

**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

---

<b>I. Purpose &amp; Scope</b>	<b>3</b>
<b>II. Introduction</b>	<b>4</b>
<b>III. Systematic Review in Pesticide Risk Assessment: Epidemiology</b>	<b>8</b>
<i>A. Problem Formulation</i>	9
<i>B. Data Collection</i>	10
<i>C. Data Evaluation</i>	11
<i>D. Data Integration: Weight of Evidence (WOE)</i>	12
<i>E. Overall conclusions, recommendations for risk assessment, statement of areas of confidence and uncertainty</i>	15
<b>IV. Reviewing Epidemiology Studies for Use in Pesticide Risk Assessment</b>	<b>16</b>
<i>A. Introduction</i>	16
<i>B. Types of Epidemiology Studies</i>	18
<i>C. Evaluating epidemiology studies for use in pesticide risk assessment</i>	21
1. Exposure Assessment	25
2. Confounding Factors	30
3. Statistical Analysis	33
4. Potential Bias in Observational Research	34
5. Interpretation of Null studies	35
6. External Validity (Generalizability)	36
<b>V. Human Incident Surveillance Data</b>	<b>37</b>
<b>VI. Summary &amp; Conclusions</b>	<b>43</b>
<b>VII. References</b>	<b>44</b>

## 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<sup>1</sup> and from the National Institute of Environmental Health Sciences (NIEHS)/EPA Children's Centers<sup>2</sup>, 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; Glymour 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

---

<sup>1</sup> <https://aghealth.nih.gov/>

<sup>2</sup> <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 21<sup>st</sup> century toxicity testing advocates a shift away from the current focus of using apical toxicity endpoints to using toxicity pathways<sup>3</sup> 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 21<sup>st</sup> 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.

---

<sup>3</sup> Toxicity pathways are cellular response pathways that, when sufficiently perturbed, are expected to result in adverse health effects.

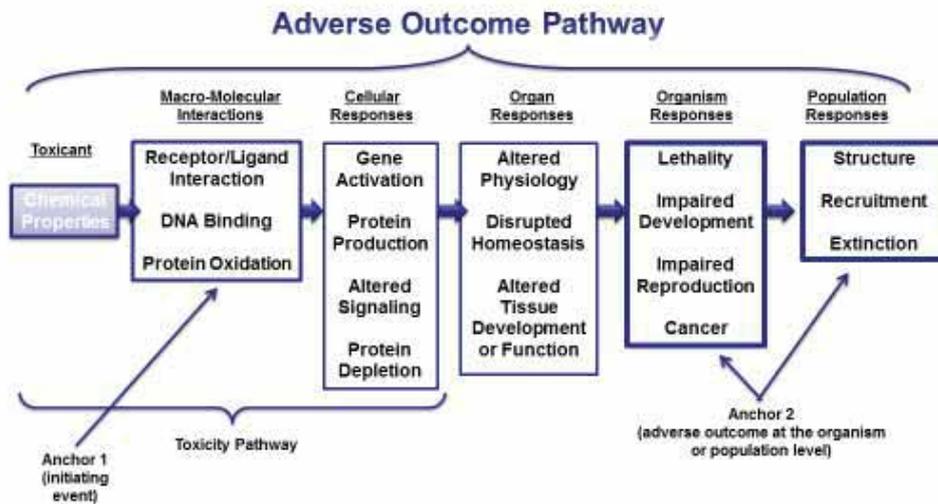
As a general principle, occupational and environmental epidemiology studies are conducted only on widely used pesticides; these pesticides also tend to have to be well-studied in the scientific literature. Thus, OPP expects in many cases where epidemiologic data are available, a significant body of literature 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).**

## Adverse Outcome Pathway



**Table 1. Key guidance documents and frameworks used by OPP**

<b>NAS</b>	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
<b>WHO/IPCS</b>	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.

<b>EPA</b>	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) <sup>4</sup>
	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
<b>OPP</b>	2001: Aggregate risk assessment
	2001 and 2002: Cumulative risk assessment
<b>OECD</b>	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<sup>5</sup> or on studies done to measure dermal or inhalation exposures in agricultural workers as they perform their activities<sup>6,7</sup>.

