PROTOCOL TO EVALUATE THE EVIDENCE FOR AN ASSOCIATION BETWEEN PERFLUOROOCTANOIC ACID (PFOA) OR PERFLUOROOCTANE SULFONATE (PFOS) EXPOSURE AND IMMUNOTOXICITY

Project Leader: Andrew A. Rooney, Office of Health Assessment and Translation (OHAT), DNTP

Summary: OHAT is conducting a systematic review to evaluate the evidence for an association between exposure to PFOA or PFOS and immunotoxicity or immune-related health effects. The evaluation is anticipated to reach hazard identification conclusions for PFOA and PFOS-associated immunotoxicity following the protocol detailed in this document.

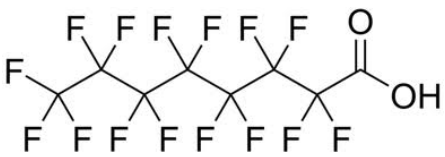
BACKGROUND AND SIGNIFICANCE

Background

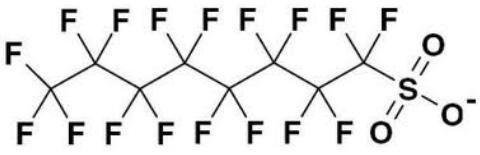
Perfluorooctanoic acid (PFOA) or perfluorooctane sulfonate (PFOS) are extremely persistent chemicals (**Figure 1**) that are widely distributed in the environment as a result of extensive use over the last 60 years in commercial and industrial applications including food packaging, lubricants, water-resistant coatings, and fire-retarding foams. They have high chemical stability and are not expected to degrade under typical environmental conditions (Lau *et al.* 2007, EFSA 2008, ATSDR 2009, US EPA 2014b). Once in surface water, apparent half-lives of PFOS and PFOA are 41 and 92 years respectively. Estimated half-lives in the human body are also long, ranging from 2 to 5 years (ATSDR 2009, Steenland *et al.* 2010, US EPA 2014b).

Toxicology studies of PFOA and PFOS in rodents have raised concerns about potential immune, developmental, reproductive, hepatic and other health effects (Lau *et al.* 2007, DeWitt *et al.* 2012, US EPA 2014b). There are much less health effects data available in humans, although some studies are available that investigated potential associations with serum PFOA and PFOS and similar outcomes including immune measures, decreased birth weight, and biomarkers of hepatotoxicity (Lau *et al.* 2007, Steenland *et al.* 2010). Through voluntary agreements, the primary manufacturer of PFOS phased out production in 2002 and PFOS is no longer manufactured in the United States (US EPA 2006, ATSDR 2009, US EPA 2009, 2014b, 2015). Similar arrangements have been made for PFOA and eight companies that manufacture PFOA have committed to eliminate emissions and product content by 2015 (US EPA 2006, ATSDR 2009, US EPA 2013, 2014b, 2015).

Figure 1: Structure of PFOA and PFOS



perfluorooctanoic acid (PFOA; CAS# 335-67-1)



perfluorooctane sulfonate (PFOS; CAS# 1763-23-1)

Although emissions have been dramatically reduced, the persistence and bioaccumulation of both PFOA and PFOS result in detectable levels in the U.S. population and therefore they remain of public health concern (US EPA 2014b). PFOA and PFOS were present in all serum samples from the general U.S. population in 1999 tested for perfluorinated compounds in the National Health and Nutrition Examination Survey (NHANES 1999-2000) (Calafat *et al.* 2007). While blood levels have declined from 1999 to 2010, PFOA (from 5.2 to 3.1µg/L geometric mean) and PFOS (from 30.4 to 9.3µg/L geometric mean) remain the two highest concentrations among perfluorinated compounds measured in the general U.S. population in the most recent National Report on Human Exposure to Environmental Chemicals for 2009-2010 (CDC 2015).

Several recent publications from 2012-2014 have linked PFOA and PFOS exposure to functional immune changes in humans that are consistent with evidence of PFOA- and PFOS-related immunotoxicity in 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 (Steenland *et al.* 2013, Looker *et al.* 2014). 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 that overlap with levels occurring in the general U.S. population (e.g., Peden-Adams *et al.* 2008, Fair *et al.* 2011, DeWitt *et al.* 2012, CDC 2015). 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., Keller *et al.* 2005, Kannan *et al.* 2006, Hart *et al.* 2009).

Significance

This OHAT evaluation is anticipated to reach hazard identification conclusions for PFOA- and PFOS-associated immunotoxicity. This evaluation will complement ongoing assessments of PFOA and PFOS toxicity being conducted by the Agency for Toxic Substances and Disease Registry (ATSDR) and EPA's Office of Water and Office of Pesticide Prevention and Toxics (US EPA 2013, 2014c). Data management will be conducted in a manner that permits sharing of data extraction files with the public and other agencies. This evaluation topic was also considered in conjunction with the ongoing assessment of perfluorinated compounds by NTP's testing program.

OVERALL OBJECTIVE AND SPECIFIC AIMS

The overall objective of this evaluation is to develop hazard identification conclusions as to whether exposure to PFOA or PFOS (or their salts) is associated with immunotoxicity or immune-related health effects by integrating levels of evidence from human, animal, and *in vitro* studies.

Specific Aims

- Identify literature reporting the effects of PFOA or PFOS exposure on immune endpoints in humans, animals (experimental and wildlife), or *in vitro* model systems.

- Extract data on potential health effects from relevant studies (data extraction files of the included studies will be shared upon release of final report).
- Assess the internal validity (risk of bias) of individual studies using pre-defined criteria.
- Synthesize the evidence using a narrative approach or meta-analysis (if appropriate) considering limitations on data integrating such as study design heterogeneity.
- Rate confidence in the body of evidence for human and animal studies separately according to one of four statements: High, Moderate, Low, or Very Low/No Evidence Available.
- Translate confidence ratings into level of evidence of health effects for human and animal studies separately according to one of four statements: High, Moderate, Low, or Inadequate.
- Combine the level of evidence ratings for human and animal data and consider the degree of support from mechanistic data to reach one of five possible hazard identification conclusions: Known, Presumed, Suspected, Not classifiable, or Not identified to be a hazard to humans.

For the evaluation of immunotoxicity associated with PFOA or PFOS exposure, we are most interested in data on primary immune outcomes (i.e., immune function and immune disease data that is more predictive of an immune-related health effect) from studies in humans, animals, or *in vitro* exposures. We are also interested in secondary immune outcomes (i.e., observational immune data or data on upstream indicators that are less predictive of immune-related health effects, but may provide supportive evidence).

Many of the studies on secondary immune outcomes will provide data that may be relevant for potential mechanisms of immune-related health effects. Mechanistic data can come from a wide variety of studies that are not intended to identify a disease phenotype. This source of experimental data includes *in vitro* and *in vivo* laboratory studies directed at cellular, biochemical, and molecular mechanisms that explain how a chemical produces particular adverse health effects.

Although it is envisioned that strong evidence for a relevant immune process from mechanistic data alone could indicate a greater potential that the substance is an immune hazard to humans, for this evaluation the mechanistic data will only be considered to inform the biological plausibility of observed outcomes from *in vivo* exposure studies in humans or animals. The mechanistic data will be collected and then grouped by the immune effects that it would be relevant for and considered in step 7, when integrating evidence to develop hazard identification conclusions. For example, observational data on total serum immunoglobulin E (IgE) or *in vitro* IgE production would support a functional measure of sensitization, but it would not support suppression of the natural killer (NK) cell response.

PECO Statement

To address our overall objective we developed PECO (**P**opulation, **E**xposure, **C**omparators and **O**utcomes) statements for the three evidence streams: human (**Table 1**), animal (**Table 2**), and *in vitro* studies (**Table 3**). The PECO statements were used as an aid to develop the specific research questions, search terms, and inclusion/exclusion criteria for the systematic review (Higgins and Green 2011).

Table 1. Human PECO (Population, Exposure, Comparator and Outcome) Statement	
PECO Element	**Evidence**
Population	Humans without restriction based on age, sex, or lifestage at exposure or outcome assessment
Exposure	Exposure to PFOA (CAS# 335-67-1) and PFOS (CAS# 1763-23-1) based on administered dose or concentration, biomonitoring data (e.g., urine, blood, or other specimens), environmental measures (e.g., air, water levels), or indirect measures such as job title
Comparators	A comparison population exposed to lower levels (or no exposure/exposure below detection levels) of PFOA or PFOS
Outcomes	**Primary outcomes:** Immune-related diseases and measures of immune function: immunosuppression (e.g., otitis, infections, or decreased vaccine antibody response); sensitization and allergic response (e.g., atopic dermatitis or asthma); autoimmunity (e.g., thyroiditis or systemic lupus erythematosus) **Secondary outcomes:** Observational immune endpoints (e.g., lymphocyte counts, lymphocyte proliferation, cytokine levels, serum antibody levels, serum autoantibody levels, or serum autoantibody levels); or immunostimulation (e.g., unintended stimulation of humoral immune function)

Table 2. Animal PECO (Population, Exposure, Comparator and Outcome) Statement	
PECO Element	**Evidence**
Population	Animals (experimental and wildlife) without restriction based on species, age, sex, or lifestage at exposure or outcome assessment
Exposure	Exposure to PFOA (CAS# 335-67-1) and PFOS (CAS# 1763-23-1) based on administered dose or concentration, biomonitoring data (e.g., urine, blood, or other specimens) or environmental measures (e.g., air, water levels)
Comparators	Comparable animal populations exposed to vehicle-only treatment in experimental animal studies or a comparison animal population exposed to lower levels (or no exposure/exposure below detection levels) of PFOA or PFOS in wildlife studies
Outcomes	**Primary outcomes:** *Disease resistance assays* (e.g., host resistance to influenza A, changes in incidence or progression in animal models of autoimmune disease) *Immune function assays following in vivo exposure to PFOA or PFOS* (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]) **Secondary outcomes:** *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)

Table 3. *In vitro* PECO (Population, Exposure, Comparator and Outcome) Statement	
PECO Element	**Evidence**
Population	Human or animal cells, tissues or model systems with *in vitro* exposure regimens
Exposure	Exposure to PFOA (CAS# 335-67-1) and PFOS (CAS# 1763-23-1) based on administered dose or concentration
Comparators	Comparable cells or tissues exposed to vehicle-only treatment or untreated controls
Outcomes	**Primary outcomes:** Immune function assays following *in vitro* exposure to PFOA or PFOS (e.g., natural killer cell [NK] activity, phagocytosis or bacterial killing by monocytes, proliferation following anti-CD3 antibody stimulation of lymphocytes) **Secondary outcomes:** Observational immune endpoints *in vitro* exposure to the test substance (e.g., general mitogen-stimulated lymphocyte proliferation, cytokine production)

The overall objective, PECO statements, and strategy to synthesize study results were based on a series of problem formulation steps that included: (1) review by technical experts with backgrounds in immunotoxicology, PFOA and PFOS, and systematic review; (2) deliberation with NTP staff and

consultation with scientists at other Federal agencies represented on the NTP Executive Committee[1]; (3) comments received on the draft protocol posted for public comment in April of 2013 (http://ntp.niehs.nih.gov/go/36501); and (4) a public review of the concept document for "Evaluation of Immunotoxicity Associated with Exposure to PFOA or PFOS" at the December 10, 2014 meeting of the NTP Board of Scientific Counselors (http://ntp.niehs.nih.gov/go/9741). More details about problem formulation can be found below in the methods section.

Key Questions

The overall objectives of the evaluation can be phrased in terms of a specific research question "What is the hazard identification conclusion as to whether exposure to PFOA or PFOS or their salts are associated with immunotoxicity or immune-related health effects?" This research question serves a focus of the evaluation to be answered by addressing the following key questions.

Table 4: Key Questions (KQ)	
KQ1	What is the hazard identification category for an association between exposure to PFOA or PFOS or their salts and immunotoxicity or immune-related health effects based on integrating levels of evidence from human and animal studies: Known, Presumed, Suspected, Not classifiable, or Not identified to be a hazard to humans?
KQ2	How does the evidence from other relevant studies (e.g., mechanistic studies) support or refute the biological plausibility of the association between exposure to PFOA or PFOS or their salts and immunotoxicity or immune-related health effects?

METHODS

Step 1. Problem Formulation

Nomination History

A draft protocol to evaluate the immune effects of PFOA and PFOA was initially developed for use as a case study to provide input for refining the OHAT systematic review and evidence integration framework. The case study was used to provide input for refining the OHAT framework, and was not intended to result in hazard identification conclusions. The protocol and literature search strategy were reviewed by technical experts with backgrounds in immunotoxicology, PFOA and PFOS, and systematic review; distributed to other government agencies through the NTP executive committee and points of contact; posted for public comment in April of 2013 (http://ntp.niehs.nih.gov/go/36501); and revised based on comments received. Although a detailed protocol was developed and peer-reviewed to outline the approach for conducting the evaluation, only subsets of the studies were used for any step in the process because the goal was to test the systematic review procedures and avoid issues with the specifics of the case study. The case-study phase was completed and the OHAT framework for

[1] Consumer Product Safety Commission (CPSC), Department of Defense (DoD), Environmental Protection Agency (EPA), Food and Drug Administration (FDA), National Cancer Institute (NCI), National Center for Environmental Health/Agency for Toxic Substances and Disease Registry (NCEH/ATSDR), National Institute of Environmental Health Sciences (NIEHS), National Institute for Occupational Safety and Health (NIOSH), Occupational Safety and Health Administration (OSHA) http://ntp.niehs.nih.gov/go/163

systematic review and evidence integration for literature-based health science evaluations was subsequently published (Rooney *et al.* 2014).

During the case-study process, we received multiple requests to complete the PFOA and PFOS case study as a full OHAT evaluation with the goal of reaching an immunotoxicity hazard identification conclusion. This evaluation topic was also considered in conjunction with the nomination and ongoing assessment of perfluorinated compounds including PFOA and PFOS by NTP's testing program.

Consideration of key scientific issues

Several key scientific issues were identified during problem formulation. A summary of those issues and how OHAT will address them in the evaluation are summarized below.

1. **The consideration of developing conclusions across the two chemicals (PFOA and PFOS).**

   NTP plans to develop conclusions separately for PFOA and for PFOS. The evidence on specific health effects will then be compared between the two chemicals when there are data on the same or related immune effects. For example, the database for PFOA and PFOS both include data on the antibody response, and the evidence for effects will be compared between the two chemicals. We are not planning a mixtures assessment or a statement regarding the immunotoxicity of the closely related class of perfluoroalkyl acids or the wider group of perfluorinated compounds based on this evidence alone.

2. **The relevance of peroxisome proliferator-activated receptor alpha (PPARα) as a mechanism for immune effects of PFOA and PFOS and species differences between animal models and humans.**

   The role of PPARα in the mechanism for immune effects will be considered when evaluating the animal immune data because of strong species differences in PPARα between rodents and humans. Some of the health effects observed in experimental animals have been linked to the ability of PFOA and PFOS to activate the peroxisome proliferator-activated receptor alpha (PPARα), and others have been shown to be independent of PPARα. For example, developmental effects of PFOA including neonatal lethality were shown to be PPARα-dependent (Abbott *et al.* 2007), while PFOS induced neonatal lethality and delayed eye opening that was independent of PPARα (Abbott *et al.* 2009). The mechanism of action for immune effects of PFOA and PFOS are not understood at this time. Targeted studies suggest that immune effects reported in laboratory animals appear to be partially or wholly independent of PPARα (DeWitt *et al.* 2009, DeWitt *et al.* 2012). This is particularly the case for suppression of the antibody response for which there is evidence that PFOA- and PFOS-associated suppression in mice are not dependent on PPARα (reviewed in DeWitt *et al.* 2012), and there are human data on PFOA- and PFOS-associated suppression in antibody response to vaccination. Studies conducted in mice, rats, and other mammalian model systems will be considered relevant for humans unless compelling evidence to the contrary is identified during the course of the evaluation.

3. **The importance of pronounced species differences in elimination rates for PFOA and PFOS between experimental animals and humans.**

   Species differences in elimination rates are important when considering dose level used in experimental animals studies. Although there is little evidence for gender differences in elimination rates in humans or non-human primates, there are gender and age differences in elimination rates in rodents (e.g., male rats have lower rates than females). Known, species,

gender, and age differences in elimination will be considered in evaluating the consistency of results reported for a given health effect. NTP recognizes that the dose or level of exposure is an important factor when considering the relevance of study findings. In the OHAT evaluation process, consideration of dose would occur after hazard identification as part of reaching a "level of concern" conclusion when the health effect is interpreted in the context of what is known regarding the extent and nature of human exposure (Shelby 2005). PFOA and PFOS have significantly lower elimination rates in humans than experimental animals resulting in long half-life values in humans (2-8 years) compared to half-life values from 10 to 20 days in monkeys and rodents (ATSDR 2009). The significantly slower elimination rates of PFOA and PFOS in humans compared to experimental animals would require pharmacokinetic adjustment to evaluate effective doses for immune effects in humans based on experimental animal evidence.

Step 2. Search For and Select Studies for Inclusion

Literature Search Strategy

Search terms were developed to identify all relevant published evidence that addresses the key questions on immunotoxicity or immune-related health effects potentially associated with exposure to PFOA or PFOS by (1) reviewing Medical Subject Headings for relevant and appropriate terms, (2) extracting key terminology from reviews and a sample of relevant primary data studies, and (3) review of PFOA search terms from a draft systematic review of developmental PFOA exposure and fetal growth (Johnson *et al.* 2013, Koustas *et al.* 2013) [note that no similar review of PFOS was located so the search for PFOS was developed using search terms from methods #1 and #2 and by analogy to the published PFOA review]. A combination of relevant subject headings and keywords were subsequently identified. A test set of relevant studies was used to ensure the search terms retrieve 100% of the test set. The following nine electronic databases will be searched using a search strategy tailored for each database (details presented in **Appendix 1**). No language restrictions or publication year limits will be imposed and the literature search will be updated for a final time approximately 90-120 days prior to peer-review.

Databases searched

- Cochrane Library
- EMBASE
- EPA ACToR (Aggregated Computational Toxicology Resource)
- EPA Chemical Data Access Tool
- EPA Docket Center
- PubChem
- PubMed
- Scopus
- Toxline
- Web of Science

Searching other resources

We will use the following methods to potentially find studies that would not be identified through the electronic searches. Studies will be evaluated using the same inclusion and exclusion criteria as used for

screening records retrieved from the electronic search. Relevant studies identified through these steps will be marked as "provided from other sources" in the study selection flow diagram.

- Hand searching the reference lists of relevant reviews, draft Federal hazard assessments (US EPA 2005, ATSDR 2009, US EPA 2014d, a), commentaries, or other non-research articles identified during the initial search. Commentaries or letters on specific studies are also reviewed to see if they contain content that should be noted during data extraction or risk-of-bias assessment of the original report.

- Hand searching the reference lists of all included studies after the full text review.

- Studies identified by the public when the initial list of included studies is posted on the OHAT website (approximately 60-90 days prior to peer-review; studies identified within 30 days of posting will be considered for inclusion) or during the public comment period when the draft Monograph is released for public comment (approximately 45-60 days prior to peer-review).

Unpublished data

NTP only includes publicly accessible, peer-reviewed information in its evaluations. If a study is identified that may be critical to the evaluation and is not peer reviewed, the NTP's practice is to obtain external peer review if the owners of the data are willing to have the study details and results made publicly accessible. The peer review would include an evaluation of the study similar to that for peer review of a journal publication. The NTP would identify and select two to three scientists knowledgeable in scientific disciplines relevant to the topic as potential peer reviewers. Persons invited to serve as peer reviewers would be screened for conflict of interest (COI) prior to confirming their service. In most instances, the peer review would be conducted by letter review. The study authors would be informed of the outcome of the peer review and given an opportunity to clarify issues or provide missing details. OHAT would consider the peer review comments regarding the scientific and technical evaluation of the unpublished study in determining whether to include the study in its evaluation. The study and its related information, if used in the OHAT evaluation, would be included in the systematic review and publicly available. OHAT would acknowledge via a note for the report that the document underwent external peer review managed by the NTP, and the names of the peer reviewers would be identified. Unpublished data from personal author communication can supplement a peer-reviewed study, as long as the information is made publicly available.

