinels to identify changes in fish ventilatory and movement patterns. Fish are held in individual chambers under flow-through conditions, and electrical signals generated by reflexes of individual fish to water conditions are continuously monitored. Ventilatory responses are measures including ventilatory rate, ventilatory depth, gill purge (cough) frequency, and whole

body movement. The ventilatory responses are compared against a reference distribution based on blue gills exposed to water that meets health and environmental standards for drinking or for industrial effluent. The system alarms when a specified percentage of the sentinel fish exhibit ventilatory response deviations from reference values beyond a statistically tuned baseline threshold. An important characteristic of this biologically based biomonitoring system is that it responds to specified levels of biological response, with criteria invariant to the nature of the toxicant. This work has been carried out for many years by Dr. W.H. van der Schalie and has been widely published, for example, van der Schalie et al. (2004).

The above biomonitoring system can be extended to provide quantitative comparisons of different water sources or of a single water source over time. Namely, the biomonitoring system can be augmented with positive and negative control water. The time to response of the test water relative to the time to response of the positive control water provides a metric to quantify the degree of toxicity of the test water.

15.7.3 Sufficient Similarity Based on Tolerance Intervals and Components of Variation

It was stated at the beginning of Sect. 15.6 on sufficient similarity that the principal distinguishing characteristic between simple or component mixtures and whole or complex mixtures is that the chemical makeup of complex mixtures cannot be fully characterized, and the uncharacterized portion can be regarded as varying in a random fashion among samples of the complex mixture separated spatially or temporally. This section illustrates how tolerance interval methodology accounting for multiple sources of random variation can be used to develop similarity regions for toxicity or health outcomes based on a reference sample of similar "sources" that can be used to characterize the similarity or dissimilarity of future "sources." These "sources" could well be complex mixtures from water treatment plant finished water, from industrial effluents, from diesel exhaust, or many other sources.

As discussed in earlier sections, the ability to carry out complex mixture risk assessment analyses at sources for which toxicity results are not available requires the availability and use of sufficiently similar sources for which toxicity results are available. The extent of natural variability among sources is characterized by establishing a reference set of sources whose toxicity levels are considered to be acceptable and assessing the variability of responses among sources within the reference set. The reference set is considered to be a random sample of sources from a homogeneous population of sources. The use of such reference sets is illustrated in Sect. 15.4 with respect to the prediction of various toxicity endpoints for new petroleum HBPS's based on the toxicity levels observed in a reference set of HBPS's and their associated ARC profiles. Variability among sources in the

reference set includes systematic factors and random factors. Systematic factors are determinable values and enter into comparisons among sources as covariates to adjust for and reduce the variability among them. They are characterized by their individual levels. For example, for comparisons among water treatment plants, systematic factors might include water source type (ground water versus surface water), season, type of treatment (chlorination versus chloramination), water stage (finished, distribution), or sampling time/distance from treatment, and concentrations of measurable constituents. Random factors cannot be controlled and are considered to be components of variation among sources. They are characterized by the variability among replicate sources within the reference set, for example, water sources within type, treatment plants, longer-term time-to-time variation within treatment plant, and short-term replication variation.

Variation among sources within the reference set can be characterized in terms of toxicological characteristics or chemical concentrations of selected components of the mixtures. As discussed earlier, it is more physically meaningful to define sufficient similarity in terms of toxicological endpoints such as mutagenic activity. This requires that these endpoints are available for the sources in the reference set as well as the new sources to be compared with those in the reference set. In the event that these toxicological endpoints are not available for the new sources, sufficient similarity can be defined in terms of chemical concentrations and would be anticipated to imply similarity of the toxicological endpoints of interest.

The variability among replicate sources within the reference set is determined and used to set bounds on acceptable values for new sources in order to consider them sufficiently similar to the reference set. Such bounds are based on "tolerance intervals," which are constructed to include a specified portion of the population (e.g., 90%) from which the reference set is drawn with high probability. Rode and Chinchilli (1988) and Krishnamoorthy and Mathew (2009) discuss the construction and application of tolerance intervals for univariate and multivariate endpoints. Tolerance interval methodology has been extended to account for multiple components of variation.

Although not involving mixtures, the illustration below of the application of tolerance intervals in toxicology with multiple components of variation is informative. In this example, an in vitro assay has been established at multiple laboratories (the reference set), and one wants to determine whether the results obtained in several additional laboratories are sufficiently similar to those from the reference set laboratories. In this example, the reference set of laboratories is analogous to the reference set of treatment plants or petroleum samples. Components of variation include those among laboratories, among tasks within laboratories, among runs within tasks, and within runs. Figure 15.8 below shows a two-sided tolerance interval to contain at least 80% of the reference set population of test runs with 95% confidence.[2]

[2]See U.S. EPA (2009) and OECD (2015) for discussion of statistical methods underlying Figure 15.8.

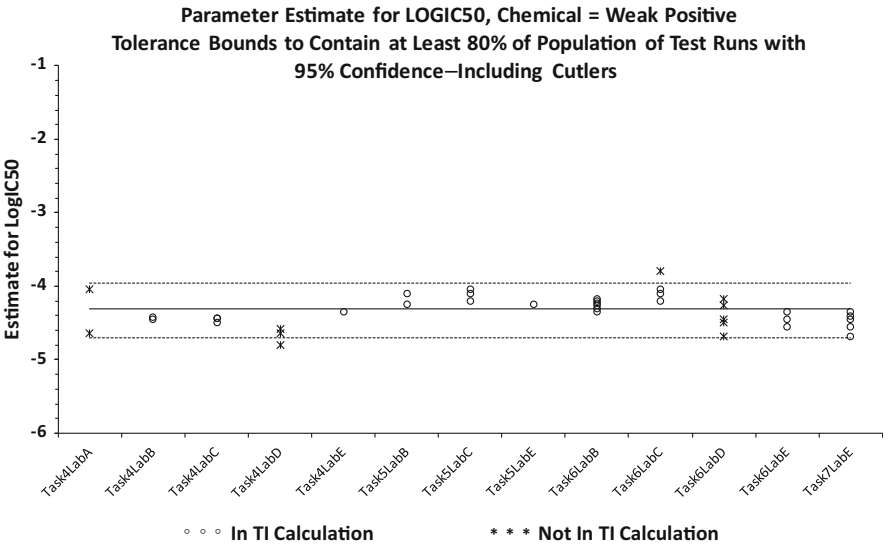

Fig. 15.8 Tolerance interval (TI) to contain at least 80% of the population of in vitro assay dose-response $\log_{10}$IC50 with 95% confidence

15.8 Conclusions

Whole mixture approaches are preferred to component approaches when assessing risks and hazards associated with exposures to potentially hazardous chemical mixtures in the environment. Approaches have been developed for conducting exposure assessments and dose-response assessments based on whole mixture data. These approaches typically treat the mixture as a single entity. Developing the underlying mixture data to implement these assessment approaches can be resource intensive. Given the resources needed to develop whole mixture toxicology or epidemiology data and considering the variability of mixtures in the environment, the use of whole mixture data in risk assessments from a mixture judged to be sufficiently similar is appealing. Addressing whether two mixtures (i.e., the tested mixture and the mixture in the environment that is of concern) are sufficiently similar is a critical question for furthering the applicability of whole mixture methods. Several biostatistical approaches for evaluating whether mixtures are sufficiently similar are presented in this chapter, but additional case studies are needed to increase the frequency with which such approaches are used in mixture risk assessments. Finally, three potential future directions are discussed. These include the following: (1) examining dose-response and sufficient similarity for whole mixtures using data generated through such high-throughput platforms, which potentially include toxicogenomics, proteomics, metabolomics, chemoinformatics, bioinformatics, and cell-based bioactivity screening assays, although, at this time, information from these sources is more useful qualitatively rather than quantitatively in risk assessments, (2) using biomonitoring

data in whole mixture assessments, and (3) using tolerance intervals and components of variation to evaluate sufficient similarity among whole mixtures.

Acknowledgments All figures adapted with permission.

Figures 15.1, 15.2 recalculated from McDonald, JD; Eide, I; Seagrave, J; Zielinska, B; Whitney, K; Lawson, DR; Mauderly, JL. (2004). Relationship between composition and toxicity of motor vehicle emission samples. Environ Health Perspect 112: 1527-1538. https://doi.org/10.1289/ehp.6976

Figures 15.3, 15.4, 15.5 adapted from Feder, PI; Ma, ZJ; Bull, RJ; Teuschler, LK; Schenck, KM; Simmons, JE; Rice, G. (2009). Evaluating sufficient similarity for disinfection by-product (DBP) mixtures: multivariate statistical procedures. Journal of toxicology and environmental health Part A 72: 468-481. http://www.ncbi.nlm.nih.gov/pubmed/19267308

Figures 15.6, 15.7 adapted from Schenck, KM; Sivaganesan, M; Rice, GE. (2009). Correlations of water quality parameters with mutagenicity of chlorinated drinking water samples. Journal of toxicology and environmental health Part A 72: 461-467. http://www.ncbi.nlm.nih.gov/pubmed/19267307

Figure 15.8 adapted from U.S. EPA (2009) OSCP Endocrine Disruptor Screening Program (EDSP). See, e.g., Integrated summary report for validation of an estrogen receptor binding assay using rat urine cytosol as source of receptor. Table 28 and Appendix 9

The authors wish to thank Drs. Richard C. Hertzberg, Cynthia V, Rider, and Jane Ellen Simmons for many helpful discussions and comments on a preliminary version of this manuscript.

References

Andra, S., C. Austin, R. Wright, and M. Arora. 2015. Reconstructing pre-natal and early childhood exposure to multi-class organic chemicals using teeth: Towards a retrospective temporal exposome. *Environment International* 83: 137–145.

ATSDR. 1996. *Toxicological profile for toxaphene (update)*. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service.

Berger, R.L., and J.C. Hsu. 1996. Mutagenicity testing of organic extracts of diesel exhaust particles after spiking with PAHs. *Statistical Science* 11: 283–319.

Bostrom, E., S. Engen, and I. Eide. 1998. Mutagenicity testing of organic extracts of diesel exhaust particles after spiking with polycyclic aromatic hydrocarbons (PAH). *Archives of Toxicology* 72: 645–649.

Bove, F., Y. Shim, and P. Zeitz. 2002. Drinking water contaminants and adverse pregnancy outcomes: a review. *Environmental Health Perspectives* 110 (Suppl 1): 61–74.

Cantor, K.P., R. Hoover, P. Hartge, T.J. Mason, D.T. Silverman, R. Altman, D.F. Austin, M.A. Child, C.R. Key, L.D. Marrett, and E. Al. 1987. Bladder cancer, drinking water source, and tap water consumption: A case-control study. *Journal of the National Cancer Institute* 79 (6): 1269–1279.

Carrico, C., C. Gennings, D.C. Wheeler, and Pam Factor-Litvak. 2015. Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. *Journal of Agricultural, Biological, and Environmental Statistics* 20 (1): 100–120. https://doi.org/10.1007/s13253-014-0180-3.

Cogliano, V.J. 1998. Assessing the cancer risk from environmental PCBs. *Environmental Health Perspectives* 106 (6): 317–323.

Colman, J., G.E. Rice, J.M. Wright, E.S. Hunter 3rd, L.K. Teuschler, J.C. Lipscomb, R.C. Hertzberg, J.E. Simmons, M. Fransen, M. Osier, and M.G. Narotsky. 2011. Identification of developmentally toxic drinking water disinfection byproducts and evaluation of data

relevant to mode of action. *Toxicology and Applied Pharmacology* 254 (2): 100–126. https://doi.org/10.1016/j.taap.2011.02.002.

Czarnota, J., C. Gennings, J.S. Colt, A.J. De Roos, J.R. Cerhan, R.K. Severson, P. Hartge, M.H. Ward, and D.C. Wheeler. 2015. Analysis of environmental chemical mixtures and non-Hodgkin lymphoma risk in the NCI-SEER NHL Study. *Environmental Health Perspectives* 123 (10): 965–970. https://doi.org/10.1289/ehp.1408630.

Donnell, D.J., A. Buja, and W. Stuetzle. 1994. Analyses of additive dependencies and concavities using smallest additive principal components. *Annals of Statistics* 22: 1635–1673.

Dybing, E., P.E. Schwarze, P. Nafstad, K. Victorin, and T.M. Penning. 2013. Polycyclic aromatic hydrocarbons in ambient air and cancer. Chapter 7. IARC Publications. Air Pollution and Cancer. IARC Scientific Publication No. 161. Editors: Kurt Straif, Aaron Cohen, and Jonathan Samet. http://www.iarc.fr/en/publications/books/sp161/161-Chapter7.pdf.

Eide, I., and H.G. Johnsen. 1998. Mixture design and multivariate analysis in mixture research. *Environmental Health Perspectives* 106 (Suppl 6): 1373–1376.

Eide, I., and G. Neverdal. 2014. Fingerprinting bio-oils from lignocellulose and comparison with fossil fuels. *Energy & Fuels* 28: 2617–2623.

Eide, I., G. Neverdal, B. Thorvaldsen, B. Grung, and O.M. Kvalheim. 2002. Toxicological evaluation of complex mixtures by pattern recognition: correlating chemical fingerprints to mutagenicity. *Environmental Health Perspectives* 110 (Suppl 6): 985–988.

Eide, I., and K. Zahlsen. 1996. Inhalation experiments with mixtures of hydrocarbons. Experimental design, statistics and interpretation of kinetics and possible interactions. *Archives of Toxicology* 70 (7): 397–404.

———. 2007. Chemical fingerprinting of biodiesel using electrospray mass spectrometry and chemometrics: Characterization, discrimination, identification, and quantification in petrodiesel. *Energy & Fuels* 21: 3702–3708.

Feder, P.I., Z.J. Ma, R.J. Bull, L.K. Teuschler, K.M. Schenck, J.E. Simmons, and G. Rice. 2009. Evaluating sufficient similarity for disinfection by-product (DBP) mixtures: multivariate statistical procedures. *Journal of Toxicology and Environmental Health. Part A* 72 (7): 468–481. https://doi.org/10.1080/15287390802608965.

Feron, V.J., F.C. Cassee, and J.P. Groten. 1998. Toxicology of chemical mixtures: International perspective. *EHP* 106 (6): 1281–1289.

Feron, V.J., and J.P. Groten. 2002. Toxicological evaluation of chemical mixtures. *Food and Chemical Toxicology* 40: 825–839.

Flowers, L., S.H. Rieth, V.J. Cogliano, G.L. Foureman, R.C. Hertzberg, E.L. Hofmann, D.L. Murphy, S. Nesnow, and R.S. Schoeny. 2002. Health assessment of polycyclic aromatic hydrocarbon mixtures: Current practices and future directions. *Polycyclic Aromatic Compounds* 22 (3–4).

Fries, G.F. 1995. Transport of organic environmental contaminants to animal products. *Reviews of Environmental Contamination and Toxicology* 141: 71–109.

Garshick, E., M.B. Schenker, A. Munoz, M. Segal, T.J. Smith, S.R. Woskie, S.K. Hammond, and F.E. Speizer. 1988. A retrospective cohort study of lung cancer and diesel exhaust exposure in railroad workers. *The American Review of Respiratory Disease* 137 (4): 820–825. https://doi.org/10.1164/ajrccm/137.4.820.

Gray, T.M., B.J. Simpson, M.J. Nicolich, F.J. Murray, A.W. Verstuyft, R.N. Roth, and R.H. McKee. 2013. Assessing the mammalian toxicity of high-boiling petroleum substances under the rubric of the HPV program. *Regulatory Toxicology and Pharmacology* 67 (2 Suppl): S4–S9. https://doi.org/10.1016/j.yrtph.2012.11.014.

Groten, J.P., E.J. Sinkeldam, T. Muys, J.B. Luten, and P.J. van Bladeren. 1991. Interaction of dietary Ca, P, Mg, Mn, Cu, Fe, Zn and Se with the accumulation and oral toxicity of cadmium in rats. *Food and Chemical Toxicology* 29 (4): 249–258.

Jackson, J.E. 1991. *A user's guide to principal components*. New York: John Wiley & Sons, Inc.

Jolliffe, I.T. 2004. *Principal component analysis*. New York: Springer-Verlag.

Kennard, R.W., and L.A. Stone. 1969. Computer aided design of experiments. *Technometrics* 11: 137–148.

Krishnamoorthy, K., and T. Mathew. 2009. *Statistical tolerance regions: theory, applications, and computation*. New York: Wiley.

Litton Bionetics. 1978. Carcinogenic evaluation in mice: toxaphene. Final report. Prepared for Hercules, Inc., Wilmington, DE. LBI Project No. 20602. Litton Bionetics, Inc., Kensington, MD.

Lorber, M., D. Cleverly, J. Schaum, L. Phillips, G. Schweer, and T. Leighton. 1994. Development and validation of an air-to-beef food chain model for dioxin-like compounds. *Science of the Total Environment* 156 (1): 39–65.

Lynberg, M., J.R. Nuckols, P. Langlois, D. Ashley, P. Singer, P. Mendola, C. Wilkes, H. Krapfl, E. Miles, V. Speight, B. Lin, L. Small, A. Miles, M. Bonin, P. Zeitz, A. Tadkod, J. Henry, and M.B. Forrester. 2001. Assessing exposure to disinfection by-products in women of reproductive age living in Corpus Christi, Texas, and Cobb county, Georgia: descriptive results and methods. *Environmental Health Perspectives* 109 (6): 597–604.

Marshall, S., C. Gennings, L.K. Teuschler, L.G. Stork, R. Tornero-Velez, K.M. Crofton, and G.E. Rice. 2013. An empirical approach to sufficient similarity: Combining exposure data and mixtures toxicology data. *Risk Analysis* 33 (9): 1582–1595. https://doi.org/10.1111/risa.12015.

Martens, H., and T. Næs. 1992. *Multivariate calibration*. New York: Wiley.

McDonald, J.D., I. Eide, J. Seagrave, B. Zielinska, K. Whitney, D.R. Lawson, and J.L. Mauderly. 2004. Relationship between composition and toxicity of motor vehicle emission samples. *Environmental Health Perspectives* 112 (15): 1527–1538. https://doi.org/10.1289/ehp.6976.

Morrison, D.F. 1976. *Multivariate statistical methods*. New York: McGraw-Hill.

Nicolich, M.J., B.J. Simpson, Murray FJa, R.N. Roth, and T.M. Gray. 2013. The development of statistical models to determine the relationship between aromatic-ring class profile and repeat-dose and developmental toxicities of high-boiling petroleum substances. *Regulatory Toxicology and Pharmacology* 67 (2): S10–S29. https://doi.org/10.1016/j.yrtph.2012.11.015.

Nieuwenhuijsen, M.J., J. Grellier, R. Smith, N. Iszatt, J. Bennett, N. Best, and M. Toledano. 2009. The epidemiology and possible mechanisms of disinfection by-products in drinking water. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 367 (1904): 4043–4076. https://doi.org/10.1098/rsta.2009.0116.

NRC. 1983. *Risk assessment in the federal government*. Washington, DC: National Academy Press.

OECD. 2015. Integrated Summary Report: Validation of Two Binding Assays Using Human Recombinant Estrogen Receptor Alpha (hrERa). OECD Environment, Health, and Safety Publications Series on Testing and Assessment. No 226. Appendix F. https://nepis.epa.gov/Exe/ZyNET.exe/P10044EB.TXT?ZyActionD=ZyDocument&Client=EPA&Index=2006+Thru+2010&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5Czyfiles%5CIndex%20Data%5C06thru10%5CTxt%5C00000008%5CP10044EB.txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C-&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=hpfr&DefSeekPage=x&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=1&ZyEntry=1&SeekPage=x&ZyPURL.

Ostby, L., S. Engen, A. Melbye, and I. Eide. 1997. Mutagenicity testing of organic extracts of diesel exhaust particles after fractionation and recombination. *Archives of Toxicology* 71 (5): 314–319.

Paustenbach, D.J. 2000. The practice of exposure assessment: a state-of-the-art review. *Journal of Toxicology and Environmental Health. Part B, Critical Reviews* 3 (3): 179–291. https://doi.org/10.1080/10937400050045264.

Pressman, J.G., S.D. Richardson, T.F. Speth, R.J. Miltner, M.G. Narotsky, E.S. Hunter 3rd, G.E. Rice, L.K. Teuschler, A. McDonald, S. Parvez, S.W. Krasner, H.S. Weinberg,

A.B. McKague, C.J. Parrett, N. Bodin, R. Chinn, C.F. Lee, and J.E. Simmons. 2010. Concentration, chlorination, and chemical analysis of drinking water for disinfection byproduct mixtures health effects research: U.S. EPA's Four Lab Study. *Environmental Science & Technology* 44 (19): 7184–7192. https://doi.org/10.1021/es9039314.

Pressman, J.G., D.L. McCurry, S. Parvez, G.E. Rice, L.K. Teuschler, R.J. Miltner, and T.F. Speth. 2012. Disinfection byproduct formation in reverse-osmosis concentrated and lyophilized natural organic matter from a drinking water source. *Water Research* 46 (16): 5343–5354. https://doi.org/10.1016/j.watres.2012.07.020.

Rahman, M.B., T. Driscoll, C. Cowie, and B.K. Armstrong. 2010. Disinfection by-products in drinking water and colorectal cancer: a meta-analysis. *International Journal of Epidemiology* 39 (3): 733–745. https://doi.org/10.1093/ije/dyp371.

Richardson, S.D., M.J. Plewa, E.D. Wagner, R. Schoeny, and D.M. Demarini. 2007. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research. *Mutation Research* 636 (1–3): 178–242. https://doi.org/10.1016/j.mrrev.2007.09.00.

Rivera-Nunez, Z., J.M. Wright, B.C. Blount, L.K. Silva, E. Jones, R.L. Chan, R.A. Pegram, P.C. Singer, and D.A. Savitz. 2012. Comparison of trihalomethanes in tap water and blood: a case study in the United States. *Environmental Health Perspectives* 120 (5): 661–667. https://doi.org/10.1289/ehp.1104347.

Rode, R.A., and V.M. Chinchilli. 1988. The use of Box-Cox transformations in the development of multivariate tolerance regions with applications to clinical chemistry. *The American Statistician* 42: 23–30.

Rook, J.J. 1974. Society for water treatment and examination. *Society for Water Treatment and Examination* 23: 234–243.

Savitz, D.A., K.W. Andrews, and L.M. Pastore. 1995. Drinking water and pregnancy outcome in central North Carolina: source, amount, and trihalomethane levels. *Environmental Health Perspectives* 103 (6): 592–596.

Schenck, K.M., M. Sivaganesan, and G.E. Rice. 2009. Correlations of water quality parameters with mutagenicity of chlorinated drinking water samples. *Journal of Toxicology and Environmental Health. Part A* 72 (7): 461–467. https://doi.org/10.1080/15287390802608940.

Simmons, J.E., S.D. Richardson, T.F. Speth, R.J. Miltner, G. Rice, K.M. Schenck, E.S. Hunter 3rd, and L.K. Teuschler. 2002. Development of a research strategy for integrated technology-based toxicological and chemical evaluation of complex mixtures of drinking water disinfection byproducts. *Environmental Health Perspectives* 110 (Suppl 6): 1013–1024.

Simmons, J.E., S.D. Richardson, L.K. Teuschler, R.J. Miltner, T.F. Speth, K.M. Schenck, E.S. Hunter 3rd, and G. Rice. 2008. Research issues underlying the four-lab study: integrated disinfection by-products mixtures research. *Journal of Toxicology and Environmental Health. Part A* 71 (17): 1125–1132. https://doi.org/10.1080/15287390802181906.

Simon, T., and R. Manning. 2006. Development of a reference dose for the persistent congeners of weathered toxaphene based on in vivo and in vitro effects related to tumor promotion. *Regulatory Toxicology and Pharmacology* 44 (3): 268–281. https://doi.org/10.1016/j.yrtph.2006.01.001.

Softeland, L., I. Eide, and P.A. Olsvik. 2009. Factorial design applied for multiple endpoint toxicity evaluation in Atlantic salmon (Salmo salar L.) hepatocytes. *Toxicology In Vitro* 23 (8): 1455–1464. https://doi.org/10.1016/j.tiv.2009.07.014.

Speth, T.F., R.J. Miltner, S.D. Richardson, and J.E. Simmons. 2008. Integrated disinfection by-products mixtures research: concentration by reverse osmosis membrane techniques of disinfection by-products from water disinfected by chlorination and ozonation/postchlorination. *Journal of Toxicology and Environmental Health. Part A* 71 (17): 1149–1164. https://doi.org/10.1080/15287390802182219.

Teuschler, L.K., G.E. Rice, C.R. Wilkes, J.C. Lipscomb, and F.W. Power. 2004. A feasibility study of cumulative risk assessment methods for drinking water disinfection by-product

mixtures. *Journal of Toxicology and Environmental Health. Part A* 67 (8–10): 755–777. https://doi.org/10.1080/15287390490428224.

U.S. EPA. 1991. *Integrated risk information system carcinogenicity assessment for toxaphene*. Washington, DC: U.S. Environmental Protection Agency.

———. 1992. *Guidelines for exposure assessment*. Washington, DC: Risk Assessment Forum.

———. 1996. *Integrated Risk Information System (IRIS) reference dose assessment for aroclor 1016*. Washington, DC: Risk Assessment Form.

———. 1998. *Guidelines for ecological risk assessment*. Washington, DC: Risk Assessment Forum.

———. 1999. *Toxaphene update: impact on fish advisories*. Washington, DC: Office of Water.

———. 2000. *Science policy council handbook: Risk characterization*. Washington, DC: U.S. Environmental Protection Agency, Science Policy Council.

———. 2002. *Exposure related dose estimating model (ERDEM) for assessing human exposure and dose*. National Exposure Research Laboratory, Research Triangle Park, NC U.S.

——— 2009. Integrated Summary Report. Validation of an estrogen receptor binding assay using rat uterine cytosol for source of receptor as a potential screen in the Endocrine Disruptor Screening Program, Tier 1 Battery. March 2009. http://www.epa.gov/endo/pubs/assayvalidation/er_ruc_isr.pdf.

———. 2012. *Benchmark dose technical guidance*. Risk Assessment Forum, Washington, DC: U.S. Environmental Protection Agency.

———. 2017. Integrated Risk Information System (IRIS) Glossary. Office of Research and Development/National Center for Environmental Assessment/Integrated Risk Information System. Last Update: August 31, 2011 https://www.epa.gov/iris.

van der Schalie, W.H., T.R. Shedd, M.W. Widder, and L.M. Brennan. 2004. Response characteristics of an aquatic biomonitor used for rapid toxicity detection. *Journal of Applied Toxicology* 24 (5): 387–394. https://doi.org/10.1002/jat.1028.

Villanueva, C.M., K.P. Cantor, J.O. Grimalt, N. Malats, D. Silverman, A. Tardon, R. Garcia-Closas, C. Serra, A. Carrato, G. Castano-Vinyals, R. Marcos, N. Rothman, F.X. Real, M. Dosemeci, and M. Kogevinas. 2007. Bladder cancer and exposure to water disinfection by-products through ingestion, bathing, showering, and swimming in pools. *American Journal of Epidemiology* 165 (2): 148–156. https://doi.org/10.1093/aje/kwj364.

Waller, K., S.H. Swan, G. DeLorenze, and B. Hopkins. 1998. Trihalomethanes in drinking water and spontaneous abortion. *Epidemiology* 9 (2): 134–140.

Weisel, C.P., and W.K. Jo. 1996. Ingestion, inhalation, and dermal exposures to chloroform and trichloroethene from tap water. *Environmental Health Perspectives* 104 (1): 48–51.

Wilkes, C. 1998. Case study. In *Exposure to contaminants in drinking water: estimating uptake through the skin and by inhalation*, ed. S. Olin, 183–224. Washington, DC: CRC Press.

Wold, S. 1978. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 20: 397–405.

Wold, S., H. Martens, and H. Wold. 1983. The multivariate calibration problem in chemistry solved by the PLS method. *Matrix Pencils* 973: 286–329.

