

Chapter 2

Epidemiology of Environmental Health

Chen-Yu Liu

Abstract There has been a growing awareness of environmental effects among human. More epidemiologic studies have been triggered toward investigating exposure effects of chemical, social or physical factors in relation to the common or complex diseases, such as heart disease, cancer, and diabetes afterward. Following recent advances in genomics and risk assessment, environmental epidemiologic studies are incorporating gene-environment interactions and epigenetic changes to explore the multidisciplinary nature of individual. The challenges for human population work in this field include the complexity of exposure biology and the small effects that are easily disturbed. Several strategies have been developed with an attempt to resolve these challenges, such as the development of exposome, the two-stage designs and Mendelian randomization. In this chapter, the conventional study design features, including particular strengths and limitations, have been summarized.

Keywords Case-control • Cohort • Confounding • Bias • Biomarkers • Critical developmental windows • Epigenetics

2.1 Introduction of Epidemiology

2.1.1 Definition

The word “epidemiology” literally meaning “the study of what is upon the people” is made up by three Greek word roots: epi, meaning upon or among, demos, meaning people, and logos, meaning study. Epidemiology is the study of the distribution of health-related events or conditions occur in specified populations and the

C.-Y. Liu (✉)
Institute of Environmental Health, College of Public Health,
National Taiwan University, Taipei, Taiwan
e-mail: chenyuliu@ntu.edu.tw

investigation of why. The application of the study is to prevent and effectively control health problems (Last 2001). It aims to characterize person, time, and space and relate to the states of human health. When the health states are not distributed uniformly across people, times, and spaces, the epidemiologist is able to generate hypotheses of the underlying causes.

While early studies in epidemiology were largely concerned about communicable diseases, more attention has been paid toward the exposure effects of chemical, social or physical factors in relation to the common or complex diseases, such as heart disease, cancer, and diabetes afterward. Both environmental and genetic factors play an important role in most complex human diseases. One of the major challenges of exploring mechanisms and treatment of complex diseases is that neither purely environmental factors, nor purely genetic factors can fully explain the observed estimates of disease incidence and progression. Epidemiologic studies of environment and health incorporation with gene-environment interactions and epigenetic changes investigations help exploring the multidisciplinary nature of individual. Discussions of epidemiologic study designs and analysis issues in the area of genes and environment interaction effects can be found in Hunter (2005), Thomas (2010), and Liu et al. (2012).

The role of epigenetics has been increasingly recognized as a mechanism of gene-environment interaction. Epigenetics refers to changes in gene function without altering DNA sequence. This field remains particularly compelling because a number of epigenetic events have been recognized as tissue-specific and reversible, this may help explain why exposures affect specific organs and the complexity of individual susceptibility among the exposed population. Because many epigenetic changes can be affected by both internal and external factors and have the ability to change gene expressions, epigenetics is a potential major mechanism of how the genomes interact with the environmental exposures. Thus, epigenetics provide a new strategy in the search for etiological factors in many environment associated diseases.

Bringing epigenetics into the studies of environmental epidemiology helps investigators recognize mechanisms linking exposure and disease outcome, and provides new opportunities to explore disease and exposure biomarkers and new strategies for disease prevention and treatment (Michels 2010). However, some challenges remain in the areas of study design and epigenetic data interpretation among human populations. In the interpretation of an epidemiologic study, it is essential to examine if there are factors other than that under investigation might explain the result. In this chapter, we focus on design issues and recent developments in studying epidemiology of environment and health. For more extensive discussion about research principles in epidemiology, the interested readers are encourage to consult textbook in the area (Koepsell and Weiss 2003; Rothman and Greenland 1998a).

2.1.2 *Epidemiologic Study Designs*

Epidemiologic study designs comprise experimental studies and observational studies. Experimental studies use randomization to assign exposures to individuals, such as clinical trials or field trials, or to communities (community trials). Experimental studies are conducted to investigate preventions and treatments for diseases. A well-conducted experimental study can provide more scientifically rigorous data, however, several difficulties such as high cost, noncompliance, and ethical issues make the conduction of experimental study really challenging.

The majority of environmental health studies are conducted by using observational designs. The choice of study design is almost always determined by the research question of interest and the feasibility. For example, the commonly used epidemiologic study designs in the search for biomarkers of genetic and environmental effects, primary including case-control, cohort, and cross-sectional studies.

