Chapter 16
Systematic Review, An Illustration of Increased Transparency in a Framework for Evaluating Immunotoxicity Associated with PFOA and PFOS Exposure

Andrew A. Rooney, Abee L. Boyles, and Vickie R. Walker

Abstract Background: Systematic review methodologies were first developed to assess the efficacy of health care interventions, but these approaches can be adapted to evaluations of environmental health questions such as immunotoxicity associated with PFOA and PFOS exposure. This structured approach provides objectivity and transparency to the process of collecting, synthesizing, and reaching conclusions based on the scientific evidence available.

Objectives: To outline the process of systematic review and evidence integration and demonstrate each step by following a single research question from start to finish. The example systematic review will evaluate the evidence that PFOA and PFOS exposure are associated with immunotoxicity – using a subset of the available evidence to illustrate concepts, not to develop hazard identification conclusions.

Methods: The Office of Health Assessment and Translation (OHAT) Approach to evaluating the scientific evidence for immunotoxicity of PFOA and PFOS is detailed in a protocol that is laid out in seven steps: scoping and problem formulation, search for and select studies for inclusion, extract data from studies, assess quality of individual studies, rate confidence in the body of evidence, translate confidence ratings into level of evidence, and integrate evidence to develop hazard identification conclusions incorporating human, animal, and mechanistic evidence.

A.A. Rooney (✉)
Office of Health Assessment and Translation, Division of the National Toxicology Program, National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH), Research Triangle Park, NC, USA

NIEHS, P.O. Box 12233, Mail Drop K2-04, Research Triangle Park, NC 27709, USA
e-mail: Andrew.Rooney@nih.gov

A.L. Boyles • V.R. Walker
Office of Health Assessment and Translation, Division of the National Toxicology Program, National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH), Research Triangle Park, NC, USA

J.C. DeWitt (ed.), *Toxicological Effects of Perfluoroalkyl and Polyfluoroalkyl Substances*, Molecular and Integrative Toxicology,
DOI 10.1007/978-3-319-15518-0_16

Results and Discussion: Eligibility criteria for identifying important exposures and outcomes were presented as the basis for assembling the relevant studies for evaluating whether or not PFOA or PFOS exposure is associated with immunotoxicity (human, n = 18; animal, n = 80; and mechanistic/in vitro assays, n = 19). A tool for assessing study quality in terms of risk of bias or internal validity was tailored to the research question – particularly for evaluating PFC exposure and assessing immunological outcomes. An example of an evidence profile is provided to illustrate the basis for confidence ratings using a hypothetical set of studies of PFOS and functional antibody response. Finally, a discussion is presented on how the hazard identification conclusions would be reached and interpreted by integrating the human, animal, and mechanistic evidence.

Conclusion: The OHAT Approach to hazard identification of health effects of PFCs is illustrated with a case study on PFOA/PFOS and immunotoxicity. Communication of the evaluation process is enhanced by using objective, reproducible methods that transparently document scientific judgments and the scientific basis for hazard identification conclusions.

Keywords Systematic review • Perfluorinated chemicals • Immunotoxicity • Risk of bias • Hazard identification • PFOA • PFOS

16.1 Introduction

The strength and reliability of hazard conclusions on the potential human health effects from environmental exposures can be hindered by inconsistent or unclear methods of how the evaluation was performed. Systematic-review methodologies provide a structure that increases transparency and objectivity in the process of collecting and synthesizing scientific evidence for literature-based evaluations. Multiple organizations have adopted (Birnbaum et al. 2013; Woodruff and Sutton 2014) or recommended (EFSA 2010; NRC 2013a, b; Rhomberg et al. 2013; US EPA 2013a) the use of systematic review methods for evaluating the association between health effects and environmental exposures. First developed and established in clinical medicine to assess data for reaching health care recommendations (AHRQ 2013; Guyatt et al. 2011; Higgins and Green 2011), systematic-review methodologies typically addressed data from clinical trials and focus on human data alone.

The data available to evaluate potential health effects from exposure to environmental chemicals comes from diverse sources and rarely include experimental trials in humans. Human data are typically from observational studies that include cohort, cross sectional, case control, and even case report study designs. Animal data, primarily from *in vivo* laboratory studies in rodents, provide a large percentage of the toxicology data used for hazard identification and risk assessment. Mechanistic or other relevant data from *in vitro* and *in vivo* studies on molecular and cellular

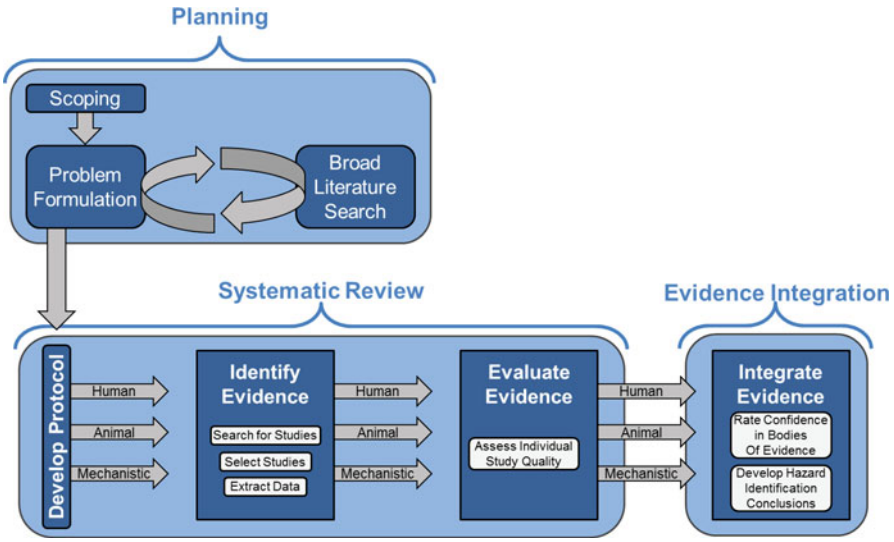

Fig. 16.1 OHAT framework for systematic review and evidence integration
An evaluation begins with the planning process and development of a detailed protocol to guide each step in the assessment. The systematic review identifies, collects and evaluates the evidence from individual studies. Evidence integration is the process where bodies of evidence and multiple lines of evidence gathered in the systematic review (human, animal, and mechanistic studies) are integrated to develop conclusions

events also inform the hazard conclusions as part of the overall database (NRC 2013b). The National Toxicology Program's (NTP) Office of Health Assessment and Translation (OHAT) developed a framework for examining environmental health questions (Fig. 16.1) using systematic review procedures that address the breadth of relevant data (e.g., human, animal, and mechanistic studies), and evidence integration procedures to consider the collective evidence in developing conclusions (Rooney et al. 2014).

16.2 Systematic Review

Use of systematic review methodology adds a level of objectivity and transparency to established principles of hazard assessment (WHO 1999) and continual improvement in communicating the basis for scientific judgments within an assessment of potential health effects of environmental chemicals (NRC 2009, 2013b). A planning process of scoping and problem formulation lays the groundwork for the systematic review by framing the specific research question to be addressed and the analytical approach: hazard assessment, risk assessment, or to identify data gaps where the scientific evidence base is small or narrow (Fig. 16.1). Then, systematic review procedures use transparent, rigorous, objective, and reproducible methodology to

identify, select, assess, and analyze results of relevant studies to complete the literature-based evaluation. These methods do not eliminate the need for scientific judgment; rather, they offer an increased level of transparency for understanding the basis of decisions and the overall confidence in the conclusions.

16.3 Evidence Integration

Evidence integration begins with identifying groups of studies with data on an outcome (or biologically related outcomes), or "bodies of evidence." Confidence ratings are then developed for the human and animal bodies of evidence separately by considering the strengths and weaknesses of the relevant studies. Ratings reflect confidence that the study findings accurately represent the true association between exposure to a substance and an effect. If the evidence base is insufficient, the evaluation can result in a summary of the data gaps. If the evidence base is sufficient, it can proceed to the last step in evidence integration – to develop hazard conclusions from the confidence ratings by integrating the animal and human evidence with consideration of the impact of mechanistic data.

This document will use an evaluation of the evidence that PFOA and PFOS exposure are associated with immunotoxicity to illustrate the systematic review and evidence integration process. While this is not intended to be a complete evaluation, the concepts and procedure presented give substance to the OHAT framework. No hazard conclusions are developed in this example because only subsets of the evidence are used to illustrate these concepts of this systematic review approach.

16.4 Systematic Review of PFOA and PFOS Immunotoxicology

16.4.1 Scoping and Problem Formulation

The foundation of an evaluation relies on focused questions that have been developed and refined through a process of scoping and problem formulation. Scoping procedures define the needs and goals of the evaluation, such as whether it will address occupational exposure or the general public, and whether the goal is to support hazard identification conclusions, a complete risk assessment, or government regulations. Outreach and consultation with subject-matter experts and interested parties, (which may include the public and stakeholders depending on the policies of the review organization), help assure that the product meets the needs of the risk manager (US EPA 1998b) and all available information is considered (especially existing analyses or reviews). Problem formulation is the process of refining the objectives of the evaluation, clearly stating the key questions to be answered, and

outlining how they will be addressed (NRC 2009). The questions define eligibility criteria for the populations, exposures, comparators, outcomes, timings, and settings of interest (PECOTS) for the evaluation (Matcher 2012; Samson and Schoeles 2012).