Wright, J.M., J. Schwartz, and D.W. Dockery. 2004. The effect of disinfection by-products and mutagenic activity on birth weight and gestational duration. *Environmental Health Perspectives* 112 (8): 920–925.

Part V
Nonchemical Stressors

Chapter 16
Consideration of Physical Stressors in Cumulative Risk Assessment

Cynthia V. Rider, Thais Morata, MaryJane K. Selgrade, and Kenneth Sexton

Abstract Physical stressors represent an important class of factors that can affect the health of humans or ecosystems and should be considered in cumulative risk assessment. Physical stressors are defined here as biological agents (e.g., bacteria, viruses) or external forces (e.g., radiation, noise) that can modify exposure and/or elicit a physiological response from the exposed organism. Physical stressors can intersect with chemical stressors in at least three ways: (1) by directly interacting with chemicals to modify exposure (e.g., photoreactions of sunlight with air pollution), (2) by interacting with the same target system as a chemical stressor to elicit joint effects (e.g., noise and chemicals can both affect the physiological mechanism leading to hearing disorders), and (3) by interacting with the target system to alter its susceptibility or response to chemical exposure (e.g., virus-initiated disease leading to hyper-responsiveness to chemical insult). In this chapter, physical stressors will be discussed in terms of their actions on biological systems, modification of exposure or effects of chemical stressors, and suggestions for incorporation into cumulative risk assessment.

Keywords Sunlight · Heat · Pathogens · Noise · Nonchemical stressors · Biological stressor

Disclaimer The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the U.S. National Institute of Occupational Safety and Health.

C. V. Rider (✉)
National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA
e-mail: cynthia.rider@nih.gov

T. Morata
National Institute for Occupational Safety and Health, Cincinnati, OH, USA

M. K. Selgrade
ICF International (Retired), Durham, NC, USA

K. Sexton
University of North Carolina (Retired), Chapel Hill, NC, USA

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_16

16.1 Introduction

The inclusion of both chemical and nonchemical stressors in human health risk assessments is required to better reflect real-world exposures and protect vulnerable populations that face disproportionate exposure to multiple stressors (Gallagher et al. 2015; Sexton 2012). This represents a significant paradigm shift from previous human health risk assessments, which overwhelmingly addressed single chemicals and only occasionally included clearly defined classes of chemicals (e.g., organophosphates (U.S. EPA 2002a)). Although there are an increasing number of examples of cumulative risk assessment efforts that include nonchemical stressors (Fox et al. 2017), challenges remain in their successful incorporation. In contrast, ecological risk assessments often include consideration of nonchemical stressors in addition to chemical stressors, albeit with a focus on protection at a population level, not an individual level (Heugens et al. 2001; U.S. EPA 1984).

The term "nonchemical stressors" is broad and includes physical factors (biological agents or external forces), as well as psychosocial influences that can modify exposure, susceptibility, or response to chemical stressors. Whereas physical stressors are often amenable to quantitative measurement (e.g., virus titers, radiation intensity), animal and mechanistic studies, and tend to have a direct impact on chemical exposure or biological response; psychosocial stressors involve perception and often have a more nuanced role in influencing biological responses to chemical stressors. Therefore, physical stressors could offer a promising starting place for including nonchemical stressors in cumulative risk assessment. Cumulative risk assessment refers to the process of estimating the risk associated with exposure to more than one stressor. In this chapter, the focus is on incorporating physical stressors into component-based cumulative risk assessment approaches (see Chap. 14). Despite the complexity involved, psychosocial stressors should also be accounted for in cumulative risk assessment and will be covered in detail in later chapters (see Chaps. 17 and 18).

Although physical stressors have not routinely been included in cumulative risk assessments, there are some important historical examples of the consideration of physical and chemical stressor interactions and their impact on human health. One such example can be found in combined radiation and chemotherapy (Rubin 1977). In a 1977 paper reviewing the topic, Rubin suggests that quantitative dose-response relationships in animal models are needed for radiation in order to protect against unnecessary long-term harm from combined radiation/chemotherapeutic treatment (Rubin 1977). A second classic example illustrating the need for considering physical stressors in cumulative risk is that of radon and smoking and their combined effects on lung cancer (Reif and Heeren 1999).

There are many, often complicated, ways for physical stressors to modify the toxicity of chemical stressors (Fig. 16.1), and vice versa. As illustrated in Fig. 16.1, physical stressors can interact with chemical stressors at the level of exposure, target, or biological response. Indeed, a single physical stressor can interact at multiple levels. For example, sunlight can modify a chemical prior to inhalation

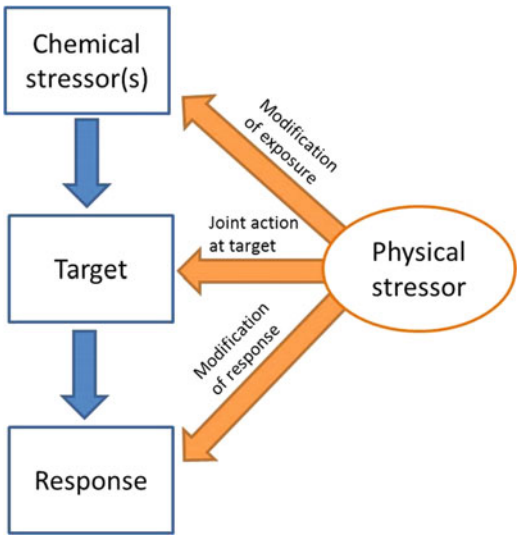

Fig. 16.1 Graphic depiction of the different mechanisms by which physical stressors can impact chemical stressor exposures and responses

exposure (see discussion below), contribute to phototoxicity of an orally administered chemical at a dermal target site, and elicit damage to skin, which could, in turn, alter absorption of or response to concurrent dermal chemical exposures. It is important to emphasize that chemicals can also modify exposure and/or response to physical stressors.

This chapter will discuss a range of physical stressors in terms of quantifying their toxicity, identifying interactions with chemical stressors, and incorporating them into cumulative risk assessment. Some examples of physical stressors are provided in Table 16.1. They fall into two major categories – biological agents and external forces. In general, the biological agents exhibit indirect effects by influencing the health status of individuals and thereby altering their response to chemical stressors (e.g., bacteria and viruses interacting with chemical stressors). The second major category is forces, which can interact directly and indirectly at both the exposure and response levels. Therefore, several different forces (e.g., sunlight, heat, noise) will be discussed to demonstrate potential issues to consider in accounting for the impact of these stressors on cumulative risk. Although there is a history of including some physical stressors in ecological risk assessment, the focus of this chapter will be on human health.

Table 16.1 Examples of potential physical stressors to human health

Biological agents/factors	Forces
Viruses	Electromagnetic radiation
	Radio waves
	Visible light
	Ultraviolet light (ionizing)
	X-rays (ionizing)
Pathogenic bacteria	Particle radiation
	α-Radiation
	β-Radiation
	γ-Radiation (ionizing)
	Neutron radiation
Allergens – pollen, dander, dust mites	Acoustic radiation
	Ultrasound
	Sound
Physical health status	Thermal radiation (heat)
Nutritional status	
Disease status	
Microbiome	
Lack of sleep	
Hypothermia	

16.2 Biological Agents: Microbial Disease and Chemical Stressors

As mentioned in the introduction, biological agents or factors interact indirectly with chemicals by changing the health status of the target subject and potentially increasing their susceptibility to chemical injury. Alternatively, chemical exposures can affect health status leading to a modified response following exposure to the biological agent. The diseases induced by different biological stressors (e.g., lack of sleep versus allergens) are varied. However, many biological stressors ultimately involve immune system modulation (Jain and Walker 2015). Microbial diseases are perhaps the most well-studied biological agents within the physical stressor spectrum and will be used to illustrate the process of exploring potential interactions with chemical stressors.

There are at least four types of mechanisms that underlie potential interactions between chemical stressors and infectious disease. The best understood is suppression of immune responses, resulting in increased incidence and/or severity of infectious disease. Alternatively, certain immune/inflammatory mediators that are activated during infection affect metabolic enzymes and transporters, resulting in increased chemical toxicity. Chemical exposure may enhance inflammation and immune pathology associated with an infection. Conversely, infection may enhance chemical-induced lesions (e.g., p53 mutations, inflammation, cell proliferation).

These mechanisms are not necessarily comprehensive, distinct, or mutually exclusive. However, they provide a starting place for evaluating interactions between infectious disease and chemical stressors. Each mechanism will be discussed in more detail below with the goal of illustrating options for incorporating biological stressors into cumulative risk assessments.

Substantial data from animal models, and more limited data in humans, suggest that a number of chemicals suppress a variety of immune responses (Selgrade 2010), which can lead to increased risk of bacterial infection. For example, select air pollutants (e.g., chloroform, toluene, ozone) have been demonstrated to decrease the function of alveolar macrophages, which are the first line of defense against some bacterial species (e.g., *Streptococcus zooepidemicus*) (Selgrade and Gilmour 2006). Research on this model provides both qualitative and quantitative approaches to describe the associated risk. Data comparing human and murine alveolar macrophages exposed to similar doses of ozone in vitro and in vivo indicate that cells from both species respond almost identically as measured by macrophage phagocytic capability. Thus, these data suggest that (1) the effects of ozone exposure on murine alveolar macrophage function are predictive of effects on human alveolar macrophage function and (2) effects of in vitro exposure of macrophages to ozone are predictive of effects that result from in vivo exposure (Selgrade et al. 1995). Application of inhalation dosimetry methods eliminates the need for the uncertainty factor that is typically applied to account for animal to human extrapolation in the absence of toxicokinetic data (U.S. EPA 2012). Furthermore, the lack of difference in sensitivity observed in studies with mice and humans eliminates the need to apply a factor for toxicodynamics. In addition to air pollution and alveolar macrophage function that demonstrate at least a qualitative relationship between immune function and disease, developmental exposures to arsenic, polychlorinated biphenyls, and cigarette smoke also have been linked to immune suppression. In the cases of arsenic (Soto-Pena et al. 2006) and PCBs (Heilmann et al. 2010), a quantitative relationship exists between exposure to the chemical and suppression of the immune system in humans. However, predicting the impact of immune suppression on the incidence or severity of infection in a population (i.e., risk) is difficult. Immunocompetence (i.e., the ability to mount a normal immune response) in a population may be represented as a bell-shaped curve including individuals with little or no immune reserve (response capacity) available (e.g., the very young and old and those who are immunocompromised by disease or medications) and very robust individuals. The proportion of the population at risk of infection depends on the level of immune competence, as well as the dose and virulence of infectious agents, with increasing risk of infection corresponding to increasing dose or virulence. In populations exposed to immunosuppressive agents, the distribution curve would be expected to shift, putting a larger portion of population at risk for disease development.

Although the effects of toxicants on host defenses against infection have received the most attention, it is also possible for infections to affect host defenses against toxicants, by interfering with metabolic enzymes and transporters. There are many examples of increased chemical/drug toxicity with infections or other

inflammation-related diseases (Morgan et al. 2008). For example, an influenza epidemic resulted in decreased clearance of theophylline, an asthma medication with a narrow therapeutic window, resulting in toxicity in children (Kraemer et al. 1982). Interestingly, many other factors can influence theophylline dose, including smoking, drugs (e.g., phenobarbital, erythromycin), and disease (e.g., heart failure, liver disease) (Kraemer et al. 1982). Other examples involve murine cytomegalovirus infection increasing the toxicity of parathion (Selgrade et al. 1984) and sodium pentobarbital-induced sleeping time and decreasing cytochrome P450 (CYP) levels in liver microsomes in mice (Catignani et al. 1989). In multiple tissue types, infection and inflammatory diseases have demonstrated downregulation of ATP binding-cassette (ABC) drug transporters involved in cellular efflux of xenobiotics (Petrovic et al. 2007). The proposed pathways underlying these effects begin when infections and other inflammatory stimuli cause the release of inflammatory cytokines from monocytes, macrophages, and stromal cells (the acute phase response), resulting in the modulation of transcription factor activities in the liver. These changes ultimately lead to a downregulation of CYP and ABC transporter genes. The production of cytokines also activates nitric oxide synthase 2 to form nitric oxide that inhibits CYP enzyme activities directly and/or leads to the downregulation of CYP proteins via destabilization (Morgan et al. 2008). Risk assessment procedures account for these enzyme-related changes by applying a tenfold intraspecies uncertainty factor (U.S. EPA 2002b). This uncertainty factor is meant to capture differences in individual susceptibility within the population. However, use of a default uncertainty factor does not accurately reflect what is known about the effects of infection/inflammation on chemical toxicity. As risk assessments begin to use adverse outcome pathways to characterize the risk associated with multicomponent mixtures (see Chap. 7), infection and inflammation may be incorporated into that process.

The third type of interaction between chemicals and infection involves chemical-mediated exacerbation of inflammation and pathology resulting from infection. Examples include effects on influenza infection by ozone (Selgrade et al. 1988), ultraviolet radiation (Ryan et al. 2000, 2002), TCDD (Burleson et al. 1996; Lawrence and Vorderstrasse 2004; Warren et al. 2000), and acrolein (Ong et al. 2012). In all cases, mortality is enhanced by exposure to the toxicant without increased viral load. Although all of these chemicals have immunosuppressive potential, reduced viral clearance resulting from immune suppression does not appear to be responsible for the increase in mortality. Instead, morbidity and mortality occur very early in infection before involvement of adaptive immunity, and surviving mice develop protective immunity that prevents subsequent reinfection (Lawrence and Vorderstrasse 2004; Ryan et al. 2000). Increased inflammatory responses appear to be responsible for observed mortality (Head and Lawrence 2009). Both pathogens and tissue damage trigger similar receptors and signaling pathways that lead to innate inflammatory responses (Kono and Rock 2008; Tolle and Standiford 2013). A systems biology approach that integrates these triggers at the pathway level is needed to account for the joint effects of these immune system modulators.

The fourth, and final, type of interaction involves enhancement of chemical induced lesions (e.g., p53 mutations, inflammation, and cell proliferation) by infection. This interaction might explain the joint effects of hepatitis B virus infection and aflatoxin on liver cancer (Kensler et al. 2010). In nested, case-control data within a cohort study of 18,000 men in Shanghai, Qian et al. (1994) demonstrated a statistically significant increase in the relative risk (95% confidence limits) of 3.4 (1.1, 10) for hepatocellular carcinoma cases with detectable urinary biomarkers for aflatoxin, 7.3 (2.2, 24.4) for individuals without evidence of aflatoxin exposure but seropositive for hepatitis B antigen, and 59.4 (16.6, 212.0) for individuals exhibiting both urinary aflatoxin markers and positive hepatitis B status. The results strongly suggest an interaction between aflatoxin and hepatitis B in the development of hepatocellular carcinoma. Hepatitis B infection and the resulting chronic inflammation may promote DNA lesions leading to P53 mutations and may promote cell proliferation, contributing to chronic hepatitis and/or cirrhosis and ultimately carcinoma. It is plausible that similar interactions may exist between other infections and toxicants that target the liver. Again, as we begin to use systems biology to work through cumulative risk, an understanding of the pathways underlying this interaction could be applied. However, in this instance, decisions to decrease risk by promoting public health interventions such as limiting exposure to aflatoxin and vaccinating against hepatitis B are recommended.

In summary, the chemical/infection interactions described here involve joint action of chemicals and immune/inflammatory responses (the biological forces). Superimposed on all of this are genetic differences which affect both susceptibility to infection and toxicity. Existing information regarding molecular pathways involved in immune activation and inflammation should be applied using a systems approach to understand the cumulative risk that results from exposure to chemicals and infectious agents.

16.3 Forces: Modification of Exposure by Sunlight

As discussed in the previous section exploring microbial disease, there are multiple pathways by which a physical force can interact with chemicals to affect health. In the case of sunlight, there is a clear primary target – the skin – which is subject to direct damage, leading to aging and cancer of the skin or photodermatosis (immune reaction to sunlight). It follows that chemicals that elicit skin toxicity could interact with sunlight to increase skin damage. While acknowledging that there are many opportunities for sunlight to interact with chemicals at a common target site, the focus of this section is not on interaction of sunlight and chemicals at the adverse outcome level but at the exposure level.

Sunlight, along with other climatic characteristics (e.g., temperature, humidity), has the potential to modify both the concentration and form of chemicals present in the air. The criteria air pollutants (ozone, particulate matter, carbon monoxide, nitrogen oxides, sulfur dioxide, and lead) are of particular interest, as these are the

six common air pollutants for which EPA is mandated by the Clean Air Act to set National Ambient Air Quality Standards. Sunlight together with temperature can trigger photochemical reactions of air pollutants. Examples of these reactions include production of ozone from hydrocarbons and nitrogen oxides, nitrogen dioxide from oxidation of nitrogen oxide, carbon monoxide from oxidation of hydrocarbons, and many well-known toxic compounds such as formaldehyde, acetaldehyde, acrolein, and other carbonyl and nitrate-containing products from oxidation of hydrocarbons and nitrogen oxides, as well as reduction of primary emitted pollutant concentrations (Finlayson-Pitts and Pitts 1999). In addition, these reactions contribute to air pollution in the form of secondary organic aerosols (SOA), which are present in fine particulate matter (PM).

Atmospheric transformation processes affect the relative composition and resulting cumulative health effects of contaminants, including criteria pollutants. It is important to understand how atmospheric transformations affect air pollution mixtures and PM composition and resulting toxicological risk. Evidence of the importance of these issues can be found in the ongoing efforts at the U.S. Environmental Protection Agency to develop a framework for addressing multipollutant risk assessment (Johns et al. 2012).

In terms of research into atmospheric pollutant mixtures, smog chambers have been used to prepare consistent, controlled mixtures of various primary pollutants to study how different conditions (e.g., natural or simulated sunlight, different temperatures, or humidity levels) change mixture composition. Smog chambers facilitate the study of sunlight and temperature effects in the absence of interference from changing weather patterns and unexpected emissions from nontarget sources. They can interface with in vitro or in vivo models to provide direct exposure in toxicity studies (e.g., direct air-to-tissue or air-liquid-interface inhalation exposures) (Lichtveld et al. 2012). This method avoids pre-collection with filters and liquids, thereby offering a significant advantage over methods that require sample preparation, which can alter component concentration ratios and toxicological responses (Lichtveld et al. 2012). There are many examples of photochemical experiments using smog chambers, such as industrial compounds and nitrogen oxide mixtures and complex mixtures of motor vehicle exhaust in urban atmospheres. These experiments often demonstrate enhanced toxicity following photochemical reactions, as measured by markers of inflammation and other biological endpoints (e.g., cytotoxicity) (Doyle et al. 2007). Confirmatory experiments can be conducted with observed secondary products to link particular species with effect. More recently, Gas In Vitro Exposure Systems (GIVES) have been used in the field to expose cells directly to ambient air (Vizuete et al. 2015). Following field exposure, cells can be evaluated for cytotoxicity and gene expression changes (Vizuete et al. 2015).

Modification of various parameters (e.g., pollutant sources, mixture composition, component concentrations, or atmospheric conditions) can also aid in interpretation of the mechanism or mode of action of air pollution determined through toxicity testing. For example, a primary pollutant mixture representing the average volatile organic compounds (VOCs) observed in 40 U.S. cities can significantly

change after 1 day of "aging," which could affect ozone concentrations, presence of co-pollutants, and toxicity. The degree of aging can be influenced by meteorological conditions (e.g., cloud cover, sunlight intensity) (Sexton et al. 2004). Based on this type of study, models can be developed to predict ozone concentration (Sexton et al. 1988). Information gained from these approaches can be used to identify toxic secondary products, which can be included in air quality simulation models for multipollutant risk assessments.

Cytotoxicity and inflammation markers are endpoints that have been used to explore the effects of sunlight and temperature on atmospheric transformations of air pollutants (Lichtveld et al. 2012; Sexton et al. 2004). However, numerous biological models and endpoints could be used in the same context. For example, novel genomic analyses of cell-based exposure to an urban mixture demonstrated transcriptional changes in a subset of genes, with increased expression alterations resulting from mixture irradiation (19–709 following a 1-day sunlight irradiation) (Rager et al. 2011). This type of study offers promise for elucidating the effects of atmospheric conditions on complex mixtures and health. Additionally, biomarkers identified in in vitro studies could be explored for their utility in an epidemiological context.

Sunlight and temperature can influence the toxicity of air pollutants. Fortunately, both of these physical stressors can be easily quantified (Jeffries et al. 1989) and incorporated into air quality simulation models to estimate air pollution exposure concentrations and distribution in a target area (Vizuete et al. 2010). Results from these simulation models can be used to estimate total exposure within a population and integrated with health information (e.g., excess deaths) (Li et al. 2010). In order to better characterize risk associated with ambient exposures, studies that incorporate sunlight and temperature should be considered. These studies capture potential transformations of primary pollutants and resulting changes in their toxicity. Without consideration of these processes, there is the potential to underestimate risk from exposure to air pollution mixtures.

16.4 Forces: Heat and Chemicals

Increasing global temperatures associated with climate change have focused attention on potential health effects associated with heat (Spector and Sheffield 2014; Patz et al. 2014; Kovats and Hajat 2008). Thermal stress encompasses temperatures that fall above (heat stress) or below (cold stress) the normal range and require a physiological response in order to maintain homeostasis of the internal body temperature (Wilson et al. 2014). While both heat stress and cold stress can impact health, heat stress will be used as an illustrative example. Heat is a complex actor – it can have a direct effect on health (e.g., heat strain, heat stroke) and exacerbate existing disease conditions both alone and in concert with chemicals, and it can interact with chemicals by modifying their absorption or effect.

A complicating factor in the study of heat stress is the differences in thermoregulation strategies between small rodents typically used in toxicity studies and humans (Gordon et al. 2014). For example, mice and rats rely on a metabolic strategy to balance heat loss and production, while humans depend more on regulation of peripheral blood flow (Gordon et al. 2014). Furthermore, laboratory animal studies are typically conducted at temperatures that are below ideal ambient temperatures for rodents, causing a mild hypothermic response that may distort their response to chemical exposure (Gordon et al. 2014). Fortunately, many epidemiology studies have addressed the role of heat in disease and will be the focus of this section.

A wide array of diseases are associated with heat stress including mental health disorders (Berry et al. 2010), reproductive system dysfunction (Strand et al. 2012), kidney disease (Tawatsupa et al. 2012; Garcia-Trabanino et al. 2015), cardiovascular disease (Braga et al. 2002; Schwartz et al. 2004), and respiratory disease (Braga et al. 2002; Michelozzi et al. 2009). It follows that heat stress has the potential to act together with chemicals and other nonchemical stressors to disrupt normal function or exacerbate disease. Incorporating heat stress into cumulative risk assessment could be motivated by co-occurrence, as in the case of agricultural workers that are occupationally exposed to both heat stress and pesticides (see case study below). Alternatively, for an assessment aimed at evaluating the cumulative risk of exposures contributing to an observed disease (effects-based risk assessment or disease-based risk assessment), heat stress could be included when it is identified as a risk factor. If heat stress is identified as a risk factor that should be included in a cumulative risk assessment, quantifying the contribution of heat stress is the next goal.

The two internationally recognized methods for rating the level of heat stress are the Wet Bulb Globe Temperature (WBGT) index (ISO 7243) and the Predicted Heat Strain (PHS) model (Alfano et al. 2014). The method for assessing the WBGT index is currently undergoing revision, but generally a combination of air temperature, black globe temperature, and natural wet bulb temperature is used to approximate heat stress. For a detailed review of the history and limitations of this method, see Budd (2008). As recommended in Budd (2008), plotting the WBGT against an adverse effect to generate a "dose"-response relationship would allow for an assessment of the heat stress dose response. Examples of heat stress dose response include heat exhaustion (Yaglou and Minard 1957) (Fig. 16.2) and heat stroke (Schickele 1947). Unfortunately, this type of dose-response data will not be available for every endpoint of interest. However, data in the literature for a general heat-related adverse outcome (e.g., heat exhaustion) could be used to identify situations where heat stress may play a significant role in shaping cumulative risk. Furthermore, it is well known that certain populations (e.g., pre-existing conditions such as asthma, elderly) are more acutely affected by heat stress. Therefore, in cumulative risk assessments where heat stress has been identified as an important risk factor, an additional uncertainty value could be applied to account for increased risk to vulnerable populations.

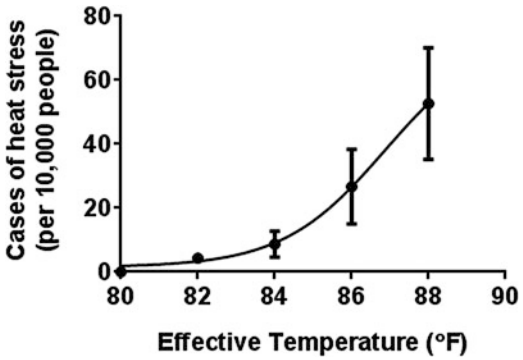

Fig. 16.2 Example of a dose-response relationship for heat stress. Heat stress data from military personnel was used to generate the dose-response relationship. Circles and bars represent the average and standard error values from three populations: Junior Platoon Leader (PLC) candidates on 6-week training, new reservists on 2-week training, and recruit trainees on 12-week training. The raw data used in the calculations described above were extracted roughly from Yaglou and Minard (1957). The solid line represents a four-parameter logistic fit to the data using GraphPad Prism

16.4.1 Case Study for Estimating Cumulative Risk from Occupational Exposure to Heat Stress and Pesticides

This case study is meant to illustrate potential considerations and decision points for conducting a cumulative risk assessment that includes both chemical and physical stressors and does not represent an accurate analysis of risk from the model stressors. Consider an agricultural community in North Carolina. The community is concerned about the combination of heat stress and pesticide exposure and would like to better understand relative contributions to overall risk in order to decide how to focus advocacy efforts. For example, they could advocate for more personal protective clothing or greater pesticide use oversight, if pesticides are driving risk, or they could advocate for implementation of cooling measures (longer breaks, greater availability of shaded areas and personal cooling devices, etc.), if heat is driving risk.

Step 1: Scoping

The first step of assessing risk from stressors present in a community is to scope the problem. This is necessarily an iterative process, as it involves identifying potential stressors and health endpoints that are of concern and exploring availability of data on exposures and outcomes. Although it would be ideal to include all relevant stressors for a particular health outcome, exposure data may limit the number and type of stressors that can be included. Additionally, the questions can be interrelated – a decision about which stressors to include could influence which risk assessment method is selected. Table 16.2 provides examples of the types of questions and potential answers that could be considered in scoping efforts related to heat and

Table 16.2 Scoping questions and answers for a risk assessment of heat stress and pesticide exposure in an agricultural community

Questions	Examples of potential answers
Is there a particular health outcome of concern?	Yes (e.g., cardiovascular disease, asthma)
	No (e.g., hospital visits, morbidity – any cause)
Who is in the population of interest?	Agricultural workers only
	Agricultural workers and their families
	Rural community in proximity to agricultural activities
What pesticides should be included?	All pesticides to which the population is exposed (exposure-based decision)
	Pesticides that are of concern based on toxicity information (disease-based decision)
	Pesticides with a particular mechanism of action (chemical class-based decision)
What other stressors should be included?	Heat only
	Heat and psychosocial stressors (e.g., socioeconomic status, exposure to violence)
	All stressors potentially linked to disease of interest
What risk assessment method is appropriate?	Dose-addition model (e.g., Hazard Index approach)
	Independent-action model

pesticide exposure. For this case study, stressors will be limited to heat and select pesticides (diazinon, parathion, chlorpyrifos, and permethrin) known to be applied on crops in the area. Three of the pesticides selected (diazinon, parathion, and chlorpyrifos) have the same mechanism of action – acetylcholinesterase inhibition – while permethrin has a different mechanism of action. All four pesticides target the nervous system and represent common agricultural exposures, while heat stress targets cardiovascular function. The set of stressors were selected based on their co-occurrence, not a common mechanism of action or target.