2.1.2.1 **Data Collection**

The data collection for observational studies may be based on existing records such as health examination records or the information collected specially for the investigators. Many lifestyle related information, such as cigarette smoking, alcohol consumption, or diet, are usually not able to obtain from the existing records. Interviews and questionnaires are especially useful in collecting this information. Exposure measurements by incorporating questionnaires, environmental sampling, personal monitoring, and biomonitoring, may also be collected in the environmental health study. Direct exposure monitoring includes personal monitoring by measuring toxics on or near the body, such as measuring air pollutants exposure levels at the breathing zone, or by sampling biological properties, such as the measurement of urinary 1-hydroxypyrene (1-OHP) as a biomarker of short-term polycyclic aromatic hydrocarbon (PAH) exposure (Jacob and Seidel 2002). *Biomarkers* of exposure are biological indicators of exogenous agents within the biological system, or other event in the biological system related to the exposure. Examples of biological samples collected for biomarker investigations include urine, blood, or other tissues of subjects.

2.1.2.2 **Types of Epidemiologic Studies**

Cross-Sectional Study

In a cross-sectional study the exposure information is usually collected simultaneously with the disease status. A cross-sectional study can be useful in estimating the health status or the disease pattern of a population at a specific time. When investigating a sudden disease outbreak, a cross-sectional study can measure several exposures and be

the first step in exploring the cause. For example, in the late 1960s, cross-sectional studies were used during the outbreak of blackfoot disease and reported associations of arsenic in drinking water and blackfoot disease (Tseng 1977; Tseng et al. 1968). It is relatively quick, inexpensively, and less time-consuming to conduct cross-sectional study. The association of multiple exposures and disease outcomes can be examined at the same time and therefore can be used as hypothesis generation. More rigorous study design such as cohort study or randomized control trial should be followed to test the hypothesis.

Case-Control Study

In a case-control study, subjects are enrolled by disease status (or other outcome variable) and historical exposure statuses of cases and controls are *retrospectively* assessed for the possible cause. Cases are individuals selected on the basis of disease. Controls are individuals without the disease when they enroll the study. Controls are selected to represent the source population where the cases come from. Its purpose is to provide information of exposure distribution in the source population as a reference so the occurrence of possible cause can be compared between cases and controls.

Cases can be either incident or prevalent cases. Incident cases are the cases enrolled in the study with disease newly diagnosed. Prevalent cases are cases already have existing disease before the enrollment of the study. The advantage of using incident cases in case-control study is to avoid the exposure-disease relationship be affected by the disease prognosis and duration. However, during rare disease research, prevalent cases may still have to be included. Investigators need to understand which kind of misleading conclusion could be introduced.

Case-Crossover Study

The case-crossover design was developed by Maclure (1991) to study effects of transient short-term exposures on acute events. In a case-crossover study, cases compare their own exposure during a risk period related to the causation of the outcome event to one or more than one exposure during control periods. Therefore, each case serves as his or her own control.

Cohort Study

A cohort study is a group of individuals who are free of disease of interest at the baseline (the start of follow-up) and are followed to determine the health outcomes. The exposure status is assessed among every subject at the baseline and possibly during the follow-up. The unexposed (or low-level exposure) group is used as

reference. Investigators compare the disease rates by exposure status. If the disease development is substantially different by exposure status, the investigator may be able to conclude the association between exposure and disease.

Cohort studies are useful in studying rare exposure since investigators can base on exposure status to select study subjects and make sure sufficient exposed subjects are enrolled.

A cohort study can be *prospective* or *retrospective* depending on if the data collection is before or after the development of health outcomes. In a prospective study, information of environmental exposures and other covariates of interest and biological samples, such as blood, are collected at the start of the study before the onset of disease or other health outcome. The study subjects are then followed over time to measure disease occurrence. Exposure status and questionnaire data as well as biological samples may be also updated during follow-up.

In a retrospective study, information is collected after the health outcome has already occurred. Exposure status and health outcome are assessed by using existing data or by subjects' recall. Investigators therefore have limited control over data collection. The existing data may not be appropriate for the study purpose and the measurements may be inaccurate, incomplete, or inconsistent between subjects.

Birth Cohort

The concept of fetal origins of adult diseases demonstrates the critical nature of exposure timing in producing later health effects (e.g., the association of maternal smoking during pregnancy and reduced fetal growth (Agrawal et al. 2010), obesity (Kristensen 1992), decreased lung function (Wiencke et al. 1991) and diabetes (Montgomery and Ekblom 2002) in the offspring). An increasing number of animal studies provide evidence of the role of environmental epigenetics both in disease susceptibility and in heritable environmentally induced transgenerational alterations in phenotype (Jirtle and Skinner 2007).