Screening Process

References retrieved from the literature search will be screened for relevance and eligibility using DistillerSR®, a web-based, systematic-review software program with structured forms and procedures to ensure standardization of the process[2]. Search results will first be consolidated in Endnote reference management software and duplicate articles will be removed prior to uploading the references into DistillerSR®.

[2]DistillerSR® (http://systematic-review.net/) is a proprietary project management tool for tracking studies through the screening process and storing data extracted from these studies using user-customized forms.

Evidence Selection Criteria

In order to be eligible for inclusion, studies must comply with the type of evidence specified by the PECO statements (**Table 1, Table 2, Table 3**). Inclusion and exclusion criteria based on the PECO statements are detailed in **Table 5**; these criteria are used to screen articles for relevance and eligibility at both the title-and-abstract and full-text screening stages. In addition to criteria defining the relevant population, exposure, comparator, and outcomes, **Table 5** defines criteria for relevant publications types (e.g., the report must contain original data). Studies that do not meet these criteria will be excluded. Some articles may be categorized as possible supportive material if they appear inappropriate for inclusion, but appear to contain relevant background information. Those studies would not provide evidence of health effects, or lack of a health effect; however, the background information could provide context or other information (e.g., exposure or metabolism data) that would be useful when evaluating confidence in bodies of evidence or integrating evidence across human, animal, and mechanistic data from the included studies.

Table 5. Inclusion and Exclusion Criteria to Determine Study Eligibility

	Inclusion Criteria	Exclusion Criteria (or blank if none)
Population (Human Studies or Experimental Model Systems)		
human	• No restrictions on sex, age, or lifestage at exposure or outcome assessment	
animal	• No restrictions on sex, age, species, or lifestage at exposure or outcome assessment	
In vitro	• Studies involving an *in vitro* exposure system and immune measures directed at cellular, biochemical, and molecular mechanisms that may explain how exposure to PFOA or PFOS produces immune effects	• Studies in non-animal organisms (e.g., plants, fungi, protists, bacteria)
Exposure		
human	• Exposure to PFOA (CAS# 335-67-1) or PFOS (CAS# 1763-23-1) or their salts based on administered dose or concentration, biomonitoring data (e.g., urine, blood, or other specimens), environmental measures (e.g., air, water levels), or indirect measures (e.g., job title)	
animal	• Exposure to PFOA or PFOS or their salts based on administered dose or concentration, bio-monitoring data (e.g., urine, blood, or other specimens), or environmental measures (e.g., water levels)	
In vitro	• Exposure to PFOA or PFOS or their salts based on administered dose or concentration	
Comparators		
human	• Humans exposed to lower levels (or no exposure/exposure below detection levels) of PFOA or PFOS	
animal	• ***For experimental studies***: study must include vehicle or untreated control group • ***For wildlife or observational studies***: animals exposed to lower levels (or no exposure/exposure below detection levels) of PFOA or PFOS	
In vitro	• Study must include vehicle or untreated control group	
Outcomes		
human	**Primary outcomes:** • Immune-related diseases and measures of immune function: ***Immunosuppression*** (e.g., otitis, infections, or decreased vaccine antibody response)	

Table 5. Inclusion and Exclusion Criteria to Determine Study Eligibility		
	Inclusion Criteria	**Exclusion Criteria (or blank if none)**
	Sensitization and allergic response (e.g., atopic dermatitis, asthma) ***Autoimmunity*** (e.g., thyroiditis or systemic lupus erythematosus)	
	Secondary outcomes: • ***Observational immune endpoints*** (e.g., lymphocyte counts, lymphocyte proliferation, cytokine levels, serum non-specific antibody levels, or serum autoantibody levels) • ***Immunostimulation*** (e.g., unintended stimulation of humoral immune function)	
animal	**Primary outcomes** ***[following in vivo exposure to PFOA or PFOS]:*** • ***Disease resistance assays*** (e.g., host resistance to influenza A or trichinella, changes in incidence or progression in animal models of autoimmune disease) • ***Immune function assays*** (e.g., antigen-specific antibody response, natural killer cell activity, delayed-type hypersensitivity response, phagocytosis by monocytes, local lymph-node assay)	• Immune tissue levels of PFOA or PFOS are not by themselves immune outcomes
	Secondary outcomes ***[following in vivo exposure to PFOA or PFOS]:*** • ***Observational immune endpoints*** (e.g., lymphoid organ weight, lymphocyte counts or subpopulations, lymphocyte proliferation, cytokine production, serum non-specific antibody levels, serum auto-antibody levels, or histological changes in immune organs) • ***Immunostimulation*** (e.g., unintended stimulation of humoral immune function)	
In vitro	**Primary outcomes** ***[following in vitro exposure to PFOA or PFOS]:*** • ***Immune* function *assays*** (e.g., natural killer cell activity, phagocytosis or bacterial killing by monocytes, proliferation following anti-CD3 antibody stimulation of spleen cells or lymphocytes	
	Secondary outcomes ***[following in vitro exposure to PFOA or PFOS]:*** • ***Observational immune endpoints*** (e.g., general mitogen-stimulated lymphocyte proliferation, cytokine production)	
Publication Type (e.g., specify any language restrictions, use of conference abstracts, etc.)		
human, animal, or *in vitro*	• Report must contain original data	• Articles with no original data (e.g., editorial or review*) • Studies published in abstract form only (grant awards con-ference abstracts), • Retracted articles
*Relevant reviews are used as background and for reference scanning.		

Primary and secondary immune outcome measures

Immunotoxicity considered in this evaluation is defined in the context of immune responses and changes in immune-related measures that reflect the four main categories of immune response: immunosuppression, immunostimulation, sensitization and allergic response, and autoimmunity. For the evaluation of immunotoxicity, primary outcomes are those with more predictive value for immunotoxicity such as disease resistance assays and functional 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, 1998) 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). The predictive value of primary and secondary outcomes is considered further in deciding whether or not to downgrade evidence for indirectness when rating the confidence in the body of evidence (**Table 10**).

Multiple publications of same data

Multiple publications with overlapping data for the same study (e.g., publications reporting subgroups, additional outcomes or exposures outside the scope of an evaluation, or longer follow-up) are identified by examining author affiliations, study designs, cohort name, enrollment criteria, and enrollment dates. If necessary, study authors will be contacted to clarify any uncertainty about the independence of two or more articles. OHAT will include all publications on the study, select one study to use as the primary, and consider the others as secondary publications with annotation as being related to the primary record during data extraction. The primary study will generally be the publication with the longest follow-up, or for studies with equivalent follow-up periods, the study with the largest number of cases or the most recent publication date. OHAT will include relevant data from all publications of the study, although if the same outcome is reported in more than one report, OHAT will include a single instance of the data (and avoid more than one, i.e., duplicate instances of the data).

Title/Abstract Review

Screeners will be trained using project-specific written instructions that reflect the criteria outlined in **Table 5** with an initial pilot phase undertaken to improve clarity of the inclusion and exclusion instructions and to improve accuracy and consistency among screeners. If changes to the inclusion criteria are made based on the pilot phase, they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes. Trained screeners from the evaluation design team will then conduct a title and abstract screen of the search results to determine whether a reference meets the inclusion or exclusion criteria. All references will be independently screened by two screeners (one of which will be the project lead, who will screen all references). Studies that are not excluded based on the title and abstract will be screened through a full-text review. In case of screening conflicts, screeners will independently review their screening results to confirm the inclusion/exclusion decision and, if needed, discuss discrepancies with the other screeners. If a true disagreement exists between screeners, the study passes to the full-text review.

Full-Text Review

After completion of the title/abstract screen, full-text articles will be retrieved[3] for those studies that either clearly meet the inclusion criteria or where eligibility to meet the inclusion criteria is unclear. Full-text review will be independently conducted by two screeners that participated in the title/abstract

[3]OHAT will initially attempt to retrieve a full-text copy of the study using an automated program, such as QUOSA, when possible, and NIH library services (NIH subscriptions and interlibrary loans). For publications not available through NIH, OHAT will search the Internet and/or may attempt to contact the corresponding author. Studies not retrieved through these mechanisms are excluded and notated as “not available.”

screening (again, one of which will be the project lead, who will screen all references). True disagreements will be resolved by discussion through consultation with other members of the evaluation design team and technical advisors.

Tracking study eligibility and reporting the flow of information

The main reason for exclusion at the full-text-review stage will be annotated and reported in a study selection flow diagram in the final report (using reporting practices outlined in Moher *et al.* 2009). The following reasons for exclusion will be documented: (1) is a review, commentary, or editorial with no original data; (2) lacks PFOS or PFOA exposure information; (3) lacks immune health outcome information; (4) only data on non-animal organisms (e.g., plants); or (5) is a conference abstract, grant application/registration, or thesis/dissertation.

Release of the list of included and excluded studies

The list of included and excluded studies will be posted on the OHAT website (http://ntp.niehs.nih.gov/go/evals posting anticipated in May 2015) once screening has been completed and prior to completion of the draft OHAT monograph.

Step 3. Data Extraction

Data Extraction Process and Data Warehousing

Data extraction will be managed with structured forms and stored in a database format using ICF International's proprietary Dose Response Analytical Generator and Organizational Network (DRAGON) software.[4] Data extraction elements are listed separately for human, animal, and *in vitro* studies in **Appendix 2-4**. The data extraction results for included studies will be visualized and made publicly available in Excel format upon publication of the final NTP Monograph using Health Assessment Workspace Collaborative (HAWC), an open source and freely available web-based interface application.[5]

The extracted data will be used to help summarize study designs and findings, facilitate assessment of risk of bias and/or conduct statistical analyses during evidence synthesis. The number of elements or collection of information on a specific element may be revised following the identification of important study details from individual studies included in the review. Data extraction will be performed by one member of the evaluation team or contract support and checked by a second member for completeness and accuracy. Any discrepancies in data extraction will be resolved by discussion or consultation with a third member of the evaluation team. Information that is inferred, converted, or estimated during data extraction will be annotated, e.g., using brackets [n=10]. OHAT will attempt to contact authors of included studies to obtain missing data considered important for evaluating key study findings (e.g., level of data required to conduct a meta-analysis). The evaluation report will note that an attempt to contact study authors was unsuccessful if study researchers do not respond to an email or phone request within one month of the attempt to contact.

[4] DRAGON (Dose Response Analytical Generator and Organizational Network) developed by ICF International (http://www.icfi.com/insights/products-and-tools/dragon-dose-response).

[5] HAWC (Health Assessment Workspace Collaborative): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals (https://hawcproject.org/portal/).

Step 4. Quality Assessment of Individual Studies

Internal validity or risk of bias will be assessed for individual studies using a tool developed by OHAT that outlines a parallel approach to evaluating risk of bias from human, animal, and *in vitro* studies to facilitate consideration of risk of bias across evidence streams with common terms and categories. The risk-of-bias tool is comprised of a common set of 11 questions that are answered based on the specific details of individual studies to develop risk-of-bias ratings (using the four options in **Table 6**) for each question. Study design determines the subset of questions that should be used to assess risk of bias for an individual study (**Table 7**). For example, the subset of risk-of-bias questions applicable to all of the experimental study designs includes a question on randomization of exposure that would not be applicable to observational study designs. Therefore, a similar set of questions are used across experimental study designs (experimental animal, *in vitro* exposure, and human controlled trials).

Table 6: Answers to the Risk-of-Bias Questions Result in One of Four Risk-of-Bias Ratings	
++	***Definitely Low*** *risk of bias*: There is direct evidence of low risk-of-bias practices
+	***Probably Low*** *risk of bias*: There is indirect evidence of low risk-of-bias practices **OR** it is deemed that deviations from low risk-of-bias practices for these criteria during the study would not appreciably bias results, including consideration of direction and magnitude of bias
– NR	***Probably High*** *risk of bias*: There is indirect evidence of high risk–of-bias practices (indicated with "-") **OR** there is insufficient information provided about relevant risk-of-bias practices (indicated with "NR" for not reported). Both symbols indicate probably high risk of bias.
––	***Definitely High*** *risk of bias*: There is direct evidence of high risk-of-bias practices

Studies are independently assessed by two assessors who answer all applicable risk-of-bias questions with one of four options in **Table 6** (answers from CLARITY Group at McMaster University 2013) following pre-specified criteria detailed in **Appendix 5**. The criteria describe aspects of study design, conduct, and reporting required to reach risk-of-bias ratings for each question and specify factors that can distinguish among ratings (e.g., what separates "definitely low" from "probably low" risk of bias). The instructions and detailed criteria are tailored to the specific evidence stream and type of human study designs. Risk of bias will be assessed at the outcome level because study design or method specifics may increase the risk of bias for some outcomes and not others within the same study.

Risk-of-Bias Assessment Process

Assessors will be trained using the criteria in **Appendix 5** with an initial pilot phase undertaken to improve clarity of criteria that distinguish between adjacent ratings and to improve consistency among assessors. All team members involved in the risk-of-bias assessment will be trained on the same set of studies and asked to identify potential ambiguities in the criteria used to assign ratings for each question. Any ambiguities and rating conflicts will be discussed relative to opportunities to refine the criteria to more clearly distinguish between adjacent ratings. If major changes to the risk-of-bias criteria

Table 7: OHAT Risk-of-Bias Questions and Applicability by Study Design

Risk-of-Bias Questions	Experimental Animal*	*In Vitro* Exposure Studies	Human Controlled Trials**	Cohort	Case-Control	Cross-Sectional***	Case Series
1. Was administered dose or exposure level adequately randomized?	X	X	X				
2. Was allocation to study groups adequately concealed?	X	X	X				
3. Did selection of study participants result in the appropriate comparison groups?				X	X	X	
4. Did study design or analysis account for important confounding and modifying variables?				X	X	X	X
5. Were experimental conditions identical across study groups?	X	X					
6. Were research personnel blinded to the study group during the study?	X	X	X				
7. Were outcome data complete without attrition or exclusion from analysis?	X	X	X	X	X	X	
8. Can we be confident in the exposure characterization?	X	X	X	X	X	X	X
9. Can we be confident in the outcome assessment (including blinding of outcome assessors)?	X	X	X	X	X	X	X
10. Were all measured outcomes reported?	X	X	X	X	X	X	X
11. Were there no other potential threats to internal validity?	X	X	X	X	X	X	X

*Experimental animal studies are controlled exposure studies. Non-human animal observational studies can be evaluated using the design features of observational human studies such as cross-sectional study design.

**Human Controlled Trials are studies in humans with controlled exposure (e.g., Randomized Controlled Trials, non-randomized experimental studies)

***Cross-sectional studies include population surveys with individual data (e.g., NHANES) and surveys with aggregate data (i.e., ecological studies).

are made based on the pilot phase (i.e., those that would likely result in revision of response), they will be documented in a protocol amendment along with the date modifications were made and the logic for the changes. It is also expected that information about confounding, exposure characterization, outcome assessment, and other important issues may be identified during or after data extraction, which can lead to further refinement of the risk-of-bias criteria (Sterne *et al.* 2014).

After assessors have independently made risk-of-bias determinations for a study across all risk-of-bias questions, the two assessors will compare their results to identify discrepancies and attempt to resolve them. Any remaining discrepancies will be considered by the project lead and, if needed, other members of the evaluation design team and/or technical advisors. The final risk-of-bias rating for each question will be recorded along with a statement of the basis for that rating. The risk-of-bias assessment of included studies will be part of the study summaries released in materials for the draft OHAT monograph that will be posted for public comment prior to peer review (anticipated for August 2015). Peer review will provide an opportunity for investigators and the public to comment on risk-of-bias

Missing Information for Risk-of-Bias Assessment

OHAT will attempt to contact authors of included studies by email to obtain missing information considered critical for evaluating risk of bias that cannot be inferred from the study. If additional information or data are received from study authors, risk-of-bias judgments will be modified to reflect the updated study information. If OHAT does not receive a response from the authors by one month of the contact attempt, a risk-of-bias response of "NR" for "not reported; probably high risk of bias" will be used and a note made in the data extraction files that an attempt to contact the authors was unsuccessful.

Step 5. Organizing and Rating Confidence in Bodies of Evidence

OHAT will consider the collection of studies on the same or closely related immune outcomes as bodies of evidence and develop overall confidence ratings in these bodies of evidence using a modification of the GRADE framework. Procedures for grouping immune outcomes, considering quantitative or narrative synthesis, and developing confidence ratings for this evaluation are described below.

Health Outcome and Endpoint Grouping

Immune endpoints will be grouped by the 3 main categories of immune response generally considered to have greater predictive value for immunotoxicity (i.e., the primary outcomes): immunosuppression, sensitization and allergic response and autoimmunity (WHO 1999b). **Table 8** lists example endpoints or assays considered primary immune outcomes for each of these categories. Secondary outcomes or those with less predictive value for immunotoxicity (e.g., cell counts, cytokine levels or other observational parameters) will be considered with the corresponding primary outcomes. For further explanation of primary and secondary outcomes, see the discussion of **Immune Health Outcomes** , **Table 10** and the inclusion and exclusion criteria for outcomes under **Table 5**.

Considerations for Pursuing a Narrative or Quantitative Evidence Synthesis

Heterogeneity within the available evidence will determine the type of evidence integration that is appropriate: either a quantitative synthesis (meta-analysis) or narrative approach for evidence integration. Where appropriate we will perform a meta-analysis. Summaries of main characteristics for each included study will be compiled and reviewed by two reviewers to determine comparability

between studies, identify data transformations necessary to ensure comparability, and determine whether heterogeneity is a concern. The main characteristics considered across all eligible studies include the following:

Human Studies

- Study design (e.g., cross-sectional, cohort)
- Details on how participants were classified into exposure groups (e.g., quartiles of exposure)
- Details on source of exposure data (e.g., questionnaire, area monitoring, biomonitoring)
- Concentrations of PFOA or PFOS for each exposure group
- Health outcome(s) reported
- Conditioning variables in the analysis (e.g., variables considered confounders)
- Type of data (e.g., continuous or dichotomous), statistics presented in paper, access to raw data
- Variation in degree of risk of bias at individual study level

Table 8. Immune Outcome Grouping

Immune Category	Example Endpoints or Assays
Immunosuppression	**Humans**: Immune related diseases (e.g., otitis, infections) **Animals**: disease resistance assays (e.g., host resistance to influenza A, trichinella) or immune function assays following *in vivo* exposure (delayed-type hypersensitivity [DTH] response, monocyte phagocytosis) ***In vitro***: immune function assays following *in vitro* exposure
	Further subgrouping of immunosuppression outcomes will be considered for functional immune assays considered to have strong predictive value for immunotoxicity (e.g., see WHO 1999b) including but not limited to: • Measures of the ***antibody response*** (e.g., vaccine antibody response in humans and T-cell dependent IgM antibody response in animals) • Measures of ***natural killer cell (NK) activity***
Sensitization and allergic response	**Humans:** sensitization-related diseases and associated measures (e.g., atopic dermatitis, asthma) **Animals:** changes in incidence or progression in animal models of sensitization (e.g., local lymph-node assay [LLNA])
Autoimmunity	**Humans:** autoimmune diseases or associated measures (e.g., thyroiditis) **Animals:** changes in incidence or progression in animal models of autoimmune disease

Animal Studies

- Experimental design (e.g., acute, chronic, multigenerational)
- Animal model used (e.g., species, strain, sex, genetic background)
- Age of animals (e.g., at start of treatment, mating, and/or pregnancy status)
- Developmental stage of animals at treatment and outcome assessment
- Dose levels, frequency of treatment, timing, duration, and exposure route
- Health outcome(s) reported
- Type of data (e.g., continuous or dichotomous), statistics presented in paper, access to raw data
- Variation in degree of risk of bias at individual study level

More detailed guidance on evaluating heterogeneity, transforming or normalizing data to ensure comparability, and the process for determining whether a meta-analysis will be pursued is provided in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673, see STEP 5). We expect to require input from topic-specific experts to help assess whether studies are too heterogeneous for meta-analysis to be appropriate. Situations where it may not be appropriate to include a study are (1) data on exposure or outcome are too different to be combined, (2) there are concerns about high risk of bias, or (3) other circumstances may indicate that averaging study results would not produce meaningful results. When it is inappropriate or not feasible to quantitatively combine results, OHAT will narratively describe or visually present findings.