Step 2. Evaluating Exposure

Evaluating chemical exposures is fairly straightforward and is the same as in a single chemical risk assessment. In this example, biomonitoring data from the published literature will be used to characterize exposure to pesticides using the following equation:

$$I_C = u_C \times \frac{e_{\text{creatinine}}}{u_{\text{creatinine}} \times \text{bw}} \times \frac{1\text{mg}}{1000\mu g}, \tag{16.1}$$

where I_C is the intake of chemical C with units of μg-g/kg body weight/day, u_C is the measured concentration of chemical C (μg-g/L urine), $u_{\text{creatinine}}$ is the measured concentration of creatinine in the urine (g creatinine/L urine), $e_{\text{creatinine}}$ is the daily creatinine excretion (g/day), and bw is body weight (kg). (See Table 16.3 for values of the four pesticides measured in urine.) For the purposes of this example, a daily creatinine excretion rate of 1.5 g/day, a concentration of creatinine in the urine of 1 g/L, and a body weight of 70 kg were used.

Table 16.3 Biomonitoring data on pesticide exposure of farmworkers in North Carolina

Pesticide	Mean ± SD (ng/ml)[a]	Max (ng/ml)[a]	Estimated mean intake (μg/kg/d) [b]	Estimated max intake (μg/kg/d) [b]	Reference dose (μg/kg/d)
Diazinon	2.76 ± 6.72	7.16	0.06	0.15	0.09[c]
Parathion	7.67 ± 42.03	457.00	0.16	9.79	6[d]
Chlorpyrifos	5.37 ± 3.70	20.70	0.12	0.44	10[e]
Permethrin	3.34 ± 4.49	30.70	0.07	0.66	50[f]

[a]Urinary data from (Raymer et al. 2014)
[b]Mean and max intake values calculated using Eq. 16.1 in text
[c]Reference dose for diazinon from Teuschler et al. (1999)
[d]Provisional reference dose calculated by EPA (https://www.epa.gov/sites/production/files/2016-09/documents/parathion.pdf)
[e]Reference dose for chlorpyrifos from Zhao et al. (2006)
[f]Reference dose for permethrin from IRIS (2017)

Table 16.4 Heat stress data for summer months in North Carolina

Week (2015)	Heat stress average over week (WGBT[a])	Heat stress max temp (WGBT[a])
June 28–July 4	86 °F	92 °F
July 5–July 11	91 °F	94 °F
July 12–July 18	88 °F	93 °F
July 19–July 25	89 °F	93 °F
July 26–Aug 1	91 °F	94 °F
Aug 2–Aug 8	90 °F	97 °F
Aug 9–Aug 15	88 °F	91 °F
Aug 16–Aug 22	89 °F	95 °F
Aug 23–Aug 29	87 °F	92 °F

[a]Temperature was used as an estimate for WGBT. Source: weather history for Raleigh-Durham International Airport (weatherunderground.com)

Physical stressors present greater challenges. As discussed throughout this chapter, there are not widely accepted measurement tools and methods for characterizing exposure to physical stressors. In this example, we will use the WBGT index as a measure of effective temperature. Next, consideration of how to capture the data is required. For example, temperatures could be presented as weekly or monthly averages. Alternatively, lowest and highest temperatures could be used to present a range of possible risk values. Finally, temperature variability has recently been shown to play a role in increased risk from heat and cold (Shi et al. 2015). (See Table 16.4 for sample data on heat stress exposure in North Carolina.) The exposure options used in the risk characterization step should be selected based on the goals of the risk assessment, i.e., capturing a worst-case exposure or the most common (e.g., average) exposure level.

Step 3: Dose-Response Analysis

Dose-response analyses for the pesticides used in this example have been conducted elsewhere (see example in Teuschler et al. 1999). These analyses are used to estimate reference values, which represent a daily exposure to a human population that is not likely to be associated with appreciable risk of harmful effects over a lifetime (Table 16.3; see Chap. 14 for more on Risk Assessment). The oral reference dose for diazinon is based on cholinesterase inhibition in the plasma of rats fed diazinon (Teuschler et al. 1999), which represents the mechanism of action involved in nervous system effects of organophosphate pesticides. The two other organophosphate pesticides, parathion (https://www.epa.gov/sites/production/files/2016-09/documents/parathion.pdf) and chlorpyrifos (Zhao et al. 2006), reference dose estimates were both based on erythrocyte cholinesterase inhibition in humans. Finally, the permethrin reference dose was based on increased liver weight in rats (IRIS 2016). A dose-response relationship for heat stress is presented in Fig. 16.2. The determination of a reference dose for a chemical differs in many ways from determining a "safe temperature" for use in a cumulative risk assessment that includes heat as a stressor. The application of uncertainty, for example, would differ. In this example, the OSHA guidelines for permissible heat exposure threshold limit values (TLVs) are used as a reference point. The OSHA TLVs for heat stress also include different work levels – continuous, 75% work to 25% rest, 50% work to 50% rest, and 25% work to 75% rest per hour. In addition, different TLVs are provided depending on the work load (light, moderate, and heavy). Assuming a 75% work to 25% rest level per hour and a heavy work load for the agricultural setting, the TLV is 78 °F (OSHA 2016). This TLV agrees with the dose-response data in Fig. 16.2 showing a lack of heat stress at temperatures of 80 °F and below.

Step 4: Risk Characterization

A Hazard Index approach can be used to combine the individual hazard quotients for the different stressors (see chapter on risk assessment for a detailed discussion of the Hazard Index). As mentioned in the exposure section, this step also involves many decision points as to which exposure data should be used. For comparison, a Hazard Index can be calculated for the lowest weekly average temperature and highest recorded temperature over the period. The Hazard Index is calculated by summing the hazard quotients for the individual stressors. The hazard quotient for each stressor is calculated by dividing the exposure by the acceptable limit (i.e., reference dose for pesticides and TLV for heat stress). A Hazard Index less than 1 indicates no expectation of adverse health effects. As the Hazard Index increases above one, there is increasing concern for adverse health effects. In this case, the Hazard Index calculation would be:

$$HI = HQ_{Heat} + HQ_{Diazinon} + HQ_{Parathion} + HQ_{Chlorpyrifos} + HQ_{Permethrin}$$

For the lowest temperature and mean pesticide exposure case, the Hazard Index would be:

$$\mathrm{HI} = \frac{86}{78} + \frac{0.06}{0.09} + \frac{0.16}{6} + \frac{0.12}{10} + \frac{0.07}{50}$$
$$\mathrm{HI} = 1.10 + 0.67 + 0.03 + 0.012 + 0.001$$
$$\mathrm{HI} = 1.81$$

In this low exposure case, it is apparent that the pesticides are contributing less to the overall Hazard Index score, and heat stress is driving the value above 1.

Alternatively, the Hazard Index for the worst-case scenario would be:

$$\mathrm{HI} = \frac{97}{78} + \frac{0.15}{0.09} + \frac{9.79}{6} + \frac{0.44}{10} + \frac{0.66}{50}$$
$$\mathrm{HI} = 1.24 + 1.67 + 1.63 + 0.04 + 0.01$$
$$\mathrm{HI} = 4.60$$

In the high-exposure case, heat stress continues to contribute to the >1 Hazard Index, but the values of diazinon and parathion represent larger contributions.

Step 5: Risk Management
The case study provided above illustrates how incorporating heat stress into an evaluation of risk can help in understanding relative contributions of different stressors. These exercises can help communities and policymakers in deciding between different risk mitigation strategies. It is interesting to note that whereas pesticide exposure can vary widely from average exposure to maximum, temperature has a less dynamic range. As more physical (and other nonchemical stressors) are incorporated into cumulative risk assessments, a better understanding of their impact will be gained.

16.5 Forces: Noise and Chemicals

Noise presents another multifaceted physical stressor that can enhance chemical toxicity indirectly via activation of stress hormones and directly by acting jointly at a common target. Diseases associated with co-occurring exposure to noise and chemicals include hearing loss, as well as respiratory and cardiovascular diseases. The indirect pathway is often referred to as "noise annoyance" because it involves perception of noise as a stress (Babisch et al. 2013). Both acute and chronic studies with noise exposure have demonstrated stress hormone responses. For example, a study with 3950 middle-aged men exploring factors contributing to incidence of ischemic heart disease found that participants who were highly annoyed by noise had a higher chance of developing heart disease (odds ratios = 1.7–3.0) (Babisch et al. 2003). Less information is available on the interaction of noise with respiratory tract disease, which represents an emerging area of study (Recio et al. 2016). Noise pollution and air pollution tend to co-occur in urban environments. Frequently heard in urban settings, noises associated with danger (e.g., sirens) have the potential to trigger a stress response even during sleep. Commonly, studies

exploring road traffic pollution have focused on either air or noise pollution, but not the combination. However, many diseases (e.g., asthma, bronchitis, skin disease) display increased incidence in areas with high noise pollution (Ising et al. 2003, 2004).

Occupational settings (e.g., construction, machine operation) can have particularly elevated and prolonged noise exposures. There is also the potential for relatively high chemical exposures in occupational settings (e.g., chemical manufacturing, oil and gas industry, agriculture). Noise exposure can cause both auditory and nonauditory effects alone and in combination with other factors. There are common features in auditory dysfunction caused by noise and some ototoxic chemicals (Fechter 1995; Johnson and Morata 2010). Degeneration of the sensory hairs in the cochlea is one of the most common findings in sensorineural hearing loss. Animal studies have demonstrated loss of hair cells from exposure to both noise and solvents with reactive oxygen species hypothesized to play a role in the hair cell damage (Henderson et al. 2006; Chen et al. 2007). Other chemicals such as metals (e.g., lead, mercury) may affect both the cochlea (Rice 1997) and the central auditory pathways (Discalzi et al. 1993; Lasky et al. 1995; Otto and Fox 1993) depending on the substance. Le Prell et al. (2007) reported that the formation of free radicals after noise trauma continued up to 10 days after cessation of the exposure, which could explain why the loss of hair cells worsens after exposure. Toxic insults on the cochlea have also been shown to continue after cessation of exposure to solvents (Johnson and Canlon 1994).

More recently, it has been reported that some aromatic solvents reduce the protective role played by the middle-ear acoustic reflex (Venet et al. 2011). A dysfunction of this reflex would increase risks to hearing by allowing higher acoustic energy levels to penetrate the inner ear (Maguin et al. 2009; Campo et al. 2007). This would make co-exposure more dangerous than exposure to noise or to styrene alone. Other chemicals such as metals (e.g., lead, mercury) and pesticides may affect the hearing function (Choi et al. 2012; Shargorodsky et al. 2011) by acting on both the cochlea (Rice 1997) and the central auditory pathways (Discalzi et al. 1993; Lasky et al. 1995; Otto and Fox 1993) depending on the substance.

Solvent exposures have the potential to affect hearing in the absence of exposure to occupational noise, or they can enhance the effects of noise on hearing loss. Carbon monoxide (CO) exposure has also been shown to potentiate noise-induced hearing loss (Rao and Fechter 2000). Fechter et al. (2000) characterized the joint effects of CO and noise on hearing loss using a benchmark-dose approach for risk assessment (U.S. EPA BMDS version 1.3). They found that an exposure of 194 ppm CO represented the lower bound of the benchmark dose that would yield a 10% increase in noise-induced hearing loss. Notably, these levels of CO are less than one order of magnitude higher than the permissible exposure level (PEL) of 50 ppm set by the Occupational Safety and Health Administration (OSHA). It is also important to note that periods of recovery following exposure did not abrogate the effects (i.e., changes from co-exposure to noise and CO were permanent) (Chen and Fechter 1999; Rao and Fechter 2000). Furthermore, the dose response was not monotonic

for noise, with the smaller noise exposure resulting in maximal hearing loss in combination with CO. In more recent work, co-exposure to styrene and different types of noise (6-h continuous noise of 85 dB Sound Pressure Level (SPL) or impulse noise of 80 dB) was evaluated, and impulse noise was found to elicit greater damage (Venet et al. 2015; Campo et al. 2014). The characteristics of the noise exposure and solvents like toluene and styrene can disrupt the natural protective mechanisms of the ear such as the middle-ear acoustic reflex. This was a demonstration of a second mechanism of solvent-induced damage, beyond the cochleotoxic mechanism described earlier. It consists of a rapid pharmacological impact on the central nervous system by the inhibition of the protective reflex (Campo et al. 2001). Finally, studies have also demonstrated that as the number of stress factors increase, the lowest observable adverse effect level (LOAEL) for hearing loss decreases.

The evidence presented above represents a small portion of the literature describing the link between occupational chemical exposure and hearing loss. To address this concern, OSHA and other groups have published comprehensive evaluations of ototoxic substances, as well as documented hazards associated with workplace exposure to noise and ototoxic chemical substances (EU-OSHA 2009; Johnson and Morata 2010). These references include qualitative information on noise and chemical interactions and highlight policies from specific countries and multinational agencies.

In an example of incorporating physical stressors into cumulative risk assessment in a quantitative manner, Evans et al. (2014) developed a case study characterizing risk for hearing impairment from combined exposures to noise and volatile organic compounds (VOCs). They used data from the 1999–2000 U.S. National Health and Nutrition Examination Survey (NHANES) to estimate VOC exposures and modeled street-level noise data (i.e., a noise map) to estimate block group-level noise categories (45–60 dB, 61–65 dB, 66–70 dB, and 71–75 dB). The cumulative risk for potential hearing loss from co-exposure to noise and VOCs was calculated using a Hazard Index approach (see the heat and pesticide case study above and Chap. 14 for a discussion of HI). Hazard Indices ranged from 0.8 (lowest noise category and 10th percentile for total VOCs) to 1.7 (highest noise category and 90th percentile for total VOCs). Although the authors noted limitations of the approach (e.g., issues with combining heterogeneous data), it did demonstrate the feasibility of combining chemical and nonchemical stressors using an established cumulative risk assessment approach. Furthermore, it identified noise as the driver of risk in the case study, which could help inform decision-makers in how to invest limited resources to provide the greatest impact to public health.

Currently, the French Institut National de Recherche et de Sécurité (INRS) is working to incorporate information on noise damage risk-criteria into the web-tool *Mixie*. The original web-tool was created by the University of Montreal and the Institut de Recherche Robert-Sauvé en Santé et en Sécurité du Travail (Vyskocil et al., 2007) to assess the risks associated with exposure to a mixture of airborne chemical substances in the workplace (http://www.irsst.qc.ca/en/publications-tools/tool/i/100037/n/mixie-mixtures-of-substances-in-the-workplace-computer-

based-tool-for-evaluating-the-chemical-risk-calculation-of-the-rm). Toxicological effects are considered additive and the multiple exposure index is used for assessing the risk encountered by people exposed to several substances present in the workplace. The sum of the fractions of measured individual exposure concentrations and their Time-Weighted Average Exposure Value (TWAEV) for each substance results in a percentage of the recommended dose of the mixture. A percentage of 100 indicated that exposures are at their recommended exposure limit according to Canadian Occupational Exposure Limits (OELs). The planned additions would incorporate information on the interaction of chemicals with noise, to alert those conducting risk assessment.

16.6 Challenges and Recommendations for Incorporating Physical Stressors into Cumulative Risk Assessment

Although it is an accepted fact that physical stressors can impact human health, they have not typically been included in cumulative risk assessment efforts, which have focused almost exclusively on chemicals. There are multiple factors that present obstacles to the inclusion of physical stressors into cumulative risk assessment. However, there are also options for working around or overcoming these challenges using available tools.

16.6.1 Challenge: Deciding Which Physical Stressors to Include in Cumulative Risk Assessments

There are numerous potential physical stressors available for consideration. For example, a risk assessment in an urban environment could reasonably include traffic noise, heat stress, microbial load, and ultraviolet radiation, among other risk-contributing factors. Determining which physical stressors to include can increase the complexity of the scoping phase. Another challenge may be that data are not available for all of the potential physical stressors that are relevant.

16.6.1.1 Recommendations

There are two considerations that could help to guide inclusion of physical stressors into cumulative risk assessment. The first consideration is which physical stressors are relevant to the goals of the risk assessment. For example, if the risk assessment is targeted toward understanding relative contributions of stressors to a particular disease outcome, only physical stressors that could plausibly contribute to the disease should be included. The same types of methods that are useful in

prioritizing chemicals for study and assessment (see Part II on prioritizing mixtures) can also be applied to physical stressors. The second consideration that will drive inclusion decisions is data availability. It should be noted that exposure to some physical stressors (e.g., radio waves, microbes) may be more difficult to measure in populations than others (e.g., temperature). However, the data needs for these physical stressors will become more apparent as researchers and risk assessors continue to work through examples.

16.6.2 Challenge: Lack of Physical Stressor Data

The lack of available "dose"-response data for physical stressors is often cited as an impediment to their incorporation into cumulative risk assessments. Typical toxicity studies with physical stressors or combinations of physical and chemical stressors tend to include one level of exposure (e.g., with or without a particular microbe). However, many of the available cumulative risk assessment approaches that have been applied to multiple chemicals require dose-response information. Additionally, there is a lack of data on interactions between physical and chemical stressors. Although examples of interaction data have been presented in this chapter, joint action of the vast majority of physical and chemical stressor combinations has not been evaluated.

16.6.2.1 Recommendations

To overcome the real deficit in traditional toxicological (i.e., dose-response) data for physical stressors and data on physical and chemical stressor interactions, research in these areas needs to be prioritized. This is likely to be an iterative process. As more case studies are developed and cumulative risk assessments that include physical stressors are performed, researchers will gain a better understanding of the types of physical stressor data that will be most useful to risk assessors, which in turn will guide study design. In addition to generating more toxicological data on physical stressors, making better use of existing databases (e.g., NHANES) and creatively using data from nontraditional sources (e.g., meteorological data combined with hospital visits for respiratory disease) are recommended. The case study by Evans et al. (2014) exploring the use of secondary data on exposure to noise and VOCs in a cumulative risk assessment for hearing loss provides an excellent template for replication with other physical/chemical stressor combinations and health outcomes.

16.6.3 Challenge: Heterogeneity of Data on Physical Versus Chemical Stressors

While doses of chemicals are typically presented in route-specific standard format (e.g., mg/kg for oral or dermal exposure and ppm for food-based or inhalation exposure), there is no consensus on "dose" units for physical stressors (e.g., for heat, the WBGT index is preferred; however, ambient temperature could be more widely available), and physical stressor "dose" units are not likely to align with chemical dose units (e.g., temperature versus chemical dose).

16.6.3.1 Recommendations

Although there is potential for confusion from combining factors across different data types, this does not pose a significant obstacle in calculating risk. Common dose units are not required in the HI approach (see Evans et al., (2014) example and the case study on heat and pesticides presented in this chapter) or the independent action approach. However, increased research attention on physical stressors and their dose-response relationships should inform selection of an appropriate "dose" measure.

16.6.4 Challenge: Categorical Differences

A common concern in cumulative risk assessment discussions is how to deal with stressors that fall under different regulatory jurisdictions. There are multiple arguments for including only stressors that fall under a single regulatory umbrella. There is the possibility of a legislative mandate guiding this decision (e.g., the 1996 Food Quality Protection Act [FQPA] charges the U.S. EPA's Office of Pesticide Programs to address pesticide mixtures). Due to the specificity of the FQPA, there is an assumption that stressors outside the scope of the legislation should be excluded from cumulative risk assessments addressing the mandate. Alternatively, the argument can be made on pragmatic grounds. For example, because there is not a clear path toward exposure reduction for some physical stressors, they should not be included in cumulative risk assessments. Finally, it could be argued that including stressors outside of the regulatory scope of an agency could result in the unintended consequence of decreasing action on chemical exposure. For example, if a cumulative risk assessment concludes that the risk from specific chemical exposures is dwarfed by a nonchemical factor such as low socioeconomic status, it might decrease political will to mitigate the chemical exposures.

16.6.4.1 Recommendations

The argument for limiting the scope of cumulative risk assessment based on legislative mandate or practical considerations brings up an often-cited goal of conducting cumulative risk assessments tailored to the question at hand. For example, if the goal is to make decisions about chemical remediation (i.e., determine which chemicals require resource-intensive cleanup efforts), limiting the risk assessment to chemicals may be in order. However, the inclusion of a broad range of potential stressors (both chemical and nonchemical) is recommended in addressing more global public health issues, such as determining stressors that are most likely to impact public health in communities-of-concern and using cumulative risk assessment information to determine how to direct limited resources to best protect public health. Furthermore, inclusion of both chemical and nonchemical stressors allows for comparison of relative contribution of the various stressors. In other words, the stressors that are likely to drive the adverse outcome of interest could be identified and targeted for intervention.

Acknowledgements This work was supported in part by the NIH, National Institute of Environmental Health Sciences.

References

Alfano, F.R.D., J. Malchaire, B.I. Palella, and G. Riccio. 2014. WBGT index revisited after 60 years of use. *Annals of Occupational Hygiene* 58: 955–970.

Babisch, W., H. Ising, and J.E. Gallacher. 2003. Health status as a potential effect modifier of the relation between noise annoyance and incidence of ischaemic heart disease. *Occupational and Environmental Medicine* 60: 739–745.

Babisch, W., G. Pershagen, J. Selander, D. Houthuijs, O. Breugelmans, E. Cadum, F. Vigna-Taglianti, K. Katsouyanni, A.S. Haralabidis, K. Dimakopoulou, P. Sourtzi, S. Floud, and A.L. Hansell. 2013. Noise annoyance - a modifier of the association between noise level and cardiovascular health? *Science of the Total Environment* 452-453: 50–57.

Berry, H.L., K. Bowen, and T. Kjellstrom. 2010. Climate change and mental health: A causal pathways framework. *International Journal of Public Health* 55: 123–132.

Braga, A.L.F., A. Zanobetti, and J. Schwartz. 2002. The effect of weather on respiratory and cardiovascular deaths in 12 US cities. *Environmental Health Perspectives* 110: 859–863.

Budd, G.M. 2008. Wet-bulb globe temperature (WBGT) - its history and its limitations. *Journal of Science and Medicine in Sport* 11: 20–32.

Burleson, G.R., H. Lebrec, Y.G. Yang, J.D. Ibanes, K.N. Pennington, and L.S. Birnbaum. 1996. Effect of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) on influenza virus host resistance in mice. *Fundamental and Applied Toxicology* 29: 40–47.

Campo, P., R. Lataye, G. Loquet, and P. Bonnet. 2001. Styrene-induced hearing loss: A membrane insult. *Hearing Research* 154: 170–180.

Campo, P., K. Maguin, and R. Lataye. 2007. Effects of aromatic solvents on acoustic reflexes mediated by central auditory pathways. *Toxicological Sciences* 99: 582–590.

Campo, P., T. Venet, A. Thomas, C. Cour, C. Brochard, and F. Cosnier. 2014. Neuropharmacological and cochleotoxic effects of styrene. Consequences on noise exposures. *Neurotoxicology and Teratology* 44: 113–120.

Catignani, J.C., M.G. Menache, and M.K. Selgrade. 1989. Increased susceptibility to pentobarbital following mouse cytomegalovirus infection: Relative roles of viral-induced interferon and viral infection of the liver. *Journal of Biochemical Toxicology* 4: 221–229.

Chen, G.D., and L.D. Fechter. 1999. Potentiation of octave-band noise induced auditory impairment by carbon monoxide. *Hearing Research* 132: 149–159.

Chen, G.D., L.H. Chi, P.J. Kostyniak, and D. Henderson. 2007. Styrene induced alterations in biomarkers of exposure and effects in the cochlea: Mechanisms of hearing loss. *Toxicological Sciences* 98: 167–177.

Choi, Y.H., H. Hu, B. Mukherjee, J. Miller, and S.K. Park. 2012. Environmental cadmium and lead exposures and hearing loss in U.S. adults: The National Health and nutrition examination survey, 1999 to 2004. *Environmental Health Perspectives* 120: 1544–1550.

Discalzi, G., D. Fabbro, F. Meliga, A. Mocellini, and F. Capellaro. 1993. Effects of occupational exposure to mercury and lead on brainstem auditory evoked potentials. *International Journal of Psychophysiology* 14: 21–25.

Doyle, M., K.G. Sexton, H. Jeffries, and I. Jaspers. 2007. Atmospheric photochemical transformations enhance 1,3-butadiene-induced inflammatory responses in human epithelial cells: The role of ozone and other photochemical degradation products. *Chemico-Biological Interactions* 166: 163–169.

EU-OSHA, E.A.F.S.A.H.A.W. 2009. *Combined exposure to noise and ototoxic substances*. Luxembourg: Office for Official Publications of the European Communities.

Evans, A.M., G. Rice, J.M. Wright, and L.K. Teuschler. 2014. Exploratory cumulative risk assessment (CRA) approaches using secondary data. *Human and Ecological Risk Assessment* 20: 704.

Fechter, L.D. 1995. Combined effects of noise and chemicals. *Occupational Medicine* 10: 609–621.

Fechter, L.D., G.D. Chen, D. Rao, and J. Larabee. 2000. Predicting exposure conditions that facilitate the potentiation of noise-induced hearing loss by carbon monoxide. *Toxicological Sciences* 58: 315–323.

Finlayson-Pitts, B.J., and J.N. Pitts, eds. 1999. *Chemistry of the upper and lower atmosphere, theory, experiments and applications*. New York: Academic Press.

Gallagher, S.S., G.E. Rice, L.J. Scarano, L.K. Teuschler, G. Bollweg, and L. Martin. 2015. Cumulative risk assessment lessons learned: A review of case studies and issue papers. *Chemosphere* 120: 697–705.

Garcia-Trabanino, R., E. Jarquin, C. Wesseling, R.J. Johnson, M. Gonzalez-Quiroz, I. Weiss, J. Glaser, J.J. Vindell, L. Stockfelt, C. Roncal, T. Harra, and L. Barregard. 2015. Heat stress, dehydration, and kidney function in sugarcane cutters in EI Salvador - a cross-shift study of workers at risk of Mesoamerican nephropathy. *Environmental Research* 142: 746–755.

Gordon, C.J., A.F.M. Johnstone, and C. Aydin. 2014. Thermal stress and toxicity. *Comprehensive Physiology* 4: 995–1016.

Head, J.L., and B.P. Lawrence. 2009. The aryl hydrocarbon receptor is a modulator of anti-viral immunity. *Biochemical Pharmacology* 77: 642–653.

Heilmann, C., E. Budtz-Jorgensen, F. Nielsen, B. Heinzow, P. Weihe, and P. Grandjean. 2010. Serum concentrations of antibodies against vaccine toxoids in children exposed perinatally to immunotoxicants. *Environmental Health Perspectives* 118: 1434–1438.

Henderson, D., E.C. Bielefeld, K.C. Harris, and B.H. Hu. 2006. The role of oxidative stress in noise-induced hearing loss. *Ear and Hearing* 27: 1–19.

Heugens, E.H.W., A.J. Hendriks, T. Dekker, N.M. Van Straalen, and W. Admiraal. 2001. A review of the effects of multiple stressors on aquatic organisms and analysis of uncertainty factors for use in risk assessment. *Critical Reviews in Toxicology* 31: 247–284.

IRIS. 2017. *Chemical assessment summary for permethrin* [Online]. U.S. EPA. Accessed 12 Dec 2017.

Ising, H., H. Lange-Asschenfeldt, G.F. Lieber, H. Weinhold, and M. Eilts. 2003. Effects of long-term exposure to street traffic exhaust on the development of skin and respiratory tract diseases in children. *Schriftenreihe des Vereins für Wasser-, Boden- und Lufthygiene* 112: 81–99.

Ising, H., H. Lange-Asschenfeldt, H.J. Moriske, J. Born, and M. Eilts. 2004. Low frequency noise and stress: Bronchitis and cortisol in children exposed chronically to traffic noise and exhaust fumes. *Noise & Health* 6: 21–28.

Jain, N., and W.A. Walker. 2015. Diet and host-microbial crosstalk in postnatal intestinal immune homeostasis. *Nature Reviews. Gastroenterology & Hepatology* 12: 14–25.