Genome-wide epigenetic reprogramming occurs at the stages of preimplantation and germ cell lineage development. Epigenome may therefore be more vulnerable to the environmental insults during these critical life periods (Sasaki and Matsui 2008; Feng et al. 2010). Epigenetic mechanisms in somatic cells also provide a potential explanation of how early life environmental exposures can program long-term effects in chronic disease susceptibility (Gluckman and Hanson 2004; Waterland and Michels 2007). In a birth cohort, pre-conceptual and prenatal exposures can be assessed. Follow-up a birth cohort allows tracking of development and variations in exposures and health outcomes over time. Birth cohort studies can measure DNA methylation profiles and chromatin states of mothers and their offspring at birth in tissues that can be easily obtained such as saliva, cord blood, and placenta. Additional serial sampling at multiple time points across life course enable the assessment of early effects, temporal variations in biomarkers, or susceptibility of exposures over time.

Nested Case-Control Design and Case-Cohort Studies

Cohort studies are less feasible in studying rare disease with long latency because of the number of subjects required and the follow-up period. A pseudo case-control design can be used to reduce the number of subjects for whom covariate data are required. A so-called ‘case-control nested within a cohort’ design (Liu et al. 2012) by matching each subject developing disease to one or more subjects without disease at the same time using incidence-density sample (Liu et al. 2012). Only cases and their matched controls require the covariate measurements (see for example, (Breslow et al. 1983; Lubin and Gail 1984; Whittemore 1981; Whittemore and McMillan 1982)).

A case-cohort design proposed by Prentice (Prentice and Pyke 1979) involves the random sample of a subcohort from the entire cohort. The covariate data are collected only among these random subcohort and all the cases. In nested case-control designs, controls are selected from individuals at risk at the times cases are identified, while in case-cohort studies, controls are a subcohort selected from the entire cohort at baseline. The subcohort in a given stratum provides a comparison set of cases occurring at a range of failure times and a basis for covariate monitoring during the cohort follow-up. Very similar designs have also been proposed by Kupper et al. (1975) and Miettinen (1982). These more efficient designs have started being used to study gene-environment interactions in cohort studies (Bureau et al. 2008).

Two-Stage Designs

One of the challenges environmental epidemiologists facing is the study of small effect that are easily affected by the confounding factors. In the situations when the exposure of interest and the disease are both rare, very large number of subjects are required and can be very expensive. One solution originally proposed by White is a two-stage design (White 1982). The first stage is the screening stage, exposure and disease information is ascertained on a large sample, but complete covariate information and maybe more refined exposure measurement is then collected on only a subsample in the second stage. The sampling fractions for stage 2 can depend jointly on disease and exposure status.

Twin and Family-Based Study

Twin- and family-based studies, such as triads of mother, father, and child, have been used to estimate the genetics basis of traits and the heritability of epigenetic profiles, including chromatin states (Rothman and Greenland 1998a) and DNA methylation (Koepsell and Weiss 2003; Hunt and White 1998; Ludwig and Weinstein 2005). Increasing gene-environment interaction studies have been conducted in the environmental health investigations. Individual variation in their sen-

sitivity to exposure is one of the aims of environmental epidemiology. Family-based studies of gene–environment interaction sometimes may be more powerful than population-based studies (Gauderman 2002). In family-based study, investigators are studying if the familial aggregation presence in the disease and if such aggregation can be explained by environmental or genetic factor. In twin study, the heritability of traits may be estimated based on comparisons between monozygotic and dizygotic twins since monozygotic twins have 100 % of their genes in common while dizygotic twins have 50 % of their genes in common. An increased disease concordance among monozygotic twins may represent a greater extent to which genetic effects contribute.

Collect DNA samples from family members can be more difficult than from unrelated cases and controls, especially for long latency or late-onset diseases. The use of relatives as controls may lead to overmatching on a range of genetic and environmental factors (Wacholder et al. 1992). Besides, the over sampling of intact families would also not be expected to represent social environments in the general population.

2.1.3 Study Design Issues

2.1.3.1 Confounding and Bias

When designing epidemiologic studies, there are a few issues required for consideration: feasibility, efficiency, expense, and potential sources of bias. In contrast to controlled clinical trials, epidemiological studies often suffer from confounding bias due to measured and unmeasured confounders (van Rein et al. 2014).