16.4.1.1 Scoping and Problem Formulation for PFOA/PFOS Immunotoxicity

The planning process for an evaluation of immunotoxicity associated with PFOA and PFOS would address the basic requirements described above including outreach to obtain input on the need for an assessment and availability of data. As part of exposure considerations, the persistence and wide environmental distribution of PFOA and PFOS would be key factors (see Chaps. 2, 3, 4, and 5 of this Book). Given the voluntary agreements by the primary manufacturers to phase out production of PFOA and PFOS is the United States by 2015 (ATSDR 2009; US EPA 2006, 2009, 2013b, 2014), the potential for future exposure would also be considered. Although emissions have been dramatically reduced, the persistence and bioaccumulation of both PFOA and PFOS still result in detectable levels in the U.S. population and therefore are of potential human health relevance (US EPA 2014).

Overview of Scientific Information on PFOA/PFOS Immunotoxicity

During problem formulation the extent of available health effect data would be outlined including whether or not the database is likely to be sufficient to develop conclusions. Several publications from 2012 to 2014 link PFOA and PFOS exposure to functional immune changes in humans that are consistent with evidence of immunotoxicity from animal studies. Immune-related health effects including suppression of the antibody response to vaccines and increased incidence of autoimmune ulcerative colitis have been reported in adults living in an area of Ohio and West Virginia where public drinking water had been contaminated with PFOA (Looker et al. 2014; Steenland et al. 2013). PFOA- and PFOS-associated antibody suppression were also described in prospective cohort studies of children in Norway (Granum et al. 2013) and the Faroe Islands (Grandjean et al. 2012).

Suppression of the antibody response in mice has been reported at blood concentrations of PFOS occurring in the general U.S. population (e.g., CDC 2009, 2014; DeWitt et al. 2012; Fair et al. 2011; Peden-Adams et al. 2008). Experimental studies of PFOA and PFOS in laboratory animals have also demonstrated exposure-related suppression of the antibody response among other immune changes including altered inflammatory response, cytokine signaling, and measures of both innate and adaptive immunity (reviewed in DeWitt et al. 2012). Wildlife studies in species ranging from loggerhead sea turtles to sea otters have also reported widespread exposure and altered immune measures associated with PFOA and PFOS (e.g., Hart et al. 2009; Kannan et al. 2006; Keller et al. 2005). Mechanistic and in vitro exposure

studies of PFOA and PFOS are primarily focused on cytokine secretion (Ahuja et al. 2009; Corsini et al. 2011, 2012; Han et al. 2012); although more predictive measures of immunotoxicity (e.g., immune function), such as natural killer cell activity have also been studied after *in vitro* exposure (Wirth et al. 2014).

Objectives and Key Questions for Evaluating PFOA/PFOS Immunotoxicity

The objective of this illustration is to develop hazard identification conclusions regarding exposure to PFOA or PFOS and potential associations with immunotoxicity or immune-related health effects. Although PFOA and PFOS are both considered in this example, conclusions would be developed separately for each chemical. The objectives would be addressed by answering key questions listed below.

- What is our confidence in the body of evidence from human studies for the association between exposure to PFOA or PFOS and immunotoxicity or immune-related health effects?
- What is our confidence in the body of evidence from animal studies for the association between exposure to PFOA or PFOS and immunotoxicity or immune-related health effects?
- How does the evidence from other relevant studies (e.g., mechanistic or in vitro studies) support or refute the biological plausibility of the association between exposure to PFOA or PFOS and immunotoxicity or immune-related health effects?

The available studies for each of the three evidence streams (human, animal, and mechanistic or other relevant studies) would be evaluated separately. Then, hazard identification conclusions for PFOA-associated immunotoxicity and PFOS-associated immunotoxicity would be developed by integrating the human and animal evidence with consideration of the impact of mechanistic or other relevant data.

16.4.2 Protocol

The evaluation is structured to answer the key questions and a detailed protocol is developed to guide the evaluation process from the literature search, through analysis, and finally the process of integrating the evidence to develop conclusions. Subject-matter experts, particularly scientists with backgrounds in exposure and relevant health effects for the chemical under review, should be consulted in establishing the protocol before proceeding with the evaluation. The protocol's "*a priori*" guidance reflects the scientific knowledge in the field and forms the basis for scientific judgments throughout the evaluation; however, if unanticipated issues arise during the evaluation the protocol can be modified. Pilot-testing and refining the procedures outlined in the protocol on a small subset of studies is recommended at multiple steps, particularly: applying inclusions/exclusion criteria, data

extraction, and risk of bias assessment of individual studies. The transparency principals of systematic review dictate that any revisions are documented and justified including when in the evaluation process the decision was made, not that initial decisions are locked.

16.4.3 Search for and Select Studies for Inclusion

Systematic review requires a comprehensive and transparent literature search strategy and a clear statement of the inclusion and exclusion criteria used to determine if a study is relevant for the evaluation. The protocol outlines the search and selection procedures in sufficient detail such that the literature retrieval could be clearly understood and reconstructed by a third party, including the basis of scientific judgments. The search strategy details the exact search terms used as well as the specifics of the literature search process including which databases will be searched, limits in language or dates of publication, and how unpublished studies will be treated. The eligibility criteria reflect the scoping and problem formulation decisions and specifies the types of human and animal studies (e.g., experimental only or also including wildlife studies), exposure metrics (e.g., potentially excluding occupational exposures or ecological studies without individual exposure measurements), and outcomes that will be used to address the key questions. The protocol also states the procedures for screening references for inclusion, resolving conflicts between reviewers, and documenting the reasons references were excluded. Screening is typically a two-step process starting at the title and abstract level to exclude references that are clearly not relevant, and then proceeding to more detailed review of the full text of studies that passed the first screen. The title and abstract of each reference are reviewed for relevance and eligibility by two screeners independently, with conflicts resolved through discussion or consultation with a third reviewer. Exclusion decisions during full text review should be documented in the form of a flow diagram (Fig. 16.2) tracking the number of references retrieved and exclusion during the screening process up to the point references are selected for data extraction (Liberati et al. 2009; Moher et al. 2009).

16.4.3.1 Searching for and Selecting PFOA/PFOS Immunotoxicity Studies

The search terms for both PFOA and PFOS exposure and immune effects for this example were identified by (1) reviewing Medical Subject Headings (MeSH) for relevant terms, (2) extracting key terminology from reviews and a sample of relevant primary data studies, and (3) consulting a review of PFOA search terms from early drafts of a systematic review of developmental PFOA exposure and fetal growth (Johnson et al. 2013, 2014; Koustas et al. 2014). Although a published search strategy of PFOS was not located, the PFOA strategy was used by analogy as the

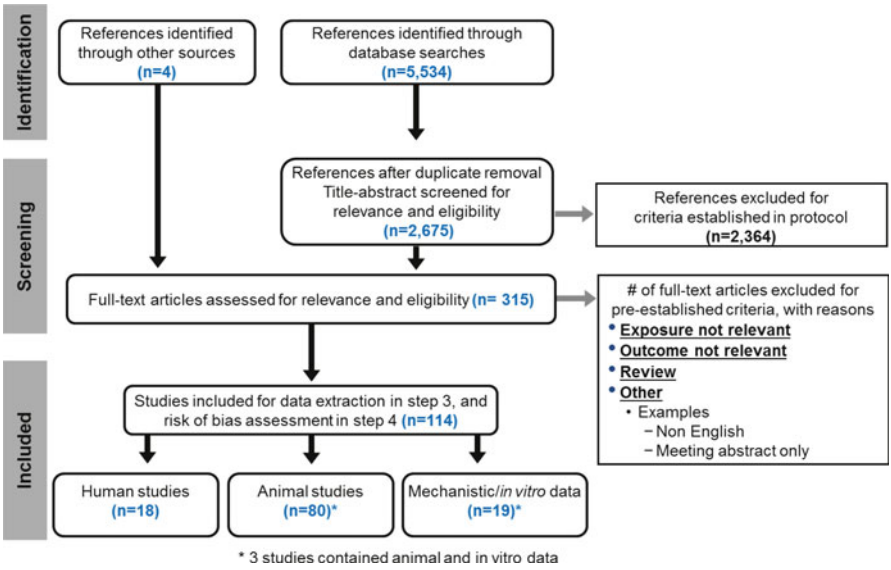

Fig. 16.2 Study tracking based on the PRISMA reporting standard
Clear documentation of the literature selection process can be accomplished with a simple flow chart following the PRISMA reporting standard for tracking references obtained from the literature search, through inclusion or exclusion for relevance or eligibility

basis for developing terms for PFOS exposure. A list of relevant subject headings and keywords were identified that combined exposure and immune or immunotoxicology terms (Table 16.1). A small set of relevant studies was used to test the search terms to ensure the strategy retrieved 100 % of previously identified "relevant" references. The search strategy presented in Table 16.1 would have to be tailored for each database.

Evaluations identify eligible studies from the PECOTS statement developed in problem formulation by clearly stating the populations, exposures, comparators, outcomes, timings, settings and important considerations such as appropriate study designs considered relevant. This example is outlined to be an evaluation of health effects for either the general population or highly exposure populations, so there would be no exclusions based occupational exposures. And as for exposure, studies with less precise exposure data are unlikely to be excluded unless there it has been established that the data are fundamentally flawed or there was a large database of epidemiological studies that have really good measures of exposure and outcome. For PFOA and PFOS immunotoxicity, there is a relatively small database of human studies, and no exclusions would be established based on exposure methods or metrics. There are a number of wildlife studies of PFOA or PFOS that include immune effects. Therefore, one of the considerations for this topic would be whether or not to include wildlife. For this example, the wildlife studies were included and would be evaluated with the other animal data as a distinct group to reflect the observational study design, rather than controlled exposure experimental studies.

