Office of Pesticide Programs' Framework for Incorporating Human Epidemiologic & Incident Data in Risk Assessments for Pesticides

December 28, 2016

Office of Pesticide Programs
US Environmental Protection Agency

Table of Contents

I. PURPOSE & SCOPE

The Environmental Protection Agency's (EPA) Office of Pesticide Programs (OPP) is a licensing program regulating pesticides in the U.S under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) and the Federal Food, Drug, and Cosmetic Act (FFDCA). As part of this program, OPP evaluates a substantial body of toxicology and exposure data to assess the effects of pesticides on human health and the environment. In evaluating human health, EPA looks first for information directly evaluating the potential for effects to people, including epidemiological data. Historically, however, few epidemiology studies have been available to inform the potential toxicity of pesticide chemicals. As such, OPP has in the past primarily relied on toxicology studies in laboratory animals to assess the hazard potential and to estimate human health risk. With the publication of numerous papers from the Agricultural Health Study[1] and from the National Institute of Environmental Health Sciences (NIEHS)/EPA Children's Centers[2], among others, the availability of epidemiology studies conducted on U.S.-relevant exposures to pesticides is increasing. Nevertheless, since the number of pesticides for which quality epidemiology data either exist or are being developed remains relatively low in the near term, experimental laboratory data will likely continue to be the primary source of data for use in quantitative risk assessment for most pesticides.

OPP's goal is to use such information -- when available -- in a scientifically robust and transparent way. To accomplish this, OPP has developed a general epidemiologic framework, as described in this document, that outlines the scientific considerations that OPP will weigh in evaluating how such studies and scientific information can be more fully integrated into risk assessments of pesticide chemicals. The current document is neither a binding regulation nor is it intended to be or serve as a reviewer's guide or manual or as a Standard Operating Procedure for assessing or using epidemiology data. Nor is it intended to be a full treatise on more modern or advanced epidemiological methods or to adequately convey the nuances and complexity that is important for interpreting these types of studies. As such, it does not discuss (or does not discuss in any detail) such important epidemiological topics as causal inference and causal diagrams (Rothman et al., 2012a; Glymor and Greenland, 2012); more recent approaches to confounder identification, assessment, and control; meta-analysis and heterogeneity and its assessment/evaluation (Borenstein et al., 2009; Greenland and O'Rourke, 2012); or sensitivity/quantitative bias analysis for epidemiologic data (Lash et al., 2009; Lash et al,. 2014; Ioannidis, 2008; Greenland and Lash, 2012; Jurek et al., 2007). All these topics, concepts, and issues can and do apply to epidemiology studies concerning pesticides, but are not covered in this OPP framework document. Instead, this document provides overall conceptual considerations concerning the evaluation and use of epidemiology studies on pesticides in

[1] https://aghealth.nih.gov/

[2] https://www.epa.gov/research-grants/niehsepa-childrens-environmental-health-and-disease-prevention-research-centers

the context of human health risk assessments to support OPP's FIFRA and FFDCA activities. An earlier version of this document was reviewed favorably by the FIFRA Scientific Advisory Panel (SAP) in February, 2010 (USEPA, 2010; FIFRA SAP, 2010). This document incorporates improvements recommended by the SAP, public comments, and the experience gained since 2010 conducting assessments on several pesticides for which epidemiological data were available, and should be considered a document that will be updated from time-to-time as we progress and on as-needed basis

II. INTRODUCTION

Two reports by the National Research Council (NRC) of the National Academy of Science (NAS), "Toxicity Testing in the 21st Century: A Vision and A Strategy (2007)" and "Science and Decisions (2009)," together provide new directions in toxicology and risk assessment. These two NRC reports advocate far reaching changes in how toxicity testing is performed, how such data are interpreted, and ultimately how regulatory decisions are made. Specifically, the 2007 report on 21st century toxicity testing advocates a shift away from the current focus of using apical toxicity endpoints to using toxicity pathways[3] to inform toxicity testing, risk assessment, and ultimately decision making. This approach is based on the rapidly evolving scientific understanding of how genes, proteins, and small molecules interact to form molecular pathways that maintain cell function in human cells. The goal for the new toxicity testing paradigm is to determine how exposure to environmental agents can perturb these pathways, thereby causing a cascade of subsequent key events leading to adverse health effects. Human information like that found in epidemiology studies, human incident databases, and biomonitoring studies, along with experimental toxicological information are expected to play a significant role in this new approach. Specifically, these types of human information provide insight into the effects caused by actual chemical exposures in humans and thus can contribute to problem formulation and hazard/risk characterization. In addition, epidemiologic and human incident data can guide additional analyses or data generations (e.g., dose and endpoint selection for use in *in vitro* and targeted *in vivo* experimental studies), identify potentially susceptible populations, identify new health effects, or confirm the existing toxicological observations.

This new vision of toxicity testing and risk assessment will involve data from multiple levels of biological organization ranging from the molecular level up to population-based surveillance with a goal of considering chemical effects from their source to the ultimate health outcome and effects on populations. Such data will come from *in vitro* and *in vivo* experimental studies along with *in silico* and modeled data. OPP's framework for incorporating epidemiology and incident data is conceptually consistent with the 2007 NRC report on 21st century toxicity testing in that both emphasize the use of the best available information from multiple data sources are compiled in a weight of the evidence (WOE) analysis.

---

[3] Toxicity pathways are cellular response pathways that, when sufficiently perturbed, are expected to result in adverse health effects.

As a general principle, occupational and environmental epidemiology studies are conducted only on widely used pesticides; these pesticides also tend to have to be well-studied in the scientific literature. Thus, OPP expects in many cases where epidemiologic data are available, a significant body of literature data on toxicology, exposure, pharmacokinetics (PK), and mode of action/adverse outcome pathway information (MOA/AOP) may also be available. Human incident data are available on a broader range of chemicals, some of which have robust databases and others which do not. In those situations, where there are significant human incident cases and little is known about the MOA/AOP or PK of a particular pesticide, the WOE analysis can be used to identify areas of new research.

OPP's approach in this framework for incorporating epidemiology and human incident data is not a new or novel approach. Instead, this approach is a reasonable, logical extension of existing tools and methods. This document relies on existing guidance documents and frameworks (Table 1) as the starting point for reviewing and evaluating epidemiology and human incident data for use in pesticide risk assessment. This framework on using epidemiology and incident data in human health risk assessment is consistent with the recommendations of the NRC in its 2009 report on *Science and Decisions*, and with the agency's recent Human Health Risk Assessment Framework (USEPA, 2014a) with respect to emphasizing the use of problem formulation as a tool for scoping, planning, and reviewing available, particularly in the context of risk management needs.

Similarly, OPP's framework is consistent with updates to the World Health Organization/International Programme on Chemical Safety MOA/human relevance framework, which highlights the importance of problem formulation and the need to integrate information at different levels of biological organization (Meek et al., 2014). The MOA/HR framework begins with identifying the series of key events that are along the causal path, that are established on weight of evidence, using principles like those described by Bradford Hill, taking into account factors such as dose-response and temporal concordance, biological plausibility, coherence and consistency (Hill, 1965). Using this analytic approach, epidemiologic findings can be evaluated in the context of other human information (including human incident findings) and experimental studies and for identifying areas of uncertainty and future research. However, it is noteworthy that the availability of a fully elucidated MOA/AOP is a not requirement for using epidemiology studies in human health risk assessment. As the agency continues to move forward in implementing the transformative approach in the 2007 and 2009 NRC reports and as OPP gains experience in integration of epidemiology and human incident information, OPP will re-evaluate and update this framework as appropriate.

Figure 1. Schematic of the adverse outcome pathway. Adapted from Ankley *et al.* (2010).

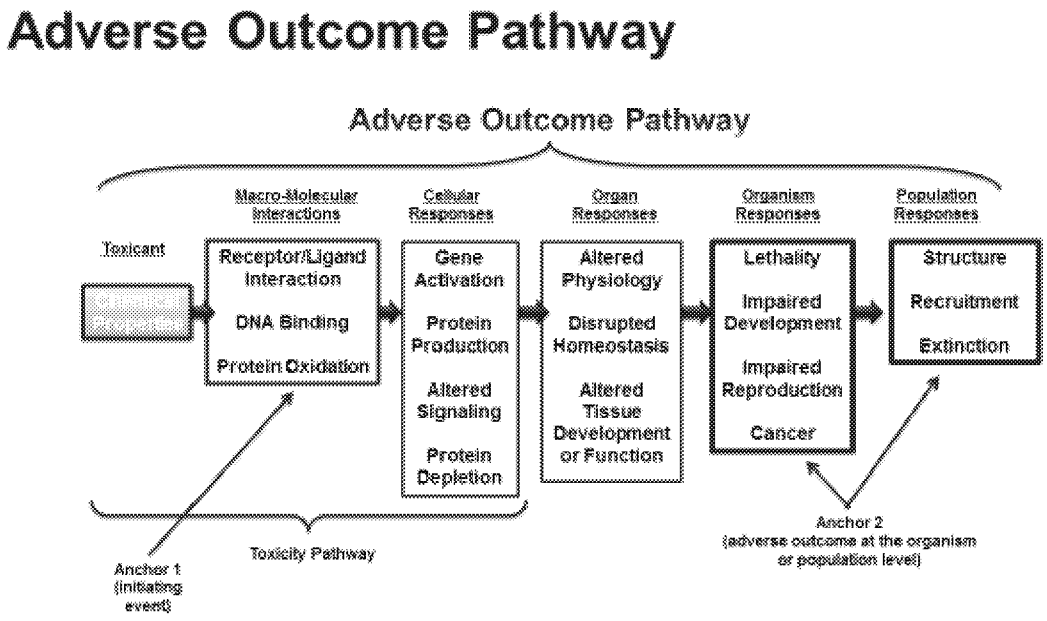


Table 1. Key guidance documents and frameworks used by OPP

NAS	1983: Risk Assessment in the Federal Government: Managing the Process
	1994: Science and Judgment
	2007: Toxicity Testing in the 21st Century
	2009: Science and Decisions: Advancing Risk Assessment
	2011: NAS report on Formaldehyde
	2014: Review of EPA's Integrated Risk Information System (IRIS) Process
WHO/IPCS	2001-2007: Mode of Action/Human Relevance Framework
	2005: Chemical Specific Adjustment Factors (CSAF)
	2014: New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis.

EPA	1991-2005: Risk Assessment Forum Guidances for Risk Assessment (e.g., guidelines for carcinogen, reproductive, developmental, neurotoxicity, ecological, and exposure assessment, guidance for benchmark dose modeling, review of reference dose and reference concentration processes)[4]
	2000: Science Policy Handbook on Risk Characterization
	2006b. Approaches for the Application of Physiologically Based Pharmacokinetic (PBPK) Models and Supporting Data in Risk Assessment
	2014a. Framework for Human Health Risk Assessment to Inform Decision Making.
	2014b. Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation
OPP	2001: Aggregate risk assessment
	2001 and 2002: Cumulative risk assessment
OECD	2013: Organisation for Economic Co-operation and Development Guidance Document On Developing And Assessing Adverse Outcome Pathways

Although there are other sources of human information, the focus of this framework is on interpreting and using ***epidemiology*** and ***human incident data*** in human risk assessment; other sources of human information are not addressed in this document in any depth. Specifically, this document does not extensively discuss research with pesticides involving intentional exposure of human subjects[5] or on studies done to measure dermal or inhalation exposures in agricultural workers as they perform their activities[6,7] .

[4] https://www.epa.gov/osa/products-and-publications-relating-risk-assessment-produced-office-science-advisor

[5] Both the conduct of such research and OPP's reliance on data from such research are governed by EPA's Rule for the Protection of Human Subjects of Research (40 CFR Part 26.) Among other things, these rules forbid research involving intentional exposure of pregnant or nursing women or of children, require prior review of proposals for new research by EPA-OPP and by the Human Studies Review Board (HSRB), and require further review by EPA-OPP and the HSRB of reports of completed research.

[6] In the last several years, OPP has extensively evaluated existing observational studies with agricultural workers in efforts to improve the data and approaches used in worker exposure assessment; those evaluations can be found elsewhere (http://www.epa.gov/scipoly/sap/meetings/2007/010907_mtg.htm)

[7] For additional information on how such worker exposure studies are conducted and used by OPP, see PPP-48 "Pesticides and human Health Risk Assessment: Policies, Processes, and Procedures "available at https://www.extension.purdue.edu/extmedia/PPP/PPP-48.pdf.

III. SYSTEMATIC REVIEW IN PESTICIDE RISK ASSESSMENT: EPIDEMIOLOGY

In recent years, the NRC has encouraged the agency to move towards systematic review processes to enhance the transparency of scientific literature reviews that support chemical-specific risk assessments to inform regulatory decision making (NRC 2011, 2014). The NRC defines systematic review as "a scientific investigation that focuses on a specific question and uses explicit, pre-specified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies" (NRC, 2014). Consistent with NRC's recommendations, the Office of Chemical Safety and Pollution Prevention (OCSPP) employs fit-for-purpose systematic reviews that rely on standard methods for collecting, evaluating and integrating the scientific data supporting our decisions.