Stratified Analyses, Meta-Regression, and Publication Bias

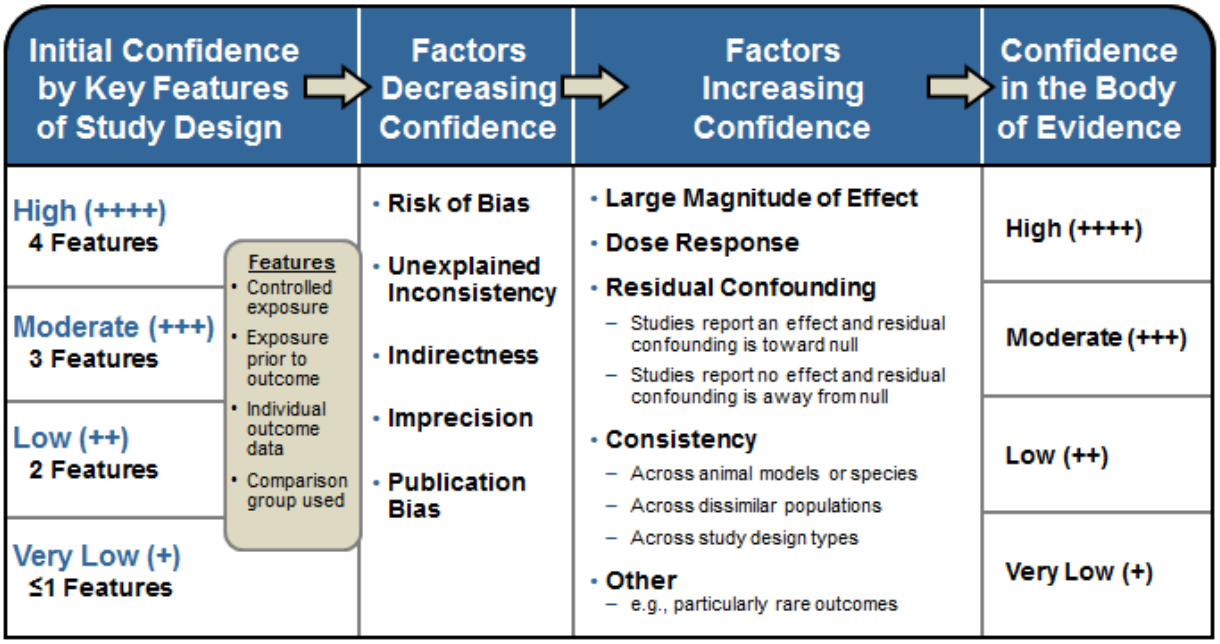
If there is significant study-level heterogeneity, then OHAT may conduct stratified analyses or multivariate meta-regression in an attempt to determine how much heterogeneity can be explained by taking into account both within- and between-study variance (Vesterinen *et al.* 2014). Multivariate meta-regression approaches are especially useful for assessing the significance of associations between study design characteristics. These approaches are considered most suitable if there are at least six to ten studies for a continuous variable and at least four studies for a categorical variable (Fu *et al.* 2011). If possible (i.e., if there are enough studies) we will assess potential publication bias by developing funnels and performing Egger regression on the estimates of effect size. In addition, if these methods suggest that publication bias is present, we will use trim and fill methods to predict the impact of the hypothetical "missing" studies (Vesterinen *et al.* 2014).

Confidence Rating: Assessment of Body of Evidence

The quality of evidence for each immune outcome will be graded using the GRADE system for rating the confidence in the body of evidence (Guyatt *et al.* 2011) as adapted by OHAT (Rooney *et al.* 2014). More detailed guidance on reaching confidence ratings in the body of evidence as "high," "moderate," "low," or "very low" is provided in the OHAT Handbook for Conducting a Literature-Based Health Assessment (http://ntp.niehs.nih.gov/go/38673, see STEP 5). In brief, available studies on a particular outcome are initially grouped by key study design features, and each grouping of studies is given an initial confidence rating by those features. This initial rating (column 1 of **Figure 2**) is downgraded for factors that decrease confidence in the results (column 2 of **Figure 2** [risk of bias, unexplained inconsistency, indirectness or lack of applicability, imprecision, and publication bias]) and upgraded for factors that increase confidence in the results (column 3 of **Figure 2** [large magnitude of effect, dose response, consistency across study designs/populations/animal models or species, consideration of residual confounding, and other factors that increase our confidence in the association or effect]).

The reasons for downgrading (or upgrading) confidence may not be due to a single domain of the body of evidence. If a decision to downgrade is borderline for two domains, the body of evidence is downgraded once in a single domain to account for both partial concerns based on considering the key drivers of the strengths or weaknesses. Similarly, the body of evidence is not downgraded twice for what is essentially the same limitation (or upgraded twice for the same asset) that could be considered applicable to more than one domain of the body of evidence. Consideration of consistency across study designs, human populations, or animal species is not included in the GRADE guidance (Guyatt *et al.* 2011); however, it is considered in the modified version of GRADE used by OHAT (Rooney *et al.* 2014).

Figure 2. Assessing Confidence in the Body of Evidence



Confidence ratings are independently assessed by federal staff on the evaluation review team, and discrepancies are resolved by consensus and consultation with technical advisors as needed. Confidence ratings are summarized in evidence profile tables (see **Table 9** for general format).

Relevance of Animal Model to Human Immune Health

- *Rats, mice, and other mammalian model systems:* No limitations of immune model systems for mammals have been identified *a priori*. Thus, studies conducted in mammalian model systems will be assumed to be relevant for humans (i.e., not downgraded for indirectness) unless compelling data to the contrary is identified during the course of the evaluation.
- *Birds, reptiles, amphibians, fish, and other non-mammalian vertebrate model systems:* Most immune cell types and immune functions are relatively consistent across vertebrate systems (Duffy and Zelikoff 2005, Rollins-Smith and Smits 2005, Rooney 2005, Salo *et al.* 2005). However, use of these model systems to address human health is not as well-established as use of the mammalian model systems (WHO 2012). For this reason, studies conducted in non-mammalian vertebrates will be downgraded one level for indirectness.
- *Invertebrate model systems:* There is a phylogenetic difference such that invertebrate immunity does not include the same level of adaptive immune function seen in vertebrates (Salo *et al.* 2005). Therefore, studies conducted in invertebrates will be downgraded two levels for indirectness.

Table 9. Evidence Profile Table Format

Example of the type of information that will be in an evidence profile for immune health outcomes

Body of Evidence	Risk of Bias	Unexplained Inconsistency	Indirectness	Imprecision	Publication Bias	Magnitude	Dose Response	Residual Confounding	Consistency Across Species/ Model	FINAL RATING
Evidence stream (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 distinguish 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**

Immune Health Outcomes

For the evaluation of immunotoxicity, primary outcomes are those with more predictive value for immunotoxicity such as disease resistance assays and functional immune parameters (see **Table 10**). Secondary outcomes are those with less predictive value for immunotoxicity such as observational parameters including cell counts or cytokine levels.

Table 10. Identification of Primary and Secondary Immune Outcomes

	Humans	Animals*	*In vitro* Assays
Primary Outcomes	**Immune-related diseases and measures of immune function** ***Immunosuppression*** (e.g., otitis, infections, or decreased vaccine antibody response); ***Sensitization and allergic response*** (e.g., atopic dermatitis or asthma); ***Autoimmunity*** (e.g., thyroiditis or systemic lupus erythematosus)	**Disease resistance assay or measures of immune function following *in vivo* expose** ***Disease resistance assays*** (e.g., host resistance to influenza A or trichinella, changes in incidence or progression in animal models of autoimmune disease) ***Immune function assays*** (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)	***Immune function assays following in vitro exposure*** (e.g., natural killer cell [NK] activity, phagocytosis or bacterial killing by monocytes, proliferation following anti-CD3 antibody stimulation of spleen cells or lymphocytes)
Secondary	***Observational immune endpoints*** (e.g., lymphocyte counts, lymphocyte proliferation, cytokine levels, or serum antibody levels) ***Immunostimulation***** (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 histological changes in immune organs)	***Observational immune endpoints*** (e.g., general mitogen-stimulated lymphocyte proliferation, cytokine production)

* Note the evaluation will consider experimental animal and observational animal studies (e.g., wildlife studies).
** Note that stimulation of the immune response is not adverse per se. It is generally agreed that stimulation of the immune system should not be disregarded (WHO 2012). Unintended immunostimulation will be considered for possible hazard if there is consistent evidence for persistent elevated immune response.

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, 1998) 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). Primary outcomes are considered to be the most direct, or applicable, to the evaluation. Secondary outcomes are relevant, but less direct and can include upstream indicators or intermediate outcomes.

- *Primary health outcomes:* The primary outcomes are most predictive of immunotoxicity and therefore there will be no downgrades for indirectness for these outcomes.
- *Secondary health outcomes:* The secondary outcomes are considered less predictive of immunotoxicity and therefore will be downgraded one level for indirectness.

PFOA or PFOS Exposure

- *Human studies:* All exposure levels and scenarios encountered in the human studies (e.g., general population, occupational settings, etc.) will be considered direct and not downgraded.
- *Dose levels used in animal studies:* There will be no downgrade for dose level used in experimental animal studies. We recognize that the level of dose or exposure is an important factor when considering the relevance of animal findings to human health. However, in OHAT's process the relevance of the dose or exposure level occurs after hazard identification as part of reaching a "level of concern" conclusion.
- *Route of administration in animal studies:* All of the most commonly used routes of administration will be considered direct for the purposes of establishing confidence ratings. We recognize that some of these exposure routes may only be relevant for certain human sub-populations. However, in OHAT's process this consideration occurs after hazard identification as part of reaching a "level of concern" conclusion.

    - Oral (no downgrade for indirectness) – Gavage, drinking water, or feeding studies are considered relevant because oral exposure through drinking water and food are considered important sources of exposure to PFOS and PFOA in humans (ATSDR 2009).
    - Dermal (no downgrade for indirectness) – Dermal exposure is considered relevant for contact with surface waters, soil, dusts, soil, and direct contact of skin with consumer products such as treated textiles (e.g., older carpet treatments) (ATSDR 2009).
    - Subcutaneous injection (no downgrade for indirectness) – Although exposure routes that bypass first metabolism are a concern for some chemicals, this is not an issue for PFOA and PFOS because they are not metabolized and therefore studies with exposure via subcutaneous injection would not be downgraded. In addition, production of PFOS has continued for limited purposes including certain medical devices for which this route of exposure may be relevant (ATSDR 2009, OECD 2013).
    - Inhalation (no downgrade for indirectness) – Inhalation studies are considered relevant because PFOA and PFOS are found in house dust (ATSDR 2009). Inhalation exposure is also relevant to occupational cohorts.
    - Intraperitoneal injection (one level downgrade for indirectness) – These studies will be downgraded one level because they are not relevant to the nature of human exposure.

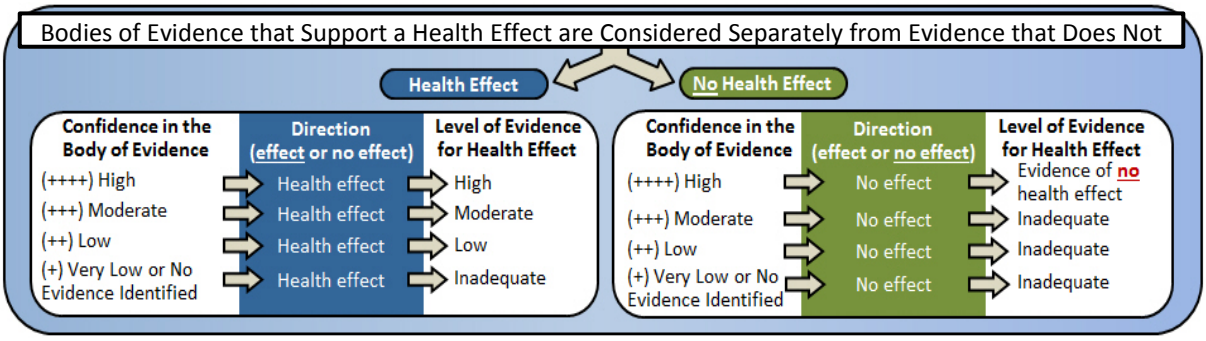
Mechanistic Studies

The framework described above only applies to human and animal studies. There is no analogous model to develop confidence ratings for other relevant data such as outcomes from *in vitro*, mechanistic, cellular or genomic studies. Thus our current approach for considering the level of support provided by other relevant data including mechanistic studies is described separately in a later section of this the document in Step 7 when integrating other relevant data (see "**Consideration of Mechanistic Data**").

Step 6. Preparation of Draft Level of Evidence Statement

The confidence ratings will be translated into draft level of evidence of health effects for each type of health outcome separately according to one of four statements: (1) High, (2) Moderate, (3) Low, or (4) Inadequate (**Figure 3**). The descriptor "evidence of no health effect" is used to indicate confidence that the substance is not associated with a health effect. Because of the inherent difficulty in proving a negative, the conclusion "evidence of no health effect" is only reached when there is high confidence in the body of evidence.

Figure 3. Translate Confidence Ratings into Evidence of Health Effect Conclusions



Evidence Descriptors	Definition
High Level of Evidence	There is high confidence in the body of evidence for an association between exposure to PFOA or PFOS and the health outcome(s).
Moderate Level of Evidence	There is moderate confidence in the body of evidence for an association between exposure to PFOA or PFOS and the health outcome(s).
Low Level of Evidence	There is low confidence in the body of evidence for an association between exposure to PFOA or PFOS and the health outcome(s), or no data are available.
Inadequate Evidence	There is insufficient evidence available to assess if exposure to PFOA or PFOS is associated with the health outcome(s).
Evidence of No Health Effect	There is high confidence in the body of evidence that exposure to PFOA or PFOS is not associated with the health outcome(s).

Step 7. Integrate Evidence to Develop Hazard Identification Conclusions

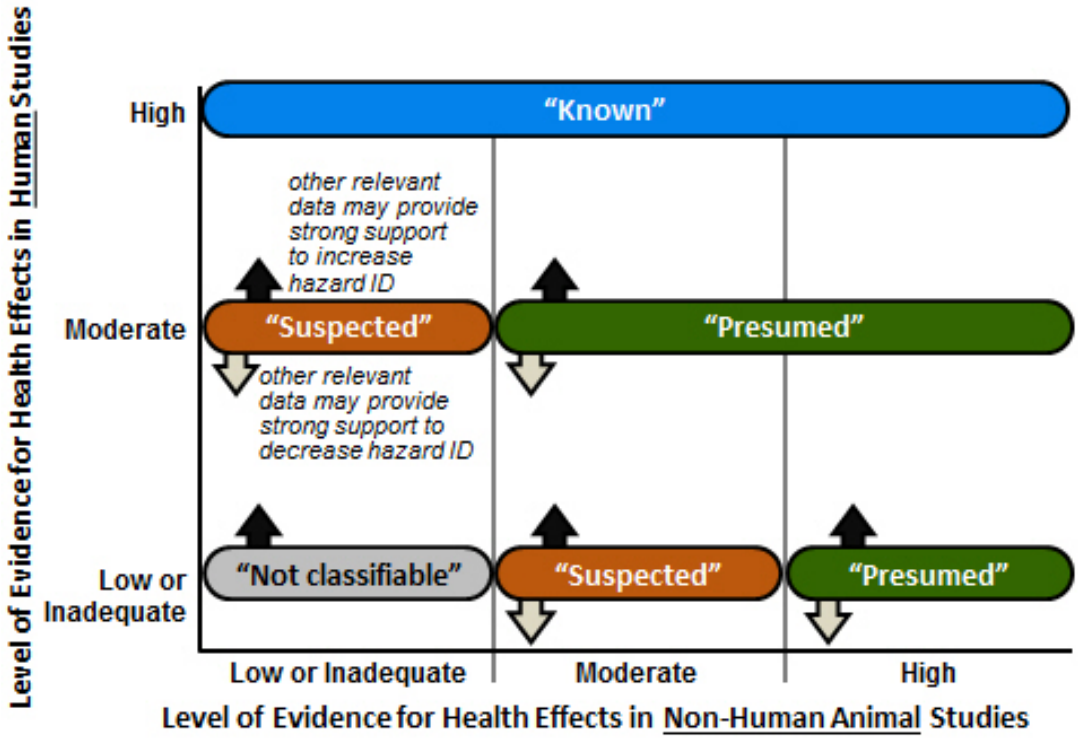
Finally, the levels of evidence ratings for human and animal data will be integrated with consideration of mechanistic data to reach one of five possible hazard identification categories: (1) Known, (2) Presumed, (3) Suspected, (4) Not classifiable, or (5) Not identified to be a hazard to humans (**Figure 4**).

Consideration of Human and Animal Data

Initial hazard identification conclusions will be reached by integrating the highest level-of-evidence conclusion for immune effect(s) on an outcome basis for the human and the animal evidence streams. Hazard identification conclusions may be reached on the groups of biologically related outcomes (using outcome groups identified in **Table 8**) as well as more specific endpoints if data are available to make more specific conclusions. If the data support an immune health effect, the level-of-evidence conclusion for human data from Step 6 for that health outcome will be considered together with the level of evidence for non-human animal data to reach one of four initial hazard identification conclusions: Known, Presumed, Suspected, or Not classifiable. If either the human or animal evidence stream is characterized as "Inadequate Evidence," then conclusions are based on the remaining evidence stream alone (which is equivalent to treating the missing evidence stream as "Low" in **Figure 4**).

If the human level of evidence rating of "Evidence of no health effect" from Step 6 is supported by a similar level of evidence rating for animal evidence for no health effect, the hazard identification conclusion would be "Not identified to be a hazard to humans."

Figure 4. Hazard Identification Scheme



Consideration of Mechanistic Data

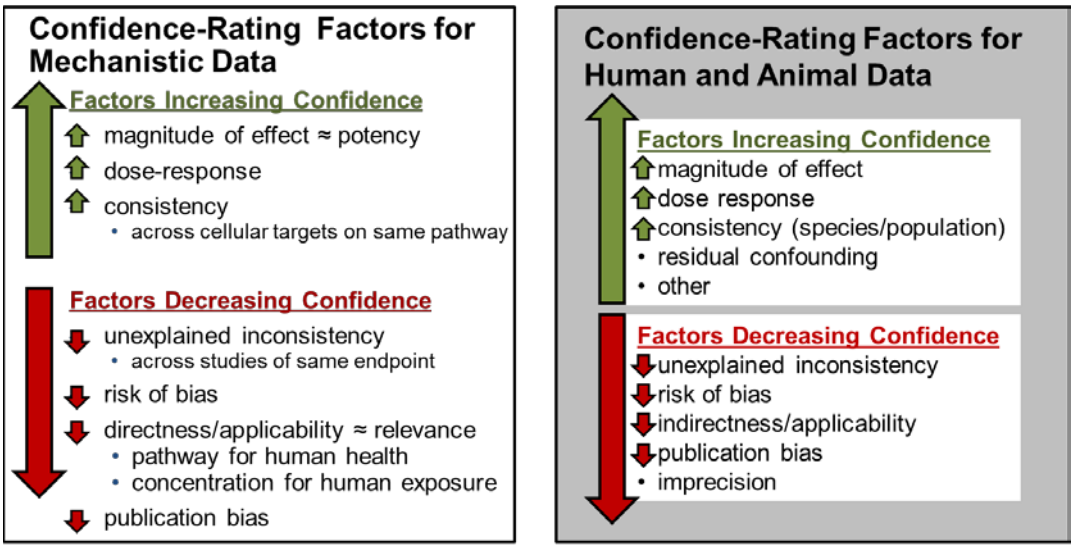
The NTP does not require mechanistic or mode-of-action data in order to reach hazard identification conclusions, although when available, this and other relevant supporting types of evidence may be used to raise (or lower) the category of the hazard identification conclusion. Mechanistic data can come from a wide variety of studies that are not intended to identify a disease phenotype. This source of experimental data includes *in vitro* and *in vivo* laboratory studies directed at cellular, biochemical, and molecular mechanisms that explain how a chemical produces particular adverse effects.