Table 16.1 PubMed search strategy

Box 1: PFOA and PFOS exposure search terms
perfluoroalkyl*[tiab] OR perfluorocaprylic[tiab] OR perfluorocarbon*[tiab] OR perfluorocarboxyl*[tiab] OR perfluorochemical*[tiab] OR (perfluorinated[tiab] AND (C8[tiab] OR carboxylic[tiab] OR chemical*[tiab] OR compound*[tiab] OR octanoic[tiab])) OR PFAA*[tiab] OR "fluorinated polymer"[tiab] OR "fluorinated polymers"[tiab] OR (fluorinated[tiab] AND (polymer[tiab] OR polymers[tiab])) OR (fluorocarbon[tiab] AND (polymer[tiab] OR polymers[tiab])) OR Fluoropolymer*[tiab] OR (fluorinated[tiab] AND telomer*[tiab]) OR fluorotelomer*[tiab] OR fluoro-telomer*[tiab] OR fluorosurfactant*[tiab] OR "FC 143"[tiab] OR FC143[tiab] OR 335-67-1 [rn] OR Pentadecafluoroctanoate*[tiab] OR Pentadecafluorooctanoate*[tiab] OR pentadecafluoroctanoic[tiab] OR pentadecafluorooctanoic[tiab] OR "pentadecafluoro-1-octanoic"[tiab] OR "pentadecafluoro-n-octanoic"[tiab] OR "perfluoro-1-heptanecarboxylic"[tiab] OR perfluorocaprylic[tiab] OR perfluoroheptanecarboxylic[tiab] OR perfluoroctanoate[tiab] OR perfluorooctanoate[tiab] OR "perfluoro octanoate"[tiab] OR "perfluorooctanoic acid"[nm] OR perfluoroctanoic[tiab] OR perfluorooctanoic[tiab] OR "perfluoro octanoic"[tiab] OR "perfluoro-n-octanoic"[tiab] OR "perfluorooctanoyl chloride"[tiab] OR PFOA[tiab] OR APFO[tiab] OR 1763-23-1[rn] OR 307-35-7[rn] OR "1-octanesulfonic acid"[tiab] OR "1-perfluorooctanesulfonic"[tiab] OR "1-perfluoroctanesulfonic"[tiab] OR "heptadecafluoro-1-octanesulfonic"[tiab] OR "heptadecafluoro-1-octane sulfonic"[tiab] OR "heptadecafluorooctanesulfonic"[tiab] OR "heptadecafluorooctane sulfonic"[tiab] OR "heptadecafluoroctane sulfonic"[tiab] OR "perfluoroalkyl sulphonate"[tiab] OR perfluoroctanesulfonate[tiab] OR perfluorooctanesulfonate[tiab] OR "perfluoroctane sulfonate"[tiab] OR "perfluorooctane sulfonate"[tiab] OR "perfluoro-n-octanesulfonic"[tiab] OR perfluoroctanesulfonic[tiab] OR perfluorooctanesulfonic[tiab] OR "perfluorooctane sulfonic acid"[nm] OR "perfluoroctane sulfonic"[tiab] OR "perfluorooctane sulfonic"[tiab] OR perfluoroctanesulphonic[tiab] OR perfluorooctanesulphonic[tiab] OR "perfluoroctane sulphonic"[tiab] OR "perfluorooctane sulphonic"[tiab] OR perfluoroctylsulfonic[tiab] OR PFOS [tiab]

Table 16.1 (continued)

Box 2: Immune/immunotoxicology search terms
immunology[sh] OR immune[tiab] OR immunocomp*[tiab] OR immunogen*[tiab] OR immunolog*[tiab] OR immunotox*[tiab] OR immunotoxins[mh] OR immunity[tiab] OR autoimmun*[tiab] OR "host resistance"[tiab] OR immunocompetence[mh] OR "immune system"[mh] OR spleen[tiab] OR splenic[tiab] OR splenocyt*[tiab] OR thymus[tiab] OR thymic[tiab] OR thymocyt*[tiab] OR leukocyt*[tiab] OR granulocyt*[tiab] OR basophil*[tiab] OR eosinophil*[tiab] OR neutrophil*[tiab] OR lymph[tiab] OR lymphoid*[tiab] OR lymphocyt*[tiab] OR "b-lymphocyte"[tiab] OR "b-lymphocytes"[tiab] OR "t-lymphocyte"[tiab] OR "t-lymphocytes"[tiab] OR "killer cell"[tiab] OR "killer cells"[tiab] OR "NK cell"[tiab] OR "NK-cell"[tiab] OR "NK-cells"[tiab] OR macrophag*[tiab] OR "mast cell"[tiab] OR "mast cells"[tiab] OR monocyt*[tiab] OR phagocyt*[tiab] OR dendrit*[tiab] OR "t-cell"[tiab] OR "t cell"[tiab] OR "t cells"[tiab] OR "t-cells"[tiab] OR "T helper"[tiab] OR "T-helper"[tiab] OR "b-cell"[tiab] OR "b cell"[tiab] OR "b cells"[tiab] OR "b-cells"[tiab] OR antibod*[tiab] OR histamine*[tiab] OR histocompatib*[tiab] OR immunoglobulins[mh] OR immunoglobulin*[tiab] OR "immunoglobulin A"[tiab] OR IgA[tiab] OR "immunoglobulin D"[tiab] OR IgD[tiab] OR "immunoglobulin E"[tiab] OR IgE[tiab] OR "immunoglobulin G"[tiab] OR IgG[tiab] OR "immunoglobulin M"[tiab] OR IgM[tiab] OR "antigens, CD"[mh] OR CD3 [tiab] OR CD4 [tiab] OR CD8 [tiab] OR CD25 [tiab] OR CD27 [tiab] OR CD28 [tiab] OR CD29 [tiab] OR CD45*[tiab] OR cytokines[mh] OR cytokine*[tiab] OR chemokine*[tiab] OR inteferon*[tiab] OR interleukin*[tiab] OR "IL-6"[tiab] OR "IL-8"[tiab] OR lymphokine*[tiab] OR monokine*[tiab] OR ("tumor necrosis"[tiab] AND (factor[tiab] OR factors[tiab])) OR "TNF alpha"[tiab] OR "TNFalpha"[tiab] OR "immune system diseases"[mh] OR autoimmun*[tiab] OR addison[tiab] OR rheumatoid[tiab] OR glomerulonephritis[tiab] OR diabetes[tiab] OR graves[tiab] OR lupus[tiab] OR thyroiditis[tiab] OR hypersensitiv*[tiab] OR sensitization OR hyperresponsiv*[tiab] OR allergy[mh] OR allerg*[tiab] OR atopy[tiab] OR atopic[tiab] OR dermatitis[tiab] OR eczema[tiab] OR otitis[tiab] OR "ear infection"[tiab] OR "ear inflammation"[tiab] OR Respiratory tract infections [mh] OR (respiratory[tiab] AND infection*[tiab]) OR asthma[tiab] OR bronchitis[tiab] OR pneumonia[tiab] OR bronchiolitis[tiab] OR rhinitis[tiab] OR sinusitis[tiab] OR wheez*[tiab] OR crackle*[tiab] OR cough[mh] OR cough*[tiab] OR dyspnea[tiab] OR gastroenteritis[tiab] OR inflammation[mh] OR inflammat*[tiab] OR pro-inflammat*[tiab] OR anti-inflamm*[tiab] OR "inflammation mediators"[mh] OR autacoid*[tiab] OR eicosanoid*[tiab] OR prostaglandin*[tiab] OR immunomodulation[mh] OR immunomodul*[tiab] OR immunotherap*[tiab] OR vaccin*[tiab] OR immuniz*[tiab] OR immunosuppress*[tiab] OR desensitiz*[tiab] OR immunoproteins[mh] OR immunoprotein*[tiab] OR "c-reactive protein"[tiab] OR CRP[tiab] OR "complement component" [tiab] OR (complement[tiab] AND (C1 OR C2 OR C3 OR C4 OR C5 OR C6 OR C7 OR C8 OR C9))

The exposure × effects search strategy for PFOA- or PFOS-associated immune effects was developed by combining exposure terms in box #1 and immunotoxicology terms from box #2

Table 16.2 lists the immune outcomes considered relevant and categorizes them as more (primary) or less (secondary) predictive for immunotoxicity (i.e., how well do the assessed outcomes predict adverse immunological effects). Primary outcomes are considered to be the most direct, or applicable, to the project. Secondary outcomes are relevant, but less direct and can include upstream indicators, intermediate outcomes, or measures biologically-related to our primary outcomes.