According to the NRC, systematic reviews "have several common elements: transparent and explicitly documented methods, consistent and critical evaluation of all relevant literature, application of a standardized approach for grading the strength of evidence, and clear and consistent summative language (NRC, 2014)." In recent years, several groups (Rooney et al., 2014; Woodruff and Sutton, 2014; Hartung, 2010) have published systematic review approaches for use in environmental health sciences. The OCSPP approach to systematic review is consistent with the principles articulated in the Cochrane Handbook for Systematic Reviews of Interventions for evidence-based medicine and with the principles of the Grading of Recommendations Assessment, Development and Evaluation (GRADE). GRADE guidelines used by systematic review approaches for environmental health sciences developed by the National Institute of Environmental Health Sciences (NIEHS) Office of Health Assessment and Translation (OHAT) (Rooney et al., 2014) and University of California, San Diego (Woodruff and Sutton, 2014). According to the *Cochrane Handbook*, the key characteristics of a systematic review are:

- a clearly stated set of objectives with pre-defined eligibility criteria for studies;
- an explicit, reproducible methodology;
- a systematic search that attempts to identify all studies that would meet the eligibility criteria;
- an assessment of the validity of the findings from the identified studies;
- a systematic presentation and synthesis of the characteristics and findings of the included studies.

Each approach mentioned above share common themes and workflow starting with a statement of scientific context (e.g., problem formulation or protocol) followed by literature review with explicit search strategy methods, analysis of study quality (often called risk of bias), evaluation of the quality of the totality of the evidence (e.g., integration) and ultimately leading to a conclusion(s). Each approach recommends transparent and pre-determined criteria for inclusion/exclusion of scientific literature, evaluation of study quality, and reporting of study quality (e.g., high, medium, low). Each approach recommends a pre-stated tool for data integration that provides the foundation for the conclusion(s).

So far, no single nomenclature has been agreed upon by the risk assessment community for systematic review and OCSPP expects terminology to evolve over time as more broad experience is gained. OCSPP considers its systematic review process and workflow as starting with problem formulation followed by data collection, data evaluation, data integration, and summary findings with critical data gaps identified. Scientific analysis is often iterative in nature as new knowledge is obtained.

A. Problem Formulation

In the NRC report *Science and Decisions-Advancing Risk Assessment,* the National Academy of Sciences (NAS) recommended to EPA that risk assessments and associated scientific analyses be developed to be useful to policy makers; in order to attain this goal, the NRC recommended that the agency more broadly use problem formulation in developing its risk assessments. In response to the NRC, the agency published the Human Health Risk Assessment Framework (USEPA, 2014) which highlights the importance of problem formulation. Problem formulation entails an initial dialogue between scientists and risk managers and provides the regulatory context for the scientific analysis and helps define the scope of an analysis. Problem formulation draws from regulatory, decision-making and policy context of the assessment, informs the technical approach to the assessment and systematically identifies the major factors to be considered. As such, the complexity and scope of each systematic review will vary among the different risk assessment contexts. In other words, an OCSPP systematic review is conducted as "fit-for-purpose" (NRC, 2009) based on the pre-determined scope and purpose determined from problem formulation.

The problem formulation involves consideration of the available information along with key gaps in data or scientific information. OPP uses problem formulation as a tool to identify exposure pathways and potential health outcomes along with the appropriate methods, data sources, and approaches for the scientific analysis. If missing data are critical to the assessment, options are discussed as to how best to obtain that information (e.g., required testing, research). The peer review process is identified and the timeline for completing the assessment is defined.

Systematic review provides a transparent tool for organizing available information and identifying gaps in information for the regulatory purpose for the analysis. As such, in problem formulation, the regulatory context of a scientific analysis is described which in turn defines the scope of and purpose for collection and evaluation of scientific literature. Some considerations in problem formulation may be related to population or life-stage, exposure pathways (e.g., route, duration, frequency), and/or health outcomes of interest identified from *in vitro* or *in vivo* laboratory studies along with epidemiology or human incident studies along with resources available and regulatory timeframe. In the context of considering epidemiology and human incident information, an initial evaluation of the study quality, study design, and uncertainties are considered.

Key scientific issues related to hazard assessment considered in problem formulation include: What are the effects associated with exposure? What are the MOA/AOPs associated with these effects? What are the temporal aspects of the effects? Are there susceptible populations and if so, who are they and what factors contribute to susceptibility? Are there differences in PK or pharmacodynamics (PD) between laboratory animals and humans? Exposure information is also evaluated in problem formulation. Key scientific issues related to exposure assessment considered in problem formulation include: How is the pesticide used? What are all of the relevant use sites of exposure? To what chemical substances will people be exposed? What are the routes, durations, and frequencies of exposures? Who may be exposed? Does the exposure pose different risks to different groups (e.g., due age or activity patterns?) In the specific case of epidemiology data, this review considers a variety of factors including, but not limited to, research hypothesis, study design (i.e., sample size, sufficient controls, quality of measurements, etc.), exposure dose/concentration, statistical analysis, and conclusions.

B. Data Collection

The data collection phase of systematic review is the collection of available information from various published and unpublished sources, such as the open scientific literature and submitted studies for pesticide registration. OPP reviews data collected under the Organisation for Economic Cooperation and Development (OECD) test guidelines, OCSPP harmonized test guidelines, and other pesticide (OPP guidelines). These guideline studies are collected primarily from in-house databases of submitted studies and are found through searches of such internal databases.

In the case of epidemiology, most studies are expected to be found in the open scientific literature. Although in some cases supplemental analyses or information may be available, dialogue with the researchers may provide additional, important information not published in the original paper in understanding and interpreting epidemiology studies. The sources of human incident information are summarized in Section IV.

Open literature search strategies use specified criteria to retrieve health effects information from the open scientific literature and unpublished sources. After identifying and selecting the most appropriate sources/databases and determining the most resource effective strategy utilizing classification codes, medical subject headings, and/or keywords, a search is conducted of the literature. Depending on the complexity of the scientific evaluation, support from a reference librarian may or may not be needed. The goal of a human health literature search is to perform a reliable and reproducible literature search by providing proper documentation of the literature search process. The following steps are conducted to retrieve relevant studies:

- The purpose of the scientific analysis and inclusion criteria are established.
- Combinations of terms/key words and/or MeSH (Medical Subject Heading) terms and their Boolean combinations (AND; OR; NOT) are used and documented.

Advanced Search and Field Search by author, title, keywords or subject heading may also be performed as needed. Knowledge of database structure, and using a separate search strategy for a specific database is helpful in retrieving relevant studies. In addition to an initial comprehensive search, periodic searches may be conducted to update the literature list.

- The search strategy is documented, including the date(s) of the search(es)to ensure that all the searches of all the databases are reproducible.
- Reference lists of retrieved articles are examined[2] for additional background and to look for articles that were not discovered in the initial search.
- After combining the retrieved articles from different databases and removing duplicates, the available titles and abstracts are screened. For some of the articles where relevance could not be determined from the title and the abstract, the article is retrieved for further review.
- Following the initial screening, articles that were not relevant (exclusion criteria) – such as opinion articles, studies not in English, and those consisting only of abstracts are excluded. Additional exclusion criteria can be identified on a case by case basis. All exclusion criteria are documented. The rest of the articles, even those that found no adverse health effects, are included for review and evaluation.

C. Data Evaluation

In the data evaluation phase, data quality is reviewed and conclusions are made about the utility of such data. Study quality reflects the overall confidence that reports findings are correct (Balshem et al., 2011). As such, study quality can include:

- reporting quality (how well or completely a study is reported);
- how credible the findings are based on the design and conduct of the study;
- and how well the study addresses the topic under review (Rooney et al., 2014).

Study quality is first considered on an individual study basis, and the quality is judged. For example, one may have stronger confidence in a well conducted case control study than a poorly conducted cohort study. Credibility of the scientific findings, often called risk of bias, is evaluated using pre-determined criteria for specific domains related to study design and conduct (See Table 2).

OPP initially developed a guidance on using the open scientific literature considerations called the "Guidance for Considering and Using Open Literature Toxicity Studies to Support Human Health Risk Assessment" (USEPA, 2012) and generally continues to follow this guidance. However, with the acceleration of systematic review in risk assessment, some aspects of the literature guidance may need updating in the future.

Conclusions about the quality of the data are made and can be described in conclusion statements or categories (e.g., acceptable/not acceptable; low, medium, high).

Specific considerations used in evaluating epidemiology studies on pesticide chemicals are provided in Section III.C below. As part of the data review, a concise written review of the study is developed. This written review describes the study design, results, conclusions, and the strengths and weaknesses of the study. The quality of the epidemiologic exposure assessment is an important factor in determining what role epidemiologic data will play in the risk assessment. As such, it is important to fully characterize the assumptions used in the epidemiologic exposure assessment and the degree to which these assumptions affect the interpretation and generalizability of the epidemiologic findings. The evaluation of the epidemiologic exposure assessment may include a consideration of past and present exposure patterns (e.g., exposed populations, pathways, routes, and levels of exposure) and may include significant changes in use patterns (e.g., risk mitigation actions or new use patterns). With regard to evaluating meta-analyses, reporting guidelines for Meta-analysis Of Observational Studies in Epidemiology (MOOSE) have been developed by Stroup et al., (2000) that are useful in evaluating the quality and interpreting meta-analysis.

D. Data Integration: Weight of Evidence (WOE)

OPP's human health characterizations involve the consideration of all available and relevant data, including but not limited to human studies/epidemiology, biomonitoring data, *in vitro* and *in vivo* experimental laboratory toxicological studies, MOA/AOP information, pharmacokinetic studies, and structure-activity relationships (SAR). Once the different types of hazard data are collected and a full evaluation of each relevant study is conducted and documented, the next step is to integrate multiple lines of evidence.

Data integration is based on the principle of reaching a judgment of the totality of the available negative and positive data for relevant hazards. OPP uses a WOE analysis for evaluating epidemiology and human incident data, such conclusions are made on the preponderance of the information rather than relying on any one study. OPP uses the modified Bradford Hill criteria like those in the MOA/human relevance framework as a tool for organizing and integrating information from different sources (Hill, 1965; U.S. EPA, 1999, 2005; Sonich-Mullin et al., 2001; Meek et al., 2003; Seed et al., 2005; OECD AOP Wiki Users Handbook[8]). It is important to note that the Hill Criteria are not intended as a check box approach but instead are points to consider when evaluating the totality of evidence. In addition, the availability of a fully elucidated MOA/AOP is a not requirement for using epidemiology studies in human health risk assessment. However, even in the absence of a fully developed MOA/AOP, collection and evaluation of mechanistic data may provide support for biological plausibility and help explain differences in tissue sensitivity, species, gender, life-stage, or other factor. The MOA/human relevance framework is a flexible tool which provides a foundation for organizing information without rigidity. It is this

[8] https://aopwiki.org/wiki/index.php/Main_Page#OECD_User_Handbook

flexibility that makes it a useful tool for a variety of purposes such as evaluating causality, integrating information across multiple lines of scientific evidence, and identifying data gaps and areas of future research. In this analysis, epidemiologic findings and human incident data can be evaluated in the context of other human information and experimental studies to evaluate biological plausibility, to identify areas of uncertainty and areas of further research. To describe how Bradford Hill aspects are considered in the WOE evaluations, OPP has used some definitions of terms as outlined in EPA's Preamble to the Integrated Science Assessments (ISAs) which serve as a scientific foundation for the review of EPA's National Ambient Air Quality Standards (NAAQS). (USEPA, 2015).

- **Key events.** In cases where the MOA/AOP are established for a particular health outcome, a clear description of each of the key events (i.e., measurable parameters) that underlie the MOA/AOP are given. Data to inform the key events may come from a combination of *in vitro* or *in vivo* data sources (human or animal). These key events can be a combination of PK and PD events. However, it noteworthy that the availability of a fully elucidated MOA/AOP is a not requirement for using epidemiology studies in human health risk assessment.

- **Biological Gradient/Exposure-Response/Dose-Response Concordance & Relationships**. The Preamble to the ISAs notes that "In the context of epidemiology, a well-characterized exposure-response relationship (e.g., increasing effects associated with greater exposure) strongly suggests cause and effect, especially when such relationships are also observed for duration of exposure (e.g., increasing effects observed following longer exposure times) (USEPA, 2015)." When the MOA/AOP is known, dose-response relationships are identified for each key event. Dose-response relationships are compared among key events. In some cases, the earlier key events may be more sensitive than later key events. In other cases, key events may share similar dose-response curves.

- **Temporal association.** Evidence of a temporal sequence between the introduction of an agent and appearance of the effect constitutes another argument in favor of causality (USEPA, 2015). The Preamble to the ISAs notes that "Strong evidence for causality can be provided through 'natural experiments' when a change in exposure is found to result in a change in occurrence or frequency of health."

This analysis considers key events which occur rapidly (e.g., metabolism to an active metabolite which could occur within minutes of exposure) and those which occur after longer durations (e.g., development of a tumor) to ensure coherence of the effects. Specific to considering epidemiology data, the temporal relationship between the exposure and health outcome may be considered.

- **Strength, consistency, and specificity**.

Consistency: An inference of causality is strengthened when a pattern of elevated risks is observed across several independent studies. The reproducibility of findings constitutes one of the strongest arguments for causality. Statistical significance is not the sole criterion by which the presence or absence of an effect is determined. If there are discordant results among investigations, possible reasons such as differences in exposure, confounding factors, and the power of the study are considered (USEPA, 2015).

Consistency of findings across studies is informed by the repeated observation of effects or associations across multiple independent studies. Further support is provided by reproducibility of findings in different populations under different circumstances. However, discordant results among independent investigations may be explained by differences in study methods, random errors, exposure, confounding factors, or study power, and thus may not be used to rule out a causal connection (USEPA, 2015).

Strength of the observed association: The finding of large, precise risks increases confidence that the association is not likely due to chance, bias, or other factors. However, it is noted that a small magnitude in an effect estimate may or may not represent a substantial effect in a population (USEPA, 2015).

Specificity of the observed association: Evidence linking a specific outcome to an exposure can provide a strong argument for causation. However, it must be recognized that rarely, if ever, do environmental exposures invariably predict the occurrence of an outcome, and that a given outcome may have multiple causes (USEPA, 2015).

- **Biological plausibility and coherence**.