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

<sup>5</sup> 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.

<sup>6</sup> 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](http://www.epa.gov/scipoly/sap/meetings/2007/010907_mtg.htm))

<sup>7</sup> 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<sup>2</sup> 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<sup>8</sup>). 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

---

<sup>8</sup> [https://aopwiki.org/wiki/index.php/Main\\_Page#OECD\\_User\\_Handbook](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., epidemiologic controlled 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.

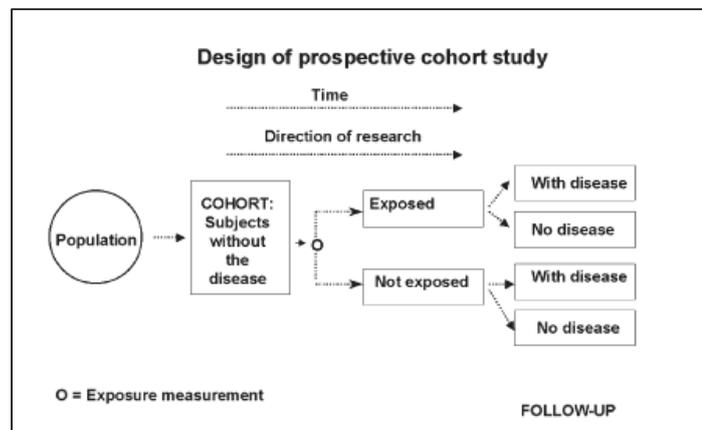
Animal studies do not replicate the length, magnitude, duration, routes of exposure and variability in exposure experienced by humans (Calderon, 2000). Human exposure often occurs through multimedia exposure pathways, including food, water, air, and indoor and outdoor environments. In contrast, controlled 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)<sup>9</sup>. 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<sup>10</sup>. 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



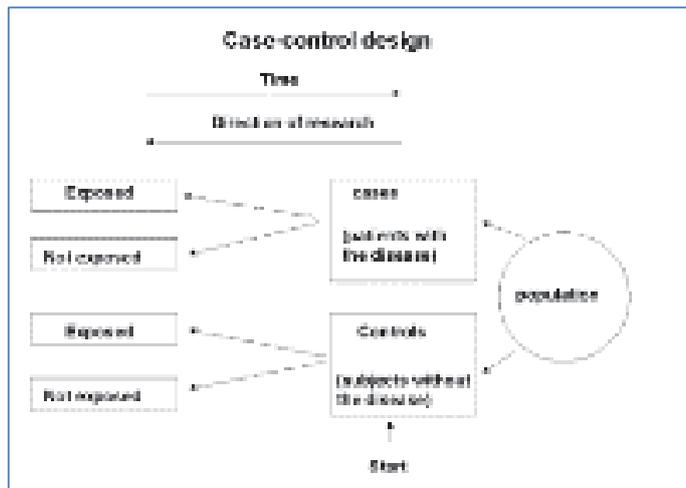
under study (which is not true for case-control studies, which are generally limited to evaluating only a single (pre-specified) disease outcome, discussed below). Cohort studies can also be performed either prospectively, like the Agricultural Health Study (AHS, <http://aghealth.nci.nih.gov/>), or retrospectively 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

<sup>9</sup> 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)

<sup>10</sup> 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<sup>11</sup>) 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 studied. For example, a cases that that have developed the disease or health outcome in an

<sup>11</sup> 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”.<sup>12</sup> 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<sup>13</sup>. 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

---

<sup>12</sup> 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.

<sup>13</sup> 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 measured 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<sup>14</sup>], 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](http://www.strobe-statement.org) and Vandembroucke 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.