Jeffries, H., K. Sexton, J. Arnold, and T. Kale. 1989. *Validation testing of new mechanisms with outdoor chamber data, volume 3: calculation of photochemical reaction photolysis rates in the unc outdoor chamber*. Washington, DC: U.S. Environmental Protection Agency. EPA/600/3-89/010c.

Johns, D.O., L.W. Stanek, K. Walker, S. Benromdhane, B. Hubbell, M. Ross, R.B. Devlin, D.L. Costa, and D.S. Greenbaum. 2012. Practical advancement of multipollutant scientific and risk assessment approaches for ambient air pollution. *Environmental Health Perspectives* 120: 1238–1242.

Johnson, A.C., and B. Canlon. 1994. Progressive hair cell loss induced by toluene exposure. *Hearing Research* 75: 201–208.

Johnson, A.C., and T.C. Morata. 2010. Occupational exposure to chemicals and hearing impairment. In *Arbete och Hälsa*, ed. The Nordic Expert Group. Gothenburg: University of Gothenburg.

Kensler, T.W., B.D. Roebuck, G.N. Wogan, and J.D. Groopman. 2010. Aflatoxin: A 50-year odyssey of mechanistic and translational toxicology. *Toxicological Sciences* 120 (Suppl 1): S28–S48.

Kono, H., and K.L. Rock. 2008. How dying cells alert the immune system to danger. *Nature Reviews. Immunology* 8: 279–289.

Kovats, R.S., and S. Hajat. 2008. Heat stress and public health: A critical review. *Annual Review of Public Health* 29: 41–55.

Kraemer, M.J., C.T. Furukawa, J.R. Koup, G.G. Shapiro, W.E. Pierson, and C.W. Bierman. 1982. Altered theophylline clearance during an influenza B outbreak. *Pediatrics* 69: 476–480.

Lasky, R.E., M.M. Maier, E.B. Snodgrass, K.E. Hecox, and N.K. Laughlin. 1995. The effects of lead on otoacoustic emissions and auditory evoked potentials in monkeys. *Neurotoxicology and Teratology* 17: 633–644.

Lawrence, B.P., and B.A. Vorderstrasse. 2004. Activation of the aryl hydrocarbon receptor diminishes the memory response to homotypic influenza virus infection but does not impair host resistance. *Toxicological Sciences* 79: 304–314.

Le Prell, C.G., D. Yamashita, S.B. Minami, T. Yamasoba, and J.M. Miller. 2007. Mechanisms of noise-induced hearing loss indicate multiple methods of prevention. *Hearing Research* 226: 22–43.

Li, Y., J.M. Gibson, P. Jat, G. Puggioni, M. Hasan, J.J. West, W. Vizuete, K. Sexton, and M. Serre. 2010. Burden of disease attributed to anthropogenic air pollution in the United Arab Emirates: Estimates based on observed air quality data. *Science of the Total Environment* 408: 5784–5793.

Lichtveld, K.M., S.M. Ebersviller, K.G. Sexton, W. Vizuete, I. Jaspers, and H.E. Jeffries. 2012. In vitro exposures in diesel exhaust atmospheres: Resuspension of PM from filters versus direct deposition of PM from air. *Environmental Science & Technology* 46: 9062–9070.

Maguin, K., P. Campo, and C. Parietti-Winkler. 2009. Toluene can perturb the neuronal voltage-dependent Ca2+ channels involved in the middle-ear reflex. *Toxicological Sciences* 107: 473–481.

Michelozzi, P., G. Accetta, M. De Sario, D. D'ippoliti, C. Marino, M. Baccini, A. Biggeri, H.R. Anderson, K. Katsouyanni, F. Ballester, L. Bisanti, E. Cadum, B. Forsberg, F. Forastiere, P.G. Goodman, A. Hojs, U. Kirchmayer, S. Medina, A. Paldy, C. Schindler, J. Sunyer, C.A. Perucci, and P.C. GRP. 2009. High temperature and hospitalizations for

cardiovascular and respiratory causes in 12 European cities. *American Journal of Respiratory and Critical Care Medicine* 179: 383–389.

Morgan, E.T., K.B. Goralski, M. Piquette-Miller, K.W. Renton, G.R. Robertson, M.R. Chaluvadi, K.A. Charles, S.J. Clarke, M. Kacevska, C. Liddle, T.A. Richardson, R. Sharma, and C.J. Sinal. 2008. Regulation of drug-metabolizing enzymes and transporters in infection, inflammation, and cancer. *Drug Metabolism and Disposition* 36: 205–216.

Ong, F.H., P.J. Henry, and P.C. Burcham. 2012. Prior exposure to acrolein accelerates pulmonary inflammation in influenza A-infected mice. *Toxicology Letters* 212: 241–251.

OSHA. 2016. *OSHA technical manual*.

Otto, D.A., and D.A. Fox. 1993. Auditory and visual dysfunction following lead exposure. *Neurotoxicology* 14: 191–207.

Patz, J.A., H. Frumkin, T. Holloway, D.J. Vimont, and A. Haines. 2014. Climate change challenges and opportunities for Global Health. *Jama-Journal of the American Medical Association* 312: 1565–1580.

Petrovic, V., S. Teng, and M. Piquette-Miller. 2007. Regulation of drug transporters during infection and inflammation. *Molecular Interventions* 7: 99–111.

Qian, G.S., R.K. Ross, M.C. Yu, J.M. Yuan, Y.T. Gao, B.E. Henderson, G.N. Wogan, and J.D. Groopman. 1994. A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in shanghai, People's Republic of China. *Cancer Epidemiology, Biomarkers and Prevention* 3: 3–10.

Rager, J.E., K. Lichtveld, S. Ebersviller, L. Smeester, I. Jaspers, K.G. Sexton, and R.C. Fry. 2011. A toxicogenomic comparison of primary and photochemically altered air pollutant mixtures. *Environmental Health Perspectives* 119: 1583–1589.

Rao, D.B., and L.D. Fechter. 2000. Increased noise severity limits potentiation of noise induced hearing loss by carbon monoxide. *Hearing Research* 150: 206–214.

Raymer, J.H., W.B. Studabaker, M. Gardner, J. Talton, S.A. Quandt, H. Chen, L.C. Michael, M. Mccombs, and T.A. Arcury. 2014. Pesticide exposures to migrant farmworkers in eastern NC: Detection of metabolites in farmworker urine associated with housing violations and camp characteristics. *American Journal of Industrial Medicine* 57: 323–337.

Recio, A., C. Linares, J.R. Banegas, and J. Diaz. 2016. Road traffic noise effects on cardiovascular, respiratory, and metabolic health: An integrative model of biological mechanisms. *Environmental Research* 146: 359–370.

Reif, A.E., and T. Heeren. 1999. Consensus on synergism between cigarette smoke and other environmental carcinogens in the causation of lung cancer. *Advances in Cancer Research* 76 (76): 161–186.

Rice, D.C. 1997. Effects of lifetime lead exposure in monkeys on detection of pure tones. *Fundamental and Applied Toxicology* 36: 112–118.

Rubin, P. 1977. Radiation toxicology - quantitative radiation pathology for predicting effects. *Cancer* 39: 729–736.

Ryan, L.K., D.L. Neldon, L.R. Bishop, M.I. Gilmour, M.J. Daniels, D.M. Sailstad, and M.J. Selgrade. 2000. Exposure to ultraviolet radiation enhances mortality and pathology associated with influenza virus infection in mice. *Photochemistry and Photobiology* 72: 497–507.

Ryan, L.K., L.R. Copeland, M.J. Daniels, E.R. Costa, and M.J. Selgrade. 2002. Proinflammatory and Th1 cytokine alterations following ultraviolet radiation enhancement of disease due to influenza infection in mice. *Toxicological Sciences* 67: 88–97.

Schickele, E. 1947. Environment and fatal heat stroke; an analysis of 157 cases occurring in the Army in the U.S. during world war II. *Military Surgeon* 100: 235–256.

Schwartz, J., J.M. Samet, and J.A. Patz. 2004. Hospital admissions for heart disease - the effects of temperature and humidity. *Epidemiology* 15: 755–761.

Selgrade, M.J.K. 2010. Immune System. In *A textbook of modern toxicology*, ed. E. Hodgson, 4th ed. New York: Wiley.

Selgrade, M.K., and M.I. Gilmour. 2006. Immunotoxicology of inhaled compounds - assessing risks of local immune suppression and hypersensitivity. *Journal of Toxicology and Environmental Health - Part A* 69: 827–844.

Selgrade, M.K., M.J. Daniels, J.W. Illing, A.L. Ralston, M.A. Grady, E. Charlet, and J.A. Graham. 1984. Increased susceptibility to parathion poisoning following murine cytomegalovirus infection. *Toxicology and Applied Pharmacology* 76: 356–364.

Selgrade, M.K., J.W. Illing, D.M. Starnes, A.G. Stead, M.G. Menache, and M.A. Stevens. 1988. Evaluation of effects of ozone exposure on influenza infection in mice using several indicators of susceptibility. *Fundamental and Applied Toxicology* 11: 169–180.

Selgrade, M.J.K., K.D. Cooper, R.B. Devlin, H. Vanloveren, R.E. Biagini, and M.I. Luster. 1995. Immunotoxicity-bridging the gap between animal research and human health-effects. *Fundamental and Applied Toxicology* 24: 13–21.

Sexton, K. 2012. Cumulative risk assessment: An overview of methodological approaches for evaluating combined health effects from exposure to multiple environmental stressors. *International Journal of Environmental Research and Public Health* 9: 370–390.

Sexton, K., J. Arnold, H. Jeffries, T. Kale, and R. Kamens. 1988. *validation data for photochemical mechanisms: Experimental results*. Washington, DC: U.S. Environmental Protection Agency. EPA/600/3-88/041.

Sexton, K.G., H.E. Jeffries, M. Jang, R.M. Kamens, M. Doyle, I. Voicu, and I. Jaspers. 2004. Photochemical products in urban mixtures enhance inflammatory responses in lung cells. *Inhalation Toxicology* 16 (Suppl 1): 107–114.

Shargorodsky, J., S.G. Curhan, E. Henderson, R. Eavey, and G.C. Curhan. 2011. Heavy metals exposure and hearing loss in US adolescents. *Archives of Otolaryngology – Head & Neck Surgery* 137: 1183–1189.

Shi, L., I. Kloog, A. Zanobetti, P. Liu, and J.D. Schwartz. 2015. Impacts of temperature and its variability on mortality in New England. *Nature Climate Change* 5: 988–991.

Soto-Pena, G.A., A.L. Luna, L. Acosta-Saavedra, P. Conde-Moo, L. Lopez-Carrillo, M.E. Cebrian, M. Bastida, E.S. Calderon-Aranda, and L. Vega. 2006. Assessment of lymphocyte subpopulations and cytokine secretion in children exposed to arsenic. *FASEB Journal* 20: 779–77+.

Spector, J.T., and P.E. Sheffield. 2014. Re-evaluating occupational heat stress in a changing climate. *Annals of Occupational Hygiene* 58: 936–942.

Strand, L.B., A.G. Barnett, and S.L. Tong. 2012. Maternal exposure to ambient temperature and the risks of preterm birth and stillbirth in Brisbane, Australia. *American Journal of Epidemiology* 175: 99–107.

Tawatsupa, B., L.L.Y. Lim, T. Kjellstrom, S.A. Seubsman, A. Sleigh, and T.C.S. Team. 2012. Association between occupational heat stress and kidney disease among 37 816 Workers in the Thai Cohort Study (TCS). *Journal of Epidemiology* 22: 251–260.

Teuschler, L.K., M.L. Dourson, W.M. Stiteler, P. Mcclure, and H. Tully. 1999. Health risk above the reference dose for multiple chemicals. *Regulatory Toxicology and Pharmacology* 30: S19–S26.

Tolle, L.B., and T.J. Standiford. 2013. Danger-associated molecular patterns (DAMPs) in acute lung injury. *Journal of Pathology* 229: 145–156.

U.S. EPA. 1984. *Estimating "concern levels" for concentrations of chemical substances in the environment*. Washington, DC: Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency.

———. 2002a. *Organophosphate pesticides: Revised cumulative risk assessment*. Washington, DC: Office of Pesticide Programs, U.S. Environmental Protection Agency.

———. 2002b. *A review of the reference dose and reference concentration process*. Washington, DC: Risk Assessment Forum, U.S. Environmental Protection Agency. EPA/630/P-02/002F.

———. 2012. *Status report: Evidence based advances in inhalation dosimetry for gases with effects in the lower respiratory tract and in the body*. Washington, DC: U.S. Environmental Protection Agency. EPA/600/R-11/067.

Venet, T., C. Rumeau, P. Campo, B. Rieger, A. Thomas, and C. Cour. 2011. Neuronal circuits involved in the middle-ear acoustic reflex. *Toxicological Sciences* 119: 146–155.

Venet, T., P. Campo, A. Thomas, C. Cour, B. Rieger, and F. Cosnier. 2015. The tonotopicity of styrene-induced hearing loss depends on the associated noise spectrum. *Neurotoxicology and Teratology* 48: 56–63.

Vizuete, W., L. Biton, H.E. Jeffries, and E. Couzo. 2010. Evaluation of relative response factor methodology for demonstrating attainment of ozone in Houston, Texas. *Journal of the Air & Waste Management Association* 60: 838–848.

Vizuete, W., K.G. Sexton, H. Nguyen, L. Smeester, K.M. Aagaard, C. Shope, B. Lefer, J.H. Flynn, S. Alvarez, M.H. Erickson, and R.C. Fry. 2015. From the field to the laboratory: Air pollutant-induced genomic effects in lung cells. *Environmental Health Insights* 9: 15–23.

Vyskocil, A., D. Drolet, C. Viau, F. Lemay, G. Lapointe, R. Tardif, G. Ruchon, M. Baril, N. Gagnon, F. Gagnon, D. Begin, and M. Gerin. 2007. A web tool for the identification of potential interactive effects of chemical mixtures. *Journal of Occupational and Environmental Hygiene* 4: 281–287.

Warren, T.K., K.A. Mitchell, and B.P. Lawrence. 2000. Exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) suppresses the humoral and cell-mediated immune responses to influenza a virus without affecting cytolytic activity in the lung. *Toxicological Sciences* 56: 114–123.

Wilson, T.E., R.E. Klabunde, and K.D. Monahan. 2014. Using thermal stress to model aspects of disease states. *Journal of Thermal Biology* 43: 24–32.

Yaglou, C.P., and D. Minard. 1957. Control of heat casualties at military training centers. *A.M.A. Archives of Industrial Health* 16: 302–316.

Zhao, Q.Y., M. Dourson, and B. Gadagbui. 2006. A review of the reference dose for chlorpyrifos. *Regulatory Toxicology and Pharmacology* 44: 111–124.

Chapter 17
Psychosocial and Chemical Stressors

Jane E. Clougherty and Jonathan I. Levy

Abstract Psychosocial stress has been consistently linked with alterations in immune, endocrine, and metabolic function, and growing evidence indicates that psychosocial stressors—including noise, poverty, and exposure to violence—may alter human susceptibility to environmental chemical exposures. As a result, there is a growing need for methods to disentangle patterns in chemical and non-chemical exposures and to quantify their independent and interacting effects on health.

Here, a framework is presented for integrating psychosocial stressors into a traditional risk assessment approach, with attention to exposure assessment for non-chemical stressors and to statistical methods for incorporation of very disparate exposures into a risk assessment. Finally, an illustrative case example is presented, to demonstrate an approach for incorporating a psychosocial stressor (here, exposure to violence, a key stressor in urban U.S. communities) into a cumulative risk assessment aiming to quantify air pollution effects on health.

Keywords Non-chemical · Psychosocial stressors · Epidemiology

17.1 Introduction/Framework

17.1.1 Psychosocial Stressors in Mixtures Analysis and Cumulative Risk Assessment

The rapidly growing interest in characterizing the combined effects of chemical and non-chemical stressors on health outcomes and in cumulative risk assessment has stemmed from a few key observations. First, significant chemical and non-chemical stressors are often spatially or demographically correlated, clustered in lower-

J. E. Clougherty (✉)
Drexel University Dornsife School of Public Health, Philadelphia, PA, USA
e-mail: jec373@drexel.edu

J. I. Levy
Boston University School of Public Health, Boston, MA, USA

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_17

income communities near highways or industrial corridors (Bullard 1993). Second, there is growing epidemiological and toxicological evidence that chronic psychological stress (oftentimes driven by poverty, exposure to violence, and other community-scale non-chemical stressors) may alter individuals' susceptibility to environmental pollution (Clougherty et al. 2007, 2010a, 2014; Clougherty and Kubzansky 2010; Virgolini et al. 2006). This heightened susceptibility may be mediated through a suite of immune, endocrine, and metabolic changes that occur under chronic stress, a condition collectively referred to as allostatic load (McEwen 1998). Together, these observations suggest that the most pollution-exposed communities also may be the most susceptible (Lipfert 2004). Thus, disentangling the effects of chemical and non-chemical exposures and identifying their potential for greater than additive effects are increasingly becoming research and policy priorities, recognized as critical toward identifying and protecting susceptible populations as well as reducing health disparities (U.S. EPA 2003). While numerous non-chemical stressors may either cause harm directly or increase vulnerability to harm by other stressors, the focus of this chapter is on the role of psychosocial stressors in cumulative risk assessments for physical or chemical environmental exposures.

A cumulative risk assessment may be motivated by observed disease patterns (effects-based risk assessment), interest in a specific set of exposures present in a given community (stressor-based risk assessment), or a defined subpopulation of concern within a community (receptor-based risk assessment). Although this third category is likely to overlap procedurally with either an effects-based or stressor-based assessment, given the nature of cumulative risk assessments, it would include a more explicit characterization of the community including elements such as demographics, geographic boundaries, and health. For any of these applications, psychosocial stressors may be hypothesized either to directly influence health outcomes or to modify health response to chemical/physical exposures. The issues at hand in accurately characterizing psychosocial stressors and their potential role in cumulative risk assessment apply equally to any assessment type, including such significant questions as:

- How does cumulative risk assessment integrate insight from epidemiology, given that (1) toxicological evidence may not fully capture human psychosocial stressors, (2) epidemiological evidence is currently lacking for the vast majority of chemicals, and (3) the vast majority of chemicals also do not currently have toxicological data sufficient for traditional risk assessment processes? Relatedly, can either toxicology or epidemiology adequately inform the distribution of stressors or vulnerability in a population, given relatively homogeneous populations in toxicology, and challenges in determining effects across differing subpopulations in epidemiology?
- How can exposures to psychosocial stressors be appropriately and jointly quantified and represented, given relatively little data beyond demographic information in most settings? Can defaults be established that are both meaningful and interpretable in a variety of settings?

- Can a cumulative risk assessment meaningfully include both chemical and psychosocial stressors, if interventions typically target chemical stressors? Is the process different if a non-chemical stressor happens to be influenced by risk management activities (e.g., if cleaning up a pollution source in a community alleviates psychosocial stress)?
- Which psychosocial stressors are potentially relevant for specific settings, chemical exposures, or outcomes?

The focus of this chapter is on the role of psychosocial stressors in human health cumulative risk assessment, with attention to a few key steps (i.e., exposure assessment, dose-response modeling). Available databases and metrics that could allow for characterization of exposure to psychosocial stressors are discussed, considering both ideal parameters and proxy measures and default assumptions that could be useful in the absence of detailed population-specific data. For dose-response modeling, strategies that could be used for epidemiological or toxicological evidence are considered, and similarities and differences from chemical mixtures are discussed. Finally, an illustrative case example is presented that emphasizes the viability of including psychosocial stressors using epidemiological evidence.

17.1.2 Background and Terminology

Psychosocial stressors (i.e., perceived stressors in our social environment) are hypothesized to lead to negative health outcomes directly, or through stress-related alterations in immune, neuroendocrine, or metabolic function, collectively referred to as 'allostatic load' (McEwen 1998). Through these multiple stress-related pathways, chronic stress may serve to damage the individual's health directly (Evans 2003) or may alter the individual's susceptibility to exposures in the physical environment, such as air pollutants (Clougherty et al. 2014) or cold viruses (Cohen et al. 1991).

The psychosocial stressor pathways can be best understood in that psychological stress, regardless of the perceived stressor (social or otherwise), results when external demands exceed an individual's perceived abilities and resources to meet those demands (Cohen et al. 1995). This may be best characterized as a three-phase process:

1. Stressor (i.e., event, condition, or stimuli which pose a challenge)
2. Appraisal (i.e., an individual's perception or interpretation of the stressor)
3. Response (e.g., psychological and physiological sequelae)

These phases are not independent, and all phases are required in order to exert a psychological or biological stress response; a stressor perceived as benign or beneficial generally produces no psychological stress response. Thus, exposure assessment for psychosocial stressors would ideally not simply catalog stressors

(e.g., life events, community violence), but would rather emphasize total perceived stress (capturing response to multiple differently appraised stressors) or negative affect (e.g., anxiety, depression) as a cumulative indicator of mental distress and psychosocial stress (Kubzansky et al. 1999; Seeman et al. 2002). In cumulative risk assessment, it can be very challenging to quantify the appraisal and response phases, especially across large populations, though these factors influence the accuracy and interpretability of many psychosocial stressors. It is a topic of ongoing research to validate associations between community-level stressors (e.g., crime) and individual perceived stress and to identify factors which strengthen or weaken associations between stressor indicators and stress responses.

17.2 Exposure Assessment

Although much discussion on psychosocial stressors in cumulative risk assessment focuses on dose-response modeling, characterizing human exposures to psychosocial stressors is a critical step. It is extremely challenging to accurately characterize exposures to chemicals, especially in a community context where there is a need to account for multiple chemicals simultaneously. Psychosocial stressors can be even more challenging and cannot be quantitatively measured or modeled using only the same methods used for chemical exposures. In large part, this is because responses to psychosocial stressors vary with individual perception (i.e., "appraisal"), often influenced by history and context. Here, we briefly address four key dimensions of exposure assessment for psychosocial stressors:

1. The need to characterize the mechanism(s) of action and hypothesized pathways of effect
2. The need to carefully assess proxy variables for psychosocial stressors, considering the level of operation (i.e., community or individual level) and validating measures
3. The need to consider correlations among psychosocial and chemical exposures, which influence the accurate development of statistical models for epidemiology and accurate interpretations of measures of association
4. The need to establish default assumptions for psychosocial stressor exposures, in the absence of population-specific exposure data

17.2.1 Characterizing Mechanism(s) of Action/Pathways of Effect

Measuring psychosocial stressors is particularly challenging because many can act through multiple pathways. Some exposures can operate (simultaneously) as both a physical and psychological stressor. For example, noise can physically damage the

inner ear but also can be a psychological stressor. The appropriate exposure assessment for the physical pathway would be direct measures of sound, while the appropriate exposure assessment for the psychosocial pathway would require measures of perceived annoyance. Although these measures may correlate, substantial exposure misclassification would occur where using physical measures alone for the psychosocial pathway. Likewise, heat may be a physical stressor, leading to heat exhaustion or hypertensive outcomes, and is also uncomfortable and therefore psychologically stressful over extended periods (one of many causes believed to contribute to greater urban violence during summer months and heat waves). Again, the optimal measure depends on the hypothesized pathway [i.e., apparent temperature (physical), or perceived discomfort (psychological)].

A related challenge is the possibility that some pollution sources may also be psychosocial stressors if, for example, residents perceive substantial risk from a local source or if the source's presence suggests to residents that their health and well-being are not valued by the larger society (Bullard 1990). As such, it is often challenging to disentangle the physical and psychological aspects of pollution which impact communities near toxic sites (Elliott et al. 1993; Eyles et al. 1993)—and, again, the appropriate measure for each pathway would differ. It has been argued that having exposure assessment be homologous with dose-response modeling is a critical feature of risk assessment (National Research Council 2009, 2011), which is especially true for psychosocial stressors. Thus, we recommend that any exposure assessment begin with a clear conceptual model for hypothesized direct and indirect pathways to health effects, which, in turn, would drive the selection of exposure measures.

Because individual stress responses vary over time, and because individuals respond differently to most community-level stressors, there is increasing interest in developing biomarkers of stress. In theory, a well-characterized biomarker could facilitate epidemiological analysis and provide a comprehensive indicator of the response of an individual to multiple stressors. Stress is, however, by definition, a non-specific condition impacting a wide range of bodily systems (Selye 1936), many of which (e.g., inflammation) are also impacted by chemical and physical exposures, and the effects of which may vary with co-exposures, comorbidities, or other facets of individual physiologic susceptibility. Further, the timing of stressor exposure can greatly influence response (e.g., acute vs. chronic stress have very different physiologic sequelae). Taken together, it is unlikely that a single biomarker—or even a resolved suite of biomarkers—will reliably and meaningfully capture all stress responses, for all periods of interest.

Nevertheless, substantial research has examined the physiology of stress and the search for "biomarkers" of both acute and chronic stress. To date, most "biomarkers" identified have been immune or neuroendocrine markers associated with physiological stress responses—e.g., cytokines and glucocorticoids (Miller et al. 2002)—and an important emphasis has been on distinguishing biomarkers of acute stress (e.g., cortisol as an indicator of hypothalamic-pituitary-adrenal (HPA)-axis activity) from those of chronic stress (e.g., endocrine disruption or NF-κB signaling as an indicator of HPA-axis regulation) (Miller et al. 2007).

Formerly, corticosteroids (e.g., cortisol) in blood or saliva were emphasized as markers of HPA-axis activity, although stress-related HPA function changes lead to cortisol dysregulation (via glucocorticoid resistance and HPA regulation), not simply increased cortisol production. As such, cortisol can be difficult to interpret and generally better indicates acute, rather than chronic, stress. Importantly, it remains unknown how acute stress response may differ under chronic stress scenarios, and thus some more recent research emphasizes indicators of glucocorticoid resistance and neuroendocrine signaling (Miller et al. 2007). Other evidence suggests that C-reactive protein (CRP) may reasonably capture chronic stress; however, CRP is a non-specific immune marker also elevated in response to air pollution and other exposures (Clougherty et al. 2010a). Although no single biomarker is appropriate for all applications (Brunner 2007), suites of physiologic parameters have been developed to represent chronic "allostatic load" in humans and include indicators of cardiovascular function, metabolism, cholesterol, glucose metabolism, HPA-axis function, and sympathetic nervous system activity (Kubzansky et al. 1999; Seeman et al. 2002). Several studies document chronic stress effects on cardiovascular risk indicators (abdominal obesity, elevated serum triglycerides, lower levels of high-density lipoprotein (HDL) cholesterol, glucose intolerance, elevated blood pressure) (Brunner 2007), known collectively as "metabolic syndrome," and this may provide a method for capturing cumulative stress effects on cardiovascular and systemic function.

It is important to recognize that biomarkers representing physiological responses to stress may provide insights that are more relevant to dose-response modeling, or to determining the mechanism of action, than to exposure measurement per se. There may be utility in characterizing some biomarkers for effects-based cumulative risk assessments, as multiple chemical and psychosocial stressors may influence cardiovascular function or other endpoints. Currently, however, these biomarkers do not necessarily inform exposure characterization. Following the stressor-appraisal-response model, commonly available exposure databases (e.g., census demographic data or community crime rates) may represent stressors, and biomarkers may capture aspects of stress response, but it is arguably individual appraisals that are most specific and germane to accurate assessment of psychosocial stress.

17.2.2 Use and Validation of Proxy Variables for Psychosocial Stressors

Many psychosocial stressors are difficult to measure or model directly, especially across all individuals in a large population. Thus, there is often value in identifying or constructing proxy variables to capture geographic patterns in stressor prevalence (e.g., community crime rates, school quality indicators), due to limited individual-level data. In many contexts, however, even characterization of relevant

community-level stressors may not be viable, and risk assessors may wish to use relatively simple aggregate-level data on socioeconomic positon (SEP) or similar proxy measures. In such cases, validation of associations between stressor indicators and a representative sample of individual stressor perceptions would be encouraged.