Coherence: An inference of causality from one line of evidence (e.g., epidemiologiccontrolled human exposure, animal, or ecological studies) may be strengthened by other lines of evidence that support a cause-and-effect interpretation of the association. There may be coherence in demonstrating effects from evidence across various fields and/or across multiple study designs or related health endpoints within one scientific line of evidence (USEPA, 2015).

When animal and human data show a similar toxic profile, both quantitatively and qualitatively, there is high confidence in the human health risk assessment. Whereas in other cases, animal and human data may show a qualitatively similar toxic profile but quantitative differences are observed. For example, a particular chemical exhibits the same MOA/AOP in animals and humans but there may be species differences in dose-response characteristics. These dose-response differences could be due to tissue dosimetry (i.e., PK) or from different response characteristics (i.e., PD). In contrast, animal and human data can, in some instances, show qualitatively dissimilar outcomes. This situation highlights the need to fully and objectively evaluate all available information in a

transparent and comprehensive manner to consider factors such as species, gender, and life-stage differences and potential susceptibilities along with study design considers and exposure potential.

Biological plausibility: An inference of causality is strengthened by results from experimental studies or other sources demonstrating biologically plausible mechanisms. A proposed mechanism, which is based on experimental evidence and which links exposure to an agent to a given effect, is an important source of support for causality (USEPA, 2015).

Similarly, information on MOA/AOP for a chemical, as one of many structural analogs, can inform decisions regarding likely causality. Structure activity relationships and information on the agent's structural analogs can provide insight into whether an association is causal (USEPA, 2015).

EPA's Cancer Guidelines (2005) indicate:

> "*evaluation of the biological plausibility of the associations observed in epidemiologic studies reflects consideration of both exposure-related factors and toxicological evidence relevant to identification of potential modes of action (MOAs). Similarly, consideration of the coherence of health effects associations reported in the epidemiologic literature reflects broad consideration of information pertaining to the nature of the biological markers evaluated in toxicologic and epidemiologic studies*. [p. 39]."

However, The Cancer Guidelines further state that "*lack of mechanistic data, however, is not a reason to reject causality* [p. 41]." As such, lack of established MOA/AOP is not necessary knowledge when using epidemiology data and epidemiology associations may still be valid even in the absence of an established MOA/AOP and may also provide insight into potential MOA/AOP.

- **Uncertainties.** Uncertainties are discussed in the WOE transparently and objectively.

E. Overall conclusions, recommendations for risk assessment, statement of areas of confidence and uncertainty

It is important to document a summary of the evidence, the procedures or methods used to weigh the evidence, the basis for the WOE conclusion or recommendation, any uncertainties and areas for further research. Recommendations are made on the role of the epidemiologic or human incident data in the risk assessment. Generally, OPP does not use human incident information for quantitative risk assessment but instead to inform risk assessment/risk management activities such as indicating a potential need for a new risk assessment or new risk management measures, evaluating the success of risk mitigation actions after they are implemented, and targeting possible enforcement activities. In

contrast to more limited role of human incident data, epidemiology studies have the potential to help inform multiple components of the risk assessment in a variety of ways. High quality studies with robust exposure assessment may be used to estimate a risk metric quantitatively. Alternatively, outcomes reported in epidemiologic studies may be compared qualitatively with those seen in *in vitro* and animal studies to evaluate the human relevance of animal findings (Hertz-Picciotto, 1995) and may be useful in assessing the biological plausibility of epidemiologic outcomes. In the final portion of the proposed WOE analysis, the overall conclusions along with statement of areas of confidence and uncertainty. This section also identifies areas of additional research. This section recommends the source of data for regulatory values and the appropriate approach for extrapolating between species (if necessary) and among humans.

IV. REVIEWING EPIDEMIOLOGY STUDIES FOR USE IN PESTICIDE RISK ASSESSMENT

A. Introduction

Epidemiology is a science that seeks to identify and evaluate relationships between exposure to chemical, physical or biological agents, and the health status of populations (Boyes et al., 2007). It has been defined as the "study of how disease is distributed in populations and the factors that influence or determine this distribution" (Gordis, 2009). More broadly, it is considered as "the study of the occurrence and distribution of health-related events, states, and processes in specified populations, including the study of the determinants influencing such processes and the application of this knowledge to control of relevant health problems" (Porta, 2014). The objective of much epidemiologic research is to obtain a valid and precise estimate of the effect of a potential cause on the occurrence of disease. A key objective of epidemiology, like other sciences, is determining cause and effect or - said differently - of identifying the etiology of a disease or health outcome and the risk factors with which it might be associated. Calderon (2000) described four major uses of such studies: 1) describe the health status of a population and discover important time trends in disease and exposure frequency; 2) explain the occurrence of diseases by identifying factors that are associated with specific diseases or trends; 3) predict the number of disease occurrences and the distribution of health states in specific populations; and 4) improving the health status of the population by identifying factors that affect environmental or human health. In the case of pesticides, epidemiology focuses on the relation between exposure and adverse health effects in the general population and in specific sub-populations, such as occupationally exposed workers or applicators.

Epidemiology studies have the potential to help inform multiple components of the risk assessment in a variety of ways. High quality studies with robust exposure assessment may be used to quantitatively estimate risk or an appropriate risk surrogate such as an odds ratio or risk ratio. However, many epidemiology studies that deal with pesticides and pesticide exposure suffer some limitations in size, scope, exposure assessment, or data analysis which prevent or otherwise impede their full use in quantitative risk assessment

(Ntzani et al., 2013). Pesticide use in the US has changed significantly over the last few decades. As the use changes, so does the exposure to workers. Changes in pesticide use have occurred due to risk mitigation actions by EPA, resistance management activities, introduction of new chemistries, and increased use of genetically modified crops. These significant changes in exposure have to be taken into account when interpreting epidemiology studies and, ultimately, the decision to use such studies in quantitative risk assessment. Even so, epidemiology studies may be used to compare with evidence from experimental animal studies to characterize assumptions used in deriving such values. In other cases, outcomes reported in epidemiologic studies may be compared qualitatively with those seen in *in vitro* and laboratory animal studies to evaluate biological plausibility or human relevance of animal findings (Hertz-Picciotto, 1995). Human information like that found in epidemiology studies are expected to potentially play a significant role in the new vision of toxicity testing recommended by the NRC (2007). Specifically, epidemiology studies can provide insight on health outcomes that may arise from real-world chemical exposures in humans and thus can contribute to problem formulation and hazard/risk characterization. Human information may guide additional studies (e.g., dose and endpoint selection for use in *in vitro* and targeted *in vivo* experimental studies); and identify novel health effects or host susceptibilities which can be investigated with future research.

When laboratory data from animal studies provide the primary source of information for hazard characterization, one potential source of uncertainty is the relevance of animal models to humans. In the absence of data to support the contrary, animal findings are assumed to be relevant to humans. Furthermore, EPA assumes that humans are more sensitive than laboratory animals in the absence of data to support the contrary. In actuality, humans may be more or less sensitive to pesticides than other animal species. Epidemiology and human incident data can provide scientific information and support to inform uncertainties associated with species extrapolation. With respect to population variability, epidemiology studies better characterize potential variability than do animal studies. Specifically, epidemiologic data include the genetic diversity, and variability inherent in human populations and thus can better account for and represent actual population response to environmental chemicals than laboratory animals (Calderon, 2000).

With respect to dose-response characterization, animal toxicology studies have the benefit that studies can be designed to cover a broad range of exposure levels. However, animal toxicology studies generally use exposures which are much larger (sometimes orders of magnitude) than those that occur in the environment. These high exposure levels in animal studies dictate the need for extrapolation from high to low doses. This extrapolation introduces added uncertainty into the risk assessment. Epidemiology studies and human incident data involve actual real-world exposures and thus high dose extrapolation may in many cases not be needed. Epidemiology studies conducted over a range of exposures (from low to high) are most useful.

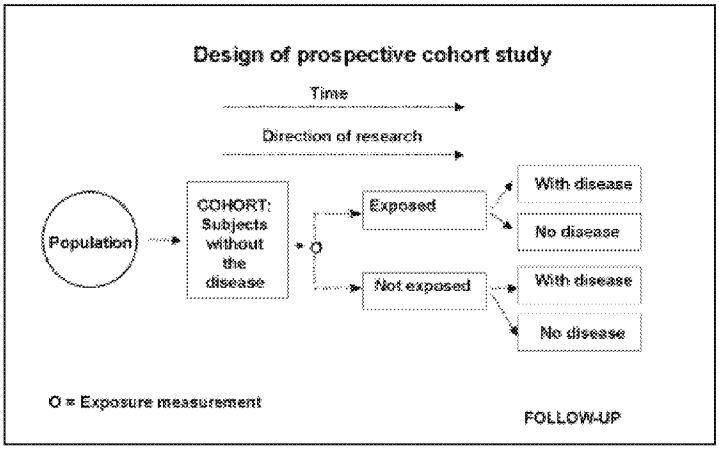
Animal studies do not replicate the length, magnitude, duration, routes of exposure and variability in exposure experienced by humans (Calderon, 2000). Human exposure often occurs through multimedia exposure pathways, including food, water, air, and indoor and outdoor environments. In contrast, controlled laboratory studies typically use a single

route of exposure. In addition, humans may experience exposure to multiple chemicals and/or non-chemical stressors simultaneously, whereas most animal studies involve a single chemical stressor. On one hand, this multi-chemical exposure in epidemiology studies can provide a challenge when attempting to attribute epidemiologic outcomes to a single pesticide chemical. On the other hand, epidemiologic research considers real-world exposures and may help, when considered along with experimental approaches, address questions associated with multiple chemical exposures which can be difficult to evaluate in an experimental setting.

B. Types of Epidemiology Studies

The major types of observational epidemiologic studies are described briefly below with consideration of their strengths and weaknesses (Lilienfeld and Lilienfeld, 1979; Mausner and Kramer, 1985; Kelsey et al., 1996; Rothman and Greenland, 2012; Paddle and Harrington, 2000; USEPA, 2005; Purdue Pesticide Programs, PPP-43).

Cohort studies begin with a group of people that share common characteristics—the cohort—and evaluate their health over an extended follow-up time period during which the occurrence of disease is recorded (see figure box from van den Brandt et al. (2002)). The common characteristic is often the presence vs. absence of "risk factors" (such as exposures)[9]. In such studies, differences in disease occurrence between the "exposed" and "non-exposed" individuals are identified and studied over time to determine differences in the rate of disease[10]. This difference in the rate of disease occurrence is then investigated to determine if the rate of disease differs between the exposed and non-exposed groups. Cohort studies have the ability to simultaneously evaluate multiple disease outcomes under study (which is not true for case-control studies, which are generally limited to evaluating only a single (pre-specified) disease outcome, discussed below). Cohort studies can also be performed either prospectively, like the Agricultural Health Study (AHS, http://aghealth.nci.nih.gov/), or retrospectively from historical records. A prospective cohort design focuses on a group of people from a current point in time through a future point in time. A retrospective cohort design focuses on a group exposed at some point in the past, and compares disease rates after exposure occurred (generally through existing

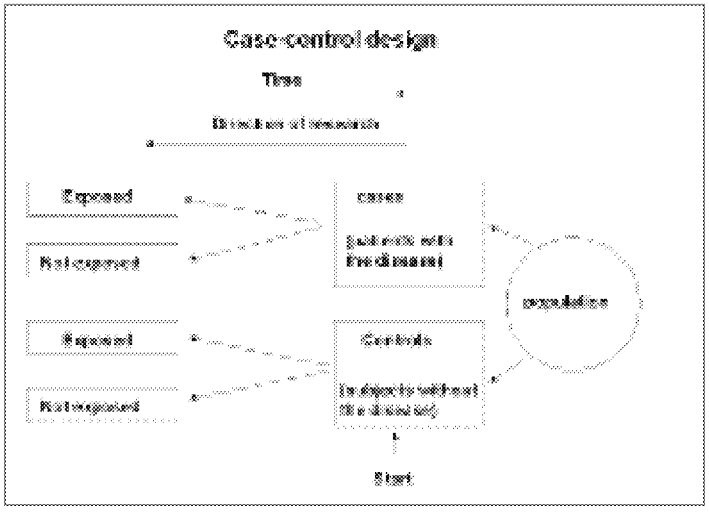


[9] While exposure is often dichotomized on an exposed vs. non-exposed basis in cohort studies, exposure can also be measured on a quantitative scale (e.g., by a continuous measure or by quantiles)

[10] Cohort studies commonly study differences in rates of disease, but these can also include other focal outcomes of interest such as birth weight, mental abilities, blood pressure, etc.

available exposure databases (or records) available on a person-by-person (individual) basis). Prospective cohort studies can be relatively lengthy and expensive to conduct, particularly for rare diseases, and require a large number of subjects to be under study. Importantly, significant resources and professional staff are required for a long period of time to collect high quality data.

Case-control studies are studies in which groups of individuals with (cases) and generally without (controls) a given disease are identified and compared with respect to (generally past[11]) exposure to determine whether those with the disease of interest are more likely or no more likely to have been exposed to the agent(s) or factor(s) of interest. That is, the analysis of case-control studies contrasts the frequency of exposure of the agent or factor in the cases with those in the controls to determine if these differ and, thus, whether there is a differential association. In case-control studies, determination of the disease status (i.e., cases with the disease; controls without) generally precedes determination of the exposure status (see figure box from van den Brandt et al. (2002)) Because disease has already occurred at the time of selection into the case-control study, this study design is particularly useful in studying uncommon diseases or diseases with long latency and can be utilized to evaluate the relation between many different exposures and a specific (pre-specified) disease outcome of interest . And because case-control studies begin with individuals who have the disease, the studies can involve fewer subjects than cohort studies and can be completed in a comparatively shorter time frame. Challenges in case-control investigations include the selection of an appropriate control group and the assessment of exposures which may have occurred long before the disease was diagnosed (Rothman, 2012; Wacholder et al. 1992a; Wacholder et al. 1992b; Wacholder et al. 1992c; Shultz and Grimes, 2002; Grimes and Schultz, 2005). Case-control studies can be particularly susceptible to "recall bias" in which diseased individuals may remember exposures or events differently (generally better) than those who serve as the controls and are healthy.