For the evaluation of immunotoxicity, primary outcomes are those with more predictive value for immunotoxicity such as disease resistance assays and functional

Table 16.2 Eligibility table for inclusion criteria and directness of immune outcomes

Humans	Animals	*In vitro* assays
Primary outcomes	**Primary outcomes**	**Primary outcomes**
Immune-related diseases and measures of immune function	Disease resistance assay or measures of immune function	*Immune function assays following in vitro exposure to the test substance* (e.g., natural killer cell [NK] activity, phagocytosis or bacterial killing by monocytes, proliferation following anti-CD3 antibody stimulation of spleen cells or lymphocytes)
Immunosuppression (e.g., otitis, infections, or decreased vaccine antibody response);	*Disease resistance assays* (e.g., host resistance to influenza A or trichinella, changes in incidence or progression in animal models of autoimmune disease)	
Sensitization and allergic response (e.g., atopic dermatitis or asthma);	*Immune function assays following in vivo exposure to the test substance* (e.g., antibody response [T-cell dependent IgM antibody response (TDAR)], natural killer cell [NK] activity, delayed-type hypersensitivity [DTH] response, phagocytosis by monocytes, local lymph-node assay [LLNA])	
Autoimmunity (e.g., thyroiditis or systemic lupus erythematosus)		
Secondary outcomes	**Secondary outcomes**	**Secondary outcomes**
Immunostimulation[a] (e.g., unintended stimulation of humoral immune function)	*Observational immune endpoints* (e.g., lymphoid organ weight, lymphocyte counts or subpopulations, lymphocyte proliferation, cytokine production, serum antibody levels, serum or tissue autoantibody levels, or histopathological changes in immune organs)	*Observational immune endpoints following in vitro exposure to the test substance* (e.g., general mitogen-stimulated lymphocyte proliferation, cytokine production)
Observational immune endpoints (e.g., lymphocyte counts, lymphocyte proliferation, cytokine levels, serum antibody levels, or serum autoantibody levels)		

An outcome eligibility table defines the relevance and eligibility for screening references on an outcome basis. Later in the evaluation process, when rating confidence in bodies of evidence, this same table identifies the applicability or directness of outcomes. Primary outcomes are those with more predictive value for immunotoxicity such as disease resistance assays and would not be downgraded for indirectness. Secondary outcomes are those with less predictive value or observational parameters such as lymphoid cell counts that would be downgraded for indirectness

[a]Note that stimulation of the immune response is not adverse per se and most vaccine preparations include adjuvants to aid in stimulation of an immune response to microbes. It is generally agreed that stimulation of the immune system should not be disregarded (WHO 2012). Unintended immunostimulation will be considered for possible hazard in the context of potency and persistence of the elevated immune response. Because evaluation of immunostimulation is less well established for health assessment, outcomes that could be evaluated under autoimmunity or sensitization will be evaluated under these more established categories when possible

immune parameters. Secondary outcomes are those with less predictive value for immunotoxicity such as observational parameters including cell counts or cytokine levels. This dichotomy separating the more and less predictive measures of immunotoxicity is consistent with testing strategies that rely on more sensitive and predictive immune assays (see Luster et al. 1992; US EPA 1996a, b, 1998a) and the NTP and WHO methods to categorize the evidence of immune system toxicity. Under these systems, measures of immune function or the ability of the immune system to respond to a challenge are weighed more heavily than observational parameters (Germolec 2009; WHO 2012). For *in vitro* studies, we are interested in immune measures that may support the biological plausibility of observed immune outcomes. For example, *in vitro* stimulation of immunoglobulin E (IgE) production would support a functional measure of sensitization or allergic response, but it would not support suppression of the natural killer response.

The health effects are also defined in the context of current understanding of the biological relatedness of outcomes and effects are "grouped" for analyzing data on related effects to reflect the four main categories of immune response: immunosuppression, immunostimulation, sensitization and allergic response, and autoimmunity. Eligible publications must include an indicator of PFOA or PFOS exposure analyzed in relation to any one of the following primary or secondary outcomes listed in Table 16.2.

16.4.4 Extract Data from Studies

The published information relevant to the evaluation from included studies is captured in a database to facilitate critical evaluation of the results, including data summary and display using separate data collection forms for human, animal, and *in vitro* studies. Procedures specified in the protocol should address quality assurance procedures such as extraction in duplicate or individual extraction followed by review.

16.4.5 Assess Quality of Individual Studies

Study quality has long been considered within environmental health assessments as an important part of synthesizing the evidence to reach conclusions (WHO 1999). However, individual study quality has not been consistently or explicitly assessed. In fact, the definition of study quality varies widely across groups, and therefore an important aspect of systematic review is to be clear where and how study quality is assessed within an evaluation. Broadly speaking, study quality can includes:

- **Reporting quality** – how thoroughly the information about a study was reported.
- **Internal validity or risk of bias** – how credible are the findings based on study design and conduct.

- **External validity or directness and applicability** – how well a study addresses the topic under review.

Internal validity or risk of bias assessment of individual studies is considered critical and is the primary study quality assessment in the OHAT method. Reporting quality is considered as part of the risk of bias assessment, as studies that do not report sufficient detail to address a risk of bias question are given a higher risk of bias rating for that question. External validity is considered when rating confidence in the body of evidence. Assessment approaches that mix these different aspects of study quality or provide a single summary score are discouraged (Balshem et al. 2011; Higgins and Green 2011; Viswanathan et al. 2012). The OHAT framework avoids these issues and addresses study quality in multiple steps in an evaluation. When major limitations for internal validity or external validity are known in advance (e.g., unreliable methods to assess exposure or health outcome), the basis for excluding those studies can be outlined as an exclusion criteria in the protocol.

The OHAT risk-of-bias tool adapts and extends guidance and specific questions from the Agency for Healthcare Research and Quality (AHRQ) methods for systematic review (Viswanathan et al. 2012). There are a number of risk-of-bias tools to address human studies that differ in specifics, but all assess some common key issues such as whether there could be systematic differences in baseline characteristics between groups (e.g., Higgins et al. 2011; Johnson et al. 2014; Viswanathan et al. 2012). The AHRQ approach was selected because it included both the key risk of bias issues found across multiple other tools and provided a "parallel approach" to address experimental and observational studies with a single set of questions. OHAT used this parallel approach to extend a common set of risk of bias questions to also address experimental animal studies which have potential sources of bias that are conceptually similar to human trials. Individual risk-of-bias questions from the OHAT tool are designated as applicable only to certain types of study designs (e.g., human controlled trials, experimental animal studies, cohort studies, case-control studies, cross-sectional studies, case series or case reports), with a subset of the questions applying to each study design (Table 16.3).

All references are independently assessed for risk of bias by two reviewers who answer all of the applicable questions with one of four rating options (definitely low, probably low, probably high, or definitely high risk of bias) (CLARITY Group at McMaster University 2013). Disagreements are resolved by reaching agreement through discussion or consultation of subject matter experts. The guidance for answering each question, and criteria to discriminate among the four ratings is outlined in extensive detail in the protocol. This guidance is specific to study design, and an example is presented in Table 16.4 for experimental animal studies. Each relevant outcome or health effect within a study is evaluated separately. While most of the risk of bias ratings will likely be the same across different outcomes, two areas of risk of bias are likely to vary by outcome: (1) potential confounding, and (2) the outcome assessment method, including the relative impact that blinding or failing to blind outcome assessors to treatment group may have had on the recorded values (e.g., white blood cell count measured by an automated cell sorter vs.

Table 16.3 OHAT internal validity or risk-of-bias questions

	Experimental animal[a]	Human controlled trials[b]	Cohort	Case-control	Cross-sectional	Case series
Selection BIAS						
Was administered dose or exposure level adequately randomized?	X	X				
Randomization requires that each human subject or animal had an equal chance of being assigned to any study group including controls (e.g., use of random number table or computer generated randomization)						
Was allocation to study groups adequately concealed?	X	X				
Allocation concealment requires that research personnel do not know which administered dose or exposure level is assigned at the start of a study. Human studies also require that allocation be concealed from human subjects prior to entering the study						
Note: (1) a question under performance bias addresses blinding of personnel and human subjects to treatment during the study; (2) a question under detection bias addresses blinding of outcome assessors						
Were the comparison groups appropriate?			X	X	X	
Comparison group appropriateness refers to having similar baseline characteristics between the groups aside from the exposures and outcomes under study						
Confounding BIAS						
Did the study design or analysis account for important confounding and modifying variables?	X	X	X	X	X	X
Note: a parallel question under detection bias addresses reliability of the measurement of confounding variables						
Did researchers adjust or control for other exposures that are anticipated to bias results?	X	X	X	X	X	X

Performance BIAS						
Were experimental conditions identical across study groups?	X					
Did researchers adhere to the study protocol?	X	X	X	X	X	X
Were the research personnel and human subjects blinded to the study group during the study?	X	X				
Blinding requires that study scientists do not know which administered dose or exposure level the human subject or animal is being given (i.e., study group). Human studies require blinding of the human subjects when possible						
Attrition/exclusion BIAS						
Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	
Attrition rates are required to be similar and uniformly low across groups with respect to withdrawal or exclusion from analysis						
Detection BIAS						
Were the outcome assessors blinded to study group or exposure level?	X	X	X	X	X	X
Blinding requires that outcome assessors do not know the study group or exposure level of the human subject or animal when the outcome was assessed						
Were confounding variables assessed consistently across groups using valid and reliable measures?	X	X	X	X	X	X
Consistent application of valid, reliable, and sensitive methods of assessing important confounding or modifying variables is required across study groups						
Note, a parallel question under selection bias addresses whether design or analysis account for confounding						
Can we be confident in the exposure characterization?	X	X	X	X	X	X
Confidence requires valid, reliable, and sensitive methods to measure exposure applied consistently across groups						

(continued)

Table 16.3 (continued)

	Experimental animal[a]	Human controlled trials[b]	Cohort	Case-control	Cross-sectional	Case series
Can we be confident in the outcome assessment?	X	X	X	X	X	X
Confidence requires valid, reliable, and sensitive methods to assess the outcome and the methods should be applied consistently across groups						
Selective reporting BIAS						
Were all measured outcomes reported?	X	X	X	X	X	X
Other						
Were there no other potential threats to internal validity (e.g., statistical methods were appropriate)?						
On a project specific basis, additional questions for other potential threats to internal validity can be added and applied to study designs as appropriate						

The OHAT risk-of-bias questions are applied to evaluate the risk of bias of studies on an outcome basis. The study design determines which questions are applicable as indicated in the table by an "X" for each question that applies to a given study design. Risk-of-bias ratings are developed by answering each applicable question with one of four options (definitely low, probably low, probably high, or definitely high risk of bias)

[a]Experimental animal studies are controlled exposure studies. Non-human animal observational studies could be evaluated using the design features of observational human studies such as cross-sectional study design

[b]Human Controlled Trials (HCTs): studies in humans with a controlled exposure, including Randomized Controlled Trials (RCTs) and non-randomized experimental studies