---

<sup>14</sup> <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 <sup>a</sup> (Adapted from Munoz-Quezada et al., 2013; LaKind et al., 2014)**

Parameter	High	Moderate	Low
<b>Exposure assessment</b>	<p>Accurate and precise quantitative relationship with external exposure, internal dose, or target dose, possibly associated with an MOA/AOP.</p> <p>If questionnaire utilized, questionnaire and/or interview answered by subjects for chemical-specific exposure</p>	<p>Evidence exists for a relationship between biomarker in a specified matrix and external exposure, internal dose, or target dose.</p> <p>Questionnaire and/or interview for chemical-specific exposure answered by subjects or proxy individuals</p>	<p>Poor surrogate</p> <p>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</p>
<b>Outcome Assessment</b>	<p>Standardized tool, validated in study population; medical record review/diagnosis confirmation by trained staff; appropriate consideration of prevalence/incidence of cases</p>	<p>Standardized tool, not validated in population, or screening tool; or, medical record review, methods unstated</p>	<p>Selected sections of test, or maternal report, other; or, maternal/paternal self-report; unclear/no consideration for whether prevalent or incident cases are appropriate</p>
<b>Confounder control</b>	<p>Good control for important confounders relevant to scientific question, and standard confounders</p>	<p>Moderately good control confounders, standard variables, not all variables relevant for scientific question</p>	<p>Multi-variable analysis not performed no adjustments; no stratification, restriction, or matching</p>
<b>Statistical Analysis</b>	<p>Appropriate to study question and design, supported by adequate sample size, maximizing use of data, reported well (not selective)</p>	<p>Acceptable methods, questionable study power (especially sub-analyses), analytic choices that lose information, not reported clearly</p>	<p>Minimal attention to statistical analyses, comparisons not performed or described clearly</p>
<b>Risk of (other) bias (selection, differential misclassification, effect size magnification, other)</b>	<p>Major sources of other potential biases not likely present, present but analyzed, unlikely to influence magnitude and direction of the risk estimate</p>	<p>Other sources of bias present, acknowledged but not addressed in study, may influence magnitude but not direction of estimate</p>	<p>Major study biases present, unacknowledged or unaddressed in study, cannot exclude other explanations for study finding</p>

<sup>a</sup> 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
<b>Indirect</b>	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
<b>Direct</b>	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 (Niewenhuisen, 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 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)).**

<b>Biomarker Consideration</b>	<b>Tier 1</b>	<b>Tier 2</b>	<b>Tier 3</b>
<b>Exposure biomarker</b>	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
<b>Effect biomarker</b>	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).
<b>Specificity</b>	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.
<b>Method sensitivity</b>	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
<b>Biomarker stability</b>	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.
<b>Sample contamination</b>	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

<b>Biomarker Consideration</b>	<b>Tier 1</b>	<b>Tier 2</b>	<b>Tier 3</b>
	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.		
<b>Method requirements</b>	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)
<b>Matrix adjustment</b>	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<sup>15</sup>.

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

---

<sup>15</sup> 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<sup>16</sup>, 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

---

<sup>16</sup> 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 that 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<sup>17</sup>. 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<sup>18</sup>. (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<sup>19,20</sup>. 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

---

<sup>17</sup> 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).

<sup>18</sup> 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.

<sup>19</sup> The old adage that “the absence of evidence does should not be interpreted as the evidence of absence” is true here.

<sup>20</sup> 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<sup>21</sup>.

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.

---

<sup>21</sup> 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.<sup>22</sup>

**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

---

<sup>22</sup> 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](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)<sup>23</sup>, as well as other incidents reported directly to EPA. OPP has compiled the pesticide related

---

<sup>23</sup> 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**.<sup>24</sup> 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

---

<sup>24</sup> 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 in 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 the 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 21<sup>st</sup> 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, 2<sup>nd</sup> 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. 4<sup>th</sup> Edition. Saunders-Elsevier, New York.

Greenland, S. 1998. "Basic Methods for Sensitivity Analysis and External Adjustment" in Rothman, KJ and Greenland, S. Modern Epidemiology. 2<sup>nd</sup> ed. Lippincott-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. 3<sup>rd</sup> 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. 3<sup>rd</sup> 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<sup>1</sup>](#), 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*. 2<sup>nd</sup> 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*, 2<sup>nd</sup> ed. Oxford University Press, New York.