Confounding occurs when a disease outcome appears to be associated with an exposure merely due to its correlation with some other risk factors of the disease. For example, age is a potential confounder in epigenetic epidemiology, since DNA methylation levels change with age and age is a risk factor for lots of diseases. At the stage of study design, confounding effects can be minimized by randomization, matching, and restriction. Confounding effects can also be addressed during data analysis by performing stratification, standardization, regression modeling. Careful examination of the criteria for a possible confounder should be done before attempting to control for confounding effects to avoid overadjustment and the introduction of new bias (Jager et al. 2008).

While the occurrence of confounding is mostly due to the nature of epidemiologic study that various characteristics unevenly distributed among humans, the occurrence of bias is an error introduced by investigator during study design or conduction. In the retrospective study design such as case-control studies, the collection of data on environmental exposures and biological samples is retrospectively after the disease diagnosis. The assessment of exposure retrospectively is fraught with potential *recall bias*. While biomarkers of exposure can reduce such bias, these measures rarely can reconstruct past exposure and may be affected by the

current disease status, which may be one of the great challenges of retrospective studies. Another potential concern is *selection bias*. A fundamental requirement of a case-control study design is that cases and controls should be selected from the same population (Rothman and Greenland 1998b). Population-based incidence cases allow investigators to maximize the generalizability of the findings. Selection bias occurs when controls do not represent the population from which the cases arose (Last et al. 2001). Case-control studies are subject to both selection bias and recall bias. For rapidly fatal diseases, since only some of the incidence cases may be available for interviewing, *survivor bias* can occur if exposure status differs by survival time.

In prospective cohort studies, because this approach has the advantage of prospective collection of environmental information and biomarkers, which both precede the disease and will be unaffected by recall bias (Albert et al. 2001). A cohort study can be affected by selection bias if the participation or follow-up is related to exposure and health outcome. Analysis of data from cohort studies can be subject to bias due to loss of follow-up. Effective follow-up should minimize selection bias by attrition, and yet, at the same time, more time and expense is required. The expense of cohort studies often limits feasibility, especially as incidence rates of most diseases are low. Even with many years of follow-up a cohort study often requires collection of an *extremely* large number of individuals before the onset of disease and a sufficient follow-up time. Hence, prospective studies are considerable challenges for diseases with low incidence rate. *Risk-based sampling* is being used to increase the power of prospective studies by enrolling first-degree relatives of probands, such as the Sister Study for breast cancer risk (Weinberg et al. 2007; Medlin 2001) or the on-going Early Autism Risk Longitudinal Investigation (EARLI) study for autism risk. For common pediatric diseases such as asthma, obesity, and some adverse birth outcomes, a prospective cohort study will be extremely valuable to identify environmental risk factors (Manolio 2009; Manolio et al. 2006). Prospective cohort studies on a national scale (Collins and Manolio 2007), by pooling data from existing prospective cohorts (Willett et al. 2007), or by collaborations of cohorts across large regions such as Birth Cohort Consortium of Asia (BiCCA) and Environmental Health Risks in European Birth Cohorts (ENRIECO) should be conducted to ensure sufficient power. The U.S. Congress, through the Children's Health Act of 2000, authorized the National Institute of Child Health and Human Development (NICHD) "to conduct a national longitudinal study of environmental influences (including physical, chemical, biological, and psychosocial) on children's health and development" (CHA 2000). The National Children's Study is a 21-year prospective cohort study of 100,000 US-born children. Environmental exposures, including chemical, physical, biological, and psychosocial exposure, will be assessed repeatedly during pregnancy and childhood in children's homes, schools, and communities. The National Children's Study will provide great opportunities to gene-environment interactions for common pediatric diseases.

In a cross-sectional study since exposure and disease status are assessed at the same time, the temporal sequence of exposure and disease status usually cannot be identified. Unless in the situation that exposure is a characteristic such as gender and

the disease developed through time, such a temporal sequence of exposure and disease development is more possible. Otherwise, for the exposure factors different by time, the causality is hard to define. In addition, like many retrospectively studies, assessment of environment exposure retrospectively is fraught with potential recall bias. Further, because cases in a cross-sectional study are existing cases at the time, cases have to survive long enough to be included and may mislead the conclusion.