[c]Cross-sectional studies include population surveys with individual data (e.g., National Health and Nutrition Examination Survey or NHANES) and population surveys with aggregate data (i.e., air pollution exposure estimated by zip code)

Table 16.4 Example risk of bias guidance

Definitely low risk of bias
There is direct evidence that animals were allocated to any study group including controls using a method with a random component. Acceptable methods of randomization include: referring to a random number table, using a computer random number generator, coin tossing, shuffling cards or envelopes, throwing dice, or drawing of lots (Higgins and Green 2011). Restricted randomization (e.g., blocked randomization) to ensure particular allocation ratios will be considered low risk of bias. Similarly, stratified randomization and minimization approaches that attempt to minimize imbalance between groups on important factors prognostic factors (e.g., body weight) will be considered acceptable. This type of approach is used by NTP, i.e., random number generator with body weight as a covariate. Please note that investigator-selection of animals from a cage is not considered random allocation because animals may not have an equal chance of being selected, e.g., investigator selecting animals with this method may inadvertently choose healthier, easier to catch, or less aggressive animals. Use of a concurrent control group is required as an indication that randomization covered all study groups.
Probably low risk of bias
There is indirect evidence that animals were allocated to any study group including controls using a method with a random component (i.e., authors state that allocation was random, without description of the method used) **OR** it is deemed that allocation without a clearly random component during the study would not appreciably bias results. For example, approaches such as biased coin or urn randomization, replacement randomization, mixed randomization, and maximal randomization may require consultation with a statistician to determine risk-of-bias rating (Higgins and Green 2011). Use of a concurrent control group is required as an indication that randomization covered all study groups.
Probably high risk of bias
There is indirect evidence that animals were allocated to study groups using a method with a non-random component **OR** there is insufficient information provided about how subjects were allocated to study groups. Non-random allocation methods may be systematic, but have the potential to allow researchers to anticipate the allocation of animals to study groups (Higgins and Green 2011). Such "quasi-random" methods include investigator-selection of animals from a cage, alternation, assignment based on shipment receipt date, date of birth, or animal number. A study with indirect evidence that there was a lack of a concurrent control group is another indication that randomization to all study groups was not conducted.
Definitely high risk of bias
There is direct evidence that animals were allocated to study groups using a non-random method including judgment of the investigator, the results of a laboratory test or a series of tests (Higgins and Green 2011). A study reporting lack of a concurrent control group is another indication that randomization to all study groups was not conducted.

Risk of bias guidance specific for experimental animal studies is outlined below for the question "Was administered dose or exposure level adequately randomized?"

behavioral observations made by trained research personnel). There is currently active methods development for risk of bias tools to address the types of evidence typically considered in environmental health – observational human, experimental animal, and *in vitro* studies. While assessing risk of bias of individual studies is critical to an environmental health assessment, the specific approach used is less important than clear documentation of the method used, along with consistent application of that method.

16.4.5.1 Assessing Risk of Bias for PFOA/PFOS Immunotoxicity Studies

Exposure, confounders, and outcome-specific modifications are the three areas of the risk of bias assessment that are likely to be the most evaluation-specific. Acceptable exposure measurements will depend on the chemical under study and the known confounders will vary by chemical and outcome. Sex and age are important confounders for evaluating immune effects because age and sex-dependent changes in immune function or observational parameters such as circulating immunoglobulin levels are common (WHO 2012). Immune-specific outcome guidance should describe the best methods and potential problems for immune assays used to measure outcomes found in the dataset. It is helpful if these criteria are described in lists or tabular form so that it can be updated quickly and shared easily to ensure the guidance is applied consistently across all studies. Inclusion of both older and newer outcome and exposure assessment methods and synonymous terms will aide reviewers to reconcile and assess the breadth of methods in the published literature, so the guidance should not only cover current terminology and methods. An abbreviated example describing discriminating risk of bias ratings for assays of antibody function (i.e., outcome assessment) are outlined below (the full guidance used is available here: http://ntp.niehs.nih.gov/ntp/ohat/evaluationprocess/appendix_2_pfoa_pfos_riskofbias.pdf).

Example

Question: Can we be confident in the outcome assessment?
Information necessary to reach a "Definitely low risk of bias" rating

- Direct evidence that immunization antigen batch/lot is the same for all treatment groups
- Direct evidence that antigen batch/lot is the same for immunization and the plating/assay

Information necessary to reach a "Probably low risk of bias" rating

- Indirect evidence that antigens used for immunizations are from the same batch/lot for all treatment groups
- Indirect evidence that antigens used for plating/assay are from the same batch/lot for all treatment groups

Information necessary to reach a "Probably high risk of bias" rating

- Indirect evidence that immunization antigens differed across treatment groups
- Indirect evidence that plating/assay antigens differed across treatment groups

Information necessary to reach a "Definitely high risk of bias" rating

- Direct evidence that immunization antigens differed across treatment groups
- Direct evidence that plating/assay antigens differed across treatment groups

16.4.6 Rate Confidence in the Body of Evidence

Groups of studies with data on an outcome (or biologically related outcomes) comprise a body of evidence that proceeds together through the evaluation process. Confidence ratings are developed separately for the human and animal bodies of evidence by considering the strengths and weaknesses of collections of studies with similar design features (Fig. 16.3). Ratings reflect confidence that the study findings accurately represent the true association between exposure to a substance and an effect. The OHAT method for rating confidence is based on the Grading of Recommendations Assessment, Development, and Evaluation Working Group (GRADE) (Guyatt et al. 2011) and AHRQ approaches (Balshem et al. 2011; Lohr 2012). This methodology is largely consistent among authoritative systematic review groups, including the Cochrane Collaboration (Schünemann et al. 2012). In the OHAT Approach ratings are developed on a 4-point scale to indicate the level of confidence in the body of evidence (High, Moderate, Low, and Very Low) consistent with the recommendations of the CLARITY Group at McMasters University (2013).

16.4.6.1 Initial Confidence Rating

For each body of evidence, studies are given an initial confidence rating by the presence or absence of four key study design features and then studies that have the same initial rating are considered together as a subgroup. The design features

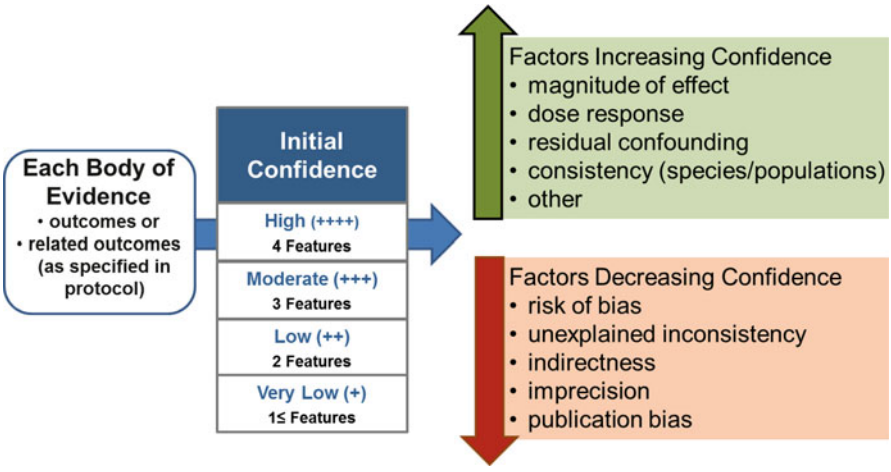

Fig. 16.3 Rating confidence in the body of evidence
Confidence ratings are developed for the human and animal bodies of evidence separately by considering the strengths and weaknesses of the collection of studies with similar design features

consider the ability of the study design to determine causality or assurance that exposure preceded and was associated with the outcome. Studies are differentiated based on whether or not: (1) exposure to the substance was controlled, (2) exposure occurred prior to the development of the outcome, (3) outcomes were assessed on the individual level not the population level, and (4) the study included a comparison group. Only experimental studies have controlled exposure, and therefore these studies will generally have all four features and be initially rated "High". Observational studies do not have controlled exposure and are differentiated by presence or absence of the three remaining study design features.

16.4.6.2 Downgrading and Upgrade Confidence Rating for Factors that Affect Confidence in the Results

The initial rating is downgraded for factors that decrease confidence or upgraded for factors that increase confidence in the results. Then, confidence across all available study designs and biologically related outcomes is assessed. The reasons for downgrading or upgrading confidence may not fit neatly into a single factor. If the decision to downgrade is borderline for two factors, the body of evidence is downgraded once to account for both partial concerns. Confidence should not be not downgraded twice for what is essentially the same limitation that may apply to more than one factor (or upgraded twice for the same asset). The protocol may specify severe factors that could downgrade confidence by two levels (typically it is moved only one level).

Five properties of the body of evidence (risk of bias, unexplained inconsistency, indirectness, imprecision, and publication bias) are considered to determine if the initial confidence rating should be downgraded. For each of the properties, a judgment is made and documented regarding whether or not there are issues that decrease the confidence rating for each property for the outcome.

- **Risk of bias of the body of evidence**: risk-of-bias assessments of individual studies developed earlier serve as the basis for an overall risk of bias conclusion for the body of evidence
- **Unexplained inconsistency**: large variability in the magnitude or direction of estimates of effect across studies, that cannot be explained by other factors (e.g., exposure assessment method, population characteristics, funding source)
- **Indirectness**: external validity (outcome, exposure, or population differs from that of the evaluation question [e.g., oral exposure studies may be downgraded as indirect evidence for an evaluation of effects from inhalation exposure]) or indirect measures of the health outcome
- **Imprecision**: lack of certainty for an estimate of effect for a specific outcome (often reflected in very wide confidence intervals around effect estimates)
- **Publication bias**: selective reporting or non-reporting of entire studies

Similarly, four properties of the body of evidence (large magnitude of effect, dose-response, residual confounding, and cross-species/population/study

consistency) are considered to determine if the confidence rating should be upgraded. Again, a judgment is made and documented regarding whether or not there are factors that increase the confidence rating for each property for the outcome.