For the evaluation of immunotoxicity associated with PFOA or PFOS exposure, we are interested in mechanistic immune measures that may support the biological plausibility of corresponding immune outcomes reported from *in vivo* studies in animals or humans. For example, observational data on total serum immunoglobulin E (IgE) or *in vitro* IgE production would support a functional measure of sensitization or allergic response, but it would not support suppression of the natural killer (NK) response.

The strength of the support or opposition presented by the other relevant data is evaluated using the guidance presented in **Figure 5**. The factors outlined for increasing or decreasing confidence in that the mechanistic data support biological plausibility are conceptually similar to those used to rate confidence in bodies of evidence for human or animal *in vivo* studies. Evaluations of the strength of evidence

provided by mechanistic data are made on an outcome-specific basis based on discussion by the evaluation team and consultation with technical advisors as needed.

Figure 5. Factors Considered in Evaluating the Support for Biological Plausibility



- If mechanistic data provide strong support for biological plausibility of the relationship between exposure and the health effect, the hazard identification conclusion may be upgraded (indicated by black "up" arrows in the Step 7 graphic in **Figure 4**) the that initially derived by considering the human and non-human animal evidence together.
- If mechanistic data provide strong opposition for biological plausibility of the relationship between exposure and the health effect, the hazard identification conclusion may be downgraded (indicated by gray "down" arrows in **Figure 4**) from that initially derived by considering the human and non-human animal evidence together.

Although it is envisioned that strong evidence for a relevant immune process from mechanistic data alone could indicate a greater potential that the substance is an immune hazard to humans, for this evaluation the mechanistic data will only be considered to inform the biological plausibility of observed outcomes from *in vivo* data.

NTP MONOGRAPH FORMAT

The NTP Monograph on the association between PFOA or PFOS exposure and immunotoxicity will include the following information:

Introduction

This section will provide a brief background on the topic.

Methodology

This section will provide a brief overview of the methodologies used in the review process, including:

- the research question
- the search strategy used to identify and retrieve studies
- the process for selecting the included studies
- the methods of data extraction
- the methods used to assess risk of bias of included studies
- the methods used to synthesize the data of included studies
- the methods used to evaluate confidence in the body of evidence
- the methods used to reach hazard identification conclusions

Results

This section will include the results from the systematic review of the evidence for an association between exposure to PFOA or PFOS and immunotoxicity or immune-related health effects. Results will be presented in tables or figures as appropriate using HAWC. The results from the included studies will be discussed by outcome. This will include a description of:

- the number of studies identified that reported the outcome
- full list of excluded studies, with reasons for exclusion documented for studies excluded at the full text review stage
- a summary of the results and risk-of-bias assessment for each included study (including files in downloadable format)
- description of results and ratings for confidence in the bodies of evidence for major immune outcomes (e.g., hypersensitivity, autoimmunity, antibody response, natural killer cell function, cytokine function or secretion) for which there are PFOA and or PFOS data using the OHAT adaption of GRADE
- evidence profiles for major immune outcomes for which there are PFOA or PFOS data
- presentation of level of evidence and draft hazard identification conclusions for major immune outcomes for which there are PFOA or PFOS data

Discussion

The discussion will provide a summary of the review findings, including a discussion of any gaps identified in the evidence and any suggestions of areas for further research. Any important limitations of the review will be described and their impact on the available evidence will be discussed.

Conclusion

This will present the conclusion of the review.

REFERENCES

Abbott BD, Wolf CJ, Schmid JE, Das KP, Zehr RD, Helfant L, Nakayama S, Lindstrom AB, Strynar MJ, Lau C. 2007. Perfluorooctanoic acid induced developmental toxicity in the mouse is dependent on expression of peroxisome proliferator activated receptor-alpha. *Toxicol Sci* 98(2): 571-581.

Abbott BD, Wolf CJ, Das KP, Zehr RD, Schmid JE, Lindstrom AB, Strynar MJ, Lau C. 2009. Developmental toxicity of perfluorooctane sulfonate (PFOS) is not dependent on expression of peroxisome proliferator activated receptor-alpha (PPAR alpha) in the mouse. *Reprod Toxicol* 27(3-4): 258-265.

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

Calafat AM, Kuklenyik Z, Reidy JA, Caudill SP, Tully JS, Needham LL. 2007. Serum concentrations of 11 polyfluoroalkyl compounds in the u.s. population: data from the national health and nutrition examination survey (NHANES). *Environ Sci Technol* 41(7): 2237-2242.

CDC (Centers for Disease Control and Prevention). 2009. *Fourth National Report on Human Exposure to Environmental Chemicals*. Atlanta, GA: U.S. Department of Health and Human Services. Available: http://www.cdc.gov/exposurereport/ [accessed 5 April 2013].

CDC (Centers for Disease Control and Prevention). 2015. *Fourth National Report on Human Exposure to Environmental Chemicals: Updated Tables, February 2015*. Atlanta, GA: U.S. Department of Health and Human Services. Available: http://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Feb2015.pdf [accessed 11 March 2015].

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 January 2013].

Dallaire F, Dewailly E, Ayotte P. 2005. Approaches to Immunotoxicology in Human Population Studies. In: *Investigative Immunotoxicology*. Tryphonas H, Fournier M, Blakeley BR, Smits JE, Brousseau P, eds. New York, NY, Taylor and Francis: 247-265.

DeWitt JC, Shnyra A, Badr MZ, Loveless SE, Hoban D, Frame SR, Cunard R, Anderson SE, Meade BJ, Peden-Adams MM, Luebke RW, Luster MI. 2009. Immunotoxicity of perfluorooctanoic acid and perfluorooctane sulfonate and the role of peroxisome proliferator-activated receptor alpha. *Crit Rev Toxicol* 39(1): 76-94.

DeWitt JC, Peden-Adams MM, Keller JM, Germolec DR. 2012. Immunotoxicity of perfluorinated compounds: recent developments. *Toxicol Pathol* 40(2): 300-311.

Duffy JE, Zelikoff JT. 2005. Approaches and Modesl for the Assessment of Chemical-Induced Immunotoxicity in Fish. In: *Investigative Immunotoxicology*. Tryphonas H, Fournier M, Blakeley BR, Smits JE, Brousseau P, eds. New York, NY, Taylor and Francis: 49-61.

EFSA (European Food Safety Authority). 2008. *Perfluorooctane sulfonate (PFOS), Perfluorooctanoic acid (PFOA) and their Salts*. The EFSA Journal 653: 1-131. Available: http://www.efsa.europa.eu/en/scdocs/doc/653.pdf.

Fair PA, Driscoll E, Mollenhauer MA, Bradshaw SG, Yun SH, Kannan K, Bossart GD, Keil DE, Peden-Adams MM. 2011. Effects of environmentally-relevant levels of perfluorooctane sulfonate on clinical parameters and immunological functions in B6C3F1 mice. *J Immunotox* 8(1): 17-29.

Fu R, Gartlehner G, Grant M, Shamliyan T, Sedrakyan A, Wilt TJ, Griffith L, Oremus M, Raina P, Ismaila A, Santaguida P, Lau J, Trikalinos TA. 2011. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol* 64(11): 1187-1197.

Germolec D. 2009. *Explanation of Levels of Evidence for Immune System Toxicity*. Research Triangle Park, NC. Available: http://ntp.niehs.nih.gov/go/9399.

Grandjean P, Andersen EW, Budtz-Jorgensen E, Nielsen F, Molbak K, Weihe P, Heilmann C. 2012. Serum vaccine antibody concentrations in children exposed to perfluorinated compounds. *JAMA : the journal of the American Medical Association* 307(4): 391-397.

Granum B, Haug LS, Namork E, Stolevik SB, Thomsen C, Aaberge IS, van Loveren H, Lovik M, Nygaard UC. 2013. Pre-natal exposure to perfluoroalkyl substances may be associated with altered vaccine antibody levels and immune-related health outcomes in early childhood. *J Immunotox*.

Guyatt GH, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schunemann HJ. 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 64(4): 383-394.

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(3): 607-614.

Higgins J, Green S. 2011. Cochrane Handbook for Systematic Reviews of Interventions, The Cochrane Collaboration.

Jandacek RJ, Rider T, Keller ER, Tso P. 2010. The effect of olestra on the absorption, excretion and storage of 2,2',5,5' tetrachlorobiphenyl; 3,3',4,4' tetrachlorobiphenyl; and perfluorooctanoic acid. *Environ Int* 36(8): 880-883.

Johnson P, Sutton P, Atchley D, Koustas E, Lam J, Robinson K, Sen S, Axelrad D, Woodruff T. 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.

Kannan K, Perrotta E, Thomas NJ. 2006. Association between perfluorinated compounds and pathological conditions in southern sea otters. *Environ Sci Technol* 40(16): 4943-4948.

Keller JM, Kannan K, Taniyasu S, Yamashita N, Day RD, Arendt MD, Segars AL, Kucklick JR. 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(23): 9101-9108.

Koustas E, Lam J, Sutton P, Johnson P, Atchley D, Robinson K, Sen S, Axelrad D, Woodruff T. 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 Non-Human Evidence - Protocol.

Lau C, Anitole K, Hodes C, Lai D, Pfahles-Hutchens A, Seed J. 2007. Perfluoroalkyl acids: a review of monitoring and toxicological findings. *Toxicol Sci* 99(2): 366-394.

Looker C, Luster MI, Calafat AM, Johnson VJ, Burleson GR, Burleson FG, Fletcher T. 2014. Influenza vaccine response in adults exposed to perfluorooctanoate and perfluorooctanesulfonate. *Toxicol Sci* 138(1): 76-88.

Luster MI, Portier C, Pait DG, White KL, Jr., Gennings C, Munson AE, Rosenthal GJ. 1992. Risk assessment in immunotoxicology. I. Sensitivity and predictability of immune tests. *Fundam Appl Toxicol* 18(2): 200-210.

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(10): 1006-1012.

Nelson JW, Hatch EE, Webster TF. 2010. Exposure to polyfluoroalkyl chemicals and cholesterol, body weight, and insulin resistance in the general U.S. population. *Environ Health Perspect* 118(2): 197-202.

OECD (Organisation for Economic Co-operation and Development). 2013. *OECD Portal on Perfluorinated Chemicals*. Available: http://www.oecd.org/ehs/pfc/ [accessed 30 January 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(1): 144-154.

Rollins-Smith LA, Smits JEG. 2005. Amphibian Models and Approaches to Immunotoxicology. In: *Investigative Immunotoxicology*. Tryphonas H, Fournier M, Blakeley BR, Smits JE, Brousseau P, eds. New York, NY, Taylor and Francis: 77-90.

Rooney AA. 2005. Reptiles: The Research Potential of an Overlooked Taxon in Immunotoxicology. In: *Investigative Immunotoxicology*. Tryphonas H, Fournier M, Blakeley BR, Smits JE, Brousseau P, eds. New York, NY, Taylor and Francis: 107-127.

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(7): 711-718.

Rooney AA, Thayer K, Boyles AL. 2015. The Office of Health Assessment and Translation's Risk of Bias Tool for In Vitro Studies. *Environ Health Perspect* submitted.

Salo H, Dautremepuits C, Smits JEG, Brousseau P, Fournier M. 2005. Immune Markers in Ecotoxicology: A comparison Across Species. In: *Investigative Immunotoxicology*. Tryphonas H, Fournier M, Blakeley BR, Smits JE, Brousseau P, eds. New York, NY, Taylor and Francis: 147-162.

Shamliyan T, Kane RL, Dickinson S. 2010. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *J Clin Epidemiol* 63(10): 1061-1070.

Shelby MD. 2005. National Toxicology Program Center for the Evaluation of Risks to Human Reproduction: guidelines for CERHR expert panel members. *Birth Defects Res B Dev Reprod Toxicol* 74(1): 9-16.

Steenland K, Fletcher T, Savitz DA. 2010. Epidemiologic evidence on the health effects of perfluorooctanoic acid (PFOA). *Environ Health Perspect* 118(8): 1100-1108.

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*.

Sterne J, Higgins J, Reeves B, on behalf of the development group for ACROBAT-NRSI. 2014. *A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0*. Available: http://www.riskofbias.info [accessed 28 September 2014].

US EPA (US Environmental Protection Agency). 1996a. *Biochemicals Test Guidelines: OPPTS 880.3550 Immunotoxicity*. EPA/712/C-96/280. Washington, DC: Office of Prevention Pesticides and Toxic Substances. 1-14. Available: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series880.htm.

US EPA. 1996b. *Biochemicals Test Guidelines: OPPTS 880.3800 Immune Response*. EPA/712/C-96/281. Washington, DC: Office of Prevention Pesticides and Toxic Substances. US Environmental Protection Agency: 1-8. Available: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series880.htm.

US EPA (US Environmental Protection Agency). 1998. *Health Effects Test Guidelines: OPPTS 870.7800 Immunotoxicity*. EPA/712/C-98/351. Washington, DC: Office of Prevention Pesticides and Toxic Substances. 1-11. Available: http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series870.htm.

US EPA (US Environmental Protection Agency). 2005. *Draft risk assessment of the potential human helath effects associated with exposure to perfluorooctanoic acid and its salts*. Washington, DC: Office of Pollution Prevention and Toxics. US Environmental Protection Agency. Available: http://www.epa.gov/opptintr/pfoa/pubs/pfoarisk.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). 2013. *Perfluorooctanoic Acid (PFOA) and Fluorinated Telomers*. Available: http://www.epa.gov/oppt/pfoa/.

US EPA (US Environmental Protection Agency). 2014a. *Draft Health Effects Document for Perfluorooctane sulfonate (PFOS)*. Office of Water. Available: https://peerreview.versar.com/epa/pfoa/pdf/Health-Effects-Document-for-Perfluorooctane-Sulfonate-(PFOS).pdf.

US EPA (US Environmental Protection Agency). 2014b. *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.

US EPA (US Environmental Protection Agency). 2014c. *Peer Review of Health Effects Documents for PFOA and PFOS*. Office of Water. Available: http://water.epa.gov/drink/standards/upload/Peer-Review-of-Health-Effects-Documents-for-PFOA-and-PFOS-Factsheet-February-2014.pdf.

US EPA (US Environmental Protection Agency). 2014d. *Draft Health Effects Document for Perfluorooctane acid (PFOA)*. Office of Water. Available: https://peerreview.versar.com/epa/pfoa/pdf/Health-Effects-Document-for-Perfluorooctanoic-Acid-(PFOA).pdf.

US EPA. 2015. *2014 Annual Progress Reports to Meet Phase-out Goals of the 2010/2015 PFOA Stewardship Program*. Available: http://epa.gov/oppt/pfoa/pubs/stewardship/preports8.html [accessed 10 March 2015].

Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL, Antonic A, Howells DW, Macleod MR. 2014. Meta-analysis of data from animal studies: a practical guide. *J Neurosci Methods* 221: 92-102.

WHO (International Programme on Chemical Safety, World Health Organization). 1996. *Principles and methods for assessing direct immunotoxicity associated with exposure to chemicals. IPCS Environmental Health Criteria 180*. IPCS Environmental Health Criteria 180. Geneva: 390. Available: http://www.inchem.org/documents/ehc/ehc/ehc180.htm.

WHO (International Programme on Chemical Safety, World Health Organization). 1999a. *Principles and methods for assessing allergic hypersensitization associated with exposure to chemicals. IPCS Environmental Health Criteria 212*. IPCS Environmental Health Criteria 212. Geneva: 312. Available: http://www.inchem.org/documents/ehc/ehc/ehc212.htm.

WHO (International Programme on Chemical Safety, World Health Organization). 1999b. *Principles for the assessment of risks to human health from exposure to chemicals. IPCS Environmental Health Criteria 210*. IPCS Environmental Health Criteria 210. Geneva: 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.

ABOUT THE PROTOCOL

Contributors

Evaluation Team

Evaluation teams are composed of federal staff and contractor staff. Contractor staff members are screened for potential conflicts of interest. Federal staff members should do a self-evaluation. Epidemiologists and toxicologists on OHAT evaluation teams should have at least three years' experience and/or training in reviewing studies, including summarizing studies and critical review (e.g., assessing study quality and interpreting findings). Experience in evaluating occupational or environmental studies is preferred. Team members should have at least a master's degree or equivalent level of experience in epidemiology, toxicology, environmental health sciences, or a related field.

Name	Affiliation
Andrew Rooney, PhD	NIEHS/NTP, Project Lead
Chad Blystone, PhD	NIEHS/NTP
Abee Boyles, PhD	NIEHS/NTP
John Cowden, PhD	US EPA, ORD/NCEA
Dori Germolec, PhD	NIEHS/NTP
Stephanie Holmgren	NIEHS/NIH
Kembra Howdeshell, PhD	NIEHS/NTP
Katie Pelch, PhD	NIEHS/NTP
Kristina Thayer, PhD	NIEHS/NTP
Kyla Taylor, MS	NIEHS/NTP
Vickie Walker	NIEHS/NTP

Contract support: Will assist in data extraction and risk-of-bias assessment

Name	Affiliation
Robyn Blain	ICF Incorporated
Pam Hartman	ICF Incorporated
Pam Ross	ICF Incorporated
Bryan Luukinen	ICF Incorporated
Addie Harris	ICF Incorporated
Susan Goldhaber	ICF Incorporated
Thy Nguyen	ICF Incorporated
Jen Fownes	ICF Incorporated
Ali Goldstone	ICF Incorporated

Peer Reviewers

The peer reviewers were outside experts selected for their experience with PFOA and PFOS, immunotoxicity, and systematic review procedures. Peer reviewers were screened for conflict of interest prior to their service and did not report any conflicts of interest. Service as a peer reviewer does not necessarily indicate that the reviewer endorses the final document.

Name	Affiliation
Jamie Dewitt, PhD	East Carolina University, Department of Pharmacology and Toxicology
Christopher Lau, PhD	US EPA, ORD/NHEERL
Tony Fletcher, PhD	London School of Hygiene and Tropical Medicine, Department of Social and Environmental Health Research
Roberta Scherer, PhD	Johns Hopkins University Bloomberg School of Public Health

Sources of Support

National Institute of Environmental Health Sciences/Division of the National Toxicology Program

Protocol History and Revisions

Date	Activity or revision
March 26, 2013:	**Draft case study protocol reviewed:** sent to peer reviewers for comment/review
April 9, 2013:	**Case study protocol posted on OHAT website:** posted as a case study used to test the OHAT Approach to Systematic Review and Evidence Integration
March 16, 2015:	**Draft evaluation protocol reviewed:** sent to peer reviewers for comment/ review (note: Tony Fletcher did not review the evaluation protocol because he was not available within a one month time frame)
June 16, 2015:	**Evaluation protocol posted on OHAT website:** Principal differences from case study protocol were: (1) updated/revised risk-of-bias tool; (2) addition of a risk-of-bias approach for in vitro studies; (3) no tiering process used to potentially exclude low quality studies

APPENDICES

Appendix 1. Literature Search Strategy

The strategy for this search is broad for the consideration of immune-related endpoints and comprehensive for PFOA or PFOS as an exposure or treatment in order to ensure inclusion of relevant papers.