In this case, analysts should also carefully describe the hypothesized pathways, differentiating to the degree possible between compositional vs. contextual variables. Compositional variables refer to measures that reveal information about the distribution of individuals within a community, whereas contextual variables reveal information about the setting in which individuals live. For example, community-level poverty measures could serve as a proxy for the likelihood of individual poverty (a *compositional variable*) or as an indicator of processes which function solely at the aggregate (or community) level (e.g., collective efficacy, social capital, disinvestment in a community). These latter community-level features which impact multiple individuals are *contextual variables*. For example, the percent of households under the poverty line could be a compositional variable reflecting individual-level likelihood of poverty or a contextual variable reflecting negative neighborhood attributes.

The impact of SEP on health has been extensively explored in many settings: using wealth or income as an index of status within and across countries (Subramanian and Kawachi 2004), in communities using measurements of perceived social standing (Singh-Manoux et al. 2005), and in workplace settings using job grade (Clougherty et al. 2010b; Marmot et al. 1991). Given numerous measures of SEP, and potential confusion about what these proxy variables capture for cumulative risk assessment, key definitions and concepts are described briefly here. SEP refers to the individual or family's relative position in a society, particularly where economic and cultural factors determine resource access, or in hierarchical societies where psychosocial goods such as social influence or security are largely determined by social stature. SEP influences human health through a highly complex mix of many social and physical factors accumulating and interacting over the life course, including diet and health behaviors, healthcare access, and working and housing conditions. Increasingly, psychosocial stress appears to be one extremely important aspect of SEP influencing health, though SEP should not be assumed to be synonymous with psychosocial stress.

Wherever possible, validation studies should be performed to ensure that an aggregate-level indicator (especially when used as a compositional variable, to proxy for individual-level data) accurately captures intra-community contrast in the construct (and pathway) of interest. For example, community crime statistics used to proxy for crime-related stress could be validated using intra-community surveys on perceived crime exposures and chronic stress, then comparing crime indicators to the individual-level measures. While such validation studies may not always be practical, looking to existing validation and multilevel studies on stressor exposures and characterization may aid in accurately interpreting potential proxy variables, at both the individual and aggregate levels.

17.2.3 Incorporating Exposure Correlations

Exposure characterization in cumulative risk assessment should explicitly consider possible correlations among stressor exposures, both positive and negative, as these may impact exposure measurement error or apparent interaction effects. For example, if proximity to major roadways is associated with increased exposures to both traffic-related air pollutants and noise, then the exposure assessments should aim to differentiate these spatially and temporally confounded exposures to the extent possible.

Correlations could exist at the aggregate level (i.e., among contextual variables) or only as a function of individual behaviors or activities (individual-level compositional variables). For psychosocial stressors associated with multiple individual or contextual factors, especially for a community-scale cumulative risk assessment, simulation approaches to characterize multivariable demographic attributes with high geographic resolution may be warranted (Levy et al. 2014). Inclusion of geographic and demographic predictors of physical or chemical exposures, time activity patterns, and other exposure-relevant behaviors would facilitate modeling of psychosocial exposures.

It is recommended that each conceptual model be as clear and simple as can reasonably capture the key exposure(s) and pathway(s) of interest—even within a cumulative risk assessment which may ultimately include many interacting exposures or a complex disease outcome. Overloading the conceptual model may obscure the specific hypothesized pathways, and lead to overly complicated (and less meaningful) "kitchen-sink" analyses.

17.2.4 Establishing Default Assumptions

For psychosocial stressors, often difficult to characterize directly in a cumulative risk assessment, default distributions can be derived from administrative data (e.g., census variables, large-scale population surveys, surveillance data, or the peer-reviewed literature, depending on the stressor of interest). Understanding the readily available factors that correlate with psychosocial stress, and thereby could serve as proxies or predictors of individual stress or stressor exposures, would be crucial in conducting assessments that are both meaningful and comparable across applications. Risk assessors would greatly benefit from an exposure factors handbook or analogous database that extended to non-chemical stressors, as has been recommended by expert committees on risk assessment (National Research Council 2009).

17.3 Dose-Response Modeling

Dose-response modeling for cumulative risk assessment presents significant methodological challenges, including but not limited to the complexities of incorporating non-chemical stressors. For psychosocial stressors, insight may be derived from epidemiological or toxicological evidence (or a combination), but the evidence must be systematically evaluated to ensure meaningful outputs.

In general, psychosocial stressors can be evaluated comparably to chemical stressors, provided that the requisite exposure data and health evidence are available. For example, the concept of "sufficient similarity" is being explored for complex mixtures of chemicals, wherein major chemical components are found in similar proportions and similarities in health effects and dose-response relationships are also considered (see Chap. 15). For psychosocial stressors, similarities in the type and magnitude of health effects may fulfill "sufficient similarity" and suggest groupings of psychosocial stressors (or psychosocial and physical/chemical stressors) that could be beneficial to the cumulative risk assessment process. Similar to chemical groupings, psychosocial stressors may be reviewed for their co-occurrence, joint action, and mode of action. Psychosocial stressors may also contribute to dose-response modeling through an improved understanding of background exposures influencing the shape of the dose-response function for other stressors. In general, existing guidelines on chemical mixtures can be modified to include psychosocial and other non-chemical stressors, potentially through a focus on common adverse outcomes, rather than on common mode of action.

17.3.1 Centrality of Epidemiological Evidence for Psychosocial Factors/Developing Dose-Response Functions Using Epidemiology

In many cases, epidemiological evidence will be the only viable strategy for incorporating psychosocial stressors into cumulative risk assessment. If the exposure assessment relied on proxy variables for SEP or demographic attributes, these proxies do not translate readily into a toxicological context. More generally, there may not be animal models to represent the types of psychosocial stressors of interest. This raises considerable challenges given that relatively few epidemiological studies are able to fully characterize effects of multiple chemical and psychosocial stressors. If adequate epidemiology is available for all stressors of interest, developing dose-response functions for cumulative risk assessment may be relatively straightforward, though several key diagnostic questions still need to be answered before dose-response functions can be fully characterized.

First, the ideal evidence would involve studies examining all risk factors simultaneously and reported dose-response functions derived from multivariable models, controlling for co-exposures and testing for effect modification. Many

epidemiological studies, however, are underpowered for such investigations or do not use statistical methods needed to discern the effects of stressors with common sociodemographic predictors or operating at multiple levels. Structural equation models have increasingly been used to evaluate the joint influence of multiple risk factors in epidemiological studies, allowing for direct and indirect effects to be simultaneously estimated (Peters et al. 2014). These and other statistical techniques for multi-exposure epidemiology require large sample sizes and are limited in the types of data they can adequately incorporate but can offer considerable insight on both proximal and distal causes.

In many situations, however, not all stressors of interest will have been included in a single epidemiological study. Extracting dose-response functions for different stressors from different studies, generally from regression models that do not include all relevant stressors, is a viable approach only where confounding is shown to be limited. While most epidemiological studies likely omit some candidate confounders, insight on the likelihood of significant correlations between exposures can be included based on external evidence and first principles. For example, two predominantly indoor pollutants, highly correlated with air exchange rates, would likely be positively correlated, whereas an ambient air pollutant and a foodborne exposure may be less correlated. Combining insights from different studies also requires judgments on the distribution of vulnerable individuals in the population and presence of potential modifiers.

Another complexity arises from epidemiological studies using socioeconomic and demographic covariates as proxies for non-chemical or lifestyle factors (including but not limited to psychosocial stress). For example, SEP may be included in a regression model linking lead with IQ decrements, with the idea that SEP could proxy for psychosocial stress, nutritional factors, presence of a stimulating home environment, or a number of other risk factors associated with neurodevelopment. Using the findings for SEP, either as a main effect or effect modifier, would require a careful judgment about what the term captures in the study population, and whether the same association is present in the population of interest for the cumulative risk assessment. Development of a detailed conceptual model that includes both proximal and distal effects on health, as described earlier, will facilitate this process.

17.3.2 Developing Dose-Response Functions Using Toxicology

Despite the centrality of epidemiological evidence for psychosocial stressors, often only toxicological information is available for many chemical stressors. For non-cancer risk assessments, where the question of cumulative exposures is rather less well-studied than for cancer risk assessments, psychosocial stressors can be considered in three different places. First, if there is direct toxicological evidence

on the psychosocial stressor illustrating a similar mode of action as a chemical stressor, it can be treated similarly to a chemical mixture. For example, in rat models of lead and chronic stress (Cory-Slechta et al. 2004, 2010), or concentrated air pollution and chronic stress (Clougherty et al. 2010a), the psychosocial stressor is considered in the same bioassay with well-characterized chemical or physical exposures. As proposed in the NRC report on cumulative risk assessment for phthalates (National Research Council 2008), dose addition can be applied in contexts other than congruent dose-response functions, allowing for a broader application. The NRC report also proposes approaches to establish benchmark dose (BMDL) values for chemical mixtures under an assumption of dose addition, which can be directly applied to psychosocial stressors as well if the analogous toxicological data are available and if the dose metrics are relevant to human populations. This is a viable approach in limited contexts where exposures to psychosocial stressors can be readily characterized in toxicological studies.

Second, psychosocial stressors can also contribute toward a general understanding of the appropriate conceptual model for chemical stressors evaluated toxicologically. In *Science and Decisions* (National Research Council 2009), the committee proposed that the functional form of a dose-response model could only be determined once a series of diagnostic questions were asked, related to the mode of action, relevant background exposures and endogenous processes, and vulnerable populations. Depending on the responses, the population dose-response function would reflect one of three conceptual models: (1) low-dose linear responses for individuals, with low-dose linear responses for the population; (2) non-linear responses for individuals with low-dose linear population responses with background dependence, and (3) non-linear responses for individuals with low-dose non-linear population responses independent from background. Historically, non-cancer responses have been considered to be of the third category (nonlinear for individuals with nonlinear population responses at low doses), although significant background exposures or other processes could be sufficient to linearize an otherwise nonlinear population dose-response function.

Practically speaking, this means that improved mechanistic knowledge for some psychosocial stressors could inform the shape of the dose-response function for those chemical stressors with adequate toxicological evidence. For example, increased risk of hypertension or elevated systolic/diastolic blood pressure has been associated with psychosocial stress, diet, and other non-chemical stressors, though perhaps not in toxicological data, in a manner necessary to follow the second approach above. The non-chemical stressors are associated with the outcome of interest and prevalent in the general population. This would imply that the toxicological evidence associating a chemical stressor with hypertension would be assumed to follow the second conceptual model above, with the point of departure (POD) used to develop a slope term and an estimated risk-specific dose. This approach is conceptually viable but has two significant challenges. First, as mentioned previously, it may be unclear whether the background processes are sufficient in magnitude to conclude that a low-dose linear model would be appropriate. Experience with some case studies would help to formalize this step. Second, one

would not be able to quantify risks attributable to non-chemical stressors incorporated in this manner. While this would be problematic in contexts where these non-chemical stressors were the targets of risk management efforts, they would only be included in this manner if there were no adequate toxicology or epidemiology, in which case they would be omitted from traditional quantitative analysis regardless of the approach.

Third, psychosocial stressors could be captured within physiologically based pharmacokinetic (PBPK) or pharmacodynamic models that provide insight about how these stressors would influence delivered dose or pharmacodynamic outcomes that could be the endpoints of cumulative risk assessments. For example, evidence has shown that chronic psychosocial stress can influence metabolism and cause hormonal changes (Agarwal and Marshall 1998), which could be incorporated into PBPK models. So, even lacking direct toxicological evidence on the influence of psychosocial stressors, these stressors could be incorporated into cumulative risk assessments via an adjustment of either the delivered doses from the toxicological study or the interpretation of the pharmacodynamic outcome.

17.3.3 Combining Insights from Epidemiology and Toxicology

In some cases, there will be epidemiological evidence for a small number of stressors, toxicological evidence for other stressors, and perhaps a subset of stressors with both toxicological and epidemiological evidence. Developing a systematic approach to incorporate psychosocial stressors in this context will therefore be key to cumulative risk assessment.

Depending on the nature of the available evidence, a hybrid of the two approaches above would be warranted. In a situation where the preponderance of the evidence is toxicological and the epidemiology is not directly applicable, the more limited epidemiological information could help establish whether the toxicants should be considered as linear or non-linear at low doses. In situations where multiple compounds are well-characterized toxicologically and at least one is well-characterized epidemiologically, approaches can be used to establish dose equivalence within toxicological studies to allow for interpretation of the epidemiological evidence. For example (Benignus et al. 2005), toxicological studies have linked both toluene and alcohol with similar neurobehavioral effects. Epidemiological evidence is robust for alcohol but not available for toluene. The toxicological studies can be used to estimate the dose of toluene that is functionally equivalent to a dose of alcohol for a defined outcome, and this could be used as a bridge to interpret the alcohol epidemiology with respect to toluene exposure. This clearly involves a number of assumptions regarding comparable dose-response function shapes, but the approach can be generalized in a variety of ways.

Developing comparable dose-response models across epidemiological and toxicological studies, in a manner that would allow for the models to be quantitatively combined, would only be possible in a limited number of situations. The adverse outcomes would need to be comparable to one another, which may be possible for some physiological measures but would be challenging for outcomes such as asthma attacks, hospitalizations, and other common epidemiological endpoints. There would also need to be detailed understanding of the vulnerability characteristics of both the human and animal populations, to ensure that adequate adjustments were made to account for the presumed greater heterogeneity in the human population. It is likely that these criteria would be met very infrequently, so that more often, cumulative risk assessment would be primarily based on either epidemiological or toxicological evidence, using the other to help inform the conceptual model or determination of mode of action.

17.4 Illustrative Case Example

To illustrate some of these approaches for incorporating psychosocial stressors, a case study example is presented, drawing on the epidemiological literature suggesting significant effect modification of associations between urban air pollutants and childhood asthma by chronic stressors prevalent in urban environments, notably exposure to violence (ETV). A process by which cumulative risk assessment could include this psychosocial stressor in the presence of chemical/ physical stressors is described, intending that the structure could extend to other stressors in a specific risk management context.

Asthma is a multifactorial illness impacted by a host of social, environmental, and genetic risk factors. As such, it serves as an appropriate case study for considering the interplay among two (or many more) risk factors—acting separately or in tandem—toward shaping patterns of asthma etiology and exacerbation in the urban environment.

This example may be conceptualized as within either a stressor-based or effects-based cumulative risk assessment (Menzie et al. 2007). For example, an analysis might consider the health benefits of multiple stressors reduced through traffic mitigation efforts and would need to take into account the modifying influence of key psychosocial stressors. Similarly, an analysis might be focused on geographic areas with elevated asthma prevalence or rates of exacerbation, determining key contributors to these patterns, in which case ETV and air pollution may be important to consider.

Air pollution and chronic stressors have been explored together in several epidemiological studies (e.g., Shankardass et al. 2009; Chen et al. 2008), and common distributions have been explored due to concerns about spatial correlations and potential confounding (e.g., traffic-related air pollution is spatially confounded by traffic-related noise) (Allen et al. 2009). ETV is explored here due, in part, to a small but growing literature on the salience of urban violence as a key chronic

stressor which may modify pollution effects on health. A study of asthmatic children in Boston public housing reported altered response to indoor allergen exposures with caregiver-reported fear of violence (Clougherty et al. 2006). A longitudinal study of childhood asthma etiology in East Boston reported significant associations with nitrogen dioxide (NO_2) exposures, but only among children with above-median prior lifetime exposure to violence (Clougherty et al. 2007). This model also informed one toxicological study exploring the effect of an aggressor stress (Social Dominance Paradigm), as a modifier of concentrated particulate air pollution (CAPs) effects on respiratory function in rats (Clougherty et al. 2010a); the authors reported substantively different responses to CAPs by stress group, with only stressed animals breathing more frequently and shallowly (e.g., hyperventilation) in response to increased CAPs exposures.

A few studies have explored other chronic stressors as modifiers of air pollution effects on asthma outcomes, though issues related to exposure measurement and the relative temporality between stress and pollution exposures have proven challenging. Aside from this small epidemiological literature suggesting a strong effect, there are several reasons why ETV is an appropriate stressor to examine in a cumulative risk assessment context:

1. Most importantly, violence is one of the few stressors that is rarely positively appraised. As described in the *Exposure Assessment* section above, under *Characterizing mechanism(s) of action/ pathways of effect*, perceived or psychosocial stress is strongly mediated through individual perception (appraisal). Unlike ETV, most other stressors may be appraised either positively or negatively (e.g., one may view losing a job as a good or bad thing, depending on whether one enjoyed the job or needed the financial benefit). A positive appraisal can render the stressor null; as such, most other stressors lend themselves to exposure misclassification. Exposure to violence, however, is almost never characterized as a positive exposure.
2. Outside of the rare instances of physical altercation, most of the impact of "exposure to violence" is through fear, hypervigilance, or stress-related pathways. (And, indeed, the experience of "fear of violence" can vary by gender, age, race, class, and other personal or community-level factors.) But, because most ETV impact occurs via psychosocial pathways, unlike other urban community stressors (e.g., housing quality), the hypothesized psychosocial pathway is relatively unconfounded by co-occurring physical impacts.
3. Analyses of spatial patterning in urban exposures suggest that ETV is not always spatially confounded by poverty and other stressors (Shmool et al. 2014). Thus, the health impacts of ETV conceivably can be disentangled from those of SEP, which is a much more complicated construct entailing a broad array of physical (e.g., diet) and psychosocial (e.g., discriminatory experience) exposures, at both the individual and community level, accumulating over the life course.
4. Crime data, albeit an imperfect community-level ETV indicator, is collected systematically by every police department nationwide, often according to standard criteria (i.e., felony crimes). Reporting bias remains a challenge—and

certainly differs by jurisdiction and type of crime—but the data are collected and publicly available nationwide.

While the details listed above are specific to ETV, the logic used to validate its inclusion is generalizable. Any psychosocial stressor under consideration would require a logical exposure metric that can be reasonably and systematically collected, evidence for a causal effect on a defined health outcome, and a systematic determination that the exposure metric is reasonably sensitive and specific for the health outcome in question.

To incorporate interaction effects in a cumulative risk assessment, we first recommend careful development of a conceptual model for the hypothesized relationships among the exposures of interest, with attention to modes of action (MOA) or common adverse outcomes. Here, our relatively simple conceptual model captures some of the key constructs of interest:

1. A psychosocial stressor (e.g., ETV) and air pollution exposures may separately influence childhood asthma etiology or exacerbation.
2. Perceived ETV (as a chronic stressor) may, through "allostatic load" pathways, alter individual susceptibility to air pollution exposures in the progression of asthma.

Notably, many more conceptual models are possible, considering the myriad of exposures that impact upon childhood asthma. Here, we restrict our analysis to the one key exposure (air pollution) and one key hypothesized psychosocial effect modifier (ETV). A typical conceptual model likely would be more complex, with explicit consideration of multiple causal pathways and both proximal and distal risk factors for health.

Here, a community-level indicator (e.g., crime rates) serves to proxy (albeit imperfectly) for individual-level perceived ETV. For analyses in which both community-level and individual-level data are available, it would be preferable to use the individual data as the primary exposure metric and to explore the community-level indicator as a predictor of the individual-level variable or as a contextual variable interacting (in a hierarchical model) with the individual-level variable. It is recommended that each conceptual model be as clear and simple as can reasonably capture the key exposure(s) and pathway(s) of primary interest.

It may be determined that solely of interest are community-level stressors, which may act primarily through psychosocial pathways and are captured reasonably through aggregate data. If so, the construct of interest (e.g., ETV) is first defined, then existing data are catalogued that may reasonably indicate the construct (e.g., felony crimes, murders, robberies, at police precinct level). For community-scale metrics and environmental exposures (e.g., air pollution), it can be valuable to apply spatial methods in Geographic Information Systems (GIS) to evaluate relative spatial distributions within and among exposures. Here, it would be valuable to understand:

1. Spatial (neighborhood-to-neighborhood) variability in crime rates—this extent of spatial "clustering" (or spatial autocorrelation) within a stressor can be

formally tested using GIS-based methods such as geographically weighted regression (GWR) or local indicators of spatial association (LISA). This analysis indicates how each exposure, separately, varies across the region of interest. In an ongoing investigation of social stressor patterning across New York City, significant spatial variation was found within all stressors examined, across multiple domains (e.g., economic stressors, crime and violence exposures, resource access, school-based stressors, etc.) (Shmool et al. 2014).

2. Spatial correlations between and among psychosocial stressors and pollutant exposures (e.g., correlation between ETV and air pollution) may be examined by comparing spatial distributions, and quantified using spatial autoregressive modeling (SAR). This analysis indicates the potential for confounding and/or effect modification between exposures. In New York City, social stressors were found to vary substantially in their spatial patterning, and not all stressors were correlated with poverty or pollution exposures (Shmool et al. 2014).

Developing the conceptual framework would normally involve initial screening-level quantification of health risks to determine whether the stressors are significant enough to merit inclusion. In this case, the relative risks of air pollution and ETV, from the epidemiological literature, are high enough, and the exposures sufficiently ubiquitous, to substantiate their inclusion. While the focus here is on epidemiological evidence, toxicological insights enhance the plausibility of the observed interaction (Virgolini et al. 2005, 2006; Cory-Slectha et al. 2004, 2010).

It is generally preferable to determine relative risks that consider both stressors simultaneously (whether as main effects or effect modifiers). Conceptually, the underlying epidemiological models would primarily be of two broad structures:

1. Direct effect of psychosocial stressor on the outcome:

$$\text{Asthma outcomes} = [\text{best metric(s) of ETV}] + \text{confounders}$$

2. Direct effect of a physical/chemical exposure on the outcome:

$$\text{Asthma outcomes} = [\text{best metric(s) of air pollution}] + \text{confounders}$$

In each case, the confounders could include the other exposure, though effect modification likely would not be considered at this stage. Importantly, the best available metrics of each exposure may differ significantly in sensitivity and specificity (if, for example, the best available metric of ETV is a community-level index, and the best available metric of air pollution is a modeled residence-specific estimate). Thus, differential exposure misclassification needs be considered, both when comparing separate models that examine two different exposures on a common health endpoint, and when merging both exposure metrics into the same epidemiological model.

At this stage, the underlying epidemiological study often would use GIS methods to visualize and assess spatial relationships among exposures (stressors)

of interest. These associations can be examined by comparing spatial maps of each stressor with that of the outcome variable (e.g., asthma hospitalizations) and quantified using SAR, to quantify these separate (unadjusted) associations. This step assumes that each stressor/exposure of interest carried forward should have a significant independent association with the outcome of interest, regardless of co-exposures or effect modifiers. In some cases, however, this may not be true; in one longitudinal study of childhood asthma etiology (Clougherty et al. 2007), significant associations between traffic-related NO_2 and asthma etiology were observed solely among children with above-median ETV. In cases of strong effect modification, the effect of the physical exposure of interest (air pollution) may be diluted to non-significance, if the sample has a high enough prevalence of low-susceptibility individuals. This concern may be alleviated through sensitivity testing wherein potential modifiers and exposures are considered iteratively prior to final exclusion. The heuristic epidemiologic model for such analyses would be:

$$\text{Asthma} = [\text{best metric(s) of ETV}] + [\text{best metric(s) of air pollution}] + \text{confounders}$$

Statistical tools, such as multiple and logistic regression and process models, can be used to explore the contributions of various stressors to the health endpoint of interest.

At this stage, GIS-based spatial approaches can be used to visualize and examine the overlay of stressors with the observed health effects. As above, maps of each exposure and outcome can be compared and formally tested using SAR or similar models, to assess the extent of spatial autocorrelation. A refinement that may be useful at this stage is examining the combined (joint) spatial distributions in ETV and air pollution (or the spatial distribution in a composite index that combines these exposures) against the spatial distribution in the health outcome of interest (e.g., maps of asthma hospitalizations). Unaccounted-for nonlinear or other complex joint distributions (such as that observed in Shmool et al. 2015) can lead to mis-specification or misinterpretation of epidemiological results, particularly when incorporating multiple exposures or interactions into the model.

Finally, an epidemiologic model would explore the possibility of interactions. Knowledge reflected in conceptual models should provide a grounding (and some limits on) the interactions considered. Clear mechanistic hypotheses indicating which stressor is hypothesized to modify each exposure are needed for useful, interpretable epidemiological analyses and related cumulative risk assessment output. Conceivably, this could lead to some stressors being considered only as effect modifiers, because no plausible mechanism exists for a main effect absent another stressor of interest. The incorporation of too many interactions, or interactions not supported by a plausible mechanistic pathway, however, can complicate the analysis, reduce statistical power, and lead to uninterpretable results (especially as the number of stressors under consideration increases). Of note, the interaction between the chemical and psychosocial stressor may not be a simple linear association, and statistical techniques that allow for multidimensional smoothed functions

of health response associated with multiple stressors may be informative. The heuristic epidemiological model that would support this interaction analysis is:

$$\begin{aligned}\text{Asthma} = \; & [\text{best metric(s) of ETV}] + [\text{best metric(s) of air pollution}] \\ & + [\text{best metric(s) of ETV}] \times [\text{best metric(s) of air pollution}] \\ & + \text{confounders}\end{aligned}$$

Epidemiological models being developed for cumulative risk assessment applications could also be designed to be responsive to the ultimate risk management decisions. For example, if the risk management strategies in question focused on air pollution exposures, it may be most salient to explicitly consider the influence of ETV as an effect modifier. Understanding the main effect of ETV would be less relevant, although it does contribute to an understanding of background rates of disease and characteristics of high-risk subpopulations. Although this is not an appreciable reduction in effort for a two-stressor analysis, an analysis of numerous stressors would benefit greatly from the analytical boundaries created through an appropriately focused set of risk management options.

Uncertainty analysis is emphasized as a key component of any cumulative risk assessment. For the epidemiology that may underlie a cumulative risk assessment, this goes beyond reported confidence intervals to include sensitivity analyses for the parameters included in the final models. It is strongly recommended that any cumulative risk assessment extract information on the sensitivity of epidemiological findings to some key assumptions. Similarly, researchers conducting epidemiological studies aiming to inform cumulative risk assessment should explicitly report uncertainties.

Because some important modifiers and predictors may be lost by omitting variables prior to testing interactions (i.e., researchers and risk assessors may miss effects that only become apparent through effect modification), some *sensitivity testing on covariate selection* is needed. This can be done by:

- Swapping order of terms/interactions tested in models
- Identifying key hypothesized predictors and modifiers carried throughout the analysis, regardless of significance
- Using automated variable selection procedures using both predictors and modifiers (e.g., regression trees)
- Using automated variable selection procedures that do not assume linearity or specific interaction structures (e.g., random forest), to identify underutilized stressor(s) for which data are available, but the relationships of such exposures with the outcome of interest have not been recognized fully in the main model

Finally, there remains significant utility in establishing that available exposure metrics accurately capture variability in the stressor(s) of interest. An effective way to do so, for aggregate-level indices (e.g., community violence rates), is to implement surveys (questionnaires) on individual's perceived stress (1) to systematically determine whether community-level indices accurately capture community-to-community variation in mean individual-level violence exposures and (2) to select

those aggregate-level metrics that best reflect individual variation in stressor exposures.

In summary, this case example illustrates that a psychosocial stressor such as exposure to violence can be incorporated reasonably into cumulative risk assessments including air pollution, as there is a biologically plausible linkage with a common adverse outcome (supported by some findings from both toxicology and epidemiology), an approach for exposure characterization that involves reasonable proxies from public databases, and empirical evidence supporting main effects and effect modification for both key stressors. Other non-chemical stressors can be evaluated and included through analogous approaches.