Nested case-control studies are an example of a hybrid design and contain the elements of a cohort and a case-control study. These designs can be useful when the analytical costs for determining pesticide exposure are too high for the entire cohort to be studies. For example, a cases that that have developed the disease or health outcome in an

[11] It is possible for case-control studies to be done prospectively in which the cases have not yet developed the disease until after the study begins under which circumstance the cases are enrolled in the study over time.

ongoing cohort study can be matched with appropriate controls from the study that have not yet developed the disease or outcome of interest at the time of the analysis. One recognized advantage of the nested case-control study (as opposed to a more standard case-control study) is that the issues of selection bias and recall bias are minimized.

Cross-sectional studies focus on the prevalence of disease (e.g., birth defects, small-for-gestational age or SGA), symptoms, biological/physical and physiologic response measurements (e.g., pulmonary function tests, blood pressure, chest X-ray, clinical examinations, liver and kidney biomarkers). A key feature of such studies is that they are observational studies which focuses on the *prevalence* as a frequency measure, with the presence or absence of disease determined at the time of sampling or over a sampling period. Prevalence is the proportion of individuals in a population that has the disease and can either be determined as a "point prevalence" or as a "period prevalence".[12] A prevalence is a proportion not a rate and thus the cross sectional studies do not involve a follow up period. Typically, the exposure status (e.g., exposed or unexposed), disease status/outcome, and demographic characteristics are determined at a point in (or over) time. The major comparison in this study design is a comparison of the prevalence of the outcome in the exposed population vs. the prevalence of that outcome in the non-exposed population, with the risk measure being the prevalence risk ratio or odds ratio. Cross-sectional studies are generally used to identify patterns or trends in disease occurrence over time or in different geographical locations, and can be conducted quickly and relatively inexpensively. However, they measure the prevalence of a disease outcome which is affected by both incidence – the rate of occurrence of new cases – and duration of the disease, and it can be difficult in any analysis to sufficiently separate these factors. Thus, they involve "survivor populations" and do not measure, evaluate, or consider those that have left the population of interest because they became ill. Another important limitation of cross-sectional studies is they do not allow one to determine whether exposure precedes the disease. As such, cross-sectional studies are unable to establish temporal relationships between disease and exposure and typically require additional studies to confirm a hypothesized causal association suggested by a cross-sectional study.

Ecologic studies examine exposure and disease patterns using information reflecting group or population-level data. In an ecologic study, the unit of analysis is a group and not an individual[13]. Here, groups of subjects are sampled, with the exposure, disease, and potential confounding factors measured at this group (or cluster) level. Groups are generally defined on a geographic, administrative, or organizations unit basis (e.g., districts, towns, counties, schools, workplaces, etc.) with all exposure, disease, or confounder measurements made or summarized at the group level rather than at the level of the individual. An ecological (group-based) study contrasts with an individual-level study in that in the former there is no information on whether the cases are the actual individuals

[12] The former involve measurements at a particular place and/or a particular time while the latter involves determinations of the proportion of cases over a given time period.

[13] Some studies can be "partially ecologic" in design in which either the exposure or the disease outcome is measured on a group level but the other variable is measures at an individual level with the researcher making inferences to the individual level.

with the exposure whereas in the latter exposure information is tied to the individual. As an example, a study of disease rates by contaminant levels in water can be ecologic with respect to evaluation of the exposure, but the health outcome or disease status may have determined on an individual basis. In these instances, the term "semi-ecological" can sometimes be used when exposure is determined at the group level but outcome is determined at the level of the individual.

Using this design, it is not possible to know whether all members of the exposed group are individually exposed (or the individual exposure levels) nor is it possible to infer individual-level effects from the group level effects that result. If the intent of the study is to direct inferences to the *group* (rather than the individual), then this is not a concern and these studies can be appropriate, particularly if measurements are constrained or difficult to perform at the individual level and exposures within the group are generally homogenous. If the intent of the study is instead to direct inferences to the individual, then this study design suffers from what is termed the ecological fallacy: the assumption that an observed relationship in an aggregated or grouped data set will reflect what would have been observed had the sampling occurred at the individual level. In addition to this ecological fallacy issue, an additional bias arises a result of the inability to appropriately control for confounding variables at the level of the individual as opposed to the group when information on confounding factors is only available at the group level.

In most cases, ecologic studies are considered as hypothesis-generating studies and best used for suggesting research hypotheses for future studies and may contribute to problem formulation. Nevertheless, it is important to assess ecological studies on the basis of the quality of their design, and useful information can be gleaned from an ecologic study if it is well-designed (FIFRA SAP, 2010). Ecologic studies alone generally do not have the ability to establish a causal association. When taken with other these studies can be useful under certain circumstances and should be noted in the hazard characterization. In particular, stable populations, clear exposure contrasts, and large differences in risk can be important factors that might increase the utility of these studies.

C. Evaluating epidemiology studies for use in pesticide risk assessment

OPP searches the peer reviewed literature for observational epidemiology studies of potential adverse acute and chronic health effects linked to chemical use. Details regarding literature search protocols and strategies are provided elsewhere. Epidemiologic research utilizing cohort, case-control, or cross-sectional study designs may provide information to OPP to strengthen OPP's understanding of the potential hazards, exposure-response characterization, exposure scenarios. or assessment methods, and – ultimately -- risk characterization (van den Brandt, 2002). In addition, compelling case reports or case series analysis may illumine a health effect or mechanism of action previously unidentified.

Generally speaking, the quality of epidemiologic research, sufficiency of documentation of the study (study design and results), and relevance to risk assessment is considered when evaluating epidemiology studies from the open literature for use in OPP's

risk assessments. It is important that these criteria are endpoint-specific as various methodological details become more or less important given the endpoint of concern. For example, it is important to understand relevant factors that influence outcome ascertainment (*e.g.*, is there a test or a biomarker available to indicate presence of an effect, or are symptoms gradual and non-specific initially leading to physician diagnosis upon advanced disease state). In addition, for environmental and occupational epidemiology studies, the quality of the exposure assessment is vitally important. Prior consideration must be given to aspects of exposure and confounder measurement to the question under consideration.

When considering individual study quality, various aspects of the design, conduct, analysis and interpretation of the epidemiology studies are important. These include:

1. Clear articulation of the hypothesis, even if the study is hypothesis-generating in nature;
2. Adequate assessment of exposure for the relevant critical windows of the health effects, the range of exposure of interest for the risk assessment target population, and the availability of a dose/exposure-response trend from the study, among other qualities of exposure assessment,
3. Reasonably valid and reliable outcome ascertainment (the correct identification of those with and without the health effect in the study population),
4. Appropriate inclusion and exclusion criteria that result in a sample population representative of the target population, and absent systematic bias,
5. Adequate measurement and analysis of potentially confounding variables, including measurement or discussion of the role of multiple pesticide exposure, or mixtures exposure in the risk estimates observed,
6. Overall characterization of potential systematic biases in the study including errors in the selection of participation and in the collection of information; this can include performing sensitivity analysis to determine the potential influence of systematic error on the risk estimates presented (*e.g.,* Greenland's formula)
7. Evaluation of the statistical power of the study to observe health effects with appropriate discussion and/or presentation of power estimates,
8. Use of appropriate statistical modeling techniques, given the study design and the nature of the outcomes under study

Other Federal and non-Federal entities have offered such guides (*e.g.*, OHAT, Navigation Guide, National Toxicology Program [NTP] Report on Carcinogens [ROC[14]], IRIS, Cochrane ACROBAT-Non-Randomized Studies of Interventions) (Sterne et al., 2015 as well as the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement for observational epidemiological studies (see www.strobe-statement.org and Vandenbroucke et al., 2007; Von Elm, 2014) As OPP gains experience with integrating epidemiology studies into human health risk assessment, relevant adjustments to its evaluation approach will be made.

[14] http://ntp.niehs.nih.gov/pubhealth/roc/index.html

Independent study evaluation is performed and documented prior to the development of evidence- tables of detailed summary tables which are informative to hazard identification and exposure response assessment. Table 2 provides a structure to the major considerations evaluated and the associated weight (low, medium, high) for each consideration. Table 2 provides a generic set of considerations and should not be considered a checklist. The specific scientific considerations appropriate for particular science analysis are adjusted on a case by case basis.

The culmination of the study evaluation process would be to provide professional/expert opinion as to the nature of the potential bias that may result from systematic errors in each specific study identified through study specific evaluations, and an assessment of overall confidence in the epidemiological database. In this way, data integration (animal, human, mechanistic, other) would be informed by level of confidence in the human epidemiological studies that inform human health effects of environmental and occupational exposures.

Table 2. Study Quality Considerations [a] (Adapted from Munoz-Quezada et al., 2013; LaKind et al., 2014)

Parameter	High	Moderate	Low
Exposure assessment	Accurate and precise quantitative relationship with external exposure, internal dose, or target dose, possibly associated with an MOA/AOP. If questionnaire utilized, questionnaire and/or interview answered by subjects for chemical-specific exposure	Evidence exists for a relationship between biomarker in a specified matrix and external exposure, internal dose, or target dose. Questionnaire and/or interview for chemical-specific exposure answered by subjects or proxy individuals	Poor surrogate Low-quality questionnaire and/or interview; information collected for groups of chemicals rather than chemical-specific; no chemical-specific exposure information collected; ever/never use of pesticides in general evaluated
Outcome Assessment	Standardized tool, validated in study population; medical record review/diagnosis confirmation by trained staff; appropriate consideration of prevalence/incidence of cases	Standardized tool, not validated in population, or screening tool; or, medical record review, methods unstated	Selected sections of test, or maternal report, other; or, maternal/paternal self-report; unclear/no consideration for whether prevalent or incident cases are appropriate
Confounder control	Good control for important confounders relevant to scientific question, and standard confounders	Moderately good control confounders, standard variables, not all variables relevant for scientific question	Multi-variable analysis not performed no adjustments; no stratification, restriction, or matching
Statistical Analysis	Appropriate to study question and design, supported by adequate sample size, maximizing use of data, reported well (not selective)	Acceptable methods, questionable study power (especially sub-analyses), analytic choices that lose information, not reported clearly	Minimal attention to statistical analyses, comparisons not performed or described clearly
Risk of (other) bias (selection, differential misclassification, effect size magnification, other)	Major sources of other potential biases not likely present, present but analyzed, unlikely to influence magnitude and direction of the risk estimate	Other sources of bias present, acknowledged but not addressed in study, may influence magnitude but not direction of estimate	Major study biases present, unacknowledged or unaddressed in study, cannot exclude other explanations for study finding

[a] Overall study quality ranking based on comprehensive assessment across the parameters.

1. Exposure Assessment

Exposure assessment can be defined as the "process of estimating or measuring the magnitude, frequency and duration of exposure to an agent, along with the number and characteristics of the population exposed. Ideally, it describes the sources, pathways, routes, and the uncertainties in the assessment. (Zartarian et al., 2005)." In environmental epidemiology, exposure assessment poses a unique challenge, particularly for toxicants that are found in low concentrations in environmental media (NRC, 1991; NRC, 1997). Given the complexity of exposure pathways, researchers have developed a number of different approaches to assess exposure, which vary in accuracy, precision, and resource requirements (Niewenhuijsen, 2003). Some of these approaches are not specific to epidemiologic research but may be used to inform exposure assessment in a variety of scientific analyses. These approaches include indirect methods, based on historical records, questionnaires, and environmental monitoring, and direct methods, based on personal monitoring and biomonitoring. A brief description of each method and its strengths and limitations is summarized below.

Table 3. Summary of indirect and direct exposure assessment methods.

Approach	Method/Tools	Example	Exposure Estimation
Indirect	Historical Records	Estimating proximity to agricultural crops using address information	Dichotomous or ordinal exposure
	Questionnaires	Determine potential for exposure based on pesticide-use responses	Dichotomous or ordinal exposure
	Environmental Monitoring	Measuring pesticide levels in community water drinking system	Dichotomous or ordinal exposure, although exposure can be estimated using modeling
Direct	Personal Monitoring	Measuring pesticide inhalation and dermal contact	Quantified exposure
	Biomonitoring	Measuring pesticide levels in blood and urine	Quantified internal dose

Historical records and questionnaires are used to characterize key characteristics which may be associated with chemical exposure. When used in epidemiologic studies, historical records and questionnaires are not typically used to predict quantitative levels of exposure. Rather, historical record information or questionnaire responses are used to assign categorical levels of exposure. Examples of historical record information that can be used to assign exposure levels includes address in proximity to an agricultural crop and employment history information on job title and history. Similarly, questionnaires can be used to determine if individuals recall using pesticides or identify individuals that perform specific job functions that increase their potential for exposure. While historical records and questionnaires can be cost-effective sources of data on potential exposure, they do have limitations. Data collected from historical records and questionnaires is only a surrogate of exposure. As a result, these

data sources may be an oversimplification of exposure and not accurately rank individual's exposure potential.

Environmental monitoring is used to characterize the levels of contaminants in environmental media, including air, water, soil, food, and home and work environments. Many state and Federal programs collect environmental monitoring data that may be useful in epidemiologic studies. Environmental monitoring is particularly useful for exposure that can be defined by geographic boundaries, such as air pollution and drinking water. As such, many epidemiologic studies have utilized ambient air monitoring data and community drinking water system data to characterize exposure to air pollution and drinking water contamination, respectively. While environmental monitoring data is useful for estimating exposures defined by geographic boundaries, it can be less reliable for the purposes of assigning individual-levels exposures, particularly when individuals live, work, and spend time in many different locations.