2.1.3.2 Measurements of Exposure

Measurements of environmental exposures have been a great challenge in epidemiologic studies due to the complex pattern of long-term exposures and the need to collect accurate and repeated individual exposure data in large populations (Morgenstern and Thomas 1993). Measurement errors, such as misclassification of exposure status, can exist regardless of study design. Misclassification of exposure generally leads to attenuation of the main effects when the error is non-differential (Carroll et al. 2006). Non-differential misclassification occurs when the misclassification of exposure does not depend on disease status or when the misclassification of disease does not depend on exposure status. Non-differential misclassification can also bias away from the null in some circumstances, including (1) if the exposure is multilevel (>2 levels), the intermediate levels of exposure could be biased away from null (Dosemeci et al. 1990; Weinberg et al. 1994); (2) if the misclassifications are correlated with other errors (Kristensen 1992; Chavance et al. 1992); (3) if the measured exposure do not change monotonically with the true exposure (Weinberg et al. 1994; ES 1991).

The use of questionnaires for exposure assessment relies on personal memory and has the potential for recall bias. Several technologies have been developed to improve measurements of environmental exposures. To incorporate qualitative and quantitative changes of environmental exposures, such as atmospheric conditions and topography, over time and space, as well as individuals' diverse demographic characteristics, lifestyles, activity patterns, geographic information systems (GIS)/global positioning system (GPS), personal monitoring, and biomonitoring are now being used in environmental epidemiology. Combined geospatial tools with statistical models allow investigators to model the transport of the pollutants from source to residence, e.g., using wind speed, temperature, and traffic density in addition to measurements from the central site, to estimate an individual-level exposure as well. Monitoring data, such as personal monitoring and measurements of biomarkers, can hold great promise for improving exposure assessment by providing objective individual-level measurements. Biomarkers take into account individual differences in absorption, distribution, and metabolism of the compound within the body and therefore can be used to reflect the effects of earlier exposures and the association between exposure and disease at the molecular level (NRC 1987; Perera and Weinstein 1982; Rothman et al. 1995). The usefulness of a biomarker is strongly depending on the specificity, sensitivity, assay reliability, accessibility, and cost (Hemstreet et al. 2001). Examples of biomarkers for exposure or effect range from

patterns of gene expressions, proteins, or metabolic profiles in cells and tissues change. Glycosylated hemoglobin for instance, a measure of chronic serum glucose, can be used to study diabetic risk factors with more power than a study focused on clinical diabetes. In spite of these potential advantages, the results of biomarker measurements sometimes can confuse the investigators a lot. Different conclusions may arise due to the differences of specimen kinds, collection and processing methods, laboratory error, and individual variation in the biomarker levels over time (Little and Sacks 2009). For instance, the correlation between epigenetic profiles in different tissues is complex and locus dependent. The tissue specificity and may be the cell type specificity in epigenetic patterns is challenging in epidemiologic studies. Population-based studies may inevitably rely on the use of easily accessible tissue, such as peripheral blood, saliva, or buccal cells. These sources may be able to reflect the environmental induced epigenetic changes or an overall indicator of the health status of a group of individuals; however, they may not accurately represent epigenetic variations of the diseased target tissue of interest (Talens et al. 2010).

Exposure Effects Across Life Stages

Measuring environment has added complexity beyond issues of measurement error or selection bias. Even measuring cumulative exposure prospectively may be insufficient. During prenatal life and childhood, critical biological events occur that establish the number, connections and proper function of cells within given tissues. As an example, environmental exposure, particularly in early development, may induce changes in gene expression modulated through DNA promoter methylation or chromatin remodeling (Fleming et al. 2008). Toxicological studies show that the central nervous system is especially vulnerable to toxic injury (Rodier 2004) and epidemiological studies clearly show an association between adverse neurodevelopment and in utero exposure to chemicals such as methyl mercury (Amin-Zaki et al. 1974; Marsh et al. 1980), PCBs (Tilson et al. 1990), while exposure later in life demonstrates less toxicity. The concept of fetal origins of adult diseases demonstrates the critical nature of exposure timing in producing later health effects (e.g., the association of maternal smoking during pregnancy and reduced fetal growth (Agrawal et al. 2010), obesity (Kristensen 1992), decreased lung function (Wiencke et al. 1991) and diabetes (Montgomery and Ekblom 2002) in the offspring). All individuals are exposed to a variety of hazardous agents and chemicals in the environment. The concept of exposome represents the totality of exposures received by a person during life and defines exposures as levels of biologically active compounds (Rappaport 2011). Because sources and levels of exposure vary by time, exposomes can be constructed by analyzing toxicants in blood specimens obtained during critical stages of life.