- **Large magnitude of effect**: an observed effect that is sufficiently large such that it is unlikely to have occurred by chance despite possible unaccounted for confounding factors
- **Dose-response**: plausible dose-response relationship demonstrated between level of exposure and outcome
- **Residual confounding**: consideration of confounding factors, including the healthy worker effect or effect modification, that would bias the effect estimate towards the null – yet an effect is still seen
- **Cross-species/population/study consistency**: consistent results reported across multiple experimental animal models or species; or across populations that differ in factors such as time, location, and/or exposure levels; or studies with different design features.

16.4.6.3 Combine Confidence Conclusions for All Study Types and Multiple Outcomes

When considering evidence across study types and multiple outcomes, conclusions are based on the evidence with the highest confidence. While confidence ratings are initially set based on key design features of the available studies for a given outcome (e.g., for experimental studies separately from observational studies), only studies with the highest confidence rating form the basis for the final confidence conclusion. At this point, consistency of results across study designs should also be considered and could contribute to an upgraded confidence conclusion across the combined body of evidence.

If the only available body of evidence receives a "Very Low" confidence rating, then the evaluator should consider whether or not to move conclusions for those outcomes forward for hazard assessment. Effectively, "Very Low" confidence can be treated the same as having no data.

After confidence conclusions are developed for a given outcome, conclusions for multiple outcomes and the entire evaluation are developed. The project-specific definition of an outcome and the grouping of biologically related outcomes used in this step follow the approach defined in the protocol; any deviations are taken with care, justified, and documented. When outcomes are sufficiently biologically related that they may inform confidence on the overall health outcome, confidence conclusions may be developed in two steps. Each outcome would first be considered separately. Then, the related outcomes would be reconsidered together for properties that relate to downgrading and upgrading the body of evidence.

16.4.6.4 Rating Confidence for PFOA/PFOS Immunotoxicity Studies

Of the 80 animal studies on PFOA and PFOS immune effects that were identified through the literature search process (see Fig. 16.1) there were 5 PFOA studies and 8 PFOS studies that reported results on antibody response data. All of the PFOS animal studies (Dong et al. 2009, 2011; Keil et al. 2008; Lefebvre et al. 2008; Peden-Adams et al. 2008; Qazi et al. 2010; Zheng et al. 2009, 2011) would be given a high initial confidence based on having the four key study design features consistent with most experimental studies. Then this 8-study body of evidence would be evaluated for the five factors that may decrease confidence and four factors that may increase confidence that the study findings accurately represent the true association between PFOS exposure and the antibody response (independent of the presence or direction of a reported effect).

An assessment of "Indirectness" will be used as an example to show how the properties of the body of evidence would be considered for each factor. Indirectness reflects both external validity and indirect measures of the health outcome. The key questions, PECOTS statement, and eligibility criteria outlined in the protocol would state the population, exposure, outcome, comparator, timing, and settings of interest for the evaluation. A strict PECOTS statement and eligibility definition could essentially eliminate all indirect evidence; however, for most datasets the eligibility criteria define the directly relevant data as well as upstream or indirect data or populations. These factors are considered in more depth when determining if studies deviated from those of most interest to the evaluation. Experimental animals are considered directly relevant to the animal evidence stream and therefore would not be downgraded. The outcomes of interest would also be defined in the protocol. Table 16.2 outlines those outcomes and specifies that functional outcomes such as the antibody response are primary outcomes. Therefore, the antibody response data would not be downgraded as they are direct measures of an outcome of interest with good predictive value for the evaluation of immunotoxicity. The following summary outlines the decision not to downgrade for indirectness for this outcome.

Indirectness Rating for PFOS Animal Antibody Data:
Rating = "Not Serious," Therefore No Downgrade

- Exposure (PFOS) and model (experimental animal studies in mice and rats) directly relevant
- Antibody response is a primary outcome with good predictive value for immunotoxicity
- SRBC IgM response by PFC or ELISA are among the best measures of antibody response

An evidence profile should be developed to summarize each of the downgrade and upgrade decisions to support and communicate the scientific judgments made to reach a confidence rating for the body of evidence. Table 16.5 illustrates how a

Table 16.5 Example evidence profile

Body of evidence	Risk of bias	Unexplained inconsistency	Indirectness	Imprecision	Publication bias	Magnitude	Dose response	Residual confounding	Consistency across species/ model	Final rating
Example of the type of information that should be in an evidence profile										
Human or animal	Serious or not serious	Serious or not serious	Serious or not serious	Serious or not serious	Detected or undetected	Large or not large	Yes or no	Yes or no	Yes or no	**Final rating**
(# Studies) initial rating	• Describe trend • Describe key questions • Describe issues	• Describe results in terms of consistency • Explain apparent inconsistency (if it can be explained)	• Discuss use of upstream indicators or populations with less relevance	• Discuss ability to disti nguish treatment from control • Describe confidence intervals	• Discuss factors that might indicate publication bias (e.g., funding, lag)	• Describe magnitude of response	• Outline evidence for or against dose response	• Address whether there is evidence that confounding would bias toward null	• Describe cross species, model, or population consistency	**High, moderate, or low**
Endpoint: functional antibody response (example "hypothetical" illustration for PFOS)										
Animal	Not serious	Not serious	Not serious	Not serious	Undetected	Not large	**Yes (increase)**	No	No	**High**
(8 PFOS studies) Initial Rating • High	• General low • Key question –Randomize= mixed low and probably high –Outcome=low • Probably high for allocation concealment	• Consistent suppression • Potential inconsistent response, but differed by: –Species (rat vs mouse), –Outcome (IgG vs IgM), –Antigen (SRBC vs KLH)	• SRBC IgM response by PFC or ELISA are among best measures of antibody response	• General small, confidence interval (CI) • Overlapping CIs between control and exposed	• No evidence of lag bias • Funding –Government –Universities –Industry	• Not sufficiently large to overcome potential bias	• Dose-response observed in multiple studies	• No evidence of confounding that would bias toward null	• All positive results from mice	Started high No serious downgrades Upgrade for dose-response Final rating would be High

simple evidence profile can be constructed. This table provides examples of the type of information required to support the basis of scientific judgments as well as hypothetical conclusions for the PFOS antibody response body of evidence. An actual profile may have many lines of evidence depending on whether the data set has multiple bodies of evidence on a given outcome (e.g., wildlife observational studies and experimental animal studies), data streams (human and animal), and outcomes (functional antibody response data and observational data such as total IgG or IgM levels).

16.4.7 Translate Confidence into Level of Evidence (Toxicity or No Toxicity)

The step to translate confidence into level of evidence is a simple step that incorporates the direction of the effect (i.e., whether the data support toxicity or no toxicity) into the confidence conclusions developed previously. The strategy uses four terms that reflect both the confidence in the body of evidence for a given outcome and the direction of effect. If data support that exposure to the substance is associated with a heath effect, the three descriptors used ("High," "Moderate," or "Low") **level of evidence** directly translate from the confidence ratings ("High," "Moderate," or "Low") **confidence in the body** of evidence. If the data support that exposure is not associated with the health effect in question, then a separate descriptor ("Evidence of No Health Effect") is used to indicate confidence that the substance is not associated with a health effect. There is inherent difficulty in proving a negative, and as such a conclusion of evidence of no health effect is only reached when there is "High" confidence in the body of evidence. A "Low" or "Moderate" level of evidence results in a conclusion of inadequate evidence to reach a conclusion.

16.4.8 Integrate Evidence to Develop Hazard Identification Conclusions

The last step in evidence integration is to develop hazard identification conclusions from the level of evidence ratings by integrating the animal and human evidence with the additional consideration of the impact of mechanistic data (Fig. 16.4). For a given health effect, the highest level of evidence from each of the evidence streams is combined in the final step of the evidence assessment process. In the absence of either human or animal data, conclusions can be developed on the remaining evidence stream by treating the missing data as a "Low" level of evidence.

The five hazard identification conclusion categories used by OHAT are "Known," "Presumed," "Suspected," "Not classifiable," and "Not identified" to be a hazard to humans. Just as confidence conclusions can be developed on individual outcomes or

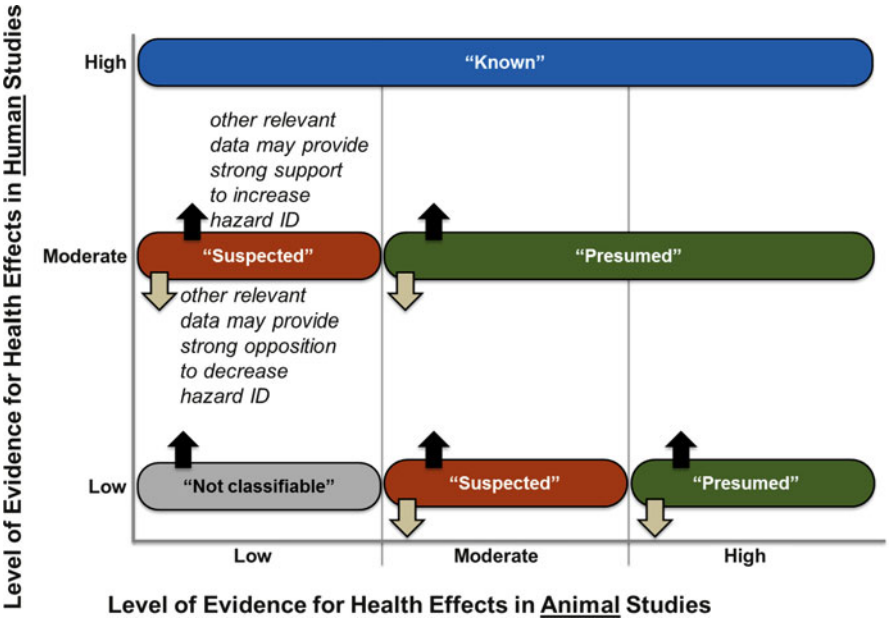

Fig. 16.4 Integrating evidence to develop hazard identification conclusions
Hazard identification conclusions reflect an integration of the level of evidence ratings from the human and animal evidence with the additional consideration of the impact of mechanistic data (Reproduced from Environmental Health Perspective: http://ehp.niehs.nih.gov/1307972/)

groups of biologically related outcomes as defined in the protocol, hazard identification conclusions may be developed as appropriate based on the objectives and the available data. The support for conclusions should be documented along with the rationale stating which outcomes were incorporated into each conclusion.