Mausner JS and Kramer S. 1985. *Epidemiology*, 2<sup>nd</sup> 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<sup>1</sup>](#), [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](http://www.nap.edu/catalog.php?record_id=18764)

Ntzani EE, Chondrogiorgi 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](http://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](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. 6<sup>th</sup> 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*. 3<sup>rd</sup> 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](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/recorddisplay.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/0/33E1AD305287588F85257D20006BE8C/\\$File/ISA\\_PREAMBLE\\_FINAL2015.PDF](https://yosemite.epa.gov/sab/sabproduct.nsf/0/33E1AD305287588F85257D20006BE8C/$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, Mulrow CD, Pocock SJ, Pool C, Schlesselman 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](#)<sup>1</sup> 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

農薬のリスク評価において  
ヒトでの疫学データと事例データを導入するための  
農薬プログラム局のフレームワーク（基本概念）

2016年12月28日

農薬プログラム局  
米国環境保護庁

---

I. 目的と範囲	3
II. 序章	4
III. 農薬のリスク評価におけるシステマティックレビュー：疫学	8
A. 問題の定式化	9
B. データ収集	10
C. データ評価	11
D. データ統合：エビデンスの重み付け (WOE)	12
E. 全体的な結論、リスク評価の推奨事項、信頼性と不確実性の分野の陳述	15
IV. 農薬のリスク評価に用いるための疫学研究のレビュー	16
A. 緒言	16
B. 疫学研究の種類	18
C. 農薬のリスク評価に用いるための疫学研究の評価	21
1. ばく露評価	25
2. 交絡因子	30
3. 統計解析	33
4. 観察研究における潜在的なバイアス	34
5. 帰無研究の解釈	35
6. 外的妥当性（一般化可能性）	36
V. ヒトでの事例調査データ	37
VI. 要約と結論	43
VII. 参考文献	44

## I. 目的と範囲

米国環境保護庁（EPA）の農薬プログラム（OPP）は、連邦殺虫剤・殺菌剤・殺鼠剤法（FIFRA）及び連邦食品・医薬品・化粧品法（FFDCA）に基づき、米国内の農薬を規制する認可プログラムである。このプログラムの一環として、OPP は毒性とばく露の多くの主要データを評価し、農薬がヒトの健康及び環境に及ぼす影響を評価している。ヒト健康影響を評価する際、EPA は、まず疫学データを含めたヒトへの影響の可能性を直接評価する情報に注目する。しかし、これまでは農薬の潜在的な毒性を報告した疫学研究はほとんどない。したがって、OPP はハザードの可能性を評価し、ヒトの健康リスクを推定するため、これまでは実験動物を用いた毒性試験に主に依存してきた。Agricultural Health Study<sup>1</sup> や National Institute of Environmental Health Sciences (NIEHS) / EPA Children's Centers<sup>2</sup> などから数多くの論文が発表されたことで、米国で実施された農薬へのばく露に関する疫学研究の有用性が高まっている。とはいえ、質の高い疫学データが存在する、あるいは充実しつつある農薬の数は、今後短期的には比較的少ないので、ほとんどの農薬の定量的リスク評価に用いるためのデータの主な情報源は試験データであることがおそらく継続するだろう。

OPP の目標は、そのような情報が入手可能な場合には、科学的に正確で明白な方法で利用することである。これを達成するために OPP は、この文書に記載されているように総合的な疫学的フレームワークを開発した。このフレームワークでは、そのような研究や科学的情報を農薬のリスク評価にどのようにして一層完全に統合することができるかを評価する際に OPP が検討する科学的に考慮すべき要点を述べている。現在の文書は、拘束力のある規則ではなく、疫学データを評価または使用するためのレビュアー向けのガイドやマニュアル、標準作業手順書であることを意図したものでも、その役割を果たすことを意図したものでもない。また、より最新で進歩した疫学的方法の完全な専門書となることも、この種の研究を解釈するために重要なニュアンスや複雑さを適切に伝えることも意図していない。したがって、因果推論及び因果関係図（Rothman ら、2009 年; Greenland and O'Rourke、2012 年）、交絡因子の特定・評価・制御に対する最新の研究法、メタ分析と異質性の評価（Borenstein ら、2009 年; Greenland and O'Rourke、2012 年）、または疫学データの感度／定量的バイアス解析（Lash ら、2009 年; Lash ら、2014 年; Ioannidis、2008 年; Greenland and Lash、2012 年; Jurek ら、2007 年）といった重要な疫学の話題は（詳細に）論じていない。これらの話題、概念及び問題点のすべては、農薬に関する疫学研究に応用できるが、この OPP フレームワーク文書では取り上げていない。代わりに、この文書では、OPP の FIFRA 及び FFDCA の機能支援のためのヒト健康リスク評価の文脈において農薬の疫学研究の評価と使用に関する全体的な概念を提供している。