Personal monitoring is used to characterize exposure at the point of contact of a body boundary. Examples of personal monitoring include the use of dosimeters to assess dermal contact with pesticides, personal air sampling devices to assess inhalation exposure, and collection of duplicate diet samples to determine pesticide levels in food. The advantage of personal monitoring is that it is likely to provide more accurate estimates of individual-level exposure than indirect methods. Personal monitoring also makes it possible to quantify exposure levels that can be useful for prioritizing the relevance of different routes of exposure. Additionally, personal monitoring can also be used to assess longitudinal exposure when repeated measurements are taken over time. While personal monitoring offers many advantages over indirect approaches, it also tends to be labor and resource intensive (Niewenhuijsen, 2003). As a result, it is not typically feasible to conduct large-scale epidemiologic studies that assess exposure using personal monitoring. Furthermore, personal monitoring is highly dependent on the measurement techniques and analytic tools used to obtain samples and it is less likely that information that characterizes exposures during the relevant time period (usually in the past) will be available. In addition, it is unlikely that the full range of exposures over the time period of interest will be captured, and sampling may not be over a sufficient time period to capture peaks and fluctuations As such, it is extremely important to consider the scientific rigor and reliability of personal monitoring methodologies that are used in epidemiologic studies, and such monitoring may need to be supplemented by other monitoring (e.g., environmental, biological, and/or interview/questionnaire data).

Biomonitoring is used to characterize exposure by measuring a chemical, its metabolite(s), or reactive product(s) in biological samples, such as blood, urine, saliva, milk, adipose, and other body tissues (Needham et al., 2007). Zartarian et al. (2005) state that "a biomarker/biological marker has been defined as an "indicator of changes or events in biological systems. Biological markers of exposure refer to cellular, biochemical, analytical, or molecular measures that are obtained from biological media such as tissues, cells, or fluids and are indicative of exposure to an agent". Thus, biomarkers can be used to assess exposure or as indicators of health effects (LaKind et al., 2014). Table 4 provides scientific considerations for evaluating the quality and relevance of biomonitoring data

collected from epidemiology studies. Assessing exposure using biomonitoring has expanded rapidly as analytical tools have become more cost-effective and more biomarkers are identified. Compared with self-reported questionnaire or interview data, biomonitoring may reduce exposure misclassification and enhance the precision of the risk estimates. Similarly, biomonitoring integrates exposures from different routes and can be used to determine the amount of exposure that is absorbed into the body (Checkoway et al., 2004). Furthermore, knowledge as to the role of the biomarker in the natural history of disease is known in certain instances, such that biomarkers may help resolve temporality of exposure issues.

While biomonitoring has many advantages over others exposure assessment methods, it also has its own limitations. In many studies, biological sample are only taken from a single point in time and may not reflect accurately reflect longitudinal patterns, particularly if exposures are highly variable. Furthermore, evaluation of biomarkers also requires an understanding of degradation and metabolism of chemicals in both the environment and human body. As such, biomarkers of exposure may differ between individuals for reasons other than exposure level. Differences in metabolism, co-morbidities such as kidney disease in relation to urinary measurements, uncertainty as to whether the biomarker measures exposure to the active ingredient or the environmental degradates may all account for apparent differences in biomarkers of exposure among individuals, and possibly between comparison groups.

Table 4. Considerations of biomonitoring data from environmental epidemiology research (Adapted from LaKind et al. (2014).

Biomarker Consideration	Tier 1	Tier 2	Tier 3
Exposure biomarker	Biomarker has accurate and precise quantitative relationship with external exposure, internal dose, or target dose.	Biomarker has an unknown quantitative relationship with external exposure, internal dose, or target dose or is poor surrogate (low accuracy and precision) for exposure/dose.	NA
Effect biomarker	Bioindicator of a key event in a MOA/AOP.	Biomarkers of effect for which the relationship to health outcome is understood	Biomarker has undetermined consequences (e.g., biomarker is not specific to a health outcome).
Specificity	Biomarker is derived from exposure to one parent chemical.	Biomarker is derived from multiple parent chemicals with similar toxicities.	Biomarker is derived from multiple parent chemicals with varying types of adverse endpoints.
Method sensitivity	Limits of detection are low enough to detect chemicals in a sufficient percentage of the samples to address the research question.	Frequency of detection too low to address the research hypothesis.	NA
Biomarker stability	Samples with a known history and documented stability data.	Samples have known losses during storage but the difference between low and high exposures can be qualitatively assessed.	Samples with either unknown history and/or no stability data for analytes of interest.
Sample contamination	Samples are contamination-free from time of collection to time of measurement (e.g., by use of	Study not using/documenting these procedures.	There are known contamination issues and no documentation that the issues were addressed

Biomarker Consideration	Tier 1	Tier 2	Tier 3
	certified analyte-free collection supplies and reference materials, and appropriate use of blanks both in the field and lab). Research includes documentation of the steps taken to provide the necessary assurance that the study data are reliable.		
Method requirements	Instrumentation that provides unambiguous identification and quantitation of the biomarker at the required sensitivity (e.g., GC-HRMS, GC-MS/MS, LC-MS/MS)	Instrumentation that allows for identification of the biomarker with a high degree of confidence and the required sensitivity (e.g., GC-MS, GC-ECD).	Instrumentation that only allows for possible quantification of the biomarker but the method has known interferants (e.g., GC-FID, spectroscopy)
Matrix adjustment	Study includes results for adjusted and non-adjusted concentrations	Study only provides results using one method (matrix-adjusted or not).	NA

FP = false positive; FN = false negative; GC-HRMS = gas chromatography/high-resolution mass spectrometry; GC-MS = gas chromatography/mass spectrometry; GC-ECD = gas chromatography-electron capture detector; GC-FID = gas chromatography-flame ionization detector], ICC = intra-class correlation coefficient ; NA = not applicable; PFP = probability of false positive

Indirect exposure assessment methods are common in retrospective studies and based on factors that are surrogates of chemical exposure. As described above, indirect exposure data cannot generally be used to estimate quantitative exposure levels without additional modeling. For example, a questionnaire can be used to determine if an individual has ever used a pesticide, but can less reliably collect data on all the environmental and behavioral factors that are needed to calculate that individual's exposure. As such, indirect exposure data are often used to classify exposure using a dichotomous exposure variable (i.e. exposed/unexposed) or ordinal exposure scale. In contrast, direct exposure assessment methods are based on data on actual individual-level exposure through personal monitoring and biomonitoring. Thus, direct methods can be used to estimate individual exposure or internal dose levels. Direct methods are more common in prospective studies, but are also used in retrospective studies when existing biological samples are available from well-defined population groups.

Quantified personal measurements, such as personal monitoring and biomonitoring, are generally considered the best source of data for estimating actual exposure levels (NRC, 1991; NRC, 1997). While this is the case, accurate qualitative measures of exposure (e.g. dichotomous and ordinal exposure metrics) from indirect methods can be just as accurate for the purpose of epidemiology. Moreover, indirect methods are often easier to interpret and may require less additional research and development to demonstrate their utility in exposure assessment.

Regardless of the approach, exposure assessment methods should be able to provide exposure estimates that are reliable and valid. In the context of epidemiology, *reliability* general refers to the ability to reproduce results and *validity* generally refers to the extent that exposure estimates reflect true exposure levels (Checkoway et al., 2004). When evaluating a particular exposure assessment's reliability and validity, it is important to consider the exposure assessment's strengths and weaknesses in the context of the study's research objectives. Less refined exposure assessment may be suitable for exploratory studies. This is because exploratory studies help raise awareness about potential hazards that can encourage investment in more focused research. Conversely, studies with more focused hypotheses can be greatly strengthened through the use of more refined exposure assessment methods. Therefore, indirect and direct exposure assessment methods represent a spectrum of tools that are complimentary and can be used at different stages of research when exploring exposure-disease relationships.

2. Confounding Factors

Confounding occurs when the relationship between the exposure and disease is to some extent attributable to the effect of a second (confounding) risk factor. This can happen when this second (i.e., confounding) risk factor is an independent, causally-associated risk factor for the disease but is also associated -- causally or non-causally -- with the exposure under analysis and does not also serve as an intermediate variable in the causal pathway between the exposure and the outcome of interest. If not properly measured and accounted

for, confounders have the ability to change the magnitude (and potentially the direction) of the estimated association between an exposure and health outcome. This can result in an over- or under-estimation of the relationship between exposure and disease because the effects of the two risk factors have not been appropriately separated, or "disentangled". As an example: a given pesticide may be associated with lung cancer in a given study, but this may be due to a confounding effect of farm tractor diesel fumes: here, this second factor – farm tractor diesel fumes – would be a confounder if it was causally associated with the disease outcome (here, lung cancer) but also associated with pesticide exposure. Confounding factors may include less intuitive lifestyle exposures such as cigarette smoking, dietary factors (e.g., high energy/calorie laden diet), and physical activity (e.g., lack of physical activity) genetics, comorbidity, medication use, alcohol consumption, etc., all of which may adversely affect health and may be statistically associated with pesticide use. In epidemiological analyses, confounding factors are measured in the study sample and typically "adjusted for" in the final risk estimate in either the design phase of the study or the analysis phase. With respect to the former, the epidemiological researcher can "restrict" the study population to individuals that share a characteristic which the researcher wishes to control; this has the result of removing the potential effect of confounding caused by that (now controlled) characteristic. A second available method – also applicable to the design phase of the study -- is for the researcher to control confounding by "matching" individuals based on the confounding variable. This ensures that the confounding variable is evenly distributed between the two comparison groups and effectively controls for this. It is important to note that the relationship between the confounder and the exposure or outcome does not need to be found to be statistically significant in order for it to have an impact on the risk estimate for the main effect[15].

At the analysis stage, one method by which confounding can be controlled is by stratification. Under this means of control, the association is measured separately under each of the (potentially) confounding variables; the separate estimates are "brought together" statistically -- if determined to be appropriate -- to produce a common odds ratio or other effect size measure by using Mantel-Haenszel approaches which weight the estimates measured in each stratum. Stratification can be difficult if there are multiple potential confounders that need to be controlled simultaneously. In such cases, confounding is typically dealt with by means of statistical modelling. (e.g., logistic regression).

It is important that careful consideration be given to confounders prior to any epidemiological studies being initiated in the field and it is important that any study adequately describe how this was done: epidemiological studies are frequently critiqued for ignoring or paying insufficient attention to potential confounders. For this reason, a sensitivity analysis can be helpful to demonstrate the potential effects that a missing or unaccounted for confounder may have on the observed effect sizes (see Gustafson and

[15] This is why it is generally considered inappropriate to "statistically test" for a confounder to determine whether the confounder needs to be adjusted for. Instead, some consider a change in the effect size of 10% or more after adjustment for (inclusion of) a potential confounder to be sufficient evidence for the confounder to be incorporated into the analysis.

McCandless, 2010). If unmeasured confounders are thought to affect the results, researchers should conduct sensitivity analyses to estimate the range of impacts and the resulting range of adjusted effect measures. Such sensitivity analyses -- generally not uniformly conducted in most published epidemiological studies – can be used when available to estimate the impact of biases and potential confounding by known but unmeasured risk factors.

Depending upon the specific exposure-disease association under study, a factor may or may not be a confounding factor that is necessary to control: in order for a substantial distortion in the effect size estimate to occur due to confounding, the confounder must be not only a relatively strong risk factor for the disease of interest[16], but also be strongly associated with the exposure of interest. Assessment of potential confounding is made on a study specific basis and – if unmeasured confounders are thought to affect the results -- researchers should conduct a sensitivity analysis to estimate the range of impacts and resulting range of adjusted effect measures. When evaluating the quality of observational epidemiology studies, OPP will consider whether relevant confounding factors are properly identified, described, measured and analyzed such that an unbiased estimate of the specific association under study can be made, and, when possible, may consider sensitivity analysis as a potential tool to assist in determining the degree to which such confounding might potentially affect the estimate of the effect size. It should be emphasized that a confounder must be a relatively strong risk factor for the disease to be strongly associated with the exposure of interest to create a substantial distortion in the risk estimate. In such cases, it is not sufficient to simply raise the possibility of confounding; one should make a persuasive argument explaining why a risk factor is likely to be a confounder, what its impact might be, and how important that impact might be to the interpretation of findings. (p. 23-25, FIFRA SAP Report, 22 April 2010)

Finally, it is important to distinguish between confounding, effect modification, synergy, and other mediating effects of covariates. Confounding is a bias that results from not controlling for a variable that is associated causally with the disease and associated – causally or non-causally -- with the exposure of interest. Epidemiologic researchers seek to minimize this bias. Effect modifiers -- on the other hand -- are variables that differentially affect the magnitude of the effect size, by strata (e.g., age, race/ethnicity, SES status, genetic polymorphisms). Effect modifiers may or may not also be confounders. Typically, they are modelled by either introducing interaction terms in multivariable models or by evaluating effect sizes by strata after stratifying the data by levels of the effect modifier. A study frequently needs to be specifically designed to evaluate effect modifiers in order to have a sufficient sample size in each population strata of interest. Epidemiologic researchers seek to understand effect modifiers (not minimize them, as they do with confounders) because they can be important in evaluating risk differences across population strata, in evaluating the association between exposure and the effect of interest, and in identifying susceptible

[16] Consideration needs to be given not only to ensuring that the confounding factor is indeed a risk factor on its own but also to ensuring not only related to the exposure of interest. Adjusting for a factor that has an association with the disease of interest wholly or partly because of its association with the exposure of interest will lead to attenuation of the exposure-disease relationship if it truly exists.

subpopulations. Effect modifiers may or may not also be confounders. For example, smoking may be a confounder in a study associating lung cancer with a pesticide often used on tobacco, but it may also be an effect modifier if the risk of exposure to this pesticide is higher among smokers than non-smokers. Synergy is often introduced as a biological or pharmacological/toxicological concept rather than an epidemiological one and relates to the ability of two chemicals, together and acting jointly, to magnify or exaggerate the effect beyond that which would be seen considering the (mathematical) sum of each chemical's effects alone. In epidemiological and statistical terms, this is often expressed as effect modification or interaction.