A schematic of how evidence is integrated into hazard identification conclusions is provided in Fig. 16.4. A "High" level of human evidence will result in a "Known" hazard identification conclusion, as animal and mechanistic data will not impact the decision if the human evidence is strong and consistent. Conversely, even if the human evidence has limitations and only has a "Moderate" level of evidence, a hazard identification conclusion of "Known" can be still be reached if there is both a "High" animal level of evidence and mechanistic evidence providing strong additional support for an association. In such case, additional experimental or longitudinal studies in humans are not necessary to identify a known hazard.

On the other end of the spectrum, it is anticipated that few systematic reviews would be initiated for topics where both the human and animal evidence is low and strong mechanistic evidence does not exist. As indicated in Fig. 16.4, such a review would result in a "Not classifiable" hazard identification conclusion.

When either the human or animal data support a conclusion of no health effect, the level-of-evidence conclusions for the two evidence streams are evaluated

together (human and animal) and the impact of mechanistic data is considered. If the human and animal level-of-evidence conclusions support no health effect, and this is not opposed by strong mechanistic data, the hazard identification conclusion is "not identified." While theoretically possible, we have not included in the schematic the hypothetical situation when the evidence streams are in direct conflict with one another (one has high confidence of no effect and the other shows evidence of an effect). Such a scenario, if possible, could be resolved by scoping and problem formulation to address this biologically implausible scenario.

When the levels of evidence are "Moderate" or "Low" for both human and animal evidence streams, mechanistic evidence has the greatest potential to influence the final hazard identification conclusions of "Presumed" or "Suspected." We anticipate that such hazard identification decisions will be accompanied by detailed description of the scientific considerations that support hazard identifications in this middle arena. Development of parallel methods to evaluate mechanistic evidence and incorporate predictive toxicology information is an area of active research.

16.4.8.1 Integrating Human, Animal and Mechanistic PFOA/PFOS Immunotoxicity Evidence

As previously indicated, only subsets of the available PFOA/PFOS immunotoxicity evidence were used in this illustration of the systematic review concepts, so no hazard identification conclusions have been developed. If the human level of evidence were to be "Moderate" or "Low" then addition mechanistic evidence would be considered. A walkthrough of the PFOS antibody response example will illustrate the process. Using the hypothetical animal data confidence conclusion of "High" confidence in the animal evidence for suppressed antibody response in Table 16.5, this data would support a "High" level of evidence for the animal data on humoral immunity because there is "high" confidence of toxicity or an effect. Using this framework, the potential hazard identification conclusions begin to emerge even before reviewing the human data. The animal data alone ("high" level-of-evidence) would support a conclusion of "suspected" to be a hazard to humans, and this would be the final hazard conclusion if there were no human data, or human data with "moderate" or "low" level-of-evidence. Human data with "high" level-of-evidence for antibody suppression would result in a hazard identification conclusion of "known."

A conclusion of "known" to be a hazard to humans could also be reached with "high" level of evidence from animal studies in the absence of human data if the mechanistic evidence provided strong support for the biological plausibility of the effect. In this case, the type of evidence providing that increased support would be *in vitro* or mechanistic studies indicating dose and temporal support for reductions in the antibody response. For example, if PFOS-associated reductions in antigen processing or presentation were consistently reported at or below the concentration of PFOA associated with antibody suppression. Consistent disruption in the antibody response across model systems, or multiple steps in the antibody response

such as interference with antigen presenting cells, as well as disruption of T-cell signaling, and reduced B-cell secretion of antibodies would provide stronger mechanistic evidence and support upgrading the hazard identification conclusions.

16.5 Summary

An approach to evaluate health effects of PFCs has been presented incorporating examples from a case study on PFOA/PFOS and immunotoxicity. The final goal of identifying the potential for hazard to human health can be reached by a systematic approach to evaluating the available literature. Interpretation of the final conclusion of such an evaluation is strengthened by using objective and reproducible methods and documenting scientific judgments in a transparent manner. The OHAT Approach to hazard identification, founded on systematic review methodology from clinical medicine, provides a scientifically rigorous approach to environmental health assessment and improves communication with the wider community of stakeholders. Future risk assessment of PFCs will also be strengthened by a clear hazard identification assessment, such as this example of PFOA/PFOS and immunotoxicity.

Acknowledgements We appreciate the helpful comments received during draft development from Dr. Kristina Thayer and Dr. Chad Blystone. We also acknowledge that this document was substantially informed by our experience with case studies developed to test the OHAT framework for systematic review and evidence integration. A case study evaluation of PFOA and PFOS immunotoxicity (http://ntp.niehs.nih.gov/go/36501) was used to provide input for refining the OHAT framework.

References

AHRQ (Agency for Healthcare Research and Quality) (2013) AHRQ training modules for the systematic reviews methods guide. Available: http://www.effectivehealthcare.ahrq.gov/index.cfm/tools-and-resources/slide-library/. Accessed 11 Oct 2013

Ahuja V, Eisenblatter M, Ignatius R, Stahlmann R (2009) Ammonium perfluorooctanoate substantially alters phenotype and cytokine secretion of human monocyte-derived dendritic cells in vitro. Immunopharmacol Immunotoxicol 31:641–646

ATSDR (Agency for Toxic Substances and Disease Registry) (2009) Draft toxcological profile for perfluoroalkyls. Division of Toxicology and Environmental Medicine/Applied Toxicology Branch. US Department of Health and Human Services, Atlanta, p 404. Available: http://www.atsdr.cdc.gov/toxprofiles/tp200.pdf

Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J et al (2011) GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 64:401–406

Birnbaum LS, Thayer KA, Bucher JR, Wolfe MS (2013) Implementing systematic review at the National Toxicology Program. Environ Health Perspect 121:A108

CDC (Centers for Disease Control and Prevention) (2009) Fourth national report on human exposure to environmental chemicals. U.S. Department of Health and Human Services, Atlanta. Available: http://www.cdc.gov/exposurereport/. Accessed 5 Apr 2013

CDC (Centers for Disease Control and Prevention) (2014) Fourth national report on human exposure to environmental chemicals: updated tables, August 2014. U.S. Department of Health and Human Services, Atlanta. Available: http://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Aug2014.pdf. Accessed 2 Sept 2014

CLARITY Group at McMaster University (2013) Tools to assess risk of bias in cohort studies, case control studies, randomized controlled trials, and longitudinal symptom research studies aimed at the general population. Available: http://www.evidencepartners.com/resources/. Accessed 15 Jan 2013

Corsini E, Avogadro A, Galbiati V, dell'Agli M, Marinovich M, Galli CL et al (2011) In vitro evaluation of the immunotoxic potential of perfluorinated compounds (PFCs). Toxicol Appl Pharmacol 250:108–116

Corsini E, Sangiovanni E, Avogadro A, Galbiati V, Viviani B, Marinovich M et al (2012) In vitro characterization of the immunotoxic potential of several perfluorinated compounds (PFCs). Toxicol Appl Pharmacol 258:248–255

DeWitt JC, Peden-Adams MM, Keller JM, Germolec DR (2012) Immunotoxicity of perfluorinated compounds: recent developments. Toxicol Pathol 40:300–311

Dong GH, Zhang YH, Zheng L, Liu W, Jin YH, He QC (2009) Chronic effects of perfluorooctanesulfonate exposure on immunotoxicity in adult male C57BL/6 mice. Arch Toxicol 83:805–815

Dong GH, Liu MM, Wang D, Zheng L, Liang ZF, Jin YH (2011) Sub-chronic effect of perfluorooctanesulfonate (PFOS) on the balance of type 1 and type 2 cytokine in adult C57BL6 mice. Arch Toxicol 85:1235–1244

EFSA (European Food Safety Authority) (2010) Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA J 90. Available: http://www.efsa.europa.eu/en/efsajournal/pub/1637.htm. Accessed 18 Jan 2012

Fair PA, Driscoll E, Mollenhauer MA, Bradshaw SG, Yun SH, Kannan K et al (2011) Effects of environmentally-relevant levels of perfluorooctane sulfonate on clinical parameters and immunological functions in B6C3F1 mice. J Immunotoxicol 8:17–29

Germolec D (2009) Explanation of levels of evidence for immune system toxicity. Research Triangle Park. Available: http://ntp.niehs.nih.gov/go/9399

Grandjean P, Andersen EW, Budtz-Jorgensen E, Nielsen F, Molbak K, Weihe P et al (2012) Serum vaccine antibody concentrations in children exposed to perfluorinated compounds. JAMA 307:391–397