Because genotypes do not vary over time and can always be presumed to precede phenotype, genotype related measurements and prognostic markers are still identified in case-control studies. Otherwise, early detection or risk predication markers are preferred in nested case-control studies with biological samples collected

pre-diagnostically. A prospective study can address timing of exposure in an unbiased manner, however, it is still challenging to assess the details of exposure timing and risk as the critical window likely differs for different phenotypes and for different exposures. These dynamic features make their interpretation in human studies challenging. Single measurement may not be reliable especially in those investigating long-term chronic effects. Incorporating different exposure assessment techniques with long-term monitoring data especially during critical periods is needed to provide an integrated view of exposure in complex exposure–disease relationships (Lioy 1995; Weis et al. 2005). Unfortunately, for most adult diseases, an unbiased reconstruction of childhood exposure is difficult or even impossible. Thus, a major limitation of adult epidemiologic research will continue to be the inability to reconstruct childhood factors that predict disease. A measure of cumulative exposure, while preferable to cross-sectional measures, cannot capture exposure during the critical developmental life stage predisposing to disease. Therefore, rather than modeling a chemical in a single dose response curve for toxicity, chemicals appear to have different dose response curves depending on the life stage at which exposure occurs. For example, in utero diethylstilbesterol exposure is associated with vaginal cancer in offspring, while mothers who took the drug do not appear to be at risk.

Another difficulty is that it is not possible to know with certainty what the critical exposure window is a priori (i.e. in utero vs. childhood vs. puberty). The difficulties in assessing the exposure effects by timing present in carefully designed observational studies and even trial results. An example is the initial report from Women's Health Initiative (WHI) randomized trial and epidemiologic data on the risk of coronary heart disease (CHD) and the menopausal hormone therapy. Large observational studies include Nurses' Health Study (NHS) suggested a reduced risk of CHD among postmenopausal hormone therapy (Hemstreet et al. 2001; Angerer et al. 2007) while WHI randomized trial found increased risk of CHD among women assigned to the menopausal hormone therapy compared to the placebo group (Chatterjee et al. 1969). Hernán et al. re-analysis of the Nurses' Health Study and concluded that most of the difference could be attributed to the age distribution at the time of initiation of hormone therapy and length of follow-up (Hernan et al. 2008).

Mendelian Randomization

Mendelian randomization is an approach, instead of studying unknown effects, using the established associations between genetic variations and exposure intermediate phenotypes to maximize causal inference. These genetic variations can mimic the modifiable exposure effects and serve as a surrogate to test the association between exposure and disease. Mendelian randomization provides an approach for making causal inferences about the exposure by using the nature of randomly assigned genotypes from parents to offspring before conception (Gray and Wheatley 1991; Davey Smith and Ebrahim 2003). However, as well with all genetic association studies, potential confounding effects by population stratifications and other limitations can still occur (Davey Smith and Ebrahim 2003; Cui et al. 2001).

Population stratification refers to the different genotype frequencies between cases and controls due to differences in ancestry rather than association between genes and disease. Careful study conduction and thorough verification remains essential before considering the causality.

2.2 Discussion

Following recent advances in genomics and risk assessment, current research has shown that both internal and external factors of human bodies are critical to disease risk and progression. Future improvements in disease prevention and treatment are anticipated. More research efforts need to be directed towards investigating the exposure biology such as identifying new biomarkers for better measuring exposures, and the roles of genome and epigenome to various environmental agents. The integration of specimen bank among stored biological materials from cohort studies and the increasing collaboration between basic bench work and epidemiology is becoming more and more important. Study design selection is strongly depending on the feasibility and cost and should be guided by the appropriateness of the research question of interest. As summarized above, each epidemiologic study design has various strengths and limitations that make it more or less suitable for particular exposure effect investigation. In the interpretation of an epidemiologic study, it is essential for the researcher to examine if there are factors other than that under investigation might explain the result.

References

- Agrawal A et al (2010) The effects of maternal smoking during pregnancy on offspring outcomes. *Prev Med* 50(1–2):13–18
- Albert PS et al (2001) Limitations of the case-only design for identifying gene-environment interactions. *Am J Epidemiol* 154(8):687–693
- Amin-Zaki L et al (1974) Intra-uterine methylmercury poisoning in Iraq. *Pediatrics* 54(5):587–595
- Angerer J, Ewers U, Wilhelm M (2007) Human biomonitoring: state of the art. *Int J Hyg Environ Health* 210(3–4):201–228
- Breslow NE, Lubin JH, Marek P, Langholtz B (1983) Multiplicative models and cohort analysis. *J Am Stat Assoc* 78:1–12
- Bureau A et al (2008) Estimating interaction between genetic and environmental risk factors: efficiency of sampling designs within a cohort. *Epidemiology* 19(1):83–93
- Carroll RJ, Crainiceanu C, Ruppert D, Stefanski LA (2006) Measurement error in nonlinear models: a modern perspective. Chapman and Hall/CRC, Boca Raton
- CHA (2000) Children's Health Act 2000 Public Law 106–310 (codified at 42 USC 201)
- Chatterjee BB et al (1969) The location of personal sampler filter heads. *Am Ind Hyg Assoc J* 30(6):643–645