17.5 Conclusions

Data on psychosocial stressors indicate important effects on health that can interact with chemical environmental exposures. For psychosocial stressors, challenges arise with exposure assessment, dose-response modeling, and risk characterization in the context of risk management. In general, the exposure assessment phase of cumulative risk assessment requires increased attention, given both the need to characterize effects of simultaneous exposure to multiple chemical stressors and the need to develop meaningful proxies of exposure to psychosocial and other non-chemical stressors that are challenging to characterize directly. Development of a strong conceptual model including proximal and distal effects on health will help in determining the appropriate constructs for the analysis. This step is key, as many psychosocial stressors can influence health through multiple pathways, and many proxies for psychosocial stressor exposures can represent multiple stressors. Dose-response modeling using epidemiological data can benefit from systematic modeling approaches tied to well-developed conceptual models and using techniques such as structural equation modeling to identify associations with proximal and distal factors. Toxicological evidence may be more limited for psychosocial stressors, but even where evidence is insufficient to incorporate a psychosocial stressor into a mixtures analysis, consideration of psychosocial stressors can contribute toward selecting a conceptual dose-response model and may be incorporated into PBPK models or other analyses related to delivered doses or pharmacodynamic outcomes. The case example illustrates that it is viable to incorporate selected psychosocial stressors into cumulative risk assessment, following a systematic logic for well-structured exposure characterization and epidemiological modeling.

References

Agarwal, S.K., and G.D. Marshall. 1998. Glucocorticoid-induced type 1/type 2 cytokine alterations in humans: A model for stress-related immune dysfunction. *Journal of Interferon & Cytokine Research* 18: 1059–1068.

Allen, R.W., H. Davies, M.A. Cohen, G. Mallach, J.D. Kaufman, and S.D. Adar. 2009. The spatial relationship between traffic-generated air pollution and noise in 2 U.S. cities. *Environmental Research* 109: 334–342.

Benignus, V.A., P.J. Bushnell, and W.K. Boyes. 2005. Toward cost-benefit analysis of acute behavioral effects of toluene in humans. *Risk Analysis* 25 (2): 447–456.

Brunner, E. 2007. Biology and health inequality. *PLoS Biology* 5 (11): e267.

Bullard, R.D. 1993. *Confronting Environmental Racism: Voices from the Grassroots*. Cambridge, MA: South End Press.

———. 1990. *Dumping in Dixie: Race, Class and Environmental Quality*. Boulder: Westview Press.

Chen, E., H.M. Scheier, R.C. Strunk, and M. Brauer. 2008. Chronic traffic-related air pollution and stress interact to predict biologic and clinical outcomes in asthma. *Environmental Health Perspectives* 116 (7): 970–975.

Clougherty, J.E., J.I. Levy, H.P. Hynes, and J.D. Spengler. 2006. A longitudinal analysis of the efficacy of environmental interventions on asthma-related quality of life and symptoms among children in urban public housing. *The Journal of Asthma* 43 (5): 335–343.

Clougherty, J.E., J.I. Levy, L.D. Kubzansky, P.B. Ryan, S.F. Suglia, M.J. Canner, and R.J. Wright. 2007. Synergistic effects of traffic-related air pollution and exposure to violence on urban asthma etiology. *Environmental Health Perspectives* 115 (8): 1140–1146.

Clougherty, J.E., C.A. Rossi, J. Lawrence, M.S. Long, E.A. Diaz, R.H. Lim, B. McEwen, P. Koutrakis, and J.J. Godleski. 2010a. Chronic social stress and susceptibility to concentrated ambient fine particles in rats. *Environmental Health Perspectives* 118 (6): 769–775.

Clougherty, J.E., K. Souza, and M.R. Cullen. 2010b. Work and its role in shaping the social gradient in health. *Annals of the New York Academy of Sciences* 1186: 102–124.

Clougherty, J.E., and L.D. Kubzansky. 2010. A framework for examining social stress and susceptibility to air pollution in respiratory health. *Environmental Health Perspectives* 117 (9): 1351–1358.

Clougherty, J.E., J.L.C. Shmool, and L.D. Kubzansky. 2014. The role of non-chemical stressors in mediating socioeconomic susceptibility to environmental chemicals. *Current Environmental Health Reports* 1 (4): 302–313.

Cohen, S., D.A. Tyrrell, and A.P. Smith. 1991. Psychological stress and susceptibility to the common cold. *New England Journal of Medicine* 325 (9): 606–612.

Cohen, S., R.C. Kessler, and G.L. Underwood. 1995. *Measuring stress: A guide for health and social scientists*. New York: Oxford University Press.

Cory-Slechta, D.A., M.B. Virgolini, M. Thiruchelvam, D.D. Weston, and M.R. Bauter. 2004. Maternal stress modulates the effects of developmental lead exposure. *Environmental Health Perspectives* 112 (6): 717–730.

Cory-Slechta, D.A., S. Stern, D. Weston, J.L. Allen, and S. Liu. 2010. Enhanced learning deficits in female rats following lifetime pb exposure combined with prenatal stress. *Toxicological Sciences* 117 (2): 427–438.

Elliott, S.J., S.M. Taylor, S. Walter, D. Stieb, J. Frank, and J. Eyles. 1993. Modelling psychosocial effects of exposure to solid waste facilities. *Social Science & Medicine* 37 (6): 791–804.

Evans, G.W. 2003. A multimethodological analysis of cumulative risk and allostatic load among rural children. *Developmental Psychology* 39 (5): 924–933.

Eyles, J., S.M. Taylor, N. Johnson, and J. Baxter. 1993. Worrying about waste: Living close to solid waste disposal facilities in southern Ontario. *Social Science & Medicine* 37 (6): 805–812.

Kubzansky, L.D., I. Kawachi, and D. Sparrow. 1999. Socioeconomic status, hostility, and risk factor clustering in the Normative Aging Study: Any help from the concept of allostatic load? *Annals of Behavioral Medicine* 21 (4): 330–338.

Levy, J.I., M.P. Fabian, and J.L. Peters. 2014. Community-wide health risk assessment using geographically resolved demographic data: A synthetic population approach. *PLoS One* 9 (1): e87144.

Lipfert, F.W. 2004. Air pollution and poverty: Does the sword cut both ways? *Journal of Epidemiology and Community Health* 58 (1): 2–3.

Marmot, M.G., G.D. Smith, S. Stansfeld, C. Patel, F. North, J. Head, I. White, E. Brunner, and A. Feeney. 1991. Health inequalities among British civil servants: The Whitehall II study. *Lancet* 337 (8754): 1387–1393.

McEwen, B.S. 1998. Stress, adaptation, and disease. Allostasis and allostatic load. *Annals of the New York Academy of Sciences* 840: 33–44.

Menzie, C.A., M.M. MacDonell, and M. Mumtaz. 2007. A phased approach for assessing combined effects from multiple stressors. *Environmental Health Perspectives* 115 (5): 807–816.

Miller, G.E., S. Cohen, and A.K. Ritchey. 2002. Chronic psychological stress and the regulation of pro-inflammatory cytokines: A glucocorticoid-resistance model. *Health Psychology* 21 (6): 531–541.

Miller, G.E., E. Chen, and E.S. Zhou. 2007. If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin* 133 (1): 25–45.

National Research Council. 2009. *Science and decisions: Advancing risk assessment*. Washington: National Research Council.

———. 2011. *Improving health in the United States: The role of health impact assessment*. Washington: National Research Council.

———. 2008. *Phthalates and cumulative risk assessment: The tasks ahead*. Washington: National Research Council.

Peters, J.L., M.P. Fabian, and J.I. Levy. 2014. Combined impact of lead, cadmium, polychlorinated biphenyls and non-chemical risk factors on blood pressure in NHANES. *Environmental Research* 132: 93–99.

Seeman, T.E., B.H. Singer, C.D. Ryff, G. Dienberg Love, and L. Levy-Storms. 2002. Social relationships, gender, and allostatic load across two age cohorts. *Psychosomatic Medicine* 64 (3): 395–406.

Selye, H. 1936. A syndrome produced by diverse nocuous agents. *Nature* 138: 32. https://doi.org/10.1038/138032a0.

Shankardass, K., R. McConnell, M. Jerrett, J. Milam, J. Richardson, and K. Berhane. 2009. Parental stress increases the effect of traffic-related air pollution on childhood asthma incidence. *PNAS* 106 (30): 12406–12411.

Shmool, J.L.C., L.D. Kubzansky, O. Dotson Newman, J. Spengler, P. Shepard, and J.E. Clougherty. 2014. Social stressors and air pollution across New York City: A spatial approach for assessing correlation among multiple exposures. *Environmental Health* 13 (91). https://doi.org/10.1186/1476-069X-13-91.

Shmool, J.L.C., J.F. Bobb, K. Ito, B. Elston, D.A. Savitz, Z. Ross, T.D. Matte, S. Johnson, F. Dominici, and J.E. Clougherty. 2015. Area-level socioeconomic deprivation, nitrogen dioxide exposure, and term birth weight in New York City. *Environmental Research* 142: 624–632. https://doi.org/10.1016/j.envres.2015.08.019.

Singh-Manoux, A., M.G. Marmot, and N.E. Adler. 2005. Does subjective social status predict health and change in health status better than objective status? *Psychosomatic Medicine* 67 (6): 855–861.

Subramanian, S.V., and I. Kawachi. 2004. Income inequality and health: What have we learned so far? *Epidemiologic Reviews* 26: 78–91.

U.S. Environmental Protection Agency. 2003. *Framework for cumulative risk assessment*. Washington: U.S. Environmental Protection Agency.

Virgolini, M.B., K. Chen, D.D. Weston, M.R. Bauter, and D.A. Cory-Slechta. 2005. Interactions of chronic lead exposure and intermittent stress: Consequences for brain catecholamine systems and associated behaviors and HPA axis function. *Toxicological Sciences* 87 (2): 469–482.

Virgolini, M.B., M.R. Bauter, D.D. Weston, and D.A. Cory-Slechta. 2006. Permanent alterations in stress responsivity in female offspring subjected to combined maternal lead exposure and/or stress. *Neurotoxicology* 27 (1): 11–21.

Chapter 18
Community-Based Cumulative Impact Assessment: California's Approach to Integrating Nonchemical Stressors into Environmental Assessment Practices

Shannon R. Murphy, Shankar B. Prasad, John B. Faust, and George V. Alexeeff

Abstract Risk assessment is complex and challenges assessors to expand its utility and bridge data gaps to better account for human health risk. Mixtures complicate the assessment landscape because cumulative chemical exposures occur at the nexus of nonchemical stressors that can influence adverse health outcomes. Traditional *risk* assessment approaches typically use comprehensive data sources and quantitative methods but have a limited capacity to account for or include nonchemical stressors. In contrast, community-based cumulative *impact* assessments utilize different types of data and apply both quantitative and semiquantitative methods. Recently, multiple approaches for cumulative impact assessment have been developed. One such example is the California Communities Environmental Health Screening Tool: CalEnviroScreen. CalEnviroScreen has been successful in evaluating the cumulative pollution burden at a census tract scale across the state, based on 12 pollution indicators. It also characterizes population vulnerabilities at the same scale, based on intrinsic and extrinsic factors (three health and four socioeconomic status indicators). The two indices are combined in a way that allows one to screen and identify communities across California at above or below various thresholds in the scale. CalEnviroScreen allows one to understand the similarities and differences between the most disadvantaged communities having similar scores. CalEnviroScreen has been instrumental in (a) identifying the disadvantaged communities across California that receive prioritized funding from Greenhouse Gas Reduction Funds derived from the cap-and-trade program, (b) prioritizing areas for targeted multimedia enforcement action, and (c) assisting California Environmental Protection Agency boards and departments with planning community engagement and outreach efforts.

S. R. Murphy (✉) · S. B. Prasad · J. B. Faust · G. V. Alexeeff
Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Sacramento, CA, USA
e-mail: Shannon.Murphy@oehha.ca.gov

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6_18

Keywords Cumulative impact · Environmental justice · Risk · Pollution · Disadvantaged communities

18.1 Introduction

California was the first state to define environmental justice in law as "the fair treatment of people of all races, cultures, and incomes with respect to the development, adoption, implementation and enforcement of environmental laws, regulations and policies" (Cal Gov Code §65040.12(e) 1999). This definition, coupled with the California Environmental Protection Agency's (CalEPA) leadership and commitment to promote environmental justice, led the Agency to recognize that understanding cumulative impacts (CI) in a specific area or within a community would be a critical first step. In 2005, CalEPA integrated the "working definition" from the CalEPA Interagency Working Group Report (CalEPA 2003), along with input from multiple stakeholders, and adopted a common working definition of CI as meaning "exposures, public health or environmental impacts from the combined emissions and discharges, in a geographic area, including environmental pollution from all sources, whether single or multi-media, routinely, accidentally, or otherwise released. Impacts will take into account sensitive populations and socio-economic factors, where applicable and to the extent data are available" (CDPR 2005; OEHHA 2010). Stakeholders included representatives from local and federal government, academia, environmental justice and community-based organizations, industry, and the general public.

Environmental Justice
Under California law "means the fair treatment of people of all races, cultures, and incomes with respect to the development, adoption, implementation, and enforcement of environmental laws and policies."
California Government Code §65040.12(e)

CalEPA's Office of Environmental Health Hazard Assessment (OEHHA) was designated as the lead in developing guidance on incorporating CI into the decision-making process. The CI analysis evaluates the intersections of multiple chemicals, multiple sources, public health, and environmental effects with characteristics of the local population that could influence an adverse health outcome. In 2010, OEHHA finalized a framework documenting the scientific evidence for disproportionate CI as a first step (OEHHA 2010). This framework described factors that make up a comprehensive measure of impacts in a community and a scientific methodology that can be pursued to evaluate CI in a given community. As early as in 2004, U.S. EPA's National Environmental Justice Advisory Council (NEJAC) had recommended a similar conceptual framework known as the "Pollution Burden Matrix" for "developing a screening tool, which would rely primarily on analyses

of existing or readily available sources of data, to identify the most burdened census tracts within a specified region" (U.S. EPA 2003; NEJAC 2004). NEJAC's Pollution Burden Matrix served as a guiding construct during OEHHA's cumulative impact framework development.

Cumulative Impacts
Exposures, public health or environmental effects from the combined emissions and discharges, in a geographic area, including environmental pollution from all sources, whether single or multimedia, routinely, accidentally, or otherwise released. Impacts will take into account sensitive populations and socioeconomic factors, where applicable and to the extent data are available.
California Environmental Protection Agency

While efforts were in progress by various institutions to evaluate and develop different approaches or methods to estimate CI in a community, the landmark California Legislation Assembly Bill 32 (AB 32) – *Global Warming Solutions Act of 2006* (Nunez, Chapter 488, Statutes of 2006) – was passed. The bill language included the term "disadvantaged communities," referred to as "communities with minority populations or low-income populations, or both," and also contained a directive to "consider the potential for direct, indirect, and cumulative emission impacts from these mechanisms, including localized impacts in communities that are already adversely impacted by air pollution" (Nunez 2006).

Although *disadvantaged community* was not defined in AB 32, subsequent legislation Senate Bill 535 (SB 535) – *Global Warming Solutions Act of 2006: Greenhouse Gas Reduction Fund* (De Leon, Chapter 830, Statutes of 2012) – provided both a clear direction and proposed factors for consideration in identifying disadvantaged communities such as those "based on geographic, socioeconomic, public health, and environmental hazard criteria, and may include, but are not limited to, either of the following:

(a) Areas disproportionately affected by environmental pollution and other hazards that can lead to negative public health effects, exposure, or environmental degradation.
(b) Areas with concentrations of people that are of low income, high unemployment, low levels of homeownership, high rent burden, sensitive populations, or low levels of educational attainment." (De Leon 2012)

Thus, in California, in addition to traditional risk assessment, a community or place-based CI assessment has been developed. This approach augments the traditional concept of "risk" with the inclusion of a broader concept, "impact." Risk indicates a largely quantifiable approach to assessment, whereas impact implies a broader scope of both quantitative and semiquantitative information, including nonchemical stressors (Alexeeff et al. 2012).

This chapter focuses on place- or community-based cumulative impact assessment in the context of integrating pollution burdens and health vulnerabilities with psychosocial nonchemical stressors. For the purposes of this chapter, any discussion of *cumulative impacts* aligns with CalEPA's definition. The scope of discussion topics includes departures from traditional risk assessment, differences between risk and impact and their assessment methodologies, environmental health and/or justice-focused screening tools, community expectations for assessors, and future directions. The central focus of this chapter is to understand how the concept of community-based cumulative impacts has been successfully integrated in CalEnviroScreen. CalEnviroScreen is used to effectively characterize and combine measures of impact that are of greatest concern and contribute to cumulative impacts in communities across the state. This approach has enabled CalEPA to target multimedia enforcement action, prioritize areas for investment in emission reduction programs, and assist CalEPA and local entities with planning community engagement and outreach efforts.

18.1.1 Traditional Risk Assessment

Traditional risk assessment (TRA) is a predominantly quantitative approach that evaluates a source and/or chemical(s) on the primary steps of hazard identification, dose-response assessment, exposure assessment, and risk characterization (NAS 1983; Faustman and Omenn 2008). This approach is widely applied and has been instrumental in identifying and reducing both human and environmental health risks by (1) evaluating sources or chemicals to estimate cancer and non-cancer risk levels, (2) controlling media-specific exposures (e.g., chemicals in drinking water), and (3) creating decision-making processes that establish risk thresholds to minimize the amount of emissions or discharges of chemicals from a specific source (U.S. EPA 1991, 1992, 1996, 2005a, b). However, the TRA approach has a limited ability to account for sensitivities of subpopulations beyond those based on physiologic characteristics, such as children and the elderly (Miller et al. 2002; Alexeeff and Marty 2008). Additionally, TRA requires specific knowledge of exposures, including chemical characterization, dose levels, and routes of exposure. An understanding of these parameters is essential to establishing health guidance values or benchmarks of harm for individual chemicals (Salmon 2010).

18.1.2 Community-Based Cumulative Impact Assessment

Community characteristics, including area-specific information (e.g., water quality, pesticide use), proximity to multiple nearby pollution sources, and socioeconomic or health vulnerability, cannot be readily incorporated into the traditional paradigm. Risk assessments conducted for regulatory purposes at individual facilities or sites

may include some area-specific considerations, including community notifications for site cleanups or facility permitting, but these factors are incorporated in a very limited context. TRA is a quantitative methodology that relies heavily on scientific data, including well-characterized exposure levels and dose-response relationships for environmental contaminants (NAS 1983). Even with robust data, these traditional approaches are useful in estimating the risk to individuals but are not well-suited to provide an estimate of cumulative impacts confronting a community in a specific location (ATEB 2008, 2009).

With the increasing concern for exposures to multiple pollutants from multiple sources, assessors are often tasked with evaluating highly complex scenarios with significant data gaps. An example of such an exposure scenario would be a mixture of chemicals emitted from a single site (e.g., oil refinery), combined with emissions from local factories and road traffic. Data gaps include poor characterization of the environmental contaminants, and little understanding of how these multiple contaminants interact with humans and the environment in a specific area, or the relative contributions of existing and emerging sources (ATEB 2008, 2009; Lee et al. 2011). Consideration of these factors, combined with vulnerability factors in the local community, such as source proximity to schools, hospitals, or elder care facilities, set the foundation for developing methodologies to perform assessments at the community level (Dunn and Alexeeff 2010).

Thus, the community-based concept establishes a framework for designing tools that allow assessors and decision-makers to identify communities that are disadvantaged with regard to environmental and personal health. Such communities include those areas and populations disproportionately burdened by pollution, as influenced by both intrinsic biological (e.g., age, genetic characteristics, preexisting health conditions, sex) and extrinsic socioeconomic factors (e.g., socioeconomic status, education, race/ethnicity, access to health care, housing) (Gee and Payne-Sturges 2004). Considering these nonchemical stressors in the context of environmental justice is a critical first step that enables regulatory agencies to evaluate and address community-based concerns and meet expectations to consider cumulative impacts in decision-making (Alexeeff et al. 2012). Additionally, engaging community members, including local decision-makers, to participate in and understand key elements of the assessment process may be essential to positive public health outcomes (Hallgren et al. 2014). Community outreach and education can facilitate communication, risk reduction strategy development, and chemical source identification (Dunn and Alexeeff 2010; McCloskey et al. 2011; Abara et al. 2014).

18.1.3 It's Impact, Not Risk

Often, the terms *risk* and *impact* are used synonymously, suggesting that they describe the same outcome. The term *risk* means a chance of injury or loss. Historically, in the two hemispheres of human and environmental health, risk entails a quantifiable approach to assessment that includes a wide spectrum of assumptions,

modeling, uncertainties, and extrapolation to fill data gaps (NAS 1983). Such assessments are useful in estimating the risk to a population, based on theoretical exposure paradigms estimated for a "central tendency exposure" for a "maximally exposed individual," and are only feasible with contaminants or chemicals that are well-characterized with respect to exposure levels and their dose-response relationships (U.S. EPA 1989). However, the data required to adequately characterize the large number of sources of environmental contaminants in a community cannot be easily generated and may not be practical in the foreseeable future (Faust 2010). These limitations have hindered agencies at the local, regional, and state levels when initiating actions to achieve environmental justice since cumulative risk cannot be ascertained in a given community or a specific area. Hence, multiple institutions are pursuing alternate approaches to evaluate CI (OEHHA 2010). *Impact* is interpreted to mean potential effects or influences of stressors or sources that do not necessarily result in an identifiable level of injury or loss, but are known to have an influence.

Risk Versus Impact
Risk indicates a largely quantifiable approach to assessment of injury or loss, whereas impact implies a broader scope of both quantitative and semiquantitative factors that enhances the risk.

18.1.4 Cumulative Impact Assessment Tools

Community-based cumulative impact assessment approaches use scientifically justifiable, quantitative, and semiquantitative methods that permit comparisons between communities or census tracts. Current methods, including CalEnviroScreen, facilitate the relative ranking of communities with scoring systems that also allow comparisons between communities with the same score to understand the relative contributions of individual indicators representing factors that influence the cumulative impact in a community. This ability to prioritize or rank communities based on cumulative impact indicators enables assessors to more effectively represent the complex relationships between health outcomes, psychosocial stressors, and environmental exposures (Alexeeff et al. 2012).

18.1.4.1 CalEnviroScreen

The California Communities Environmental Health Screening Tool, abbreviated CalEnviroScreen, was developed by the California Environmental Protection Agency's (CalEPA) Office of Environmental Health Hazard Assessment (OEHHA) as a science-based tool for evaluating the cumulative impacts of multiple pollutants and stressors in communities (Alexeeff and Mataka 2014). The working tool reflects stakeholder input and the collaborative efforts of OEHHA and the

Cumulative Impacts and Precautionary Approaches Work Group, a collective of representatives from the private, academic, nongovernmental, and government sectors (CalEPA 2005; OEHHA 2014a).

In support of CalEPA's environmental justice mission, CalEnviroScreen assists the Agency and its departments by identifying those communities disproportionately burdened by cumulative impacts. Identifying these vulnerable communities helps the Agency and its departments to support the fair treatment of all Californians. CalEnviroScreen analyses:

- Aid decision-makers in making determinations about administering environmental justice grants.
- Inform targeted environmental law compliance and enforcement initiatives.
- Provide insight on potential implications of department activities and decisions.
- Help decision-makers prioritize site-cleanup activities and identify opportunities for sustainable economic development in heavily impacted neighborhoods (OEHHA 2014a).

Beyond its valuable uses in CalEPA, CalEnviroScreen potentially could be adapted by local and regional governments to include more precise data sets, for example, those from air and water districts or transit agencies, to facilitate community planning, engagement, and outreach efforts. CalEnviroScreen interactive maps are available on OEHHA's website. Results can be filtered by location, individual indicator, or class of indicators (i.e., pollution burden or population characteristics).

CalEPA describes CalEnviroScreen as a model that "is place-based and provides information for the entire State of California on a geographic basis." The geographic scale selected is intended to be useful for a wide range of decisions" (OEHHA 2014a). The model is comprised of two key components and four subcategories as follows: pollution burden (exposures and environmental effects) and population characteristics (sensitive populations and socioeconomic factors) (see Fig. 18.1). A suite of statewide indicators that describe pollution burden and population characteristics are assigned to each subcategory. CalEnviroScreen is a fairly simple model with a limited set of indicators. Each indicator in a given area is assigned a score that is weighted according to a scoring system. The sum of pollution indicator scores (maximum value of 10) is multiplied by the sum of population characteristic

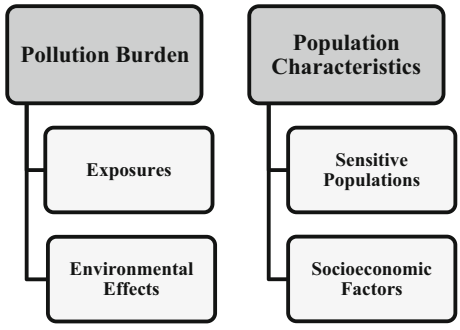

Fig. 18.1 CalEPA CalEnviroScreen tool components (OEHHA 2014a)

indicator scores (maximum value of 10) to produce a final CalEnviroScreen score with a maximum of 100. This score permits ranking of all places evaluated throughout the state relative to each other, a concept that will be discussed in more detail later in the chapter (OEHHA 2014a).

18.1.4.2 Additional Environmental Health Screening Methods

Considering cumulative impacts at the local or regional level is a practice that is gaining popularity among many decision-makers because most planning and permitting decisions take place on a local scale (Johnson Thornton et al. 2013; Corburn 2015). CI assessment leads to more informed decision-making by adding another layer of information to traditional risk assessment. Decision-makers at the statewide, regional, and community levels can utilize environmental health screening methods to guide their decision process and weigh potential impacts within a specific area or community. In the following section, we briefly describe additional approaches used to assess community-based cumulative impacts.

Environmental Justice Screening Method (EJSM) The University of Southern California Program on Environmental and Regional Equity (PERE) received a research contract from the California Air Resources Board, to develop an Environmental Justice Screening Method (EJSM) (Sadd et al. 2011). The EJSM is described as a screening approach and not as a tool because of its flexibility to include or exclude indicators or metrics, such as climate vulnerability or drinking water quality, in a given scenario (Pastor et al. 2013). EJSM incorporates data from approximately 30 metrics to generate geographic information system (GIS)-based maps of communities at the census tract scale, similar to CalEnviroScreen (Sadd et al. 2014).

The mapping approach utilizes spatial polygons that denote land use within a neighborhood such as residences, schools, health-care facilities, and playgrounds. The metrics are categorized and scored on a scale of 1 to 5 in consideration of (1) proximity to hazards, such as chrome platers and industrial emission sites; (2) air quality and estimated health risk measures, such as relative cancer risk or ambient concentration rates of ozone and particulate matter; and (3) social vulnerability measures such as poverty, race, age, home ownership rate, and birth outcomes within a community (English 2013; Sadd et al. 2011). EJSM scoring differs from CalEnviroScreen because it does not have a multiplier in the model and all indicators are weighed equally. GIS maps for the eight EJSM California regions with versions for both cumulative impact scores and select component layers are publicly available on PERE's website.

Cumulative Environment Vulnerabilities Assessment (CEVA) The University of California Davis Center for Regional Change (CRC) developed CEVA as a screening tool with the primary aim of providing a suitable framework for evaluating place-based cumulative environmental hazards that can effectively support decision-makers and environmental justice advocates in developing policy and allocating resources that

assist environmentally vulnerable communities (Huang and London 2012). Similar to the EJSM, CEVA distributes pollution and population metrics into three indices or categories labeled as (1) environmental hazards that include toxic release inventory sites and refineries; (2) social vulnerability that includes locations of health-care facilities, race, and education level; and (3) health effects that include low birth weight and asthma hospitalization rates (Huang and London 2012). Each index generates a score with the higher scores indicating those communities within a census block group that are most vulnerable to adverse environmental or hazard effects. CEVA, as with the earlier versions of CalEnviroScreen, utilizes data at the ZIP code scale for some measures. Interactive Regional Opportunity Index maps are available on CRC's website to assist decision-makers in identifying regions with disproportionately disadvantaged communities.