Database	Search Terms
COCHRANE LIBRARY	Perfluoroalkyl* OR perfluorocaprylic OR perfluorocarbon* OR perfluorocarboxyl* OR perfluorochemical* OR (perfluorinated AND (C8 OR carboxylic OR chemical* OR compound* OR octanoic)) OR PFAA* OR "fluorinated polymer" OR "fluorinated polymers" OR (fluorinated AND (polymer OR polymers)) OR (fluorocarbon AND (polymer OR polymers)) OR Fluoropolymer* OR (fluorinated AND telomer*) OR fluorotelomer* OR fluoro-telomer* OR fluorosurfactant* OR "FC 143" OR FC143 OR Pentadecafluoroctanoate* OR Pentadecafluorooctanoate* OR pentadecafluoroctanoic OR pentadecafluorooctanoic OR "pentadecafluoro-1-octanoic" OR "pentadecafluoro-n-octanoic" OR "perfluoro-1-heptanecarboxylic" OR perfluorocaprylic OR perfluoroheptanecarboxylic OR perfluoroctanoate OR perfluorooctanoate OR "perfluoro octanoate" OR "perfluorooctanoic acid" OR perfluoroctanoic OR perfluorooctanoic OR "perfluoro octanoic" OR "perfluoro-n-octanoic" OR "perfluorooctanoyl chloride" OR PFOA OR APFO OR "1-octanesulfonic acid" OR "1-perfluorooctanesulfonic" OR "1-perfluoroctanesulfonic" OR "heptadecafluoro-1-octanesulfonic" OR "heptadecafluoro-1-octane sulfonic" OR "heptadecafluorooctanesulfonic" OR "heptadecafluorooctane sulfonic" OR "heptadecafluoroctane sulfonic" OR "perfluoroalkyl sulphonate" OR perfluoroctanesulfonate OR perfluorooctanesulfonate OR "perfluoroctane sulfonate" OR "perfluorooctane sulfonate" OR "perfluoro-n-octanesulfonic" OR perfluoroctanesulfonic OR perfluorooctanesulfonic OR "perfluoroctane sulfonic" OR "perfluorooctane sulfonic" OR perfluoroctanesulphonic OR perfluorooctanesulphonic OR "perfluoroctane sulphonic" OR "perfluorooctane sulphonic" OR perfluoroctylsulfonic OR PFOS
EMBASE	Perfluoroalkyl* OR perfluorocaprylic OR perfluorocarbon* OR perfluorocarboxyl* OR perfluorochemical* OR (perfluorinated NEXT/4 (C8 OR carboxylic OR chemical OR chemicals OR compound OR compounds OR octanoic)) OR PFAA* OR "fluorinated polymer" OR "fluorinated polymers" OR (fluorinated NEXT/4 (polymer*)) OR (fluorocarbon NEXT/4 (polymer*)) OR Fluoropolymer* OR (fluorinated NEXT/4 telomer*) OR fluorotelomer* OR fluoro NEXT/0 telomer* OR fluorosurfactant* OR "FC 143" OR FC143 OR Pentadecafluoroctanoate* OR Pentadecafluorooctanoate* OR pentadecafluoroctanoic OR pentadecafluorooctanoic OR "pentadecafluoro-1-octanoic" OR "pentadecafluoro-n-octanoic" OR "perfluoro-1-heptanecarboxylic" OR perfluorocaprylic OR perfluoroheptanecarboxylic OR perfluoroctanoate OR perfluorooctanoate OR "perfluoro octanoate" OR "perfluorooctanoic acid" OR perfluoroctanoic OR perfluorooctanoic OR "perfluoro octanoic" OR "perfluoro-n-octanoic" OR "perfluorooctanoyl chloride" OR PFOA OR APFO OR "1-octanesulfonic acid" OR "1-perfluorooctanesulfonic" OR "1-perfluoroctanesulfonic" OR "heptadecafluoro-1-octanesulfonic" OR "heptadecafluoro-1-octane sulfonic" OR "heptadecafluorooctanesulfonic" OR "heptadecafluorooctane sulfonic" OR "heptadecafluoroctane sulfonic" OR "perfluoroalkyl sulphonate" OR

Database	Search Terms
	perfluoroctanesulfonate OR perfluorooctanesulfonate OR “perfluoroctane sulfonate” OR “perfluorooctane sulfonate” OR “perfluoro-n-octanesulfonic” OR perfluoroctanesulfonic OR perfluorooctanesulfonic OR “perfluoroctane sulfonic” OR “perfluorooctane sulfonic” OR perfluoroctanesulphonic OR perfluorooctanesulphonic OR “perfluoroctane sulphonic” OR “perfluorooctane sulphonic” OR perfluoroctylsulfonic OR PFOS OR 307-35-7 OR 1763-23-1 OR 335-67-1 AND immune OR immunocomp* OR immunogen* OR immunolog* OR immunotox OR immunity OR autoimmun* OR "host resistance" OR spleen OR splenic OR splenocyt* OR thymus OR thymic OR thymocyt* OR leukocyt* OR granulocyt* OR basophil* OR eosinophil* OR neutrophil* OR lymph OR lymphoid* OR lymphocyt* OR "b-lymphocyte" OR "b-lymphocytes" OR "t-lymphocyte" OR "t-lymphocytes" OR "killer cell" OR "killer cells" OR "NK cell" OR "NK-cell" OR "NK-cells" OR macrophag* OR "mast cell" OR "mast cells" OR monocyt* OR phagocyt* OR dendrit* OR "t-cell" OR "t cell" OR "t cells" OR "t-cells" OR "T helper" OR "T-helper" OR "b-cell" OR "b cell" OR "b cells" OR "b-cells" OR antibod* OR histamine* OR histocompatib* OR immunoglobulin* OR "immunoglobulin A" OR IgA OR "immunoglobulin D" OR IgD OR "immunoglobulin E" OR IgE OR "immunoglobulin G" OR IgG OR "immunoglobulin M" OR IgM OR antigen OR antigens OR CD3 OR CD4 OR CD8 OR CD25 OR CD27 OR CD28 OR CD29 OR CD45* OR cytokine* OR chemokine* OR inteferon* OR interleukin* OR "IL-6" OR "IL-8" OR lymphokine* OR monokine* OR ("tumor necrosis" NEXT/0 factor*) OR "TNF alpha" OR "TNFalpha" OR autoimmun* OR addison OR rheumatoid OR glomerulonephritis OR diabetes OR graves OR lupus OR thyroiditis OR hypersensitiv* OR sensitization OR hyperresponsiv* OR allerg* OR atopy OR atopic OR dermatitis OR eczema OR otitis OR “ear infection” OR “ear inflammation” OR (respiratory NEXT/2 infection*) OR asthma OR bronchitis OR pneumonia OR bronchiolitis OR rhinitis OR sinusitis OR wheez* OR crackle* OR cough* OR dyspnea OR gastroenteritis OR inflammat* OR pro NEXT/0 inflammat* OR anti NEXT/0 inflamm* OR autacoid* OR eicosanoid* OR prostaglandin* OR immunomodul* OR immunotherap* OR vaccin* OR immuniz* OR immunosuppress* OR desensitiz* OR immunoprotein* OR "c-reactive protein" OR CRP OR "complement component" OR (complement NEXT/2 (C1 OR C2 OR C3 OR C4 OR C5 OR C6 OR C7 OR C8 OR C9))
EPA ACToR	Select “Search on CAS Numbers” Enter each CAS number on a new line: 307-35-7 1763-23-1 335-67-1
EPA Chemical Data Access Tool	307-35-7 OR 1763-23-1 OR 335-67-1
EPA Docket Center	PFOA, immune perfluorooctanoic acid, immune perfluorooctane sulfonate, immune PFOS, immune
PubChem	307-35-7 OR 1763-23-1 OR 335-67-1

Database	Search Terms
PUBMED	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] AND 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

Database	Search Terms
	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))
SCOPUS	TITLE(Perfluoroalkyl* OR perfluorocaprylic OR perfluorocarbon* OR perfluorocarboxyl* OR perfluorochemical* OR (perfluorinated AND (C8 OR carboxylic OR chemical* OR compound* OR octanoic)) OR PFAA* OR “fluorinated polymer” OR “fluorinated polymers” OR (fluorinated AND (polymer OR polymers)) OR (fluorocarbon AND (polymer OR polymers)) OR Fluoropolymer* OR (fluorinated AND telomer*) OR fluorotelomer* OR fluoro-telomer* OR fluorosurfactant* OR “FC 143” OR FC143 OR Pentadecafluoroctanoate* OR Pentadecafluorooctanoate* OR pentadecafluoroctanoic OR pentadecafluorooctanoic OR “pentadecafluoro-1-octanoic” OR “pentadecafluoro-n-octanoic” OR “perfluoro-1-heptanecarboxylic” OR perfluorocaprylic OR perfluoroheptanecarboxylic OR perfluoroctanoate OR perfluorooctanoate OR “perfluoro octanoate” OR “perfluorooctanoic acid” OR perfluoroctanoic OR perfluorooctanoic OR “perfluoro octanoic” OR “perfluoro-n-octanoic” OR “perfluorooctanoyl chloride” OR PFOA OR APFO OR “1-octanesulfonic acid” OR “1-perfluorooctanesulfonic” OR “1-perfluoroctanesulfonic” OR “heptadecafluoro-1-octanesulfonic” OR “heptadecafluoro-1-octane sulfonic” OR “heptadecafluorooctanesulfonic” OR “heptadecafluorooctane sulfonic” OR “heptadecafluoroctane sulfonic” OR “perfluoroalkyl sulphonate” OR perfluoroctanesulfonate OR perfluorooctanesulfonate OR “perfluoroctane sulfonate” OR “perfluorooctane sulfonate” OR “perfluoro-n-octanesulfonic” OR perfluoroctanesulfonic OR perfluorooctanesulfonic OR “perfluoroctane sulfonic” OR “perfluorooctane sulfonic” OR perfluoroctanesulphonic OR perfluorooctanesulphonic OR “perfluoroctane sulphonic” OR “perfluorooctane sulphonic” OR perfluoroctylsulfonic OR PFOS) OR CASREGNUMBER(335-67-1) OR CASREGNUMBER(1763-23-1) OR ABS(Perfluoroalkyl* OR perfluorocaprylic OR perfluorocarbon* OR perfluorocarboxyl* OR perfluorochemical* OR (perfluorinated AND (C8 OR carboxylic OR chemical* OR compound* OR octanoic)) OR PFAA* OR “fluorinated polymer” OR “fluorinated polymers” OR (fluorinated AND (polymer OR polymers)) OR (fluorocarbon AND (polymer OR polymers)) OR Fluoropolymer* OR (fluorinated AND telomer*) OR fluorotelomer* OR fluoro-telomer* OR fluorosurfactant* OR “FC 143” OR FC143 OR Pentadecafluoroctanoate* OR

Database	Search Terms
	Pentadecafluorooctanoate* OR pentadecafluoroctanoic OR pentadecafluorooctanoic OR “pentadecafluoro-1-octanoic” OR “pentadecafluoro-n-octanoic” OR “perfluoro-1-heptanecarboxylic” OR perfluorocaprylic OR perfluoroheptanecarboxylic OR perfluoroctanoate OR perfluorooctanoate OR “perfluoro octanoate” OR “perfluorooctanoic acid” OR perfluoroctanoic OR perfluorooctanoic OR “perfluoro octanoic” OR “perfluoro-n-octanoic” OR “perfluorooctanoyl chloride” OR PFOA OR APFO OR “1-octanesulfonic acid” OR “1-perfluorooctanesulfonic” OR “1-perfluoroctanesulfonic” OR “heptadecafluoro-1-octanesulfonic” OR “heptadecafluoro-1-octane sulfonic” OR “heptadecafluorooctanesulfonic” OR “heptadecafluorooctane sulfonic” OR “heptadecafluoroctane sulfonic” OR “perfluoroalkyl sulphonate” OR perfluoroctanesulfonate OR perfluorooctanesulfonate OR “perfluoroctane sulfonate” OR “perfluorooctane sulfonate” OR “perfluoro-n-octanesulfonic” OR perfluoroctanesulfonic OR perfluorooctanesulfonic OR “perfluoroctane sulfonic” OR “perfluorooctane sulfonic” OR perfluoroctanesulphonic OR perfluorooctanesulphonic OR “perfluoroctane sulphonic” OR “perfluorooctane sulphonic” OR perfluoroctylsulfonic OR PFOS) OR CASREGNUMBER(335-67-1) OR CASREGNUMBER(1763-23-1) AND TITLE(immune OR immunocomp* OR immunogen* OR immunolog* OR immunoto* OR immunotoxins OR immunity OR autoimmun* OR "host resistance" OR spleen OR splenic OR splenocyt* OR thymus OR thymic OR thymocyt* OR leukocyt* OR granulocyt* OR basophil* OR eosinophil* OR neutrophil* OR lymph OR lymphoid* OR lymphocyt* OR "b-lymphocyte" OR "b-lymphocytes" OR "t-lymphocyte" OR "t-lymphocytes" OR "killer cell" OR "killer cells" OR "NK cell" OR "NK-cell" OR "NK-cells" OR macrophag* OR "mast cell" OR "mast cells" OR monocyt* OR phagocyt* OR dendrit* OR "t-cell" OR "t cell" OR "t cells" OR "t-cells" OR "T helper" OR "T-helper" OR "b-cell" OR "b cell" OR "b cells" OR "b-cells" OR antibod* OR histamine* OR histocompatib* OR immunoglobulin* OR "immunoglobulin A" OR IgA OR "immunoglobulin D" OR IgD OR "immunoglobulin E" OR IgE OR "immunoglobulin G" OR IgG OR "immunoglobulin M" OR IgM OR antigen OR antigens OR CD3 OR CD4 OR CD8 OR CD25 OR CD27 OR CD28 OR CD29 OR CD45* OR cytokine* OR chemokine* OR inteferon* OR interleukin* OR "IL-6" OR "IL-8" OR lymphokine* OR monokine* OR ("tumor necrosis" AND factor*) OR "TNF alpha" OR "TNFalpha" OR autoimmun* OR addison OR rheumatoid OR glomerulonephritis OR diabetes OR graves OR lupus OR thyroiditis OR hypersensitiv* OR sensitization OR hyperresponsiv* OR allerg* OR atopy OR atopic OR dermatitis OR eczema OR otitis OR “ear infection” OR “ear inflammation” OR (respiratory AND infection*) OR asthma OR bronchitis OR pneumonia OR bronchiolitis OR rhinitis OR sinusitis OR wheez* OR crackle* OR cough* OR dyspnea OR gastroenteritis OR inflammat* OR pro-inflammat* OR anti-inflamm* OR autacoid* OR eicosanoid* OR prostaglandin* OR immunomodul* OR immunotherap* OR vaccin* OR immuniz* OR immunosuppress* OR desensitiz* OR immunoprotein* OR "c-reactive protein" OR CRP OR "complement component" OR (complement AND (C1 OR C2 OR C3 OR C4 OR C5 OR C6 OR C7 OR C8 OR C9))) OR ABS(immune OR immunocomp* OR immunogen* OR immunolog* OR immunotox OR immunity OR autoimmun* OR "host resistance" OR spleen OR splenic OR splenocyt* OR thymus OR thymic OR thymocyt* OR leukocyt* OR granulocyt* OR basophil* OR eosinophil* OR neutrophil* OR lymph OR lymphoid* OR lymphocyt* OR "b-lymphocyte" OR "b-lymphocytes" OR "t-lymphocyte" OR "t-lymphocytes" OR "killer cell" OR "killer cells" OR "NK cell" OR "NK-cell" OR "NK-cells" OR macrophag*

Database	Search Terms
	OR "mast cell" OR "mast cells" OR monocyt* OR phagocyt* OR dendrit* OR "t-cell" OR "t cell" OR "t cells" OR "t-cells" OR "T helper" OR "T-helper" OR "b-cell" OR "b cell" OR "b cells" OR "b-cells" OR antibod* OR histamine* OR histocompatib* OR immunoglobulin* OR "immunoglobulin A" OR IgA OR "immunoglobulin D" OR IgD OR "immunoglobulin E" OR IgE OR "immunoglobulin G" OR IgG OR "immunoglobulin M" OR IgM OR antigen OR antigens OR CD3 OR CD4 OR CD8 OR CD25 OR CD27 OR CD28 OR CD29 OR CD45* OR cytokine* OR chemokine* OR inteferon* OR interleukin* OR "IL-6" OR "IL-8" OR lymphokine* OR monokine* OR ("tumor necrosis" AND factor*) OR "TNF alpha" OR "TNFalpha" OR autoimmun* OR addison OR rheumatoid OR glomerulonephritis OR diabetes OR graves OR lupus OR thyroiditis OR hypersensitiv* OR sensitization OR hyperresponsiv* OR allerg* OR atopy OR atopic OR dermatitis OR eczema OR otitis OR "ear infection" OR "ear inflammation" OR (respiratory AND infection*) OR asthma OR bronchitis OR pneumonia OR bronchiolitis OR rhinitis OR sinusitis OR wheez* OR crackle* OR cough* OR dyspnea OR gastroenteritis OR inflammat* OR pro-inflammat* OR anti-inflamm* OR autacoid* OR eicosanoid* OR prostaglandin* OR immunomodul* OR immunotherap* OR vaccin* OR immuniz* OR immunosuppress* OR desensitiz* OR immunoprotein* OR "c-reactive protein" OR CRP OR "complement component" OR (complement AND (C1 OR C2 OR C3 OR C4 OR C5 OR C6 OR C7 OR C8 OR C9)))
Toxline	NOTE: Searching on the immune terms will only retrieve the first 50,000 records (Toxline's display limit). Attempts to break the search up into separate searches is possible, but even a search on the term 'immunology' alone will hit the maximum. Perfluoroalkyl* OR perfluorocaprylic OR perfluorocarbon* OR perfluorocarboxyl* OR perfluorochemical* OR (perfluorinated AND (C8 OR carboxylic OR chemical OR chemicals OR compound OR compounds OR octanoic)) OR PFAA* OR "fluorinated polymer" OR "fluorinated polymers" OR (fluorinated AND (polymer OR polymers)) OR (fluorocarbon AND (polymer OR polymers)) OR Fluoropolymer* OR (fluorinated AND telomer*) OR fluorotelomer* OR fluoro-telomer* OR fluorosurfactant* OR "FC 143" OR FC143 OR Pentadecafluoroctanoate* OR Pentadecafluorooctanoate* OR pentadecafluoroctanoic OR pentadecafluorooctanoic OR "pentadecafluoro-1-octanoic" OR "pentadecafluoro-n-octanoic" OR "perfluoro-1-heptanecarboxylic" OR perfluorocaprylic OR perfluoroheptanecarboxylic OR perfluoroctanoate OR perfluorooctanoate OR "perfluoro octanoate" OR "perfluorooctanoic acid" OR perfluoroctanoic OR perfluorooctanoic OR "perfluoro octanoic" OR "perfluoro-n-octanoic" OR "perfluorooctanoyl chloride" OR PFOA OR APFO OR "1-octanesulfonic acid" OR "1-perfluorooctanesulfonic" OR "1-perfluoroctanesulfonic" OR "heptadecafluoro-1-octanesulfonic" OR "heptadecafluoro-1-octane sulfonic" OR "heptadecafluorooctanesulfonic" OR "heptadecafluorooctane sulfonic" OR "heptadecafluoroctane sulfonic" OR "perfluoroalkyl sulphonate" OR perfluoroctanesulfonate OR perfluorooctanesulfonate OR "perfluoroctane sulfonate" OR "perfluorooctane sulfonate" OR "perfluoro-n-octanesulfonic" OR perfluoroctanesulfonic OR perfluorooctanesulfonic OR "perfluoroctane sulfonic" OR "perfluorooctane sulfonic" OR perfluoroctanesulphonic OR perfluorooctanesulphonic OR "perfluoroctane sulphonic" OR "perfluorooctane sulphonic" OR perfluoroctylsulfonic OR PFOS OR 335-67-1 OR 1763-23-1
WEB OF SCIENCE	Perfluoroalkyl OR perfluoroalkyls OR perfluorocaprylic OR perfluorocarbon OR perfluorocarbons OR perfluorocarboxyl* OR perfluorochemical* OR perfluorocarboxyls OR perfluorochemicals OR PFAA* OR "fluorinated polymer" OR