- Chavance M, Dellatolas G, Lellouch J (1992) Correlated nondifferential misclassifications of disease and exposure: application to a cross-sectional study of the relation between handedness and immune disorders. *Int J Epidemiol* 21(3):537–546
- Collins FS, Manolio TA (2007) Merging and emerging cohorts: necessary but not sufficient. *Nature* 445(7125):259
- Cui Y et al (2001) Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. *Science* 293(5533):1289–1292
- Davey Smith G, Ebrahim S (2003) ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 32(1):1–22
- Dosemeci M, Wacholder S, Lubin JH (1990) Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 132(4):746–748
- ES G (1991) Re: “Does nondifferential misclassification of exposure always bias a true effect toward the null value?” (Letter). *Am J Epidemiol* 134(4):440–1
- Feng S, Jacobsen SE, Reik W (2010) Epigenetic reprogramming in plant and animal development. *Science* 330(6004):622–627
- Fleming JL, Huang TH, Toland AE (2008) The role of parental and grandparental epigenetic alterations in familial cancer risk. *Cancer Res* 68(22):9116–9121
- Gauderman WJ (2002) Sample size requirements for matched case-control studies of gene-environment interaction. *Stat Med* 21(1):35–50
- Gluckman PD, Hanson MA (2004) Developmental origins of disease paradigm: a mechanistic and evolutionary perspective. *Pediatr Res* 56(3):311–317
- Gray R, Wheatley K (1991) How to avoid bias when comparing bone marrow transplantation with chemotherapy. *Bone Marrow Transplant* 7(Suppl 3):9–12
- Hemstreet GP 3rd et al (2001) Biomarker risk assessment and bladder cancer detection in a cohort exposed to benzidine. *J Natl Cancer Inst* 93(6):427–436
- Hernan MA et al (2008) Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 19(6):766–779
- Hunt JR, White E (1998) Retaining and tracking cohort study members. *Epidemiol Rev* 20(1):57–70
- Hunter DJ (2005) Gene-environment interactions in human diseases. *Nat Rev Genet* 6(4):287–298
- Jacob J, Seidel A (2002) Biomonitoring of polycyclic aromatic hydrocarbons in human urine. *J Chromatogr B Analyt Technol Biomed Life Sci* 778(1–2):31–47
- Jager KJ et al (2008) Confounding: what it is and how to deal with it. *Kidney Int* 73(3):256–260
- Jirtle RL, Skinner MK (2007) Environmental epigenomics and disease susceptibility. *Nat Rev Genet* 8(4):253–262
- Last JM (2001) *A dictionary of epidemiology*, 4th edn. Oxford University Press, Oxford
- Koepsell TD, Weiss NS (2003) *Epidemiologic methods: studying the occurrence of illness*. Oxford University Press, New York
- Kristensen P (1992) Bias from nondifferential but dependent misclassification of exposure and outcome. *Epidemiology* 3(3):210–215
- Kupper LL, McMichael AJ, Spirtas R (1975) A hybrid epidemiologic study design useful in estimating relative risk. *J Am Stat Assoc* 70:524–528
- Last J, Spasoff R, Harris S (2001) *A dictionary of epidemiology*. Oxford University Press, New York
- Lioy PJ (1995) Measurement methods for human exposure analysis. *Environ Health Perspect* 103(Suppl 3):35–43
- Little RR, Sacks DB (2009) HbA1c: how do we measure it and what does it mean? *Curr Opin Endocrinol Diabetes Obes* 16(2):113–118
- Liu CY et al (2012) Design and analysis issues in gene and environment studies. *Environ Health* 11:93