3. Statistical Analysis

Epidemiologic studies are designed to measure an association between a specific exposure and a disease. When evaluating the quality of pesticide epidemiology studies, OPP will also consider the statistical methods used. Specifically, OPP will consider the extent to which the analytic methods described in the study are appropriate to the research question; the completeness of the description of the statistical methods utilized; the appropriateness of the methods for identification, assessment and adjustment of potentially confounding variables in the exposure-disease relation; and, the description, extent of, and presentation of any sub-group analyses which may have been performed (including whether statistical corrections for multiple comparisons have been made).

Epidemiologic investigations typically utilize statistical modeling to estimate risk (e.g. generalized linear models such as logistic (for odds ratios) or Poisson (for count data) regression. To do so, researchers must consider not only the relevant main exposure and outcome variables, but also consider relevant confounding factors, and whether the association under investigation may differ by level of these factors, i.e., effect modification or interaction (Szklo et al., 2004). Upon identification of a potentially confounding variable -- one that substantively changes the magnitude and/or direction of the association under study -- adjustment through regression modeling can help to isolate the risk estimate of interest, i.e., the association under study. In addition, OPP will evaluate the stratification of the association by the level of the potential effect modifier under study or evaluation of statistical interaction. If the magnitude and direction of the association of interest differs greatly by level of a third variable, then the stratified results should be considered primary.

When performing statistical modeling when the outcome is rare or the sample size is relatively small, it is important to be cautious about including too many covariates in the model. Any resulting effect size estimate may be too high or too low and is unlikely to reflect the true estimate of effect. Such issues due to rare events or low sample sizes are also possible when conditional methods are used (e.g., conditional logistic regression when the design includes matching of the comparison group under study): if too few discordant pairs (or discordant sets) are observed, the estimated effect size may also be unreliable. Thus: while controlling for confounders and other covariates is important, the assessor must take care not to over-control or end up with too few degrees of freedom to produce a

reliable test. In these cases, it may be more important to seek parsimonious models that adjust for only a smaller number of the most influential confounders and other covariates so that the effective sample size remains adequate.

Finally, it is important in any statistical modeling exercise to consider statistical significance in the context of clinical/biological/scientific significance of the result. It may be that some results are statistically significant but unimportant in a clinical/biological/ scientific context. The reverse can be true: it may be that results are not statistically significant but may be important in a clinical/biological/scientific context. The former may suggest a sample size that is larger than necessary while the latter may suggest one than is smaller than needed. The latter case may be important from a public health perspective and warrant further exploration, especially when the association is strong (despite it being imprecise)

4. Potential Bias in Observational Research

Bias is a systematic error in the design or conduct of a study that gives rise to study results that are systematically different from the (unobserved) true situation. This contrasts with random errors which relate to sampling variability and precision (or, equivalently, confidence bounds) around the effect size measure, but which do not "drive" or "push" the result in one particular direction (e.g., either toward or away from the null).

Bias is a reflection of methodological imperfections in the design or conduct of the study and should be addressed or discussed by researchers as part of their analysis. There are a number of ways that bias can be introduced into a study: studies may be biased in the way in which participants are selected into the study (selection bias), or the way in which information about exposure and disease status is collected (information bias, including recall bias discussed earlier for case-control studies). One example of a common occupational selection bias is the "healthy worker effect" which can create an important bias in occupational epidemiology studies, leading to bias toward the null, and even below (creating the interpretation that the exposure is "protective") No study is totally devoid of bias and one should consider the extent to which authors of published studies described potential bias in the study, and how (if at all) they attempted to address it and characterize it in the study. Bias can result from differential or non-differential misclassification (Greenland, 1998). Differential misclassification (bias) means that misclassification has occurred in a way that depends on the values of other variables, while non-differential misclassification (bias) refers to misclassifications that do not depend on the value of other variables. Misclassification biases – either differential or non-differential – depend on the sensitivity and specificity of the study's methods used to categorize such exposures and can have a predictable effect on the direction of bias under certain (limited) conditions: this ability to characterize the direction of the bias based on knowledge of the study methods and analyses can be useful to the regulatory decision-maker since it may allow the decision maker to determine the extent to which, if any, the epidemiological effect sizes being considered (e.g., OR, RR) are likely underestimates or overestimates of the true effect

size[17]. It is not atypical to find degrees of misclassification in the range of 10 to 20 percent and it can be helpful in reviewing epidemiological studies to consider a form of sensitivity (or "what if") analysis which evaluates such a degree of misclassification -- and whether it is differential or non-differential – and the degree to which such misclassification might impact the odds ratio or relative risk with respect to both magnitude and direction[18]. (p.25, FIFRA EPA SAP report, 22 April, 2010). As mentioned earlier with respect to confounding, such quantitative sensitivity analysis is only rarely performed or practiced in published epidemiology studies, with bias instead more typically evaluated in a narrative manner without any quantitative assessment of its potential magnitude and the effect it may have on the epidemiological effect size estimates (Jurek at al., 2006). This may be due – in part -- to a general lack of availability of computational tools for such analysis by epidemiologists or their unfamiliarity with them. Such tools are becoming increasingly available and may be valuable in developing more rigorous quantitative methods for evaluation of potential biases.

5. Interpretation of Null studies

"Null" studies -- or well-conducted studies which report no association between exposure to the pesticide and an adverse health outcome -- will be evaluated carefully for their potential usefulness in human health risk assessment. The study may report a null result either because the investigated association indeed does not in reality exist, or because the study was conducted failed to detect an association at a given predetermined level of significance. This latter result –the failure to detect an association -- should not necessarily be interpreted to mean that no association exists, but rather as simply one was not found in the particular study[19,20]. To evaluate which of these two conditions may be correct when reviewing "null" studies, one should consider other research reported concerning the same or similar research question, the manner in which exposure and outcome were assessed, the extent to which exposure misclassification may have biased the study to the null, the statistical methods used including the identification and analysis of confounding variables in the association, the extent to which the exposure is below a threshold at which an effect would occur or be detected, as well as the power of the study and its ability to detect an effect size of substantive interest. Statistical power refers to the probability that researchers may correctly identify that there is a difference between the two comparison groups, i.e., there is an association between exposure and disease, when in

---

[17] The direction of bias that results from the degree of non-differential misclassification will also depend on the categorization of exposure (either dichotomous or polytomous).

[18] Such sensitivity analyses might be especially recommended for exposure misclassification biases which in many cases are expected to result in more substantive effects on the effect size estimate than those from confounding.

[19] The old adage that "the absence of evidence does should not be interpreted as the evidence of absence" is true here.

[20] See also the American Statistical Association's Statement on Statistical Significance and P-values at https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf

fact there is in fact a true difference (or association). Studies that are "low powered" may falsely conclude there is no association, when an association actually exists[21].

Finally, it is important to consider the effects of publication bias in any systematic review of the literature with respect to interpretation of null studies. The term publication bias refers to the tendency for the available published literature to disproportionately exclude such null studies. Studies that demonstrate such a "null" association between a disease or health outcome can be as equally informative as those that do provided that the study in question meets the quality criteria established as part of the epidemiological review process. These may include such factors as study design; the existence of an *a priori* hypothesis vs. an exploratory analysis; sample size and statistical power to detect an effect size of interest; proper ascertainment of outcome *vis-à-vis* sensitivity and specificity; the quality of the exposure assessment and the potential for differential and non-differential misclassification; adequacy of the measurement of key potential confounders and other forms of bias (information, selection, etc.); and evaluation of effect modifiers; appropriate statistical analyses, including consideration of and possible correction for multiple comparisons that a unsupported by a priori hypotheses, biological plausibility, or other supporting information.

6. *External Validity (Generalizability)*

As noted above, *validity* generally refers to the extent that exposure estimates reflect true exposure levels (Checkoway et al., 2004). *External validity*, or *generalizability*, refers to the ability to extend the epidemiologic study results derived from a sample of the population (e.g., pesticide applicators) to other populations (e.g., all agricultural workers). To assess external validity, comparison of characteristics in the sample to the larger population (if known) can be made. Such evaluation should include not only demographic factors, but also whether exposures (e.g., dose, timing, duration) are similar and whether important effect modifiers (e.g., sensitivity of vulnerable populations) were considered. Generalizability is of particular importance because it is important to understand whether and how individual study results may be applied to the larger group or targeted sub-groups in regulatory risk assessment. For example, the AHS has reported statistical associations between some cancer and non-cancer health outcomes for some pesticide chemicals. OPP has an interest in evaluating the extent to which the reported findings may apply to pesticide applicators in states other than North Carolina and Iowa or to farm workers who primarily do post-application activities.

[21] Studies that are low-powered but find statistically significant effects may also be subject to the phenomenon of effect size magnification and this can be important to investigate as well. (Ioannidis, 2008).

V. HUMAN INCIDENT SURVEILLANCE DATA

Generally speaking, epidemiology studies on pesticides such as those described above focus on lower exposures (over a longer time period) that are less likely to result in acute clinical symptoms. OPP is also interested in exposures that are higher and occur over shorter-intervals (often on an acute "one-time" basis). This "human incident," or poisoning data can be useful for evaluating short term, high exposure scenarios that can be readily attributed to the pesticide in question.

OPP uses such "human incident information" for several purposes. Most broadly, the program uses incident data to inform risk assessment/risk management activities; this forms an integral part of our registration review activities under our Pesticide Registration Improvement Act (PRIA) responsibilities. To this end, OPP evaluates human incident data for trends over time and examines patterns in the severity and frequency of different pesticide exposures. In some cases, incident information can indicate need for additional information or additional risk management measures. Incident information can also help assess the success of risk mitigation actions after they are implemented, and incident information is an important part of OPP's performance accountability system to ensure the effectiveness of risk management actions that OPP has taken to protect human health and the environment. Lastly, incident information can be useful in providing real world use information with respect to usage practices and also in potentially targeting enforcement or educational activities, where appropriate.

OPP obtains this information from a variety of sources. Sources of human incident data include both (human) **medical case reports** appearing in the medical and toxicological literature as well as information from a variety of national **toxico-surveillance activities** for acute pesticide poisonings which are considered jointly to aid acute and chronic hazard identification and as an integral part of the risk assessment process.[22]

Medical case reports (first-hand accounts written by physicians) or medical case series (a compendium of medical case reports across individuals that share common source or symptomology) are valuable tools for analyzing all available evidence of health effects, and to complement the findings of animal studies and epidemiological studies. In addition, they can identify unusual or novel occurrences of an adverse health effects plausibly associated with use of a specific pesticide providing "advance notice" to the agency for toxico-vigilance purposes. Published case reports for pesticides typically describe the effects from an atypical (high exposure/dose, illegal, off-label) acute or short-term exposure. The reports are often anecdotal and can be highly selective in nature. They can, however, can be particularly valuable in identifying previously unidentified toxic effects in humans and in learning about the effects, health outcomes, and medical sequelae following high exposures. They frequently have more detailed medical information (including sequelae), detailed follow-up, and generally higher quality and/or quantitative

[22] OPP is aware of efforts by IPSC to consider human incident data in risk assessment. http://www.who.int/ipcs/publications/methods/human_data/en/index.html

information about dose. If similarities are seen across multiple medical case studies or patterns emerge – in symptoms, exposure scenarios or usage practices -- these can provide valuable information for the risk assessment process and strengthen any findings. Medical case studies and series that include quantitative exposure information can be compared to exposure estimates in the risk assessment (which are based on labeled application rates and surrogate exposure information) to characterize margins of exposure expected from typical use, when appropriate.

The following considerations are evaluated in assessing medical case reports and medical case series:

- A detailed history of exposure (when, how, how much); time of onset of adverse effects; and signs and symptoms of the patient, are reported.
- Information on the product/chemical/pesticide, such as name, pesticide label, registration number, etc.
- Patient information (e.g. age, race, sex); underlying health conditions and use of any medications that can produce similar signs and symptoms; relevant medical history; and the presence of any risk factors.
- Description of events and how the diagnosis was made.
- Management and treatment of the patient, and laboratory data (before, during and after the therapy), including blood levels of pesticides and chemicals.
- Whether the medical report is reliable, reasonable and whether it is consistent with current knowledge, including other research, reviews and guidelines.
- Clinical course of the event and patient outcome (e.g. patient recovered and discharged from hospital; condition of patient after the discharge, any chronic health effects or premature death related to the pesticide or chemical exposure).

In addition to using medical case reports/series as a source of real-world exposure and toxicological information, OPP also engages in toxico-surveillance activities using a variety of pesticide poisoning incident databases are also available. Specifically, OPP has access to the following five human incident data sources: the *OPP Incident Data System* (IDS); the American Association of Poison Control Centers (PCC) summary reports from their *National Poison Data System* (NPDS); data from the EPA-funded *National Pesticide Information Center* (NPIC), currently at Oregon State University; the Centers for Disease Control and Prevention/National Institute for Occupational Safety and Health *Sentinel Event Notification System for Occupational Risk-Pesticides* (NIOSH SENSOR-Pesticides) and the *California Pesticide Illness Surveillance Program* (PISP). Each of these are described, in turn below:

- **OPP Incident Data System (IDS)** is maintained by OPP and incorporates data submitted by registrants under FIFRA section 6(a)(2)[23], as well as other incidents reported directly to EPA. OPP has compiled the pesticide related

[23] Under FIFRA 6(a)(2), pesticide registrants are required to notify EPA if and when they become aware of "factual information regarding unreasonable adverse effects on the environment of the pesticide."

incident reports in the IDS since 1992. The IDS includes reports of alleged human health incidents from various sources, including mandatory FIFRA Section 6 (a) (2) reports from registrants, other federal and state health and environmental agencies and individual consumers. IDS include information on incidents involving humans, plants, wild and domestic animals where there is a claim of an adverse effect. The vast majority of IDS reports are received by the agency in paper format. IDS entries act as a "pointers" to copies of original reports retained on microfilm and scanned images in OPP's Information Service Center.