Database	Search Terms
	"fluorinated polymers" OR Fluoropolymer* OR (fluorinated AND telomer*) OR fluorotelomer* OR fluoro-telomer* OR fluorosurfactant* OR "FC 143" OR FC143 OR Pentadecafluoroctanoate* OR Pentadecafluorooctanoate* OR pentadecafluoroctanoic OR pentadecafluorooctanoic OR "pentadecafluoro-1-octanoic" OR "pentadecafluoro-n-octanoic" OR "perfluoro-1-heptanecarboxylic" OR perfluorocaprylic OR perfluoroheptanecarboxylic OR perfluoroctanoate OR perfluorooctanoate OR "perfluoro octanoate" OR "perfluorooctanoic acid" OR perfluoroctanoic OR perfluorooctanoic OR "perfluoro octanoic" OR "perfluoro-n-octanoic" OR "perfluorooctanoyl chloride" OR PFOA OR APFO OR "1-octanesulfonic acid" OR "1-perfluorooctanesulfonic" OR "1-perfluoroctanesulfonic" OR "heptadecafluoro-1-octanesulfonic" OR "heptadecafluoro-1-octane sulfonic" OR "heptadecafluorooctanesulfonic" OR "heptadecafluorooctane sulfonic" OR "heptadecafluoroctane sulfonic" OR "perfluoroalkyl sulphonate" OR perfluoroctanesulfonate OR perfluorooctanesulfonate OR "perfluoroctane sulfonate" OR "perfluorooctane sulfonate" OR "perfluoro-n-octanesulfonic" OR perfluoroctanesulfonic OR perfluorooctanesulfonic OR "perfluoroctane sulfonic" OR "perfluorooctane sulfonic" OR perfluoroctanesulphonic OR perfluorooctanesulphonic OR "perfluoroctane sulphonic" OR "perfluorooctane sulphonic" OR perfluoroctylsulfonic OR PFOS AND immune OR immunocomp* OR immunogen* OR immunolog* OR immunotox OR immunity OR autoimmun* OR "host resistance" OR spleen OR splenic OR splenocyt* OR thymus OR thymic OR thymocyt* OR leukocyt* OR granulocyt* OR basophil* OR eosinophil* OR neutrophil* OR lymph OR lymphoid* OR lymphocyt* OR "b-lymphocyte" OR "b-lymphocytes" OR "t-lymphocyte" OR "t-lymphocytes" OR "killer cell" OR "killer cells" OR "NK cell" OR "NK-cell" OR "NK-cells" OR macrophag* OR "mast cell" OR "mast cells" OR monocyt* OR phagocyt* OR dendrit* OR "t-cell" OR "t cell" OR "t cells" OR "t-cells" OR "T helper" OR "T-helper" OR "b-cell" OR "b cell" OR "b cells" OR "b-cells" OR antibod* OR histamine* OR histocompatib* OR immunoglobulin* OR "immunoglobulin A" OR IgA OR "immunoglobulin D" OR IgD OR "immunoglobulin E" OR IgE OR "immunoglobulin G" OR IgG OR "immunoglobulin M" OR IgM OR antigen OR antigens OR CD3 OR CD4 OR CD8 OR CD25 OR CD27 OR CD28 OR CD29 OR CD45* OR cytokine* OR chemokine* OR inteferon* OR interleukin* OR "IL-6" OR "IL-8" OR lymphokine* OR monokine* OR ("tumor necrosis" AND factor*) OR "TNF alpha" OR "TNFalpha" OR autoimmun* OR addison OR rheumatoid OR glomerulonephritis OR diabetes OR graves OR lupus OR thyroiditis OR hypersensitiv* OR sensitization OR hyperresponsiv* OR allerg* OR atopy OR atopic OR dermatitis OR eczema OR otitis OR "ear infection" OR "ear inflammation" OR (respiratory AND infection*) OR asthma OR bronchitis OR pneumonia OR bronchiolitis OR rhinitis OR sinusitis OR wheez* OR crackle* OR cough* OR dyspnea OR gastroenteritis OR inflammat* OR pro-inflammat* OR anti-inflamm* OR autacoid* OR eicosanoid* OR prostaglandin* OR immunomodul* OR immunotherap* OR vaccin* OR immuniz* OR immunosuppress* OR desensitiz* OR immunoprotein* OR "c-reactive protein" OR CRP OR "complement component" OR (complement AND (C1 OR C2 OR C3 OR C4 OR C5 OR C6 OR C7 OR C8 OR C9))

Appendix 2. Data Extraction Elements for Human Studies

HUMAN	
Funding	Funding source(s)
	Reporting of conflict of interest (COI) by authors (*reporting bias)
Subjects	Study population name/description
	Dates of study and sampling time frame
	Geography (country, region, state, etc.)
	Demographics (sex, race/ethnicity, age or lifestage at exposure and at outcome assessment)
	Number of subjects (target, enrolled, n per group in analysis, and participation/follow-up rates) (*missing data bias)
	Inclusion/exclusion criteria/recruitment strategy (*selection bias)
	Description of reference group (*selection bias)
Methods	Study design (e.g., prospective or retrospective cohort, nested case-control study, cross-sectional, population-based case-control study, intervention, case report, etc.)
	Length of follow-up (*information bias)
	Health outcome category, e.g., cardiovascular
	Health outcome, e.g., blood pressure (*reporting bias)
	Diagnostic or methods used to measure health outcome (*information bias)
	Confounders or modifying factors and how considered in analysis (e.g., included in final model, considered for inclusion but determined not needed (*confounding bias)
	Substance name and CAS number
	Exposure assessment (e.g., blood, urine, hair, air, drinking water, job classification, residence, administered treatment in controlled study, etc.) (*information bias)
	Methodological details for exposure assessment (e.g., HPLC-MS/MS, limit of detection) (*information bias)
	Statistical methods (*information bias)
Results	Exposure levels (e.g., mean, median, measures of variance as presented in paper, such as SD, SEM, 75th/90th/95th percentile, minimum/maximum); range of exposure levels, number of exposed cases
	Statistical findings (e.g., adjusted β, standardized mean difference, adjusted odds ratio, standardized mortality ratio, relative risk, etc.) or description of qualitative results. When possible, OHAT will convert measures of effect to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data are expressed as mean difference, standardized mean difference, and percent control response. Categorical data are typically expressed as odds ratio, relative risk (RR, also called risk ratio), or β values, depending on what metric is most commonly reported in the included studies and on OHAT's ability to obtain information for effect conversions from the study or through author query.

HUMAN	
	If not presented in the study, statistical power can be assessed during data extraction using an approach that can detect a 10% to 20% change from response by control or referent group for continuous data, or a relative risk or odds ratio of 1.5 to 2 for categorical data, using the prevalence of exposure or prevalence of outcome in the control or referent group to determine sample size. For categorical data where the sample sizes of exposed and control or referent groups differ, the sample size of the exposed group will be used to determine the relative power category. Recommended sample sizes to achieve 80% power for a given effect size, i.e., 10% or 20% change from control, will be compared to sample sizes used in the study to categorize statistical power as "appears to be adequately powered" (sample size for 80% power met), somewhat underpowered (sample size is 75% to < 100% of number required for 80% power), "underpowered" (sample size is 50% to < 75% of number required for 80% power), or "severely underpowered" (sample size is < 50% of number required for 80% power).
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias

Appendix 3. Data Extraction Elements for Animal Studies

ANIMAL	
Funding	Funding source(s)
	Reporting of COI by authors (*reporting bias)
Animal Model	Sex
	Species
	Strain
	Source of animals
	Age or lifestage at start of dosing and at health outcome assessment
	Diet and husbandry information (e.g., diet name/source)
Treatment	Chemical name and CAS number
	Source of chemical
	Purity of chemical (*information bias)
	Dose levels or concentration (as presented and converted to mg/kg bw/d when possible)
	Other dose-related details, such as whether administered dose level was verified by measurement, information on internal dosimetry (*information bias)
	Vehicle used for exposed animals
	Route of administration (e.g., oral, inhalation, dermal, injection)
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, days per week)
Methods	Study design (e.g., single treatment, acute, subchronic (e.g., 90 days in a rodent), chronic, multigenerational, developmental, other)
	Guideline compliance (i.e., use of EPA, OECD, NTP or another guideline for study design, conducted under GLP guideline conditions, non-GLP but consistent with guideline study, non-guideline peer-reviewed publication)
	Number of animals per group (and dams per group in developmental studies) (*missing data bias)
	Randomization procedure, allocation concealment, blinding during outcome assessment (*selection bias)
	Method to control for litter effects in developmental studies (*information bias)

ANIMAL	
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls – was expected response observed? (*information bias)
	Endpoint health category (e.g., reproductive)
	Endpoint (e.g., infertility)
	Diagnostic or method to measure endpoint (*information bias)
	Statistical methods (*information bias)
Results	Measures of effect at each dose or concentration level (e.g., mean, median, frequency, and measures of precision or variance) or description of qualitative results. When possible, OHAT will convert measures of effect to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data will be expressed as mean difference, standardized mean difference, and percent control response. Categorical data will be expressed as relative risk (RR, also called risk ratio).
	No Observed Effect Level (NOEL), Lowest Observed Effect Level (LOEL), benchmark dose (BMD) analysis, statistical significance of other dose levels, or other estimates of effect presented in paper. **Note:** The NOEL and LOEL are highly influenced by study design, do not give any quantitative information about the relationship between dose and response, and can be subject to author's interpretation (e.g., a statistically significant effect may not be considered biologically important). Also, a NOEL does not necessarily mean zero response. Ideally, the response rate at specific dose levels is used as the primary measure to characterize the response.
	If not presented in the study, statistical power can be assessed during data extraction using an approach that assesses the ability to detect a 10% to 20% change from control group's response for continuous data, or a relative risk or odds ratio of 1.5 to 2 for categorical data, using the outcome frequency in the control group to determine sample size. Recommended sample sizes to achieve 80% power for a given effect size, i.e., 10% or 20% change from control, will be compared to sample sizes used in the study to categorize statistical power as "appears to be adequately powered" (sample size for 80% power met), "somewhat underpowered" (sample size is 75% to < 100% of number required for 80% power), "underpowered" (sample size is 50% to < 75% of number required for 80% power), or "severely underpowered" (sample size is < 50% of number required for 80% power).
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)
	Data on internal concentration, toxicokinetics, or toxicodynamics (when reported)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias

Appendix 4. Data Extraction Elements for *In Vitro* Studies

In vitro	
Funding	Funding source(s)
	Reporting of COI by authors (*reporting bias)
Cell/Tissue Model	Cell line, cell type, or tissue
	Source of cells/tissue (and validation of identity)
	Sex of human/animal origin
	Species
	Strain
Treatment	Chemical name and CAS number
	Concentration levels (as presented and converted to µM when possible)
	Source of chemical

In vitro	
	Purity of chemical (*information bias)
	Vehicle used for experimental/control conditions
	Duration and frequency of dosing (e.g., hours, days, weeks when administration was ended, days per week)
Methods	Guideline compliance (i.e., use of EPA, OECD, NTP or another guideline for study design, conducted under GLP guideline conditions, non-GLP but consistent with guideline study, non-guideline peer-reviewed publication)
	Randomization procedure, allocation concealment, blinding during outcome assessment (*selection bias)
	Number of replicates per group (*information bias)
	Percent serum/plasma in medium
	Use of negative controls and whether controls were untreated, vehicle-treated, or both
	Report on data from positive controls – was expected response observed? (*information bias)
	Endpoint health category (e.g., immune)
	Endpoint or assay target (e.g., IL-2 cytokine levels)
	Name and source of assay kit
	Diagnostic or method to measure endpoint (e.g., reporter gene)(*information bias)
	Statistical methods (*information bias)
	Measures of effect at each dose or concentration level (e.g., mean, median, frequency, and measures of precision or variance) or description of qualitative results. When possible, OHAT will convert measures of effect to a common metric with associated 95% confidence intervals (CI). Most often, measures of effect for continuous data will be expressed as mean difference, standardized mean difference, and percent control response. Categorical data will be expressed as relative risk (RR, also called risk ratio).
Results	No Observed Effect Concentration (NOEC), Lowest Observed Effect Concentration (LOEC), statistical significance of other concentration levels, AC50, or other estimates of effect presented in paper. **Note:** The NOEC and LOEC are highly influenced by study design, do not give any quantitative information about the relationship between dose and response, and can be subject to author's interpretation (e.g., a statistically significant effect may not be considered biologically important). Also, a NOEC does not necessarily mean zero response.
	Observations on dose response (e.g., trend analysis, description of whether dose-response shape appears to be monotonic, non-monotonic)
Other	Documentation of author queries, use of digital rulers to estimate data values from figures, exposure unit, and statistical result conversions, etc.

Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias

Appendix 5. Risk-of-Bias Criteria

The OHAT risk-of-bias tool for human and animal studies (version date January 2015 and available at http://ntp.niehs.nih.gov/go/38673) reflects OHAT's current best practices and provides the detailed discussion and instructions for the risk-of-bias practices used in this evaluation. The OHAT tool uses a single set of questions (also called "elements" or "domains") to assess risk of bias across various study types to facilitate consideration of conceptually similar potential sources of bias across the human and animal evidence streams with a common terminology. Individual risk-of-bias questions are designated as only applicable to certain study designs (e.g., cohort studies or experimental animal studies), and a subset of the questions apply to each study design (**Table 6**).

The eight questions relevant to experimental or controlled-exposure studies were used as the basis for development of an OHAT *in vitro* risk-of-bias tool. This tool will be applied to studies using cells or tissues from humans or other animals with an *in vitro* exposure regime; in contrast to *in vivo* exposures that are already addressed by risk-of-bias tools for experimental animal studies or controlled human exposures. A manuscript detailing the *in vitro* risk-of-bias method to be used in this evaluation is currently under peer review for publication (Rooney *et al.* 2015). Comments received during the manuscript review process will be considered for potential revisions to the risk-of-bias method used to evaluate *in vitro* studies.

The specific criteria used to assess risk of bias for this evaluation are outlined below for Human/observational studies, experimental animal studies, and *in vitro* studies. Based on literature searches done for the case study we do not expect any controlled exposure studies in humans (i.e., human controlled trials) and therefore have not included risk-of-bias criteria for that study design. If relevant human controlled trials of PFOA or PFOS are identified, the criteria from the January 2015 OHAT risk–of-bias tool will be used to evaluate risk of bias.

Observational Studies (Human studies or wildlife animal studies)

Cohort studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?

Definitely Low Risk of Bias (++)
• Direct evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4), • **Note:** Immune-specific exclusion criteria should be discussed including the presence or history of infectious or autoimmune diseases other than the outcome of interest or diagnosis of the outcome of interest before participation in the study.

Probably Low Risk of Bias (+)
• Indirect evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **OR** differences between groups would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates, • **OR** there is insufficient information provided about the comparison group including a different rate of non-response without an explanation (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates.

4. Did study design or analysis account for important confounding and modifying variables?

Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered. • **Note:** The following variables should be considered as potential confounders and/or effect measure modifiers for the relationship between PFOA or PFOS exposure and immune outcomes: age, sex, race/ethnicity, smoking, body mass index, alcohol consumption, and variables that represent socioeconomic status (e.g., educational level, household income) based on prior reports of associations with PFOA and PFOS exposure levels (Calafat *et al.* 2007, Nelson *et al.* 2010) and immune outcomes (WHO 1996, Dallaire *et al.* 2005) • **Note:** The following variables should be considered as potential effect measure modifiers of the relationship between PFOA or PFOS exposure and allergy/asthma or cold incidence: maternal/paternal asthma or allergy, age at start of daycare, anti-inflammatory medication and diet/nutritional status based on established practices in immunotoxicology (WHO 1999a, Dallaire *et al.* 2005) • **Note:** The following variables should be considered as potential effect measure modifiers for the relationship between PFOA or PFOS exposure and antibody response to vaccination: previous history of vaccination, anti-inflammatory medication, acute stress (WHO 1996, Dallaire *et al.* 2005) • **Note:** Exposure to other known or suspected immunotoxicants will be considered as confounders (e.g., lead or polychlorinated biphenyls (PCBs); perfluorononanoate (PFNA) and perfluorohexane sulfonate (PFHxS) will be considered as potential confounders because there is some evidence that these (non-PFOA and non-PFOS) perfluorinated compounds (PFCs) may be associated with immune effects; PFOA and PFOS may serve as confounders for potential effects of the other compound respectively • **Note:** if reported, consumption of olestra will be considered as there is evidence that olestra decreased the absorption of PFOA from the gastrointestinal tract of mice (Jandacek *et al.* 2010)
Probably Low Risk of Bias (+)
• Indirect evidence that appropriate adjustments were made, • **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results,

- **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements,
- **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research),
- **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for,
- **OR** it is deemed that co-exposures present would not appreciably bias results.
- **Note:** this includes insufficient information provided on co-exposures in general population studies.

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that the distribution of important covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses,
- **OR** there is insufficient information provided about the distribution of known confounders (record "NR" as basis for answer),
- **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity,
- **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer),
- **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for,
- **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).

Definitely High Risk of Bias (--)

- Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses,
- **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements,
- **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)

- Direct evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study.
- **Note**: Acceptable handling of subject attrition includes: very little missing outcome data; reasons for missing subjects unlikely to be related to outcome (for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups,
- **OR** missing data have been imputed using appropriate methods and characteristics of subjects lost to follow up or with unavailable records are described in identical way and are not significantly different from those of the study participants.

Probably Low Risk of Bias (+)

- Indirect evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study,
- **OR** it is deemed that the proportion lost to follow-up would not appreciably bias results. This would include reports of no statistical differences in characteristics of subjects lost to follow up or with unavailable records from those of the study participants. Generally, the higher the ratio of participants with missing data to participants with events, the greater potential there is for bias. For studies with a long duration of follow-up,

some withdrawals for such reasons are inevitable.

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed,
- **OR** there is insufficient information provided about numbers of subjects lost to follow-up (record "NR" as basis for answer).

Definitely High Risk of Bias (--)

- Direct evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed.
- **Note**: Unacceptable handling of subject attrition includes: reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across study groups; or potentially inappropriate application of imputation.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)

- Direct evidence that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure (e.g., measurement of PFOA or PFOS in drinking water or measurement of PFOA or PFOS in blood, serum, or plasma),
- **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods,
- **AND** exposure was assessed in a relevant time-window for development of the outcome. Current exposure measures will be considered relevant for any immune outcome for two reasons: (1) they are likely to be relevant for most immune outcomes other than disease endpoints with a significant lag time such as autoimmune disease, and (2) because of the long half-life value of PFOA and PFOS in humans (2-8 years) such that current exposure levels may be indicative of past exposures,
- **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes,
- **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished,
- **AND** the study used spiked samples to confirm assay performance.
- **Note:** Measurement of serum or whole-blood PFOA or PFOS is the standard accepted biomarker of exposure (not urine or feces) in humans using quantitative techniques such as liquid chromatography-triple quadrupole mass spectrometry (LC-MS/MS) and high pressure liquid chromatography with tandem mass spectrometry (HPLC/MS) (ATSDR 2009, CDC 2009, US EPA 2014d, a, CDC 2015).
- **Note:** It is understood that serum levels of PFOA and PFOS may indicate exposure to PFOA or PFOS or exposure to other PFCs that may break down into PFOA or PFOS. For this evaluation we will use the available exposure metric as an indication of exposure to PFOA or PFOS.

Probably Low Risk of Bias (+)

- Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure),
- **OR** exposure was assessed using indirect measures (e.g., drinking water levels and residency, questionnaire or occupational exposure assessment by a certified industrial hygienist) that have been validated or empirically shown to be consistent with methods that directly measure exposure (i.e., inter-methods validation: one method vs. another),
- **AND** exposure was assessed in a relevant time-window for development of the outcome. Current exposure measures will be considered relevant for any immune outcome for two reasons: (1) they are likely to be relevant for most immune outcomes other than disease endpoints with a significant lag time such as autoimmune disease, and (2) because of the long half-life value of PFOA and PFOS in humans (2-8 years) such that current exposure levels may be indicative of past exposures,
- **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or

never exposed),
- **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.

Probably High Risk of Bias (-) or (NR)
- Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure
- **OR** there is evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., questionnaire, job-exposure matrix or self-report without validation) (record "NR" as basis for answer),
- **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).