- Lubin JH, Gail MH (1984) Biased selection of controls for case-control analyses of cohort studies. *Biometrics* 40(1):63–75
- Ludwig JA, Weinstein JN (2005) Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* 5(11):845–856
- Maclure M (1991) The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 133(2):144–153
- Manolio TA (2009) Cohort studies and the genetics of complex disease. *Nat Genet* 41(1):5–6
- Manolio TA, Bailey-Wilson JE, Collins FS (2006) Genes, environment and the value of prospective cohort studies. *Nat Rev Genet* 7(10):812–820
- Marsh DO et al (1980) Fetal methylmercury poisoning: clinical and toxicological data on 29 cases. *Ann Neurol* 7(4):348–353
- Medlin J (2001) Sister study hopes to answer breast cancer questions. *Environ Health Perspect* 109(8):A368–A369
- Michels KB (2010) The promises and challenges of epigenetic epidemiology. *Exp Gerontol* 45(4):297–301
- Miettinen O (1982) Design options in epidemiologic research. An update. *Scand J Work Environ Health* 8(Suppl 1):7–14
- Montgomery SM, Ekbom A (2002) Smoking during pregnancy and diabetes mellitus in a British longitudinal birth cohort. *BMJ* 324(7328):26–27
- Morgenstern H, Thomas D (1993) Principles of study design in environmental epidemiology. *Environ Health Perspect* 101(Suppl 4):23–38
- NRC (1987) Biological markers in environmental health research. *Environ Health Perspect* 74:3–9
- Perera FP, Weinstein IB (1982) Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. *J Chron Dis* 35(7):581–600
- Prentice RL, Pyke R (1979) Logistic disease incidence models and case-control studies. *Biometrika* 66:403–411
- Rappaport SM (2011) Implications of the exposome for exposure science. *J Expo Sci Environ Epidemiol* 21(1):5–9
- Rodier PM (2004) Environmental causes of central nervous system maldevelopment. *Pediatrics* 113(4 Suppl):1076–1083
- Rothman KJ, Greenland S (1998a) *Modern epidemiology*, 2nd edn. Lippincott-Raven Press, Philadelphia
- Rothman KJ, Greenland S (1998b) Case-control studies. In: Rothman KJ, Greenland S (eds) *Modern epidemiology*, 2nd edn. Lippincott-Raven, Philadelphia, pp 93–114
- Rothman N, Stewart WF, Schulte PA (1995) Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. *Cancer Epidemiol Biomarkers Prev* 4(4):301–311
- Sasaki H, Matsui Y (2008) Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* 9(2):129–140
- Talens RP et al (2010) Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. *FASEB J* 24(9):3135–3144
- Thomas D (2010) Gene – environment-wide association studies: emerging approaches. *Nat Rev Genet* 11(4):259–272
- Tilson HA, Jacobson JL, Rogan WJ (1990) Polychlorinated biphenyls and the developing nervous system: cross-species comparisons. *Neurotoxicol Teratol* 12(3):239–248
- Tseng WP (1977) Effects and dose – response relationships of skin cancer and blackfoot disease with arsenic. *Environ Health Perspect* 19:109–119
- Tseng WP et al (1968) Prevalence of skin cancer in an endemic area of chronic arsenicism in Taiwan. *J Natl Cancer Inst* 40(3):453–463
- van Rein N et al (2014) Suspected survivor bias in case-control studies: stratify on survival time and use a negative control. *J Clin Epidemiol* 67(2):232–235

- Wacholder S et al (1992) Selection of controls in case-control studies. II. Types of controls. *Am J Epidemiol* 135(9):1029–1041
- Waterland RA, Michels KB (2007) Epigenetic epidemiology of the developmental origins hypothesis. *Annu Rev Nutr* 27:363–388
- Weinberg CR, Umbach DM, Greenland S (1994) When will nondifferential misclassification of an exposure preserve the direction of a trend? *Am J Epidemiol* 140(6):565–571
- Weinberg CR et al (2007) Using risk-based sampling to enrich cohorts for endpoints, genes, and exposures. *Am J Epidemiol* 166(4):447–455
- Weis BK et al (2005) Personalized exposure assessment: promising approaches for human environmental health research. *Environ Health Perspect* 113(7):840–848
- White JE (1982) A two stage design for the study of the relationship between a rare exposure and a rare disease. *Am J Epidemiol* 115(1):119–128
- Whittemore AS (1981) The efficiency of synthetic retrospective studies. *Biom J* 23:73–78
- Whittemore AS, McMillan A (1982) Analyzing occupational cohort data: application to U.S. uranium miners. In: Prentice RL, Whittemore AS (eds) *Environmental epidemiology: risk assessment*. SIAM, Philadelphia, pp 65–81
- Wiencke JK et al (1991) Individual susceptibility to induced chromosome damage and its implications for detecting genotoxic exposures in human populations. *Cancer Res* 51(19):5266–5269
- Willett WC et al (2007) Merging and emerging cohorts: not worth the wait. *Nature* 445(7125):257–258