Similar to CalEnviroScreen and EJSM, CEVA generates a spatial analysis that illustrates place-based findings that allow communities to be ranked relative to one another. CEVA's goal was to account for "both the highest concentrations of cumulative environmental hazards and the fewest social, economic and political resources to prevent, mitigate, or adapt to the conditions" (Huang and London 2012). CEVA was initially developed with a focus on Central California and Eastern Coachella Valley communities selected for their diversity in agriculture, socioeconomic status, education, language, political influence, and hazard sources (London et al. 2011, 2013).

EJSCREEN The U.S. Environmental Protection Agency created the EJSCREEN tool to assist EPA staff and managers in considering environmental justice issues. EJSCREEN uses nationwide data sets and methods to "screen for areas that may be candidates for additional consideration, analysis, or outreach as the agency develops programs, policies and activities that may affect communities" (U.S. EPA 2014). Similarly, EJSCREEN uses information at the census block group or user-defined area level and considers both demographic and environmental indicators. EJSCREEN generates an EJ index or summary of demographic information combined with a single environmental indicator (e.g., air toxics respiratory hazard). These indices generate maps, charts, and reports using a web interface. EJSCREEN contains many different environmental indicators, but only one environmental indicator is evaluated at a time in a given scenario, limiting its capacity for evaluating cumulative impacts from multiple environmental indicators. EJSCREEN is publically available, and its interactive tool can be accessed at www.epa.gov/ejscreen.

U.S. EPA continues to provide guidance for national, state, and local agencies for considering and implementing environmental justice actions in planning and decision-making. Entities such as the Federal Interagency Working Group on Environmental Justice National Environmental Policy Act (NEPA) Committee, whose members represent federal agencies subject to NEPA, aspire to design and optimize best practices for addressing environmental justice issues (U.S. EPA 2013). In addition to California, New Jersey and other states are building on U.S. EPA's example to form commissions and develop tools to facilitate the consideration and

implementation of environmental justice conscious policies, such as the New Jersey Smart Growth and Environmental Justice State Planning Commission and the interactive NJ-GeoWeb environmental information tool (New Jersey 2014, 2016).

18.1.4.3 Limitations of Screening Tools

Overall, environmental health screening approaches demonstrate how the data from multiple sources can be combined and characterized to make comparisons between different geographic areas and provide helpful insights into identifying "disadvantaged communities." Evaluating information at the census tract scale, both in the context of cumulative and individual metrics, allows decision-makers to consider area- or community-specific actions that would reduce the pollution burden or decrease the vulnerability in a community. Inherent limitations to these approaches vary with the degree of accuracy, precision, and uncertainty associated with the data for each of the indicators. As tools improve and more robust data sets become available, it may be possible to reduce uncertainty by applying additional statistical analyses. This concept is similarly applied to traditional risk assessment whereby more sophisticated tools, such as benchmark dose modeling of dose-response that facilitates understanding of response levels at low doses, continue to improve and overcome current analytical limitations (U.S. EPA 2012). Other approaches to characterizing these limitations should also be explored. One current limitation is that some areas in a state or a county may have more and better quality data than others, requiring approximation or modeling to fill the data gaps. An example of this would be drinking water quality monitoring data. Densely populated areas tend to have more sophisticated drinking water systems with enhanced monitoring and quality control measures to detect contamination. More rudimentary systems or individual well sites often serve less populated areas and have very limited capabilities for monitoring drinking water contamination. Several California governmental agencies maintain databases that provide and inform decision-making tools like CalEnviroScreen.

In spite of these constraints, evidence suggests that impact assessment tools are highly beneficial in distinguishing higher-impacted from lower-impacted communities, in identifying factors that are the primary contributors to the community's cumulative impact, and in assisting regulatory agencies in allocation of resources and more effective prioritization of area-specific mitigation efforts. Evaluation of the accuracy of these tools and the value of the results is ongoing. One example of this is with EJSM and the Los Angeles Collaborative for Environmental Health and Justice (LACEHJ 2010). This cooperative of community organizations and academic researchers serves as a "frontline" team that assesses the merits and limitations of applying the Environmental Justice Screening Method in communities throughout the Greater Los Angeles Area (Sadd et al. 2011, 2014). CalEPA and OEHHA continue to hone and evaluate CalEnviroScreen, soliciting stakeholder input throughout the process. In this chapter, the focus is on CalEnviroScreen as a model screening tool because it encompasses a robust number of indicators, includes

communities across the state, and is currently used by decision-makers within the California government.

18.2 CalEnviroScreen: California Communities Environmental Health Screening Tool

18.2.1 Design Factors and Considerations: Modeling Environmental Justice Concepts

CalEnviroScreen is a tool that combines multiple sets of data on pollutants and stressors in a geographic area to screen for places with the highest cumulative burdens. The tool creates one combined measure, the CalEnviroScreen score, for visualizing geographies in California that are most burdened. This combined index, as well as the underlying data sets, is made publicly available through OEHHA's website. Users of the tool can view the information as both static and interactive web maps and can download the results in various formats. The tool is not updated continuously but rather represents a snapshot of the data at the time of the release. Each version of CalEnviroScreen is the product of extensive public input and reflects the concerns of many stakeholders in California, including community-based organizations and the general public. Users, however, cannot add data to CalEnviroScreen after a version is released, but can submit feedback on additional data sets or gaps that may be addressed in the next revision. OEHHA updates CalEnviroScreen as additional relevant, statewide data sets emerge.

The early CalEnviroScreen versions (1.0 and 1.1) utilized data organized by ZIP code and included fewer indicators. The 2.0 version analysis[1], released in October 2014, contains additional indicators and now analyzes community data at the census tract scale because census tract data (approximately 8000 tracts in California) provides a finer scale of resolution for many California regions (U.S. Census Bureau 2010). Tracts are comprised of multiple block groups that contain several blocks each, with a block being the smallest geographic unit for which population data are available. In California, not all census blocks are populated. Independent of the version, CalEnviroScreen (OEHHA 2014a):

- "Produces a relative, rather than absolute, measure of impact.
- Provides a baseline assessment and methodology that can be expanded upon and updated periodically as important additional information becomes available.

[1] A subsequent version of CalEnviroscreen (3.0) with additional indicators and some modifications has been released since this chapter was authored. CalEnviroScreen 3.0 can be accessed at https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-30.

- Demonstrates a practical and scientific methodology for evaluating multiple pollution sources and stressors that takes into account a community's vulnerability to pollution."

The next section expands on the CalEnviroScreen methodology, including indicator selection, indicator scoring, and the relative ranking scheme. These indicators model *stressors* or factors that contribute to the pollution burden or vulnerability within a community. Indicator selection and the data that accompany these indicators determine the total CalEnviroScreen score. The final CalEnviroScreen scores provide the basis for the ranking scheme that ultimately models the community-based cumulative impact.

18.2.2 Indicator Selection: Translating Environmental Justice Concepts into Operation

18.2.2.1 Indicator and Component Scoring

CalEnviroScreen indicators are selected based on two general considerations, (1) "information that will best represent statewide pollution burden and population characteristics" and (2) "the availability and quality of such information at the necessary geographic scale statewide" (OEHHA 2014a). These indicators are proxies for the characteristics they model. CalEnviroScreen models California communities at the census tract scale, so indicator data should be available statewide and translate to census tracts. This approach poses considerable challenges for assessors to evaluate those regions with significant data gaps for a potential indicator of interest. Hence, it is important to select data sets that are as accurate, complete, and current as possible at the state level.

The following is an overview of the indicator selection and scoring process (OEHHA 2014a):

1. "Identify potential indicators for each component.
2. Find sources of data to support indicator development.
3. Select and develop indicator, assigning a value for each geographic unit.
4. Assign a percentile for each indicator for each geographic unit, based on the rank-order of the value.
5. Generate maps to visualize data.
6. Derive scores for pollution burden and population characteristics components.
7. Derive the overall CalEnviroScreen score by combining the component scores.
8. Generate maps to visualize overall results."

CalEnviroScreen is applied to the entire state, but it is worth emphasizing that modeled data sets provide a "broad environmental snapshot of a given region" (OEHHA 2014a). A specific indicator, such as toxic cleanup sites, may be a robust marker of pollution burden, but any given region may not have any toxic cleanup

sites. In such cases, this indicator is scored as zero. Alternatively, when there are not enough data to conclusively identify the presence or absence of an indicator in a specific area, such as the lack of an air monitoring station within a certain distance, it is removed from the calculation, and no score is assigned for that indicator. Next, census tract indicator raw values above zero are ordered from highest to lowest values. These ordered values are used to calculate a percentile for all areas that have a score.

Generally speaking, the percentile indicator for a select geographic area describes the percentage of California with lower values for that indicator. For example, a 75th percentile for that indicator or suite of indicators means a select geographic area is higher or more impacted compared to 75% of all other geographic areas in California. The magnitude of difference between two or more areas cannot be calculated from the difference in percentiles because of the shape of the distribution of the data. For example, the difference between the 75th and 50th percentile may not be the same as the difference between the 50th and the 25th percentile.

Pollution Burden Indicators Gathering information about direct environmental exposures poses a significant challenge as such data sets are limited and not readily available on a statewide level. Evaluating how individuals or populations come in contact with chemicals from air, water, food, or soil sources can be indirectly modeled by considering data sets relating to pollution sources, releases, and environmental concentrations. CalEnviroScreen takes this approach and includes seven *exposure indicators*: ozone concentrations in air, $PM_{2.5}$ concentrations in air (particulate matter or particles with a diameter measuring less than 2.5 microns), diesel particulate matter emissions, certain high-hazard/high-volatility pesticide use, toxic releases from facilities, traffic density, and drinking water contaminants (see Table 18.1).

When evaluating environmental effects, it is important to consider several concepts. Effects reflect a process, whether immediate or delayed, and can include environmental degradation, ecological system changes, and human lifestyle or activity changes for individuals or populations (Fan et al. 2010; Howd 2010). Communities and the environment can experience a myriad of effects when physical, biological, and chemical pollutants are released into the environment (Alexeeff et al. 2012). These effects vary by the nature, degree, and prevalence of environmental harm. Whether directly impacted through contact exposure or indirectly affected by shifts in routine practices, including restricted swimming or fishing in local waterways or changes in local traffic patterns, environmental effects can lead to elevated stress that results in adverse human health impacts (Gee and Payne-Sturges 2004). CalEnviroScreen incorporates the following five indicators to model *environmental effects*: toxic cleanup sites, groundwater threats from leaking underground storage sites and cleanups, hazardous waste facilities and generators, impaired water bodies, and solid waste sites and facilities (see Table 18.1).

Population Characteristic Indicators The process of identifying sensitive populations with increased vulnerability to the effects of pollution can be

Table 18.1 CalEPA CalEnviroScreen pollution burden indicators

	Indicator	Description
Exposures	$PM_{2.5}$ concentrations	Annual mean concentration of $PM_{2.5}$ over 3 years (2009–2011)
	Ozone concentrations	Daily maximum 8-h ozone concentration over the California 8-h standard (0.070 ppm), averaged over 3 years (2009 to 2011)
	Diesel PM emissions	Diesel PM emissions from on-road and non-road sources for a 2010 summer day in July (kg/day)
	Drinking water contaminants	Drinking water contaminant index for selected contaminants
	Pesticide use	Pounds of selected active pesticide ingredients used in production-agriculture per square mile
	Toxic releases from facilities	Toxicity-weighted concentrations of modeled chemical releases to air from facilities
	Traffic density	Vehicle-kilometers per hour divided by total road length (kilometers) within 150 meters of the census tract boundary
Environmental effects	Cleanup sites	Sum of weighted DTSC* cleanup sites
	Groundwater threats	Sum of weighted SWRCB** groundwater cleanup sites
	Hazardous waste facilities and generators	Sum of weighted permitted hazardous waste facilities and large quantity hazardous waste generators
	Impaired water bodies	Sum of number of pollutants from water bodies designated as impaired
	Solid waste sites and facilities	Sum of weighted solid waste facilities

*Data acquired from the Department of Toxic Substances Control
**Data acquired from the State Water Resources Control Board

challenging. Within a given area, factors such as health status and age can predispose individuals to adverse health outcomes and vary widely, independent of pollution (August et al. 2012; English 2013). CalEnviroScreen incorporates three indicators that may suggest increased health vulnerabilities associated with toxic chemical exposures. Robust data sets are available statewide for the following three *sensitive population indicators*: prevalence of children and elderly populations, asthma emergency department visit rates, and the rate of low-birth-weight infants (see Table 18.2).

Emerging research supports the finding that socioeconomic status, including education level and employment status, is a significant factor in gauging the vulnerability of populations to pollutants (LACEHJ 2010; English 2013). Language barriers, prevalence of individuals with less than a high school education, and disproportionate unemployment rates can reduce a community's ability to adapt to or cope with pollution (LACEHJ 2010; Ramey et al. 2015). CalEnviroScreen integrates four socioeconomic factors that link pollution with adverse health impacts. *Socioeconomic factor indicators* include educational attainment, linguistic isolation, poverty, and unemployment (see Table 18.2).

Table 18.2 CalEPA CalEnviroScreen population characteristic indicators

	Indicator	Description
Sensitive populations	Age (children and elderly)	Percentage of the population under age 10 or over age 65
	Asthma emergency department visit rate	Age-adjusted rate of emergency department visits for asthma per 10,000, spatially modeled (2007–2009)
	Low-birth-weight rates	Percentage of low-birth-weight infants under 2500 grams, spatially modeled (2006–2009)
Socioeconomic factors	Educational attainment	Percentage of the population over age 25 with less than a high school education
	Linguistic isolation	Percentage of households in which no one age 14 and over speaks English "very well" or speaks English only
	Poverty	Percentage of residents below two times the national poverty level
	Unemployment	Population over age 16 that is unemployed and eligible for the labor force

Collectively, CalEnviroScreen integrates these seven exposures and five environmental effect indicators to model relative pollution burden impacts and three sensitive population and four socioeconomic factor indicators to model relative population characteristics. The methodology and rationale for each specific indicator are described in detail in the CalEnviroScreen document *California Communities Environmental Health Screening Tool, Version 2.0 (CalEnviroScreen 2.0) Guidance and Screening Tool* (OEHHA 2014a).

18.2.2.1.1 CalEnviroScreen Score and Maps

The final CalEnviroScreen score is the product of the indicator value of the pollution burden and the indicator value of the population characteristics (see Fig. 18.2). The pollution burden component is composed of seven exposure and five environmental effect indicators. The environmental effect indicator values are multiplied by one-half (noted by an asterisk *) to weight them half as much as the exposure indicators because exposure sources generally contribute more than environmental effects to total pollution impact (OEHHA 2014a) (see Fig. 18.2 and Table 18.1). The population characteristic component is comprised of three sensitive population and four socioeconomic factor indicators with all seven indicators weighted equally (see Fig. 18.2 and Table 18.2).

The final scores for both components are calculated as follows (OEHHA 2014a):

1. Average the percentiles for all individual indicators in a group (group: exposure and environmental effects). Environmental effects are weighted half as much as exposure indicators, making the pollution burden a weighted average.

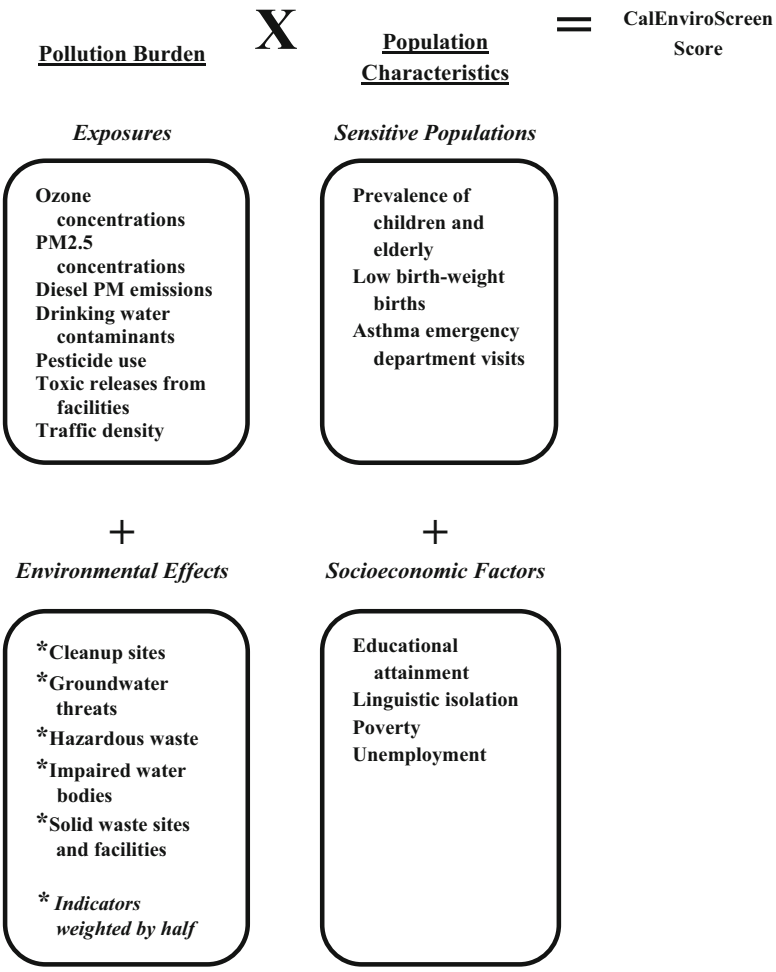

Fig. 18.2 CalEPA CalEnviroScreen model

2. Pollution burden and population characteristic percentile averages are scaled with a maximum value of 10 and a range of 0.1–10. Each average is divided by the maximum value observed in the state and multiplied by 10. Scaling ensures that the pollution and population components contribute equally to the final CalEnviroScreen score.
3. The final CalEnviroScreen score for an area is calculated as the final pollution burden score multiplied by the final population score with a possible total of 100. This final CalEnviroScreen score for each area is then used to rank all the areas from highest to lowest, based on their overall score. The percentile for the overall score is calculated. Geographic maps are generated to illustrate the percentiles for

all census tracts statewide. Highest ranking percentiles are generally brightly colored to indicate the area of greatest impact.

18.2.2.1.2 Uncertainty and Error

Even with careful data set selection, assessors must account for uncertainties. Such uncertainties can develop for any number of reasons, including database gaps or inaccuracies, changing environmental conditions over time, and the limited capacity of selected indicators to measure outcomes or exposures of interest. Despite these uncertainties, CalEnviroScreen remains a powerful tool in identifying those communities most adversely impacted due to its ranking function, particularly when modeling data sets where the highest or lowest 15–20% of communities is of great interest.

Identifying Community Profiles and Key Drivers By taking a look at the individual component and indicator scores, one can understand the similarities and differences between two communities having similar scores. Communities can have nearly equivalent overall scores but be comprised of vastly different scoring for pollution burden and population characteristic profiles. For example, a census tract in Lamont, near Bakersfield in the Central Valley, and a census tract in Long Beach in the Los Angeles region, both have overall scores of 48, placing them among the top 10 percent of the most impacted census tracts in California (see Fig. 18.3 and Tables 18.3 and 18.4).

The Lamont tract has very high scores for ozone, particulate matter, drinking water contaminants, and pesticides while scoring only moderately among the other pollution burden indicators. In contrast, the Long Beach tract has very high scores for diesel, toxic releases, traffic density, groundwater threats, and impaired water bodies while scoring only moderately for indicators for which Lamont scored highly. The Long Beach tract scores slightly higher for the overall pollution score, while the Lamont tract scores higher for the overall vulnerable population and socioeconomic score. The two components combine to yield a very similar overall score, meaning that the two tracts are viewed as equally disadvantaged in CalEnviroScreen.

A third census tract in Richmond near the San Francisco Bay Area, compared here, demonstrates that despite scoring slightly lower in the overall pollution score when compared to the other two tracts, a very high population characteristic component still yields a relatively high overall score. The Richmond tract scores highly in the diesel indicator as well as for several environmental effect indicators while scoring extremely high in the vulnerable population and socioeconomic indicators. The overall score of 45 (compared to the other two with a score of 48) places the tract among the top 15 percent of the most impacted census tracts in California (Fig. 18.3).

Fig. 18.3 Census tracts with similar CalEnviroScreen scores

18.2.3 Applying Community-Based Concepts to Decision-Making

CalEnviroScreen was developed through a highly public and interactive process that aligns well with the U.S. EPA's *Guidance on Considering Environmental Justice*

Table 18.3 Identifying major drivers from CalEnviroScreen scores in three census tracts

Location	Lamont, Kern County	Long Beach, Los Angeles County	Richmond, Contra Costa County
Census tract	(6029006401)	(6037572201)	(6013379000)
Population	8,320	6,197	6,117
CalEnviroScreen score	*48.14* (91–95th percentile)	*47.93* (91–95th percentile)	*45.49* (86–90th percentile)
Pollution burden	*Medium-high* (78th percentile)	*Very-high* (92nd percentile)	*Medium* (57th percentile)
Population characteristics	*Very-high* (90th percentile)	*Medium-high* (74th percentile)	*Very-high* (98th percentile)
Main drivers (≥80th percentile)	Ozone, $PM_{2.5}$, drinking water, pesticides, education, linguistic isolation, poverty	Diesel, toxic releases, traffic density, groundwater threats, impaired water, asthma, low birth weight	Diesel, cleanup sites, groundwater threats, hazardous waste, impaired water, asthma, low birth weight, education, poverty

During the Development of Regulatory Actions (Alexeeff and Mataka 2014; U.S. EPA 2015). CalEPA and OEHHA held multiple meetings with stakeholders which included community and environmental justice organizations, academia, other government agencies, and industry groups, then released interim CalEnviroScreen drafts for public comment, and conducted a dozen workshops that solicited extensive written and oral comment feedback (OEHHA 2013; OEHHA 2014b). During the conceptual phase of CalEnviroScreen's development, CalEPA and OEHHA began devising general principles that gauge and strategize efforts in the context of assessing chemical hazards from multiple sources within communities. Many studies, including individual community-based studies, served as a training ground for honing both the principles and practices of community-based cumulative impact assessment (Dunn and Alexeeff 2010).

18.2.3.1 Community-Based Studies in Decision-Making

Four general principles were derived from examining several case studies and are discussed in detail by Dunn and Alexeeff (2010). These principles can be summarized as follows: (1) consider exposure patterns and cultural practices, (2) identify populations with increased susceptibility, (3) understand the cumulative impacts, and (4) involve the community in all phases of an assessment (Dunn and Alexeeff 2010). These guiding principles were drawn from an evaluation of four diverse case studies. These included a study of traffic-related air pollution and childhood respiratory diseases around San Francisco Bay Area schools, sport fishing advisories related to chemical contamination of fish in general and in specific water bodies throughout California, a risk assessment of a chromium "hot spot" in a poor Latino

Table 18.4 Percentile ranking of individual indicators in three census tracts

LOCATION	Lamont, Kern County	Long Beach, Los Angeles County	Richmond, Contra Costa County	
Ozone	**95**	0	0	POLLUTION BURDEN
$PM_{2.5}$	**99**	67	17	
Diesel	21	**93**	**86**	
Drinking Water	**100**	24	3	
Pesticides	**95**	29	0	
Toxic Releases	25	**95**	77	
Traffic Density	9	**86**	46	
Cleanup Sites	76	56	**98**	
Groundwater Threats	16	**88**	**90**	
Hazardous Waste	0	39	**89**	
Impaired Water	0	**90**	**86**	
Solid Waste	0	73	0	
Age	66	73	73	POPULATION CHARACTERISTICS
Asthma	36	**81**	**98**	
Low Birth Weight	71	**94**	**98**	
Low Education	**98**	55	**82**	
Linguistic Isolation	**94**	44	75	
Poverty	**94**	40	**80**	
Unemployment	60	60	77	

community of San Diego known as Barrio Logan, and a study of lead exposure in Latino children throughout San Diego County.

In the sport fishing advisories case, OEHHA approached the development of advisories with an awareness of cultural practices that may increase the risks of exposure and adverse health effects that arise from eating fish contaminated with chemicals, including methylmercury and polychlorinated biphenyls (PCBs)

(OEHHA 2001). Fish consumption is much greater among some minority populations, namely, Southeast Asians, and low-income subsistence sport fishers, groups that rely on sport fishing as a major source of dietary protein (Dunn and Alexeeff 2010). These groups may also engage in practices that increase the risk for exposure, including consumption of the entire fish (OEHHA 2001). Subpopulations within these communities, including children and pregnant women, were identified as susceptible subpopulations with increased risk for adverse effects from exposure to multiple contaminants, such as methylmercury and PCBs, due to their harmful and cumulative effects on neural development (U.S. EPA 2004; Davis et al. 2012).

OEHHA evaluated the potential harmful effects from exposure to contaminants common in fish along with the health benefits of eating fish. OEHHA developed fish advisories for California sport fishers that provided guidance on fish cooking and preparation methods and recommendations for fish consumption, especially for sensitive groups such as children and pregnant women. To further enhance community outreach, OEHHA created signs and pamphlets in multiple languages to better inform communities that may not otherwise be aware of potential adverse health effects associated with eating contaminated fish. These recommendations help individuals reduce their exposure risk by modifying consumption practices based on the species, size, and number of fish consumed. The principles derived from this case study and the consideration of additional factors that influence the vulnerability within a population contributed significantly to the development of CalEnviroScreen, particularly when selecting population characteristic indicators.

18.2.3.2 CalEnviroScreen in State Regulatory Activities

Each case study highlights not only the diversity of exposure sources but also the complex factors that affect individual communities. Cultural practices and lifestyle, not just how close a population is to a pollution source, influence how and to what extent individuals within a community can be exposed (CDC 2002). Understanding the biological characteristics or types of preexisting conditions that increase the vulnerability of certain individuals to adverse pollution impacts helps to identify susceptible subpopulations within a community (de Fur et al. 2007; Medina-Ramon and Schwartz 2008; Zanobetti and Schwartz 2011). These concepts contributed to the development of CalEnviroScreen and helped focus its original purpose which was to assist CalEPA departments and the state of California in carrying out its environmental justice mission, and to continue to be a useful tool for this end.