While IDS includes both occupational and non-occupational incidents, the majority of incidents reported relate to non-occupational/residential scenarios The reports are obtained from across the U.S. and most incidents have all relevant product information (such as the EPA Registration Number) recorded. As IDS is populated mostly by information provided by pesticide registrants under their FIFRA 6(a)(2) reporting requirements, the agency has relatively high confidence in the identification of the specific product which is involved. Severity rankings are included for each incident (as specified by CFR §159.184). Symptom information is sometimes included in the narrative portion of the incident, but this information is usually not validated/confirmed by a healthcare professional. IDS also includes narrative information on exposure scenario and hazard information. Many companies use standardized, industry-developed Voluntary Incident Reporting Forms.

OPP collects and evaluates the data from the IDS and identifies potential patterns with respect to the extent and severity of the health effects due to pesticides exposure. While IDS reports are broad in scope and can in some cases contain detailed information, the system does not necessarily consistently capture detailed information about incident events, such as occupational exposure circumstances or medical outcome.

In addition, most cases data going into IDS is not validated or verified, though some reports are collected from calls to contract poison control centers. Nevertheless, incident information can provide an important post-marketing feedback loop to the agency following initial registration of the product: IDS incidents of a severe nature, or a suggested pattern or trend among less severe incidents can signal the agency to further investigate a particular chemical or product. Because IDS has such extensive coverage, it can assist in providing temporal trend information and determining whether risk mitigation has helped reduce potential pesticide exposure and decreased the number of potential incidents reported to IDS. Overall, IDS provides good information about national trends and frequency of incidents for pesticides and can provide valuable insights into the hazard and/or exposure potential of a pesticide.

- **The National Poison Data System (NPDS)** -- formerly called the Toxic Effects Surveillance System (TESS) -- is maintained by the American Association of Poison Control Centers (AAPCC) and is supported with funding from several federal agencies. NPDS is a computerized information system with geographically specific and near real-time reporting. Although the main mission of Poison Control Centers is in helping callers respond to emergencies, NPDS data can help identify emerging problems in chemical product safety. Hotlines at 61 PCC's nationwide are open 24/7, 365 days a year and are staffed by specially trained nurses, pharmacists, and other clinical health care specialists to provide poisoning information. Using computer assisted data entry, standardized protocols, and strict data entry criteria, local callers report incidents. These reported incidents are retained locally and are updated in summary form to the national database maintained by AAPCC. Information calls are tallied separately and not counted as incidents. The PCC system covers nearly all the US and its territories and has undergone major computer enhancements since 2001.

  NPDS includes mainly non-occupational incidents. NPDS does not include narrative information and the product information may not be complete. NPDS provides severity rankings and symptom information that are designated/recorded by trained specialists, and the agency has relatively high confidence in this information. NPDS also provides some information on the likelihood of the adverse effect being a result of the reported exposure. Overall, NPDS provides good information about national trends, frequency of incidents for pesticides, as well as the hazard potential for particular pesticides. However, resource limitations permit the agency to only access AAPCC summary reports published each year (e.g., see http://www.aapcc.org/annual-reports/) and these serve as a supplement to other data sources for which the agency has more complete access.

- **The National Pesticide Information Center (NPIC)** (http://npic.orst.edu/index.html) is funded by EPA to serve as a source of objective, science-based pesticide information in response to inquiries and to respond to incidents. NPIC functions nationally during weekday business hours and is a cooperative effort between Oregon State University (currently) and EPA; it is intended to serve as a source of objective, science-based pesticide information and to respond to inquiries from the public and to incidents. Similar to Poison Control Centers, NPIC's primary purpose is not to collect incident data (about 10% of NPIC's annual calls are considered "incident" related), but rather to provide information to inquirers on a wide range of pesticide topics, and direct them to other sources for pesticide incident investigation and emergency treatment. Nevertheless, NPIC does collect information about incidents (approximately 4000 incidents per year) from inquirers and records that information in a database. NPIC is a source of national incident information, but generally receives fewer reports than IDS. Regardless, if a high frequency is observed in IDS for a given pesticide or

product, NPIC provides a source of information that can prove valuable in determining consistency across national data sets.

As with IDS and PCC, the incidents in NPIC are mainly non-occupational. NPIC incidents include narratives and product information when the caller provides the information. Although the scope is national, there are significantly fewer incidents reported to NPIC than to NPDS or IDS but considerably more information is provided and the agency can request custom reports on an as-needed basis. Hazard information includes severity rankings, route of exposure and symptoms – which are recorded by trained personnel. NPIC also provides information on how likely the link between exposure and adverse effect is (which they call a certainty index). NPIC also publishes annual reports and analyses in the open literature which are valuable resources.

- ❑ The Center for Disease Control and Prevention National Institute for Occupational Health (CDC/NIOSH) manages a pesticide surveillance program and database entitled the **Sentinel Event Notification System for Occupational Risk (SENSOR)-Pesticides.**[24] This database includes pesticide illness case reports in 12 states from 1998-2013. Participating states are: California, Florida, Iowa, Louisiana, Michigan, Nebraska, New Mexico, New York, North Carolina, Oregon, Texas and Washington. The participating states for a given year vary depending on state and federal funding for pesticide surveillance.

  Cases of pesticide-related illnesses in the SENSOR-Pesticides database are ascertained from a variety of sources, including: reports from local Poison Control Centers, state Department of Labor workers' compensation claims when reported by physicians, reports from state Departments of Agriculture, and physician reports to state Departments of Health. Although both occupational and non-occupational incidents are included in the database, the SENSOR coordinators primarily focus their follow-up case investigation efforts on the occupational pesticide incidents. The SENSOR coordinator at the state Department of Health will follow-up with cases and work to obtain medical records in order to verify exposure scenario, symptoms, severity, and health outcome. Using standardized protocol and case definitions, SENSOR coordinators at state Departments of Health enter the incident interview description provided by the case, medical report, physician and patient into the SENSOR data system.

  All SENSOR-Pesticides cases must report a minimum of two health effects in order to be included in the aggregate database that EPA uses for incident

[24] SENSOR-Pesticides webpage: http://www.cdc.gov/niosh/topics/pesticides/overview.html

analyses. Evidence for each case is evaluated, based on the NIOSH case classification matrix, for its causal relationship between exposure and illness. 98% of SENSOR-Pesticides cases are classified as definite, probable, or possible, and 2% of the cases are classified as suspicious. Unlikely, asymptomatic, and unrelated cases, as well as those with insufficient information, are not included in the SENSOR-Pesticides database.

Overall, SENSOR-Pesticides provides very useful information on both occupational and non-occupational incidents, and sometimes valuable insights into the hazard and/or exposure potential of a pesticide. SENSOR-Pesticides also conducts analyses of its own data and publishes these in the Morbidity and Mortality Weekly. Unlike the aforementioned databases and although it contains both non-occupational/residential and occupational incidents, SENSOR's has traditionally focused on occupational pesticide incidents, and is of particular value in providing that information. SENSOR-Pesticides data from 1998-2011 is available online at: http://wwwn.cdc.gov/Niosh-whc/Home/Pesticides.

- **The California Pesticide Illness Surveillance Program (PISP)** is maintained by the State of California. This database documents pesticide-related illnesses and injuries. Case reports are received from physicians and via workers' compensation records. The local County Agricultural Commissioner investigates the circumstances of the exposure. Medical records and investigative findings are then evaluated by California's Department of Pesticide Regulation (DPR) technical experts and entered into an illness registry. All reported pesticide illnesses in the California PISP program are investigated by the county agricultural commissioners, and the DPR evaluates the reports and compiles them into a database, which is used to improve the state's program to protect workers and others from the adverse effects of pesticide exposure (http://apps.cdpr.ca.gov/calpiq/).

Currently, OPP evaluates human incident data on a chemical-specific basis. Incidents from each database are analyzed for hazard potential (deaths, frequency of more severe incidents, and patterns/trends of reported symptoms) and exposure potential (frequency of incidents/ trends over time, patterns/trends of exposure scenarios, of factors affecting exposure or of products). When evaluating human incident data from the above databases, OPP considers several general criteria. OPP considers the relative severity and frequency of symptoms. Additionally, OPP generally has greater confidence in reports in which temporal association can be verified or are at least plausible. Lastly, other factors that are used to evaluate human incident data include evidence of an exposure response association, consistency in reported health effects, biological plausibility of reported health effects, elimination of alternative causes of health effect such as pharmaceutical use, and the specificity of the observed symptoms and health effects. Additionally, narratives of more severe incidents are often evaluated for any temporal association between time-of-exposure and effects reported to determine whether an association is supported by the circumstances. For example, a heart attack in an elderly individual that occurs three

months following an indoor pesticide application may be determined not to be a likely causal association. On the other hand, a severe incident occurring at or shortly after the time of exposure with symptoms consistent with known symptomology for the pesticide class and that occurs without prior medical history may suggest that causal inference is more justified.

In sum, then, incident data -- consisting of both medical case reports/case series appearing the medical and human toxicological literature and toxico-surveillance data derived from the databases that EPA either maintains, funds, or accesses -- can provide useful, complementary information that assists OPP in evaluating the real-world risks of pesticides.

VI. SUMMARY & CONCLUSIONS

This framework describes important factors in reviewing epidemiology and human incident data and describes a proposed WOE analysis for incorporating such data in pesticide human health risk assessment. OPP uses the best available data across multiple lines of evidence and from *in vitro*, *in vivo*, and *in silico* data sources. OPP uses a WOE approach when integrating data from multiple sources to take into account for quality, consistency, relevancy, coherence and biological plausibility using modified Bradford Hill criteria as an organizational tool. Application of WOE analysis is an integrative and interpretive process routinely used by EPA according to in scientific analysis outlined in its risk assessment guidelines. The WOE analysis also evaluates the quality of the combined data set and is consistent with the level of effort and complexity that is appropriate for a particular scientific assessment (U.S. EPA, 2002). OPP acknowledges that toxicology and risk assessment are currently undergoing transformational changes towards implementing the new vision of 21st century toxicity testing. As these transformation changes occur, OPP will update this approach as appropriate.

VII. REFERENCES

American Statistical Association. 2016 "AMERICAN STATISTICAL ASSOCIATION RELEASES STATEMENT ON STATISTICAL SIGNIFICANCE AND P-VALUES" March 7. Available at: https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf

Ankley, GT, Bennett RS, Erickson RJ et al. 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29(3):730–741.

Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, and Guyatt GH. 2011. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011 Apr;64(4):401-6. doi: 10.1016/j.jclinepi.2010.07.015. Epub 2011 Jan 5.

Blair A, Tarone R, Sandler D, Lynch C, Rowland A, Wintersteen W, Steen W, Dosemeci M, and Alavanja M. 2002. Reliability of reporting on lifestyle and agricultural factors by a sample of participants in the agricultural health study from Iowa. Ann Epidemiol. Oct 1;10(7):478.

Borenstein M, Hedges LV, Higgins JPT, and Rothstein HR. 2009. Introduction to Meta-analysis. John Wiley and Sons, Chichester, UK.

Boyes WK, Moser VC, Geller AM, Benignus VA, Bushnell PJ, and Kamel F. 2007. Integrating epidemiology and toxicology in neurotoxicity risk assessment. Hum. Exp. Toxicol. 26(4):283-93.

Calderon RL 2000. Measuring risks in humans: the promise and practice of epidemiology. Food and Chemical Toxicology. 38:S59-S63.

Carlile DJ, Zomorodi, K , and Houston, JB. 1997. Scaling factors to relate drug metabolic clearance in hepatic microsomes, isolated hepatocytes, and the intact liver: studies with induced livers involving diazepam. Drug Metab. Dispos. 25(8):903-911.

Checkoway H, Pearce, N, and Kriebel D. 2004. Research Methods in Occupational Epidemiology, 2nd Edition. Oxford University Press, New York.

Clark LH, Setzer RW, and Barton, HA. 2004. Framework for evaluation of physiologically-based pharmacokinetic models for use in safety or risk assessment. Risk Anal. 24(6):1697-1717.

FIFRA Scientific Advisory Panel. (2010). February 2 - 4, 2010: Incorporation of Epidemiology and Human Incident Data into Human Risk Assessment.

Glymor, MM and Greenland, S. 2012. "Causal Diagrams" in Rothman, KJ, Greenland, S, Poole, C, Lash, TL. Modern epidemiology. 3rd ed. Lippincott Williams & Wilkins, Philadelphia. pp. 183-212.

Gordis, L. 2009. Epidemiology. 4th Edition. Saunders-Elsevier, New York.

Greenland, S. 1998. "Basic Methods for Sensitivity Analysis and External Adjustment" in Rothman, KJ and Greenland, S. Modern Epidemiology. 2nd ed. Lippencott-Raven Publishers, Philadelphia. pp.343-357.

Greenland, S. and Lash T. 2012. "Bias Analysis" in Rothman KJ, Greenland S, Poole C, Lash TL. Modern epidemiology. 3rd ed. Lippincott Williams & Wilkins, Philadelphia. pp. 345-380.