Definitely High Risk of Bias (--)
- Direct evidence that the exposure was assessed using methods with poor validity,
- **OR** evidence of exposure misclassification (e.g., differential recall of self-reported exposure).

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
- Direct evidence that the immune outcome was assessed using well-established methods (e.g., gold standard)
- **AND** subjects had been followed for the same length of time in all study groups,
- **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group or exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes.
- **NOTE:** Well-established methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits, commercial laboratories, or standard assays such as ELISAs for IgG with sufficiently low variation and limits of detection to allow discrimination between groups (or evidence that the assay could have detected a difference based on responses to a positive control); doctor diagnosis of asthma or incidence data obtained from medical records; incidence of doctor-diagnosed otitis by trained interviewers; obtained from registries (Shamliyan *et al.* 2010).

Probably Low Risk of Bias (+)
- Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard),
- **AND** subjects had been followed for the same length of time in all study groups
- **OR** it is deemed that the outcome assessment methods used would not appreciably bias results,
- **AND** there is indirect evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes,
- **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures,
- **NOTE:** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes such as and mining of data collected for other purposes. Proxy reporting (e.g., parental reporting of days sick or doctor-diagnosis) of immune disease, colds, etc. should be considered on a case-by-case basis with consideration of whether or not there is empirical evidence as to the reliability of proxy reporting for that outcome.

Probably High Risk of Bias (-) or (NR)
- Indirect evidence that the outcome assessment method is an insensitive instrument (e.g., a questionnaire used to assess outcomes with no information on validation),
- **OR** the length of follow up differed by study group,
- **OR** there is indirect evidence that it was possible for outcome assessors (including study subjects if outcomes were self-reported) to infer the study group prior to reporting outcomes,
- **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for

answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of follow up differed by study group, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors (including study subjects if outcomes were self-reported), including no blinding or incomplete blinding.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no PFOA- or PFOS-specific additions to the risk-of-bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Cross Sectional and Case Series Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?[NA to Case series]

Definitely Low Risk of Bias (++)
• Direct evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4), • **Note:** Immune-specific exclusion criteria should be discussed including the presence or history of infectious or autoimmune diseases other than the outcome of interest or diagnosis of the outcome of interest before participation in the study.
Probably Low Risk of Bias (+)
• Indirect evidence that subjects (both exposed and non-exposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates, • **OR** differences between groups would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates, • **OR** there is insufficient information provided about the comparison group including a different rate of non-response without an explanation (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that subjects (both exposed and non-exposed) were not similar, recruited within very different time frames, or had the very different participation/response rates.

4. Did study design or analysis account for important confounding and modifying variables?

Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments or explicit considerations were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching, adjustment in multivariate model, stratification, propensity scoring, or other methods that were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. In occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered. • **Note:** The following variables should be considered as potential confounders and/or effect measure modifiers for the relationship between PFOA or PFOS exposure and immune outcomes: age, sex, race/ethnicity, smoking, body mass index, alcohol consumption, and variables that represent socioeconomic status (e.g., educational level, household income) based on prior reports of associations with PFOA and PFOS exposure levels (Calafat *et al.* 2007, Nelson *et al.* 2010) and immune outcomes (WHO 1996, Dallaire *et al.* 2005) • **Note:** The following variables should be considered as potential effect measure modifiers of the relationship between PFOA or PFOS exposure and allergy/asthma or cold incidence: maternal/paternal asthma or

allergy, age at start of daycare, anti-inflammatory medication and diet/nutritional status based on established practices in immunotoxicology (WHO 1999a, Dallaire *et al.* 2005)

- **Note:** The following variables should be considered as potential effect measure modifiers for the relationship between PFOA or PFOS exposure and antibody response to vaccination: previous history of vaccination, anti-inflammatory medication, acute stress (WHO 1996, Dallaire *et al.* 2005)
- **Note:** Exposure to other known or suspected immunotoxicants will be considered as confounders (e.g., lead or polychlorinated biphenyls (PCBs); perfluorononanoate (PFNA) and perfluorohexane sulfonate (PFHxS) will be considered as potential confounders because there is some evidence that these (non-PFOA and non-PFOS) perfluorinated compounds (PFCs) may be associated with immune effects; PFOA and PFOS may serve as confounders for potential effects of the other compound respectively
- **Note:** if reported, consumption of olestra will be considered as there is evidence that olestra decreased the absorption of PFOA from the gastrointestinal tract of mice (Jandacek *et al.* 2010)

Probably Low Risk of Bias (+)

- Indirect evidence that appropriate adjustments were made,
- **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results,
- **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements,
- **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research),
- **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for,
- **OR** it is deemed that co-exposures present would not appreciably bias results.
- **Note:** this includes insufficient information provided on co-exposures in general population studies.

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that the distribution of important covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses,
- **OR** there is insufficient information provided about the distribution of known confounders (record "NR" as basis for answer),
- **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity,
- **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer),
- **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for,
- **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).

Definitely High Risk of Bias (--)

- Direct evidence that the distribution of important covariates and known confounders differed between the groups, confounding was demonstrated, and was not appropriately adjusted for in the final analyses,
- **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements,
- **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across the primary study groups, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably Low Risk of Bias (+)
• Indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that exclusion of subjects from analyses was not adequately addressed, • **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that exclusion of subjects from analyses was not adequately addressed. • **Note**: Unacceptable handling of subject exclusion from analyses includes: reason for exclusion likely to be related to true outcome, with either imbalance in numbers or reasons for exclusion across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure (e.g., measurement of PFOA or PFOS in drinking water or measurement of PFOA or PFOS in blood, serum, or plasma), • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods, • **AND** exposure was assessed in a relevant time-window for development of the outcome. Current exposure measures will be considered relevant for any immune outcome for two reasons: (1) they are likely to be relevant for most immune outcomes other than disease endpoints with a significant lag time such as autoimmune disease, and (2) because of the long half-life value of PFOA and PFOS in humans (2-8 years) such that current exposure levels may be indicative of past exposures, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished, • **AND** the study used spiked samples to confirm assay performance. • **Note:** Measurement of serum or whole-blood PFOA or PFOS is the standard accepted biomarker of exposure (not urine or feces) in humans using quantitative techniques such as liquid chromatography-triple quadrupole mass spectrometry (LC-MS/MS) and high pressure liquid chromatography with tandem mass spectrometry (HPLC/MS) (ATSDR 2009, CDC 2009, US EPA 2014d, a, CDC 2015). • **Note:** It is understood that serum levels of PFOA and PFOS may indicate exposure to PFOA or PFOS or exposure to other PFCs that may break down into PFOA or PFOS. For this evaluation we will use the available exposure metric as an indication of exposure to PFOA or PFOS.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), • **OR** exposure was assessed using indirect measures (e.g., drinking water levels and residency, questionnaire or occupational exposure assessment by a certified industrial hygienist) that have been validated or empirically shown to be consistent with methods that directly measure exposure (i.e., inter-methods validation: one method vs. another),

- **AND** exposure was assessed in a relevant time-window for development of the outcome. Current exposure measures will be considered relevant for any immune outcome for two reasons: (1) they are likely to be relevant for most immune outcomes other than disease endpoints with a significant lag time such as autoimmune disease, and (2) because of the long half-life value of PFOA and PFOS in humans (2-8 years) such that current exposure levels may be indicative of past exposures,
- **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed),
- **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure
- **OR** there is evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., a job-exposure matrix or self-report without validation) (record "NR" as basis for answer),
- **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).

Definitely High Risk of Bias (--)

- Direct evidence that the exposure was assessed using methods with poor validity,
- **OR** evidence of exposure misclassification (e.g., differential recall of self-reported exposure).

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)

- Direct evidence that the immune outcome was assessed using well-established methods (the gold standard),
- **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes.
- **NOTE** Well-established assessment methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits, commercial laboratories, or standard assays such as ELISAs for IgG with sufficiently low variation and limits of detection to allow discrimination between groups (or evidence that the assay could have detected a difference based on responses to a positive control); doctor diagnosis of asthma or incidence data obtained from medical records; obtained from registries (Shamliyan *et al.* 2010).

Probably Low Risk of Bias (+)

- Indirect evidence that the outcome was assessed using acceptable methods,
- **OR** it is deemed that the outcome assessment methods used would not appreciably bias results,
- **AND** there is indirect evidence that the outcome assessors were adequately blinded to the exposure level, and it is unlikely that they could have broken the blinding prior to reporting outcomes,
- **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results (including that subjects self-reporting outcomes were likely not aware of reported links between the exposure and outcome lack of blinding is unlikely to bias a particular outcome).
- **NOTE:** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes such as asthma and mining of data collected for other purposes. Proxy reporting (e.g., parental reporting of days sick or doctor-diagnosis) of immune disease, colds, etc. should be considered on a case-by-case basis with consideration of whether or not there is empirical evidence as to the reliability of proxy reporting for that outcome.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were likely aware of reported links between the exposure and outcome), • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** there is direct evidence that outcome assessors were aware of the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were aware of reported links between the exposure and outcome).

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no PFOA- or PFOS-specific additions to the risk-of-bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Case Control Studies

1. Was administered dose or exposure level adequately randomized? [NA]

2. Was allocation to study groups adequately concealed? [NA]

3. Did selection of study participants result in the appropriate comparison groups?

Definitely Low Risk of Bias (++)
• Direct evidence that cases and controls were similar (e.g., recruited from the same eligible population including being of similar age, gender, ethnicity, and eligibility criteria other than outcome of interest as appropriate), recruited within the same time frame, and controls are described as having no history of the outcome, • **Note:** A study will be considered low risk of bias if baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables (see question #4), • **Note:** Immune-specific exclusion criteria should be discussed including the presence or history of infectious or autoimmune diseases other than the outcome of interest or diagnosis of the outcome of interest before participation in the study.
Probably Low Risk of Bias (+)
• Indirect evidence that cases and controls were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age), recruited within the same time frame, and controls are described as having no history of the outcome, • **OR** it is deemed differences between cases and controls would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames, • **OR** there is insufficient information provided about the appropriateness of controls including rate of response reported for cases only (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames.

4. Did study design or analysis account for important confounding and modifying variables?

Definitely Low Risk of Bias (++)
• Direct evidence that appropriate adjustments were made for the variables listed below as potential confounders and/or effect measure modifiers in the final analyses through the use of statistical models to reduce research-specific bias including standardization, matching of cases and controls, adjustment in multivariate model, stratification, propensity scoring, or other methods were appropriately justified, • **AND** there is direct evidence that primary covariates and confounders were assessed using valid and reliable measurements, • **AND** there is direct evidence that other exposures anticipated to bias results were not present or were appropriately measured and adjusted for. • **Note:** The following variables should be considered as potential confounders and/or effect measure modifiers for the relationship between PFOA or PFOS exposure and immune outcomes: age, sex, race/ethnicity, smoking, body mass index, alcohol consumption, and variables that represent socioeconomic status (e.g., educational level, household income) based on prior reports of associations with PFOA and PFOS exposure levels (Calafat *et al.* 2007, Nelson *et al.* 2010) and immune outcomes (WHO 1996, Dallaire *et al.* 2005) • **Note:** The following variables should be considered as potential effect measure modifiers of the relationship between PFOA or PFOS exposure and allergy/asthma or cold incidence: maternal/paternal asthma or allergy, age at start of daycare, anti-inflammatory medication and diet/nutritional status based on established practices in immunotoxicology (WHO 1999a, Dallaire *et al.* 2005) • **Note:** The following variables should be considered as potential effect measure modifiers for the relationship between PFOA or PFOS exposure and antibody response to vaccination: previous history of vaccination,

anti-inflammatory medication, acute stress (WHO 1996, Dallaire *et al.* 2005)

- **Note:** Exposure to other known or suspected immunotoxicants will be considered as confounders (e.g., lead or polychlorinated biphenyls (PCBs); perfluorononanoate (PFNA) and perfluorohexane sulfonate (PFHxS) will be considered as potential confounders because there is some evidence that these (non-PFOA and non-PFOS) perfluorinated compounds (PFCs) may be associated with immune effects; PFOA and PFOS may serve as confounders for potential effects of the other compound respectively
- **Note:** if reported, consumption of olestra will be considered as there is evidence that olestra decreased the absorption of PFOA from the gastrointestinal tract of mice (Jandacek *et al.* 2010)

Probably Low Risk of Bias (+)

- Indirect evidence that appropriate adjustments were made,
- **OR** it is deemed that not considering or only considering a partial list of covariates or confounders in the final analyses would not appreciably bias results,
- **AND** there is evidence (direct or indirect) that covariates and confounders considered were assessed using valid and reliable measurements,
- **OR** it is deemed that the measures used would not appreciably bias results (i.e., the authors justified the validity of the measures from previously published research),
- **AND** there is evidence (direct or indirect) that other co-exposures anticipated to bias results were not present or were appropriately adjusted for,
- **OR** it is deemed that co-exposures present would not appreciably bias results.
- **Note:** this includes insufficient information provided on co-exposures in general population studies.

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that the distribution of important covariates and known confounders differed between cases and controls and was not investigated further,
- **OR** there is insufficient information provided about the distribution of known confounders in cases and controls (record "NR" as basis for answer),
- **OR** there is indirect evidence that covariates and confounders considered were assessed using measurements of unknown validity,
- **OR** there is insufficient information provided about the measurement techniques used to assess covariates and confounders considered (record "NR" as basis for answer),
- **OR** there is indirect evidence that there was an unbalanced provision of additional co-exposures across cases and controls, which were not appropriately adjusted for,
- **OR** there is insufficient information provided about co-exposures in occupational studies or studies of contaminated sites where high exposures to other chemical exposures would have been reasonably anticipated (record "NR" as basis for answer).

Definitely High Risk of Bias (--)

- Direct evidence that the distribution of important covariates and known confounders differed between cases and controls, confounding was demonstrated, but was not appropriately adjusted for in the final analyses,
- **OR** there is direct evidence that covariates and confounders considered were assessed using non valid measurements,
- **OR** there is direct evidence that there was an unbalanced provision of additional co-exposures across cases and controls, which were not appropriately adjusted for.

5. Were experimental conditions identical across study groups? [NA]

6. Were the research personnel blinded to the study group during the study? [NA]

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)

- Direct evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.

Probably Low Risk of Bias (+)
• Indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that exclusion of subjects from analyses was not adequately addressed, • **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that exclusion of subjects from analyses was not adequately addressed. • **Note**: Unacceptable handling of subject exclusion from analyses includes: reason for exclusion likely to be related to true outcome, with either imbalance in numbers or reasons for exclusion across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure (e.g., measurement of PFOA or PFOS in drinking water or measurement of PFOA or PFOS in blood, serum, or plasma), • **OR** exposure was assessed using less-established methods that directly measure exposure and are validated against well-established methods. • **AND** exposure was assessed in a relevant time-window for development of the outcome. Current exposure measures will be considered relevant for any immune outcome for two reasons: (1) they are likely to be relevant for most immune outcomes other than disease endpoints with a significant lag time such as autoimmune disease, and (2) because of the long half-life value of PFOA and PFOS in humans (2-8 years) such that current exposure levels may be indicative of past exposures, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes, • **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished, • **AND** the study used spiked samples to confirm assay performance. • **Note:** Measurement of serum or whole-blood PFOA or PFOS is the standard accepted biomarker of exposure (not urine or feces) in humans using quantitative techniques such as liquid chromatography-triple quadrupole mass spectrometry (LC-MS/MS) and high pressure liquid chromatography with tandem mass spectrometry (HPLC/MS) (ATSDR 2009, CDC 2009, US EPA 2014d, a, CDC 2015). • **Note:** It is understood that serum levels of PFOA and PFOS may indicate exposure to PFOA or PFOS or exposure to other PFCs that may break down into PFOA or PFOS. For this evaluation we will use the available exposure metric as an indication of exposure to PFOA or PFOS.
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure was consistently assessed using well-established methods that directly measure exposure), • **OR** exposure was assessed using indirect measures (e.g., drinking water levels and residency, questionnaire or occupational exposure assessment by a certified industrial hygienist) that have been validated or empirically shown to be consistent with methods that directly measure exposure (i.e., inter-methods validation: one method vs. another), • **AND** exposure was assessed in a relevant time-window for development of the outcome. Current exposure measures will be considered relevant for any immune outcome for two reasons: (1) they are likely to be relevant for most immune outcomes other than disease endpoints with a significant lag time such as autoimmune disease, and (2) because of the long half-life value of PFOA and PFOS in humans (2-8 years) such that current exposure levels may be indicative of past exposures, • **AND** there is sufficient range or variation in exposure measurements across groups to potentially identify associations with health outcomes (at a minimum from high exposure or ever exposed from low exposure or never exposed),

- **AND** there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure,
- **OR** there is direct evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., a job-exposure matrix or self-report without validation) (record "NR" as basis for answer),
- **OR** there is insufficient information provided about the exposure assessment, including validity and reliability, but no evidence for concern about the method used (record "NR" as basis for answer).

Definitely High Risk of Bias (--)

- Direct evidence that the exposure was assessed using methods with poor validity,
- **OR** evidence of exposure misclassification (e.g., differential recall of self-reported exposure).

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)

- Direct evidence that the immune outcome was assessed in cases (i.e., case definition) and controls using well-established methods (the gold standard),
- **AND** subjects had been followed for the same length of time in all study groups,
- **AND** there is direct evidence that the outcome assessors (including study subjects, if outcomes were self-reported) were adequately blinded to the exposure level when outcome was assessed in cases (i.e., case definition) and controls.
- **NOTE** Well-established methods will depend on the outcome, but examples of such methods may include: doctor diagnosis of asthma or doctor diagnosis obtained from medical records.

Probably Low Risk of Bias (+)

- Indirect evidence that the outcome was assessed in cases (i.e., case definition) and controls using acceptable methods),
- **AND** subjects had been followed for the same length of time in all study groups,
- **OR** it is deemed that the outcome assessment methods used would not appreciably bias results,
- **AND** there is indirect evidence that the outcome assessors were adequately blinded to the exposure level when reporting outcomes,
- **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results (including that subjects self-reporting outcomes were likely not aware of reported links between the exposure and outcome or lack of blinding is unlikely to bias a particular outcome).
- **NOTE** Acceptable, but not ideal assessment methods will depend on the outcome, but examples of such methods may include proxy reporting of outcomes such as asthma and mining of data collected for other purposes. Proxy reporting of immune disease should be considered on a case-by-case basis with consideration of whether or not there is empirical evidence as to the reliability of proxy reporting for that outcome.

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that the outcome was assessed in cases (i.e., case definition) using an insensitive instrument,
- **OR** there is insufficient information provided about how cases were identified (record "NR" as basis for answer).
- **OR** there is indirect evidence that it was possible for outcome assessors to infer the exposure level prior to reporting outcomes (including that subjects self-reporting outcomes were likely aware of reported links between the exposure and outcome),
- **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).

Definitely High Risk of Bias (--)

- Direct evidence that the outcome was assessed in cases (i.e., case definition) using an insensitive instrument,
- **OR** there is direct evidence that outcome assessors were aware of the exposure level prior to reporting

outcomes (including that subjects self-reporting outcomes were aware of reported links between the exposure and outcome).

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no PFOA- or PFOS-specific additions to the risk-of-bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

Experimental Animal Studies

1. Was administered dose or exposure level adequately randomized?

Definitely Low Risk of Bias (++)
• Direct evidence that animals were allocated to any study group including controls using a method with a random component, • **AND** there is direct evidence that the study used a concurrent control group as an indication that randomization covered all study groups, • **Note:** Acceptable methods of randomization include: referring to a random number table, using a computer random number generator, coin tossing, or shuffling cards (Higgins and Green 2011). • **Note:** Restricted randomization (e.g., blocked randomization) to ensure particular allocation ratios will be considered low bias. Similarly, stratified randomization approaches that attempt to minimize imbalance between groups on important prognostic factors (e.g., body weight) will be considered acceptable.
Probably Low Risk of Bias (+)
• Indirect evidence that animals were allocated to any study group including controls using a method with a random component (i.e., authors state random allocation, without description of method), • **AND** evidence that the study used a concurrent control group as an indication that randomization covered all study groups, • **OR** it is deemed that allocation without a clearly random component would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that animals were allocated to study groups using a method with a non-random component, • **OR** indirect evidence that there was a lack of a concurrent control group, • **OR** there is insufficient information provided about how animals were allocated to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• 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, • **OR** direct evidence that there was a lack of a concurrent control group.