Granum B, Haug LS, Namork E, Stolevik SB, Thomsen C, Aaberge IS et al (2013) Pre-natal exposure to perfluoroalkyl substances may be associated with altered vaccine antibody levels and immune-related health outcomes in early childhood. J Immunotoxicol 10:373–379

Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J et al (2011) GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 64:383–394

Han Z, Liu Y, Wu D, Zhu Z, Lü C (2012) Immunotoxicity and hepatotoxicity of PFOS and PFOA in tilapia (Oreochromis niloticus). Chin J Geochem 31:424–430

Hart K, Gill VA, Kannan K (2009) Temporal trends (1992–2007) of perfluorinated chemicals in Northern Sea Otters (Enhydra lutris kenyoni) from South-Central Alaska. Arch Environ Contam Toxicol 56:607–614

Higgins J, Green S (2011) Cochrane handbook for systematic reviews of interventions. Part version 5.1.0 [updated March 2011]: The Cochrane Collaboration

Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD et al (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 343:d5928

Johnson P, Sutton P, Atchley D, Koustas E, Lam J, Robinson K et al (2013) Applying the navigation guide: case study #1. The impact of developmental exposure to perfluorooctanoic acid (PFOA) on fetal growth. A systematic review of the human evidence - protocol. Available: http://prhe.ucsf.edu/prhe/pdfs/PFOA%20Human%20Protocol.pdf

Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S et al (2014) The navigation guide-evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. Environ Health Perspect 122:1028–1039

Kannan K, Perrotta E, Thomas NJ (2006) Association between perfluorinated compounds and pathological conditions in southern sea otters. Environ Sci Technol 40:4943–4948

Keil DE, Mehlmann T, Butterworth L, Peden-Adams MM (2008) Gestational exposure to perfluorooctane sulfonate suppresses immune function in B6C3F1 mice. Toxicol Sci 103:77–85

Keller JM, Kannan K, Taniyasu S, Yamashita N, Day RD, Arendt MD et al (2005) Perfluorinated compounds in the plasma of loggerhead and Kemp's ridley sea turtles from the southeastern coast of the United States. Environ Sci Technol 39:9101–9108

Koustas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S et al (2014) The navigation guide-evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. Environ Health Perspect 122:1028–1039

Lefebvre DE, Curran I, Armstrong C, Coady L, Parenteau M, Liston V et al (2008) Immunomodulatory effects of dietary potassium perfluorooctane sulfonate (PFOS) exposure in adult Sprague-Dawley rats. J Toxicol Environ Health A 71:1516–1525

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP et al (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 62:e1–e34

Lohr KN (2012) Grading the strength of evidence. In: The Agency for Healthcare Research and Quality (AHRQ) training modules for systematic reviews methods guide. Available: http://www.effectivehealthcare.ahrq.gov/index.cfm/slides/?pageAction=displaySlides&tk=18. Accessed 13 July 2012

Looker C, Luster MI, Calafat AM, Johnson VJ, Burleson GR, Burleson FG et al (2014) Influenza vaccine response in adults exposed to perfluorooctanoate and perfluorooctanesulfonate. Toxicol Sci 138:76–88

Luster MI, Portier C, Pait DG, White KL Jr, Gennings C, Munson AE et al (1992) Risk assessment in immunotoxicology. I. Sensitivity and predictability of immune tests. Fundam Appl Toxicol 18:200–210

Matcher DB (2012) Introduction to the methods guide for medical test reviews. In: Chang SM, Matchar DB (eds) Methods guide for medical test reviews. Agency for Healthcare Research and Quality, Rockville, pp 1.1–11

Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 62:1006–1012

NRC (National Research Council) (2009) Science and decisions: advancing risk assessment. 9780309120463. Press TNA, Washington, DC. Available: http://www.nap.edu/openbook.php?record_id=12209

NRC (National Research Council) (2013a) Critical aspects of EPA's IRIS assessment of inorganic arsenic – interim report (The National Academies Press). Press TNA, p 117. Available: http://www.nap.edu/catalog.php?record_id=18594. Accessed 7 Nov 2013

NRC (National Research Council) (2013b) Review of the IRIS process. Available: http://www8.nationalacademies.org/cp/projectview.aspx?key=49458. Accessed 1 Nov 2013

Peden-Adams MM, Keller JM, Eudaly JG, Berger J, Gilkeson GS, Keil DE (2008) Suppression of humoral immunity in mice following exposure to perfluorooctane sulfonate. Toxicol Sci 104:144–154

Qazi MR, Abedi MR, Nelson BD, DePierre JW, Abedi-Valugerdi M (2010) Dietary exposure to perfluorooctanoate or perfluorooctane sulfonate induces hypertrophy in centrilobular hepatocytes and alters the hepatic immune status in mice. Int Immunopharmacol 10:1420–1427

Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C et al (2013) A survey of frameworks for best practices in weight-of-evidence analyses. Crit Rev Toxicol 43:753–784

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA (2014) Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect 122:711–718

Samson D, Schoeles KM (2012) Developing the topic and structuring systematic reviews of medical tests: utility of PICOTS, analytic frameworks, decision trees, and other frameworks. In: Chang SM, Matchar DB (eds) Methods guide for medical test reviews. Agency for Healthcare Research and Quality, Rockville, pp 2.1–14

Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P et al (2012) Chapter 12: Interpreting results and drawing conclusions. In: Higgins JPT, Green S (eds) Cochrane handbook for systematic reviews of interventions version 510 [updated March 2011]. The Cochrane Collaboration

Steenland K, Zhao L, Winquist A, Parks C (2013) Ulcerative colitis and perfluorooctanoic acid (PFOA) in a highly exposed population of community residents and workers in the mid-Ohio valley. Environ Health Perspect 121:900–905

US EPA (1996a) Biochemicals test guidelines: OPPTS 880.3800 immune response. EPA/712/C-96/281. Office of Prevention Pesticides and Toxic Substances. US Environmental Protection Agency, Washington, DC, pp. 1–8. Available: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series880.htm

US EPA (US Environmental Protection Agency) (1996b) Biochemicals test guidelines: OPPTS 880.3550 immunotoxicity. EPA/712/C-96/280. Office of Prevention Pesticides and Toxic Substances, Washington, DC, pp 1–14. Available: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series880.htm

US EPA (US Environmental Protection Agency) (1998a) Health effects test guidelines: OPPTS 870.7800 immunotoxicity. EPA/712/C-98/351. Office of Prevention Pesticides and Toxic Substances, Washington, DC, pp 1–11. Available: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series870.htm

US EPA (US Environmental Protection Agency) (1998b) Guidelines for ecological risk assessment. EPA/630/R-95/002F. Office of Prevention Pesticides and Toxic Substances, Washington, DC, pp 1–188. Available: http://www.epa.gov/raf/publications/pdfs/ECOTXTBX.PDF

US EPA (US Environmental Protection Agency) (2006) 2010/2015 PFOA Stewardship Program. US Environmental Protection Agency. Available: http://www.epa.gov/oppt/pfoa/pubs/stewardship/index.html

US EPA (US Environmental Protection Agency) (2009) Long-chain Perfluorinated Chemicals (PFCs) action plan. Available: http://www.epa.gov/opptintr/existingchemicals/pubs/actionplans/pfcs.html

US EPA (US Environmental Protection Agency) (2013a) Materials submitted to the National Research Council part I: status of implementation of recommendations. Environmental Protection Agency: Integrated Risk Information System Program. Available: http://www.epa.gov/iris/pdfs/IRIS%20Program%20Materials%20to%20NRC_Part%201.pdf. Accessed 22 Feb 2013

US EPA (US Environmental Protection Agency) (2013b) Perfluorooctanoic acid (PFOA) and fluorinated telomers. Available: http://www.epa.gov/oppt/pfoa/

US EPA (US Environmental Protection Agency) (2014) Emerging contaminants – perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA). Available: http://www2.epa.gov/sites/production/files/2014-04/documents/factsheet_contaminant_pfos_pfoa_march2014.pdf

WHO (International Programme on Chemical Safety, World Health Organization) (1999) Principles for the assessment of risks to human health from exposure to chemicals. IPCS Environmental Health Criteria 210 (IPCS Environmental Health Criteria 210). Geneva, p 312. Available: http://www.inchem.org/documents/ehc/ehc/ehc210.htm

WHO (International Programme on Chemical Safety, World Health Organization) (2012) Guidance for immunotoxicity risk assessment for chemicals (IPCS Harmonization Project No 10). Geneva. Available: http://www.inchem.org/documents/harmproj/harmproj/harmproj10.pdf

Viswanathan M, Ansari M, Berkman ND, Chang S, Hartling L, McPheeters LM et al (Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews)

(2012) Assessing the risk of bias of individual studies when comparing medical interventions. Agency for Healthcare Research and Quality (AHRQ) (Publication No. 12-EHC047-EF). Available: http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=998. Accessed 3 Jan 2013

Wirth JR, Peden-Adams MM, White ND, Bossart GD, Fair PA (2014) In vitro PFOS exposure on immune endpoints in bottlenose dolphins (Tursiops truncatus) and mice. J Appl Toxicol 34:658–666

Woodruff TJ, Sutton P (2014) The navigation guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect 122:1007–1014

Zheng L, Dong GH, Jin YH, He QC (2009) Immunotoxic changes associated with a 7-day oral exposure to perfluorooctanesulfonate (PFOS) in adult male C57BL/6 mice. Arch Toxicol 83:679–689

Zheng L, Dong GH, Zhang YH, Liang ZF, Jin YH, He QC (2011) Type 1 and Type 2 cytokines imbalance in adult male C57BL/6 mice following a 7-day oral exposure to perfluorooctanesulfonate (PFOS). J Immunotoxicol 8:30–38