In addition, as discussed in Alexeeff and Mataka (2014), CalEnviroScreen is a valuable resource in many additional ways. One important way CalEnviroScreen is being used is to identify disadvantaged communities for allocation of cap-and-trade funds generated under the Global Warming Solutions Act of 2006 (De Leon 2012). Of the total monies allocated from the Greenhouse Gas Reduction Fund, 25 percent "must go to projects that provide a benefit to disadvantaged communities," and a "minimum of 10 percent of the funds must be for projects located directly within disadvantaged communities" (CalEPA 2014). Another use of CalEnviroScreen is by

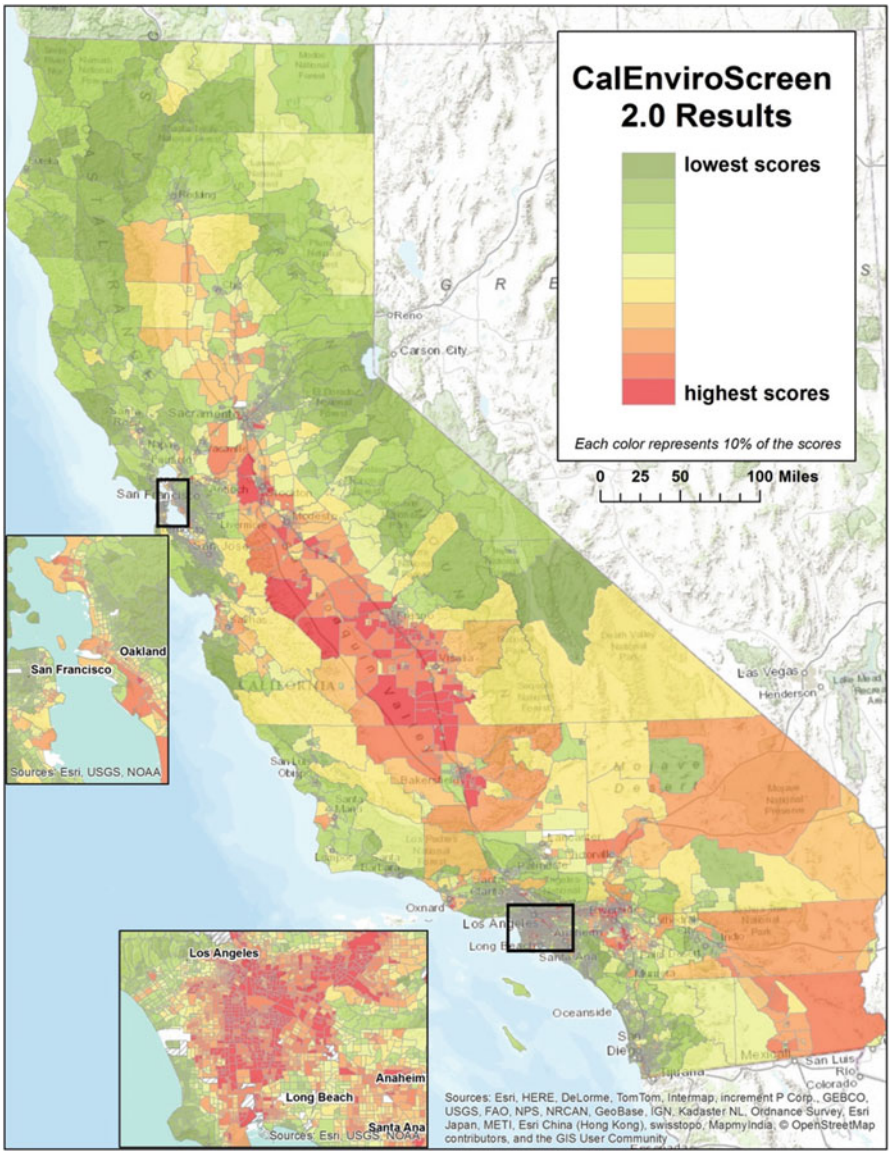

Fig. 18.4 CalEPA CalEnviroScreen statewide results

other state entities, including the Strategic Growth Council, who use CalEnviroScreen results to select communities where resources allotted under sustainable community grant funding can be most effectively distributed. An example of statewide CalEnviroScreen results is illustrated in Fig. 18.4. CalEnviroScreen facilitates collaboration between CalEPA departments, like OEHHA and the Air

Resources Board (ARB), in adapting monitoring and health benefit programs for those communities disproportionately impacted. These communities are highlighted by a specific indicator, such as air pollution hot spots (Dunn and Alexeeff 2010). Ultimately, the success of the state's application of CalEnviroScreen has led to inquiries from smaller regulatory entities about how the tool can be further scaled to provide relevant information to more effectively assist in decision-making at the city or county level. For example, decision-makers for the Greater Los Angeles Area can use CalEnviroScreen results for both population characteristics and pollution burden, such as those presented in Fig. 18.5, to identify which smaller communities may warrant a more refined scale of analysis.

18.3 Challenges and Next Steps: Future Directions in Community-Based Cumulative Impact Assessment

Risk assessment as currently practiced in environmental regulatory programs at the federal, state, and local levels is typically designed to evaluate a single contaminant or source, in one media type, and is based on the concept of risk thresholds that are considered either safe or acceptable (NAS 1983). An acceptable risk level is often set as a target and considers several factors associated with meeting the target level (ATEB 2008, 2009). These factors include evaluating the pollution control technology available in the foreseeable future, potential costs to the owners of the source, and costs subsequently passed on to the consumer. Thus, traditional quantitative risk assessment and the practices and policies that develop in response to assessment findings play a decisive role in our society. Risk assessment is evolving, particularly at the federal level. U.S. EPA, with guidance from the National Research Council Committee on Improving Risk Analysis, is broadening traditional concepts to improve both the utility and technical approaches used in risk analysis (NAS 2009). One key shift is to involve input from stakeholders early in the planning process. However, these expanded approaches still focus primarily on risk, not impact.

In contrast to the traditional assessment paradigm, people face scenarios with exposure to multiple contaminants from multiple sources. The resulting risks and impacts are also influenced by nonchemical factors and require additional approaches to integrate both chemical and nonchemical stressors. The relative ranking of communities, expressed as percentiles in CalEnviroScreen, provides a snapshot of the existing conditions, not a measure of potential risk. Increased pollution burden and poor socioeconomic status frequently go hand in hand. However, the underlying causes for this collinearity can differ significantly in different parts of state.

To design adequate and effective strategies to reduce the observed disparities, it is important to evaluate the causes that influence variability, depending on the location. Urban sprawl and zoning flaws can contribute to the formation of source clusters and resource limitations in some neighborhoods (LACEHJ 2010; Schwartz et al. 2015).

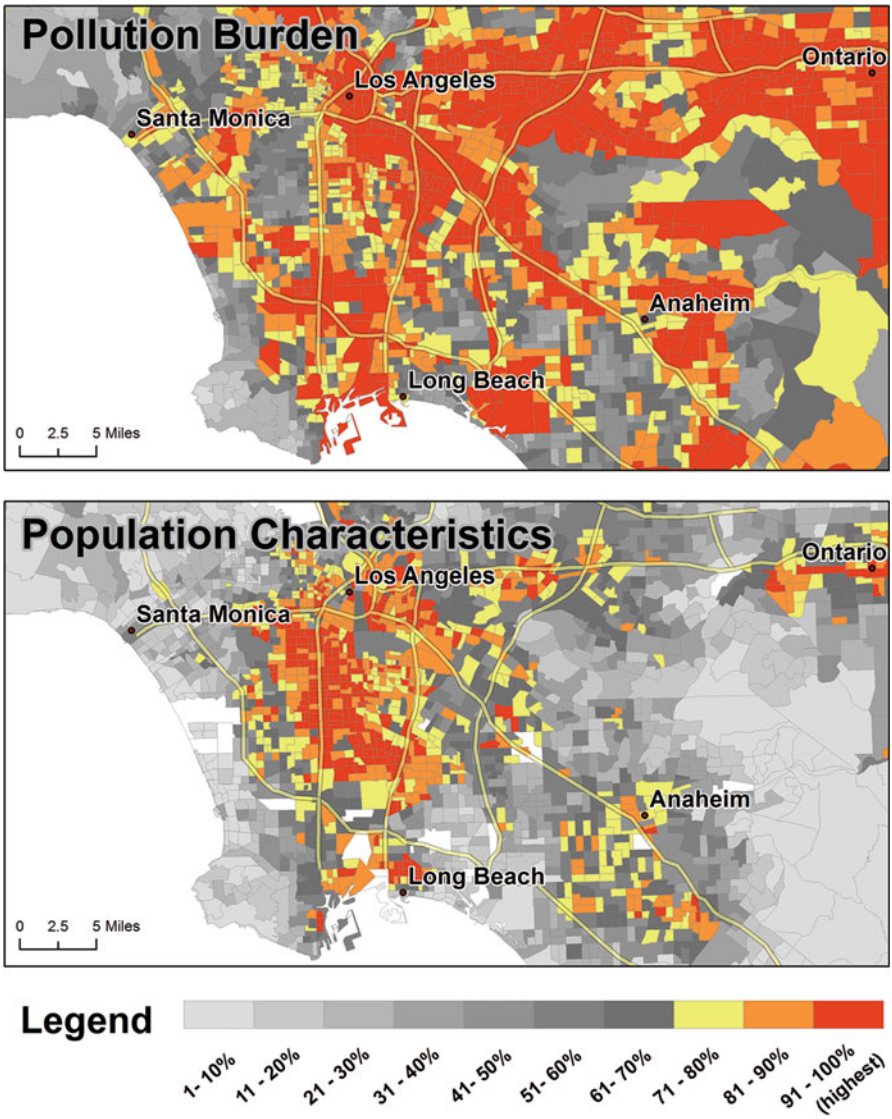

Fig. 18.5 CalEPA CalEnviroScreen results for (**a**) pollution burden and (**b**) population characteristics for the Greater Los Angeles Area

For example, past efforts to minimize expenses involved in freeway and major roadway expansion or building new roadways to meet the transportation needs of urban sprawl could have been responsible for the observed increased risk to people living near these structures. Similarly, major economic fluctuations could have led to gentrification that brings lower socioeconomic status population segments closer to sources (Pohanka and Fitzgerald 2004; Porebski et al. 2014; Shmool et al. 2015). In some multi-source clusters, individual sources may comply with the "safe" or "acceptable" set emission/discharge limit, but collectively the area or the community could be exceeding these safe or acceptable levels of risk or impacts (Batterman et al. 2014).

Although people living in such communities have demanded that cumulative impact assessments be included and considered in the context of siting, permitting, zoning, and other decision-making processes, researchers have stated that both regulatory agencies and legislative bodies have yet to take specific actions to move in that direction (Johnson Thornton et al. 2013; Corburn 2015). In addition, reluctance among some business and industry groups to support a move toward CI assessments often stems from the contention that such "redlining" could economically isolate or harm those communities (Pager and Shepherd 2008; Tso et al. 2011; Gase et al. 2014). In some instances, reconciling the realities of cumulative impacts with the potential scale of economic impacts involved to take remedial action seems to pose major challenges for any near-term legislative or regulatory action in the current political climate at the federal level. Yet, the bold step taken by the California legislature to incentivize investment in these disadvantaged communities with the allocation of cap-and-trade funds is noteworthy.

Methodological challenges that face CalEnviroScreen and other environmental health screening tools include (1) the influence of the number of indicators that are proxies for sources or media and how those are modeled, (2) capturing the strength of skewed data sets that are often associated with pollution levels and population characteristics, (3) evaluating how regional variations in cost of living may affect estimates of socioeconomic vulnerability, and (4) providing a format of quantitative information to track area-specific changes over time. Assessors confront these challenges in attempts to meet the expectations of communities throughout California. Through improved data quality, better statistical approaches, the addition of valid indicators, and constructive feedback from CalEnviroScreen users, this tool can be further modified and adapted to increase its utility.

The momentum to include cumulative impact assessment in the decision-making process is building across the country, and more methods are likely to evolve in the near future. Many are of the view that CI assessment at a local or regional level is critical since most of the growth planning, siting, and permitting decisions take place at the regional or local level (Johnson Thornton et al. 2013; Corburn 2015). A list of actions that can be considered in cumulatively impacted areas could include (1) requiring alternate buffer zone limits for new buildings from sources like refineries, landfills, oil and gas operations, agricultural lands, major roadways, and ports; (2) including permit conditions that limit the days, timing, or methods of pesticide application to reduce drift exposure; and (3) modifying area-specific risk

thresholds for new and existing sources, if necessary (Prasad and Murphy 2016). Thus, CI assessment provides an additional layer of information to traditional risk assessment, leading to more informed decision-making.

References

Abara, W., W. Sacoby, J. Vena, et al. 2014. Engaging a chemical disaster community: Lessons from Graniteville. *International Journal of Environmental Research and Public Health* 11: 5684–5697.

Alexeeff, A., and M. Marty. 2008. Protection of infants, children and other sensitive subpopulations. In *Risk assessment for chemicals in drinking water*, ed. R.A. Howd and A.M. Fan. Hoboken NJ: John Wiley.

Alexeeff, G.V., and A.Y. Mataka. 2014. CalEnviroScreen: A pathway to address environmental justice issues in California. *EM Mag* 1: 27–31.

Alexeeff, G.V., J.B. Faust, L.M. August, et al. 2012. A screening method for assessing cumulative impacts. *International Journal of Environmental Research and Public Health* 9: 648–659.

ATEB: Air Toxicology and Epidemiology Branch. 2008. *Air toxics "hot spots" risk assessment guidelines: Technical support document for the derivation of noncancer reference exposure levels*. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. http://oehha.ca.gov/media/downloads/crnr/noncancertsdfinal.pdf. Accessed 19 January 2018.

ATEB: Air Toxicology and Epidemiology Branch. 2009. *Air toxics "hot spots" risk assessment guidelines: Technical support document for cancer potency factors*. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. http://oehha.ca.gov/media/downloads/crnr/tsdcancerpotency.pdf . Accessed 19 January 2018.

August, L.M., J.B. Faust, L. Cushing, et al. 2012. Methodological considerations in screening for cumulative environmental health impacts: Lessons learned from a pilot study in California. *International Journal of Environmental Research and Public Health* 9 (9): 3069–3084.

Batterman, S., F. Su, S. Li, et al. 2014. Personal exposure to mixtures of volatile organic compounds: Modeling and further analysis of the RIOPA data. *Research report Health Effects Institute* 181: 3–63.

CalEPA. 2003. Recommendations of the California Protection Agency (Cal/EPA) Advisory Committee on Environmental Justice to the Cal/EPA Interagency Working Group on Environmental Justice: Final Report. California Environmental Agency. https://calepa.ca.gov/wpcontent/uploads/sites/34/2016/10/EnvJustice-Documents-2003yr-FinalReport.pdf. Accessed 12 Dec 2017.

———. 2005. Cal/EPA Environmental Justice Advisory Committee (CEJAC) Adopted Working Definitions of Cumulative Impacts and Precautionary Approaches. CEJAC Meeting 15 Feb 2005. California Environmental Agency. http://oehha.ca.gov/media/downloads/calenviroscreen/document/workingdefs052808.pdf. Accessed 19 January 2018.

———. 2014. Designation of disadvantaged communities pursuant to Senate Bill 535 (De Leon). California Environmental Protection Agency. https://pdfs.semanticscholar.org/b82a/a9e8695d3a8ff41ff7732a4e35cd4531c08a.pdf. Accessed 12 Dec 2017.

California Government Code. 1999 §65040.12(e).

CDC: Centers for Disease Control. 2002. Childhood lead poisoning associated with tamarind candy and folk remedies- California, 199-2000. *Centers for Disease Control and Prevention, Morb Mortal Wkly Rep* 51: 684–686.

CDPR. 2005. DPR Environmental Justice Update. CDPR Press Release 23 Feb 2005. California Environmental Agency Department of Pesticide Regulation. http://www.cdpr.ca.gov/docs/envjust/envjust_update.htm. Accessed 19 January 2018.

Corburn, J. 2015. City planning as preventive medicine. *Preventive Medicine* 77: 48–51.
Davis, J.A., J.R.M. Ross, S.N. Bezalel, et al. 2012. Contaminants in fish from the California coast, 2009–2010: Summary report on a two-year screening survey. Surface Water Ambient Monitoring Program (SWAMP), California State Water Resources Control Board, California Environmental Protection Agency. http://www.waterboards.ca.gov/water_issues/programs/swamp/docs/coast_study/bog2012may/coast2012report.pdf. Accessed 19 January 2018.
De Leon, K. 2012. Senate Bill 535: California Global Warming Solutions Act of 2006: Greenhouse Gas Reduction Fund. California State Senate, Chapter 830, Statutes of 2012. http://www.leginfo.ca.gov/pub/11-12/bill/sen/sb_0501-0550/sb_535_bill_20120930_chaptered.html. Accessed 19 January 2018.
Dunn, A.J., and G.V. Alexeeff. 2010. Beyond risk assessment: Principles for assessing community impacts. *International Journal of Toxicology* 29 (1): 78–87.
English, P. 2013. Racial and income disparities in relation to a proposed climate change vulnerability screening method for California. *The International Journal of Climate Change: Impacts and Responses* 4 (2): 1–18.
Fan, A.M., G.V. Alexeeff, and S.B. Harris. 2010. Cumulative risks and cumulative impacts of environmental chemical exposures. *International Journal of Toxicology* 29 (1): 57.
Faust, J.B. 2010. Perspectives on cumulative risks and impacts. *International Journal of Toxicology* 29 (1): 58–64.
Faustman, E.M., and G.S. Omenn. 2008. Risk Assessment. In *Casarett and Doull's toxicology: The basic science of poisons*, ed. C.D. Klaassen, 7th ed., 107–128. New York: McGraw-Hill.
deFur, Peter L., Gary W. Evans, Elaine A. Cohen Hubal, Amy D. Kyle, Rachel A. Morello-Frosch, and David R. Williams. 2007. Vulnerability as a function of individual and group resources in cumulative risk assessment. *Environmental Health Perspectives* 115 (5): 817–824.
Gase, L.N., T. Kuo, S. Teutsch, et al. 2014. Estimating the costs and benefits of providing free public transit passes to students in Los Angeles County: Lessons learned in applying a health lens to decision-making. *International Journal of Environmental Research and Public Health* 11: 11384–11397.
Gee, G.C., and D.C. Payne-Sturges. 2004. Environmental health disparities: A framework integrating psychosocial and environmental concepts. *Environmental Health Perspectives* 112: 1645–1653.
Hallgren, E.A., P.A. McElfish, and J. Rubon-Chutaro. 2014. Barriers and opportunities: A community-based participatory research study of health beliefs related to diabetes in a US Marshallese community. *The Diabetes Educator* 41: 86–94. https://doi.org/10.1177/0145721714559131.
Howd, R.A. 2010. Considering changes in exposure and sensitivity in an early life cumulative risk assessment. *International Journal of Toxicology* 29 (1): 71–77.
Huang, G., and J.K. London. 2012. Cumulative environmental vulnerability and environmental justice in California's San Joaquin Valley. *International Journal of Environmental Research and Public Health* 9: 1593–1608.
Johnson Thornton, R.L., A. Greiner, C.M. Fichtenberg, et al. 2013. Achieving a healthy zoning policy in Baltimore: Results of a health impact assessment of the TransForm Baltimore zoning code rewrite. *Public Health Reports* 128 (Suppl 3): 87–103.
LACEHJ: Los Angeles Collaborative for Environmental Health and Justice. 2010. *Hidden hazards: A call to action for healthy, livable communities*. Liberty Hill Foundation: Los Angeles, CA.
Lee, S.J., L. Mehler, J. Beckman, et al. 2011. Acute pesticide illnesses associated with off-target pesticide drift from agricultural applications: 11 states, 1998–2006. *Environmental Health Perspectives* 119 (8): 1162–1169.
London, J.K., G. Huang, and T. Zagofsky. 2011. Land of Risk/Land of Opportunity: Cumulative Environmental Vulnerabilities in California's San Joaquin Valley. UC Davis Center for Regional Change. https://regionalchange.ucdavis.edu/news/ceva-san-joaquin-valley. Accessed 19 January 2018.

London, J.K., T. Greenfield, and T. Zagofsky, 2013. Revealing the Invisible Coachella Valley: Putting Cumulative Environmental Vulnerabilities on the Map. UC Davis Center for Regional Change and the California Institute for Rural Studies. https://regionalchange.ucdavis.edu/project/ceva-eastern-coachella-valley. Accessed 19 January 2018.

McCloskey, D.J., S. Aguilar-Gaxiola, J.L. Michener, et al. 2011. Principles of Community Engagement. Silberberg M et al (eds.) Clinical and Translational Science Awards Consortium, Community Engagement Key Function Committee Task Force Report, 2nd edn. National Institutes of Health Publication No. 11–7782. http://www.atsdr.cdc.gov/communityengagement/. Accessed 19 January 2018.

Medina-Ramon, M., and J. Schwartz. 2008. Who is more vulnerable to die from ozone air pollution? *Epidemiology* 19 (5): 672–679.

Miller, M., M.A. Marty, A. Arcus, et al. 2002. Differences between children and adults: Implications for risk assessment at California EPA. *International Journal of Toxicology* 21: 403–418.

NAS: National Academy of Sciences. 1983. *Risk assessment in the Federal Government: Managing the process*. Washington, DC: Natl Academy Press.

NAS: National Academy of Sciences. 2009. *Science and decisions: Advancing risk assessment*. Washington, DC: Natl Academy Press, National Research Council.

NEJAC: National Environmental Justice Advisory Council. 2004. Ensuring Risk Reduction in Communities with Multiple Stressors: Environmental Justice and Cumulative Risks/Impacts. Eady V et al (eds) National Environmental Justice Advisory Council, Cumulative Risks/Impacts Work Group Report. https://www.epa.gov/sites/production/files/2015-04/documents/ensuringriskreducationnejac.pdf. Accessed 19 January 2018.

New Jersey. 2014. NJ-GeoWeb. Department of Environmental Protection, Office of Environmental Justice, Bureau of GIS. http://www.nj.gov/dep/gis/geowebsplash.htm. Accessed 19 January 2018.

———. 2016. State of New Jersey, Department of Environmental Protection, Office of Environmental Justice, Smart Growth and Environmental Justice Division. http://www.nj.gov/dep/ej/resources.html#smartgrowth. Accessed 19 January 2018.

Nunez, F. 2006. Assembly Bill 32: California Global Warming Solutions Act of 2006. California State Assembly, Chapter 488, Statutes of 2006. http://www.leginfo.ca.gov/pub/05-06/bill/asm/ab_0001-0050/ab_32_bill_20060927_chaptered.pdf. Accessed 19 January 2018.

OEHHA. 2001. California sport fish consumption advisories: Important facts to know if you can eat the fish you catch in California. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. http://seafood.oregonstate.edu/.pdf%20Links/Sport-fish-consumption-advisories.pdf. Accessed 19 January 2018.

OEHHA. 2010. Cumulative impacts: Building a scientific foundation. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. http://oehha.ca.gov/media/downloads/calenviroscreen/report/cireport123110.pdf. Accessed 19 January 2018.

———. 2013. Responses to major comments on the CalEnviroScreen public review drafts, April 2013. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. http://oehha.ca.gov/media/downloads/calenviroscreen/document/042313resptocom.pdf. Accessed 19 January 2018.

———. 2014a. California Communities Environmental Health Screening Tool, Version 2.0 (CalEnviroScreen 2.0): Guidance and Screening Tool. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. http://oehha.ca.gov/media/downloads/report/ces20finalreportupdateoct2014.pdf. Accessed 19 January 2018.

———. 2014b. Summary of Major Changes in CalEnviroScreen 2.0, October 2014. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. http://oehha.ca.gov/media/CES20SummaryMajorChanges.pdf. Accessed 19 January 2018.

Pager, D., and H. Shepherd. 2008. The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology* 34: 181–209.

Pastor, M., R. Morello-Frosch, and J. Sadd. 2013. Screening for justice: Proactive spatial approaches for environmental disparities. *Environmental Managers* 8: 14–17.

Peter L. deFur, Gary W. Evans, Elaine A. Cohen Hubal, Amy D. Kyle, Rachel A. Morello-Frosch, David R. Williams, (2007) Vulnerability as a Function of Individual and Group Resources in Cumulative Risk Assessment. Environmental Health Perspectives 115 (5):817-824

Pohanka, M., and S. Fitzgerald. 2004. Urban sprawl and you: How sprawl adversely affects worker health. *Journal of the American Association of Occupational Health Nurses* 52 (6): 242–246.

Porebski, G., M. Wozniak, E. Czarnobilska, et al. 2014. Residential proximity to major roadways is associated with increased prevalence of allergic respiratory symptoms in children. *Annals of Agricultural and Environmental Medicine* 21 (4): 760–766.

Prasad, S., and S. Murphy. 2016. Moving from a 'risk' to 'impact' paradigm is critical to achieve environmental justice. *Environmental Managers* 66: 24–28.

Ramey, S.L., P. Schafer, J.L. DeClerque, et al. 2015. The preconception stress and resiliency pathways model: A multi-level framework on maternal, paternal, and child health disparities derived by community-based participatory research. *Maternal and Child Health Journal* 19: 707–719.

Sadd, J.L., M. Pastor, R. Morello-Frosch, et al. 2011. Playing it safe: Assessing cumulative impact and social vulnerability through an environmental justice screening method in the South Coast Air Basin, California. *International Journal of Environmental Research and Public Health* 8 (5): 1441–1459.

Sadd, J.L., R. Morello-Frosch, M. Pastor, et al. 2014. The truth, the whole truth, and nothing but the ground-truth: Methods to advance environmental justice and researcher-community partnerships. *Health, Education and Behavior* 4 (3): 281–290.

Salmon, A.G. 2010. Do standard risk assessment procedures adequately account for cumulative risks? An exploration of the possibilities using California's air toxics hot spots guidelines. *International Journal of Toxicology* 29 (1): 65–70.

Schwartz, N.A., C.A. von Glascoe, V. Torres, et al. 2015. "Where they (live, work and) spray": Pesticide exposure, childhood asthma and environmental justice Mexican-American farmworkers. *Health and Place* 32: 83–92.

Shmool, J.L., M.A. Yonas, O.D. Newman, et al. 2015. Identifying perceived neighborhood stressors across diverse communities in New York City. *American Journal of Community Psychology* 56 (1–2): 145–155.

Tso, P., A.J. Culyer, M. Brouwers, et al. 2011. Developing a decision aid to guide public sector health policy decisions: A study protocol. *Implementation Science* 6: 46.

U.S. Census Bureau. 2010. Census Tallies of Census Tracts, Block Groups and Blocks. https://www.census.gov/geo/maps-data/data/tallies/tractblock.html. Accessed 19 January 2018.

U.S. EPA. (1989). *Risk Assessment Guidance for Superfund (RAGS) Part A*. Washington, DC: U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2015-09/documents/rags_a.pdf. Accessed 19 January 2018.

U.S. EPA. 1991. Guidelines for Developmental Toxicity Risk Assessment. Risk Assessment Forum, Office of Research and Development, US Environmental Protection Agency, Washington, DC. EPA/600/FR-91/001. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=23162. Accessed 19 January 2018.

———. 1992. Guidelines for Exposure Assessment. Risk Assessment Forum, Office of Research and Development, US Environmental Protection Agency, Washington, DC. EPA/600/Z-92/001.

———. 1996. Guidelines for Reproductive Toxicity Risk Assessment. Risk Assessment Forum, Office of Research and Development, US Environmental Protection Agency, Washington, DC. EPA/630/R-96/009.

———. 2003. Framework for Cumulative Risk Assessment. Barnes DG, Firestone M, Forman DL et al (eds) United States Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, Risk Assessment Forum Report, US EPA Publication No. EPA/630/P-02/001F. https://www.epa.gov/sites/production/files/2014-11/documents/frmwrk_cum_risk_assmnt.pdf. Accessed 19 January 2018.

———. 2004. What you need to know about mercury in fish and shellfish. 2004 EPA and FDA advice for: women who might become pregnant, women who are pregnant, nursing mothers,

young children. Washington, DC: U.S. Environmental Protection Agency. EPA-823-R-04-005. http://www.foodprotect.org/issues/packets/2010Packet/attachments/I_006__all.pdf. Accessed 12 Dec 2017.

———. 2005a. Guidelines for Carcinogen Risk Assessment. Risk Assessment Forum, Office of Research and Development, US Environmental Protection Agency, Washington, DC. EPA/630/P-03/001Fs.

———. 2005b. Supplemental Guidance for Assessing Susceptibility from Early Life Exposure to Carcinogens. Risk Assessment Forum, Office of Research and Development, US Environmental Protection Agency, Washington, DC. EPA/630/P-03/003F.

———. 2012. Benchmark Dose Technical Guidance. Risk Assessment Forum, Office of Research and Development, US Environmental Protection Agency, Washington, DC. EPA/100-R-12-001.

———. 2013. Federal Interagency Working Group on Environmental Justice National Environmental Policy Act (NEPA)/EJ Resource Compendium. https://www.epa.gov/sites/production/files/2015-02/documents/nepa-ej-compendium.pdf. Accessed 19 January 2018.

———. 2014. *EJSCREEN fact sheet*. Washington, DC: U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2016-07/documents/ejscreen_fact_sheet.pdf. Accessed 19 January 2018.

———. 2015. *Action development process: Guidance on considering environmental justice during the development of regulatory actions*. Washington, DC: U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2015-06/documents/considering-ej-in-rulemaking-guide-final.pdf. Accessed 19 January 2018.

Zanobetti, A., and J. Schwartz. 2011. Ozone and survival in four cohorts with potentially predisposing diseases. *American Journal of Respiratory and Critical Care Medicine* 184 (7): 836–841.

Index

C. V. Rider, J. E. Simmons (eds.), *Chemical Mixtures and Combined Chemical and Nonchemical Stressors*, https://doi.org/10.1007/978-3-319-56234-6

N