Greenland, S and O'Rourke, K. 2012. "Meta-analysis" in Rothman, KJ, Greenland S, Poole C, and Lash, TL. Modern epidemiology. 3rd ed. Lippincott Williams & Wilkins, Philadelphia. pp. 652-682.

Grimes, DA and Schultz, KF. 2005. Compared to What? Finding controls for case-control studies. Lancet 365: 1429-1433.

Gustafson P[1], and McCandless LC. 2010. Probabilistic approaches to better quantifying the results of epidemiologic studies. Int J Environ Res Public Health. 2010 Apr;7(4):1520-39. doi: 10.3390/ijerph7041520..

Hartung T. 2010. Evidence-based toxicology - the toolbox of validation for the 21st century? ALTEX. 2010;27(4):253-63.

Hertz-Picciotto I. 1995. Epidemiology and quantitative risk assessment: a bridge from science to policy. American Journal of Public Health. 85(4): 484-491.

Hill AB. 1965. The Environment and Disease: Association or Causation? President's Address. Proceedings of the Royal Society of Medicine 58:293-300

Hoppin JA, Yucel F, Dosemeci M, and Sandler DP. 2002. Accuracy of self-reported pesticide use duration information for licensed pesticide applicators in the Agricultural Health Study. Journal of Exposure Analysis and Environmental Epidemiology, 12: 313-318.

IPCS (2005). Chemical-specific adjustment factors for interspecies differences and human variability: guidance document for use of data in dose/concentration response assessment. Harmonization Project Document 2. World Health Organisation, International Programme on Chemical Safety, Geneva, Switzerland.

Ioannidis JP. 2008. Why Most Discovered True Associations Are Inflated. Epidemiology. 19(5): 640-8.

Jurek AM, Maldonado G, Greenland G, and Church TR. 2006. Exposure-measurement Error is Frequently Ignored When Interpreting Epidemiological Study Results. Europ. J. Epid. 21: 871-876.

Kelsey JL, Whittemore AS, Evans AS, and Thompson WD. 1996. Methods in Observational Epidemiology. 2nd ed. Oxford University Press, New York.

LaKind JS, Sobus JR, Goodman M, Barr DB, Fürst P, Albertini RJ, Arbuckle TE, Schoeters G, Tan YM, Teeguarden J, Tornero-Velez R, and Weisel CP. 2014. A proposal for assessing study quality: Biomonitoring, Environmental Epidemiology, and Short-lived Chemicals (BEES-C) instrument. Environ Int. Dec;73:195-207. doi: 10.1016/j.envint.2014.07.011.

Lash, TL, Fox, MP, and Fink, AK. 2009. Applying Quantitative Bias Analysis to Epidemiologic Data. Springer, New York.

Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, and Greenland S. 2014. Good practices for quantitative bias analysis. International Journal of Epidemiology p. 1-17.

Lilienfeld AM and Lilienfeld D. 1979. Foundations of epidemiology, 2nd ed. Oxford University Press, New York.

Mausner JS and Kramer S. 1985. Epidemiology, 2nd ed. W.B. Saunders, Philadelphia.

Meek, ME, Bucher, JR, Cohen, SM et al. 2003. A framework for human relevance analysis of information on carcinogenic modes of action. Crit. Rev. Toxicol. 33:591-653.

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S, Seed J, and Vickers, C. 2014. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol. Jan;34(1):1-18.

Muñoz-Quezada MT[1], Lucero BA, Barr DB, Steenland K, Levy K, Ryan PB, Iglesias V, Alvarado S, Concha C, Rojas E, and Vega C. 2013. Neurodevelopmental effects in children associated with exposure to organophosphate pesticides: a systematic review. Neurotoxicology. Dec;39:158-68. doi: 10.1016/j.neuro.2013.09.003.

Needham LL, Calafat AM, and Barr DB. 2007. Uses and issues of biomonitoring. Int. J. Hyg. Environ. Health. 210: 229-238.

Nieuwenhuijsen MJ. 2003. Exposure Assessment in Occupational and Environmental Epidemiology. Oxford University Press, New York.

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. National Academy Press, Washington, DC.

NRC (National Research Council). 1991. Environmental Epidemiology, Volume 1: Public Health and Hazardous Wastes. National Academy Press, Washington, DC.

NRC (National Research Council). 1994. Science and Judgment in Risk Assessment. National Academy Press, Washington, DC.

NRC (National Research Council). 1997. Environmental Epidemiology, Volume 2: Use of the Gray Literature and Other Data in Environmental Epidemiology. National Academy Press, Washington, DC.

NRC (National Research Council). 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. National Academy Press, Washington, DC.

NRC (National Research Council). 2009: Science and Decisions: Advancing Risk Assessment. National Academy Press, Washington, DC.

NRC (National Research Council). 2011. Review of the Environmental Protection Agency's draft IRIS assessment of formaldehyde. National Academies Press, Washington, DC. http://www.nap.edu/catalog/13142.html

NRC (National Research Council). 2014. Review of EPA's Integrated Risk Information System (IRIS) process. The National Academies Press, Washington, DC. http://www.nap.edu/catalog.php?record_id=18764

Ntzani EE, Chondrogiori MNG, Evangelou E and Tzoulaki I. 2013. Literature review of epidemiological studies linking exposure to pesticides and health effects. External Scientific Report. EFSA supporting publication 2013-EN-497. 159 pp. Available online at www.efsa.europa.eu/publications.

Organisation for Economic Co-operation and Development. 2013. GUIDANCE DOCUMENT ON DEVELOPING AND ASSESSING ADVERSE OUTCOME PATHWAYS, Series on Testing and Assessment, No. 184, ENV/JM/MONO(2013)6, April 17, 2013. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en

Paddle GM, and Harrington JM. 2000. Environmental epidemiology--strengths and weaknesses. Int Arch Occup Environ Health. 73:7-14.

Porta MJM. 2014. A Dictionary of Epidemiology. 6th ed. Oxford University Press, New York.

Purdue Pesticides Programs. 2003. Pesticides and Epidemiology: Unraveling Disease Patterns. Purdue University Cooperative Extension Service. http://www.btny.purdue.edu/Pubs/PPP/PPP-43.pdf.

Rothman KJ and Greenland S. 2012. Modern epidemiology. 3rd ed. Lippincott Williams & Wilkins, Philadelphia.

Rothman, KJ, Greenland S, Poole C, and Lash TL. 2012a. "Causation and Causal Inference" in Rothman, KJ, Greenland S, Poole C, and Lash TL. Modern epidemiology. 3rd ed. Lippincott Williams & Wilkins, Philadelphia. pp. 5-31.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, and Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect. Jul;122(7):711-8. doi: 10.1289/ehp.1307972.Seed, J., E.W.

Carney, RA, Corley, et al. 2005. Overview: Using mode of action and lifestage information to evaluate the human relevance of animal toxicity data. Crit. Rev. Toxicol. 35(8-9):664-672.

Schultz, KF and Grimes DA. 2002. Case-control studies: research in reverse. Lancet: 359:431-434.

Sonich-Mullin C, Fielder R, Wiltse J, et al. 2001. IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regul Toxicol Pharmacol. 34:146-152.

Sterne JAC, Higgins JPT, and Reeves, BC. on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for non-randomized studies of interventions (ACROBAT-NRSI), Version 1.0. 0, 24 September 2014." *www. riskofbias. info* (2015)

Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Bennie D, Moher D, Becker BJ, Sipe TA, and Thacker SB for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. 2000. J. American Medical Association 283(15): 2008-2012.

Szklo M and Nieto FJ. 2004. Epidemiology: Beyond the Basics. Jones and Bartlett Publishers, Boston, MA.

U.S. Environmental Protection Agency. (1999). Guidelines for carcinogen risk assessment. Risk Assessment Forum. SAB review draft. Washington, DC: U.S. Environmental Protection Agency. www.epa.gov/ncea/raf/crasab.htm.

U.S. EPA (U.S. Environmental Protection Agency). 2000. Science Policy Council Handbook: Risk Characterization. U.S. Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment, Washington, DC. EPA/100/B-00/002. Available at http://www.epa.gov/iris/backgr-d.htm.

U.S. EPA (U.S. Environmental Protection Agency). 2001. General Principles For Performing Aggregate Exposure And Risk Assessments Washington, DC. Available at https://www.epa.gov/sites/production/files/2015-07/documents/aggregate.pdf

U.S. Environmental Protection Agency. 2001a. "Guidance on Cumulative Risk Assessment of Pesticide Chemicals that Have a Common Mechanism of Toxicity." Office of Pesticide Programs, Office of Prevention, Pesticides, and Toxic Substances, U.S. Environmental Protection Agency. Washington, DC.

U.S. EPA (U.S. Environmental Protection Agency). 2002a. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity for Information Disseminated by the Environmental Protection Agency. Office of Environmental Information, Washington, DC. EPA/260/R-02/008. Available at http://www.epa.gov/quality/informationguidelines/documents/EPA_InfoQualityGuidelines.pdf.

U.S. EPA (U.S. Environmental Protection Agency). 2002b. A Review of the Reference Dose and Reference Concentration Processes. December. Risk Assessment Forum. Washington, DC. EPA/630/P-02/002F.

U.S. Environmental Protection Agency. 2002c. "Guidance on Cumulative Risk Assessment of Pesticide Chemicals That Have a Common Mechanism of Toxicity"; January 14, 2002.
U.S. EPA (U.S. Environmental Protection Agency). 2004. An Examination of EPA Risk Assessment Principles & Practices. Staff Paper Prepared for the U.S. Environmental Protection Agency by members of the Risk Assessment Task Force. Office of the Science Advisor. U.S. Environmental Protection Agency, Washington, DC. EPA/100/B-04/001.

U.S. EPA (U.S. Environmental Protection Agency). 2005. Guidelines for Carcinogen Risk Assessment. U.S. Environmental Protection Agency, Risk Assessment Forum, Washington, DC. EPA/630/P-03/001F. Federal Register 70(66):17765-17817. Available at http://www.epa.gov/raf.

U.S. EPA. (U.S. Environmental Protection Agency). 2006a. Harmonization in Interspecies Extrapolation: Use of $BW^{3/4}$ as Default Method in Derivation of the Oral RfD (External Review Draft). U.S. Environmental Protection Agency, Washington, DC. EPA/630/R-06/001. Available at http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=148525.

U.S. EPA (U.S. Environmental Protection Agency). 2006b. Approaches for the Application of Physiologically Based Pharmacokinetic (PBPK) Models and Supporting Data in Risk Assessment (Final Report). U.S. Environmental Protection Agency, Washington, DC. EPA/600/R-05/043F.

U.S. EPA (U.S. Environmental Protection Agency). 2009. Scientific Issues Associated with Field Volatilization of Conventional Pesticides. U.S. Environmental Protection Agency, Washington, DC. OPP Regulatory Public Docket EPA-HQ-OPP-2009-0687.

U.S. EPA (U.S. Environmental Protection Agency). 2010. Draft Framework for Incorporating Human Epidemiologic and Incident Data in Health Risk Assessment, January 7, 2010.

U.S. EPA (U.S. Environmental Protection Agency). 2012. Guidance for Considering and Using Open Literature Toxicity Studies to Support Human Health Risk Assessment U.S. Environmental Protection Agency, Washington, DC. Office of Pesticide Programs. Available at: https://www.epa.gov/sites/production/files/2015-07/documents/lit-studies.pdf

U.S. EPA (U.S. Environmental Protection Agency). 2014a. Framework for Human Health Risk Assessment to Inform Decision Making. https://www.epa.gov/sites/production/files/2014-12/documents/hhra-framework-final-2014.pdf

U.S. EPA (U.S. Environmental Protection Agency). 2014b. Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation. http://www2.epa.gov/osa/guidance-applying-quantitative-data-develop-data-derived-extrapolation-factors-interspecies-and

U.S. EPA (U.S. Environmental Protection Agency). 2015. Preamble to the Integrated Science Assessments. National Center for Environmental Assessment, RTP Division, Office of Research and Development, USEPA. https://yosemite.epa.gov/sab/sabproduct.nsf/0/33E1AD305287588F85257D20006BE8CC/$File/ISA_PREAMBLE_FINAL2015.PDF

van den Brandt P, Voorrips L, Hertz-Picciotto I, Shuker D, Boeing H, Speijers G, Guittard C, Kleiner J, Knowles M, Wolk A, and Goldbohm A. 2002. The contribution of epidemiology. Food Chem Toxicol. Feb-Mar;40(2-3): 387-424.

Vandenbroucke JP, Van Elm E, Altman DG, Gotzsche PC, Mulroew CD, Pockock SJ, Pool C, Schlesseman JJ, and Egger, M. 2007. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. Ann. Int. Med. Vol 147(8): W163-193

Von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, and Vandenbroucke JP. 2014. The Strengthening of the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. Int. J. Surgery 12: 1495-99.

Wacholder S, McLaughlin JK, Silverman DT, and Mandel JS. 1992a. Selection of Controls in Case-Control Studies I. Principles. American J. Epid. 135(9): 1019-1028.

Wacholder S, McLaughlin JK, Silverman DT, and Mandel JS. 1992b. Selection of Controls in Case-Control Studies II. Types of Controls. American J. Epid. 135(9): 1029-1041.

Wacholder, S, McLaughlin, JK, Silverman, DT, and Mandel, JS. 1992c. Selection of Controls in Case-Control Studies III. Design Options. American J. Epid. 135(9): 1042-1050.

Woodruff TJ[1] and Sutton P. 2014. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect. Oct;122(10):1007-14. doi: 10.1289/ehp.1307175.

Zartarian V., Bahadori T, and McKone T. 2005. Adoption of an official ISEA glossary. Journal of Exposure Analysis and Environmental Epidemiology 15:1-5