2. Was allocation to study groups adequately concealed?

Definitely Low Risk of Bias (++)
• Direct evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to, and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable. • **Note:** Acceptable methods used to ensure allocation concealment include sequentially numbered treatment containers of identical appearance or equivalent methods.
Probably Low Risk of Bias (+)
• Indirect evidence that at the time of assigning study groups the research personnel did not know what group animals were allocated to and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable, • **OR** it is deemed that lack of adequate allocation concealment would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable, • **OR** there is *insufficient* information provided about allocation to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that at the time of assigning study groups it was possible for the research personnel to know what group animals were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable.

3. Did selection of study participants result in the appropriate comparison groups? [NA]

4. Did study design or analysis account for important confounding and modifying variables? [NA]

5. Were experimental conditions identical across study groups?

Definitely Low Risk of Bias (++)
• Direct evidence that same vehicle was used in control and experimental animals, • **AND** direct evidence that non-treatment-related experimental conditions were identical across study groups (i.e., the study report explicitly provides this level of detail).
Probably Low Risk of Bias (+)
• Indirect evidence that the same vehicle was used in control and experimental animals, • **OR** it is deemed that the vehicle used would not appreciably bias results, • **AND** identical non-treatment-related experimental conditions are assumed if authors did not report differences in housing or husbandry.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the vehicle differed between control and experimental animals, • **OR** authors did not report the vehicle used (record "NR" as basis for answer), • **OR** there is indirect evidence that non-treatment-related experimental conditions were not comparable between study groups.
Definitely High Risk of Bias (--)
• Direct evidence from the study report that control animals were untreated, or treated with a different vehicle than experimental animals, • **OR** there is direct evidence that non-treatment-related experimental conditions were not comparable between study groups.

6. Were the research personnel blinded to the study group during the study?

Definitely Low Risk of Bias (++)
• Direct evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study. Methods used to ensure blinding include central allocation; sequentially numbered treatment containers of identical appearance; sequentially numbered animal cages; or equivalent methods.
Probably Low Risk of Bias (+)
• Indirect evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study, • **OR** it is deemed that lack of adequate blinding during the study would not appreciably bias results. This would include cases where blinding was not possible but research personnel took steps to minimize potential bias, such as restricting the knowledge of study group to veterinary or supervisory personnel monitoring for overt toxicity, or randomized husbandry or handling practices (e.g., placement in the animal room, necropsy order, etc.).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the research personnel were not adequately blinded to study group, • **OR** there is insufficient information provided about blinding to study group during the study (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the research personnel were not adequately blinded to study group.

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study.

- **Note**: Acceptable handling of attrition includes: very little missing outcome data; reasons for missing animals unlikely to be related to outcome (or for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups; missing outcomes is not enough to impact the effect estimate.
- **OR** missing data have been imputed using appropriate methods (insuring that characteristics of animals are not significantly different from animals retained in the analysis).

Probably Low Risk of Bias (+)

- Indirect evidence that loss of animals was adequately addressed and reasons were documented when animals were removed from a study,
- **OR** it is deemed that the proportion lost would not appreciably bias results. This would include reports of no statistical differences in characteristics of animals removed from the study from those remaining in the study.

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that loss of animals was unacceptably large and not adequately addressed,
- **OR** there is insufficient information provided about loss of animals (record "NR" as basis for answer).

Definitely High Risk of Bias (--)

- Direct evidence that loss of animals was unacceptably large and not adequately addressed.
- **Note**: Unacceptable handling of attrition or exclusion includes: reason for loss is likely to be related to true outcome, with either imbalance in numbers or reasons for loss across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)

- Direct evidence that the exposure to PFOA, PFOS, or their salts (including purity but not stability because PFOA and PFOS are extremely stable compounds that are even stable under environmental conditions) was independently characterized and purity confirmed generally as ≥98%, (and compliance with the treatment, if applicable)
- **AND** that exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups,
- **AND** for dietary or drinking water studies that information is provided on consumption or internal dose metrics to confirm expected exposure levels sufficiently to allow discrimination between exposure groups,
- **AND** if internal dose metrics are available, there is evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished,
- **AND** if internal dose metrics are available, the study used spiked samples to confirm assay performance.
- **Note:** if internal dose measurements are made, measurement of serum or whole-blood PFOA or PFOS is the standard accepted biomarker of exposure (preferred over urine or feces) using quantitative techniques such as liquid chromatography-triple quadrupole mass spectrometry (LC-MS/MS) and high pressure liquid chromatography with tandem mass spectrometry (HPLC/MS) (ATSDR 2009, CDC 2009, US EPA 2014d, a, CDC 2015).

Probably Low Risk of Bias (+)

- Indirect evidence that the exposure to PFOA, PFOS, or their salts (including purity but not stability because PFOA and PFOS are extremely stable compounds that are even stable under environmental conditions) was appropriately characterized and purity confirmed generally as ≥98% (i.e., the supplier of the chemical provides documentation of the purity of the chemical),
- **OR** direct evidence that purity was independently confirmed as ≥95% and it is deemed that impurities of up to 5% would not appreciably bias results,
- **AND** that exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups,
- **AND** for dietary or drinking water studies no information is provided on consumption or internal dose metrics,
- **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.

Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure (including purity of the test substance and compliance with the treatment, if applicable) was assessed using poorly validated methods, • **OR** there is insufficient information provided about the validity of the exposure assessment method, but no evidence for concern (record "NR" as basis for answer), • **AND** if internal dose metrics are available, there is indirect evidence that most of the exposure data measurements are below the limit of quantitation for the assay such that different exposure groups cannot be distinguished.
Definitely High Risk of Bias (--)
• Direct evidence that the exposure (including purity of the test substance and compliance with the treatment, if applicable) was assessed using poorly validated methods.

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the outcome was assessed using well-established methods (e.g., gold standard) • **AND** assessed at the same length of time after initial exposure in all study groups, • **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes. • **NOTE** Well-established methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits, commercial laboratories with experience in the assay, or standard assays such as ELISAs for IgG and with sufficiently low variation and limits of detection to allow discrimination of responses between treatment groups (or direct evidence that the assay could have detected a difference based on responses to a positive control).
Probably Low Risk of Bias (+)
• Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard), • **AND** assessed at the same length of time after initial exposure in all study groups, • **OR** it is deemed that the outcome assessment methods used would not appreciably bias results, • **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes, • **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures. • **NOTE** For some outcomes, particularly histopathology assessment, outcome assessors are not blind to study group as they require comparison to the control to appropriately judge the outcome, but additional measures such as multiple levels of independent review by trained pathologists can minimize potential bias. • **NOTE** Acceptable assessment methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits with some variation, but ability to discriminate between the high dose treatment and control group (or indirect evidence that the assay could have detected a difference based on responses to a positive control).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of time after initial exposure differed by study group, • **OR** there is indirect evidence that it was possible for outcome assessors to infer the study group prior to reporting outcomes without sufficient quality control measures, • **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).

Definitely High Risk of Bias (--)
• Direct evidence that the outcome assessment method is an insensitive instrument, • **OR** the length of time after initial exposure differed by study group, • **OR** there is direct evidence for lack of adequate blinding of outcome assessors, including no blinding or incomplete blinding without quality control measures.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.
Probably Low Risk of Bias (+)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, • **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported, • **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results, • **OR** there is insufficient information provided about selective outcome reporting (record "NR" as answer basis).
Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no PFOA- or PFOS-specific additions to the risk-of-bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.

In vitro Studies

1. Was administered dose or exposure level adequately randomized?

Definitely Low Risk of Bias (++)
• Direct evidence that cells were allocated to any study group including controls using a method with a random component, • **AND** direct evidence that the study used a concurrent control group as an indication that randomization covered all study groups, • **OR** all cells in culture come from a homogenous cell suspension recently collected from cell culture vessels following appropriate cell culture techniques (e.g., US EPA 2014a). • **Note:** Acceptable methods of randomization include: referring to a random number table, using a computer random number generator, coin tossing, or shuffling cards (Higgins and Green 2011).
Probably Low Risk of Bias (+)
• Indirect evidence that cells were allocated to any study group including controls using a method with a random component (i.e., authors state random allocation, without description of the method), • **AND** evidence that the study used a concurrent control group as an indication that randomization covered all study groups, • **OR** it is deemed that allocation without a clearly random component would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that cells were allocated to study groups using a method with a non-random component, • **OR** indirect evidence that there was a lack of a concurrent control group, • **OR** there is insufficient information provided about how cells were allocated to study groups (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that cells were allocated to study groups using a method with a non-random component including judgment of the investigator or the results of laboratory tests, • **OR** direct evidence that there was a lack of a concurrent control group.

2. Was allocation to study groups adequately concealed?

Definitely Low Risk of Bias (++)
• Direct evidence that at the time of assigning study groups the research personnel did not know what group cells were allocated to, and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable. • **Note:** Acceptable methods used to ensure allocation concealment include sequentially numbered treatment containers of identical appearance or equivalent methods.
Probably Low Risk of Bias (+)
• Indirect evidence that at the time of assigning study groups the research personnel did not know what group cells were allocated to and it is unlikely that they could have broken the blinding of allocation until after assignment was complete and irrevocable, • **OR** it is deemed that lack of adequate allocation concealment would not appreciably bias results. This may also be the case for *in vitro* studies with very low potential differences between cells that comprise the different groups, e.g., cells pipetted from a homogeneous cell suspension (single or mixed cell types) recently collected from cell culture vessels by accepted methods.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that at *the* time of assigning study groups it was possible for the research personnel to know what group cells were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable, • **OR** there is *insufficient* information provided about allocation to study groups (record "NR" as basis for answer).

Definitely High Risk of Bias (--)
• Direct evidence that at the time of assigning study groups it was possible for the research personnel to know what group cells were allocated to, or it is likely that they could have broken the blinding of allocation before assignment was complete and irrevocable.

3. Did selection of study participants result in the appropriate comparison groups? [NA]

4. Did study design or analysis account for important confounding and modifying variables? [NA]

5. Were experimental conditions identical across study groups?

Definitely Low Risk of Bias (++)
• Direct evidence that culture conditions included identical concentrations of any solvents (e.g., DMSO) used in getting the treatment compound into solution, • **AND** the same media was used for control and experimental cells particularly for biological materials such as serum which must be from the same lot, • **AND** appropriate adjustments were made such as normalization to blank/media controls, cell numbers in culture, use of positive and negative control responses in acceptance criteria, or others, • **AND** non-treatment-related experimental conditions were identical across study groups (i.e., the study report explicitly provides this level of detail).
Probably Low Risk of Bias (+)
• Indirect evidence that culture conditions included identical concentrations of any solvents (e.g., DMSO) used in getting the treatment compound into solution, • **AND** the same media was used for control and experimental cells, • **OR** it is deemed that the media used would not appreciably bias results, • **AND** appropriate adjustments were made such as normalization to blank/media controls, cell numbers in culture, use of positive and negative control responses in acceptance criteria, or others, • **OR** it is deemed that not considering or only considering a partial list of covariates in the final analyses would not appreciably bias results, • **AND** as described above, identical non-treatment-related experimental conditions are assumed if authors did not report differences in culture conditions or handling.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the concentration of solvents used in getting the treatment compound into solution differed between control and experimental cells, • **OR** there is indirect evidence that the media differed between control and experimental cells, • **OR** there is insufficient information provided on maintaining identical concentrations of solvents (record “NR” as basis for answer), • **OR** there is indirect evidence that appropriate adjustments were not made such as failing to normalize to blank/media controls, adjust for cell numbers in culture, use positive and negative control responses in acceptance criteria, or others, • **OR** there is insufficient information provided about analysis of relevant covariates (record “NR” as basis for answer), • **OR** there is indirect evidence that non-treatment-related experimental conditions were not comparable between study groups.

Definitely High Risk of Bias (--)
• Direct evidence from the study report that the concentration of solvents used in getting the treatment compound into solution differed between control and experimental cells, • **OR** there is direct evidence that the media (or biological components such as serum) differed between control and experimental cells, • **OR** there is direct evidence that appropriate adjustments were not made such as failing to normalize to blank/media controls, adjust for cell numbers in culture, use positive and negative control responses in acceptance criteria, or other relevant covariates, • **OR** there is direct evidence that non-treatment-related experimental conditions were not comparable between study groups.

6. Were the research personnel blinded to the study group during the study?

Definitely Low Risk of Bias (++)
• Direct evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study. Methods used to ensure blinding include central allocation, sequentially numbered treatment containers of identical appearance; sequentially numbered culture plates, or equivalent, • **OR** the use of robotic testing systems during the study that are deemed to eliminate the opportunity for performance bias to influence results.
Probably Low Risk of Bias (+)
• Indirect evidence that the research personnel were adequately blinded to study group, and it is unlikely that they could have broken the blinding during the study, • **OR** it is deemed that lack of adequate blinding during the study would not appreciably bias results (e.g., minimal possibility of researchers to handle cells or plates after treatment due to primarily automated procedures).
Probably High
• Indirect evidence that the research personnel were not adequately blinded to study group, • **OR** there is insufficient information provided about blinding to study group during the study (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the research personnel were not adequately blinded to study group.

7. Were outcome data complete without attrition or exclusion from analysis?

Definitely Low Risk of Bias (++)
• Direct evidence that loss of cells was adequately addressed and reasons were documented when wells or plates were removed from a study (e.g., visual observation of contamination, cells missing from wells due to pipetting error, visual morphological changes in cells unexplainable based on surrounding wells, documented removal of statistical outliers). • **Note**: Acceptable handling of attrition includes: very little missing outcome data; reasons for missing cells unlikely to be related to outcome (or for viability data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups; missing outcomes is not enough to impact the effect.
Probably Low Risk of Bias (+)
• Indirect evidence that loss of cells was adequately addressed and reasons were documented when wells or plates were removed from a study, • **OR** it is deemed that the proportion lost would not appreciably bias results.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that loss of wells or culture plates was unacceptably large and not adequately addressed, • **OR** there is insufficient information provided about loss of cells (record "NR" as basis for answer).

Definitely High Risk of Bias (--)
• Direct evidence that loss of cells, wells, or plates was unacceptably large and not adequately addressed. • **Note**: Unacceptable handling of attrition or exclusion includes: reason for loss is likely to be related to true outcome, with either imbalance in numbers or reasons for loss across study groups.

8. Can we be confident in the exposure characterization?

Definitely Low Risk of Bias (++)
• Direct evidence that the exposure to PFOA, PFOS, or their salts (including purity but not stability because PFOA and PFOS are extremely stable compounds that are even stable under environmental and biological conditions) was independently characterized and purity confirmed generally as ≥98%, • **AND** that exposure was consistently administered (i.e., with the same method and time-frame) across treatment groups, • **AND** solubility and volatility to the test substance have been addressed with appropriate methods (no special methods required as PFOA and PFOS are not considered volatile compounds). • **AND** if assay media were examined for actual exposure concentrations, there is direct evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished, • **AND** if assay media were examined for actual exposure concentrations, the study used spiked samples to confirm assay performance. • **Note:** quantitative techniques such as liquid chromatography-triple quadrupole mass spectrometry (LC-MS/MS) and high pressure liquid chromatography with tandem mass spectrometry (HPLC/MS) are among the standard assays for determining PFOA and PFOS concentrations (ATSDR 2009, CDC 2009, US EPA 2014d, a, CDC 2015).
Probably Low Risk of Bias (+)
• Indirect evidence that the exposure to PFOA, PFOS, or their salts (including purity but not stability because PFOA and PFOS are extremely stable compounds that are even stable under environmental and biological conditions) was appropriately characterized and purity confirmed generally as ≥98% (i.e., the supplier of the chemical provides documentation of the purity of the chemical), • **OR** direct evidence that purity was independently confirmed as ≥95% and it is deemed that impurities of up to 5% would not appreciably bias results, • **AND** solubility of the test substance have been adequately addressed • **AND** if assay media were examined for actual exposure concentrations, there is indirect evidence that most of the exposure data measurements are above the limit of quantitation for the assay such that different exposure groups can be distinguished.
Probably High Risk of Bias (-) or (NR)
• Indirect evidence that the exposure to PFOA, PFOS, or their salts (including purity) was assessed using poorly validated methods, • **OR** there is insufficient information provided about the validity of the exposure assessment method, but no direct evidence for concern (record "NR" as basis for answer).
Definitely High Risk of Bias (--)
• Direct evidence that the exposure to PFOA, PFOS, or their salts (including purity) was assessed using poorly validated methods, • **OR** solubility of the test substance were not appropriately controlled.

9. Can we be confident in the outcome assessment?

Definitely Low Risk of Bias (++)
• Direct evidence that the outcome was assessed using well-established methods (the gold standard), • **AND** assessed at the same length of time after initial exposure in all study groups, • **AND** there is direct evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes.

- **NOTE** Well-established methods will depend on the outcome, but examples of such methods may include: objectively measured cytokine concentrations with diagnostic methods using commercial kits, commercial laboratories with experience in the assay, or standard assays such as ELISAs for IgG and with sufficiently low variation and limits of detection to allow discrimination of responses between treatment groups (or direct evidence that the assay could have detected a difference based on responses to a positive control).

Probably Low Risk of Bias (+)

- Indirect evidence that the outcome was assessed using acceptable methods (i.e., deemed valid and reliable but not the gold standard),
- **AND** assessed at the same length of time after initial exposure in all study groups,
- **OR** it is deemed that the outcome assessment methods used would not appreciably bias results,
- **AND** there is indirect evidence that the outcome assessors were adequately blinded to the study group, and it is unlikely that they could have broken the blinding prior to reporting outcomes,
- **OR** it is deemed that lack of adequate blinding of outcome assessors would not appreciably bias results, which is more likely to apply to objective outcome measures.
- **NOTE** Acceptable assessment methods will depend on the outcome, but examples of such methods may include: objectively measured antibody or cytokine concentrations with diagnostic methods using commercial kits with some variation, but ability to discriminate between the high dose treatment and control group (or indirect evidence that the assay could have detected a difference based on responses to a positive control).

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that the outcome assessment method is an insensitive instrument,
- **OR** the length of time after initial exposure differed by study group,
- **OR** there is indirect evidence that it was possible for outcome assessors to infer the study group prior to reporting outcomes without sufficient quality control measures,
- **OR** there is insufficient information provided about blinding of outcome assessors (record "NR" as basis for answer).

Definitely High Risk of Bias (--)

- Direct evidence that the outcome assessment method is an insensitive instrument,
- **OR** the length of time after initial exposure differed by study group,
- **OR** there is direct evidence for lack of adequate blinding of outcome assessors, including no blinding or incomplete blinding without quality control measures.

10. Were all measured outcomes reported?

Definitely Low Risk of Bias (++)

- Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction and analyses had been planned in advance.

Probably Low Risk of Bias (+)

- Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported,
- **OR** analyses that had not been planned in advance (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and deemed that unplanned analyses were appropriate and selective reporting would not appreciably bias results (e.g., appropriate analyses of an unexpected effect). This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).

Probably High Risk of Bias (-) or (NR)

- Indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported,
- **OR** and there is indirect evidence that unplanned analyses were included that may appreciably bias results,
- **OR** there is insufficient information provided about selective outcome reporting (record "NR" as basis for answer).

Definitely High Risk of Bias (--)
• Direct evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified, or that unplanned analyses were included that would appreciably bias results.

11. Were there no other potential threats to internal validity?

There are no PFOA- or PFOS-specific additions to the risk-of-bias questions for this evaluation. This question will be used to examine individual studies for appropriate statistical methods (e.g., confirmation of homogeneity of variance for ANOVA and other statistical tests that require normally distributed data). It will also be used for risk-of-bias considerations that do not fit under the other questions.