---

<sup>1</sup> <https://aghealth.nih.gov/>

<sup>2</sup> <https://www.epa.gov/research-grants/niehsepa-childrens-environmental-health-and-disease-prevention-research-centers>

本文書の以前のバージョンは、2010年2月にFIFRA科学諮問パネル(SAP)によって好意的にレビューされた(USEPA、2010年、FIFRA SAP、2010年)。この文書は、SAPが推奨する改善点、パブリックコメント及び2010年以降に疫学データが利用可能であった農薬の評価から得られた知識が盛り込まれており、進捗及び必要に応じて随時更新される文書と考えるべきである。

## II. 序章

全米科学アカデミー(NAS)の米国研究評議会(NRC)による2つの報告書、「21世紀の毒性試験：ビジョンと戦略(2007年)」と「科学と政策調査(2009年)」は、共に毒性学とリスク評価の新たな方向性を示している。これら2つのNRC報告書は、毒性試験の実施方法、データの解釈方法、そして最終的に規制上の政策調査の方法に大きな変更を提唱している。具体的には、21世紀の毒性試験に関する2007年の報告書では、毒性試験、リスク評価、そして最終的な政策調査に情報を提供するために、現在の先端毒性エンドポイントの使用に焦点を当てたものから、毒性発現経路<sup>3</sup>を使用するようにシフトすることを提唱している。このアプローチは、遺伝子、タンパク質、低分子がどのように相互作用してヒト細胞の機能を維持する分子経路を形成しているかという急速に進化する科学的理解に基づいている。新しい毒性試験のパラダイムの目標は、環境因子へのばく露がこれらの経路をどのように乱すかを明らかにし、それによって有害な健康影響につながる重要な事象を連鎖的に引き起こす原因を明らかにすることである。この新しいアプローチでは、疫学研究、ヒトでの事例データベース、バイオモニタリング研究に見られるようなヒトの情報と実験毒性学的情報が重要な役割を果たすことが期待されている。具体的には、このようなヒト情報は、実際の化学物質ばく露による影響についての見識を提供し、問題の定式化やハザード/リスクの特性評価に貢献することができる。さらに、疫学データやヒトでの事例データは、追加の解析やデータ生成(例えば、*in vitro*試験や目標とする*in vivo*試験で使用するための用量やエンドポイントの選択)の指針となり、影響を受けやすい集団を特定したり、新たな健康影響を特定したり、既存の毒性学的結果を確認したりすることができる。

この毒性試験とリスク評価の新しいビジョンには、生物学的体系の多種多様なレベル(分子レベルから化学物質の影響をその発生源から最終的な健康影響や集団への影響まで考慮することを目的とした集団を基にした調査レベルまで)のデータが含まれる。このようなデータには、*in vitro*試験や*in vivo*試験から得られるデータのほかに、*in silico*モデルデータがある。疫学データと事例データを取り入れるOPPのフレームワークは、21世紀の毒性試験に関する2007年のNRCの報告書と概念的に一致しており、どちらも多種多様なデータソースから得られる最良の情報の用途はエビデンスの重み付け(WOE)解析にまとめて使用することを強調している。

---

<sup>3</sup> 毒性発現経路とは、十分に攪乱された場合に有害な健康影響を及ぼすことが予想される細胞応答経路のことである。

一般的な原則として、産業環境疫学研究は広く使用されている農薬のみを対象に実施されており、これらの農薬も科学的文献で十分に研究されている必要がある。したがって、疫学データが利用可能な多くの場合、毒性、ばく露、薬物動態（PK）、作用機序／有害性発現経路（MOA／AOP）の情報に関する主要な文献データも利用可能であると OPP は予想している。ヒトでの事例データは、より広範囲の化学物質について入手可能であるが、中には正確なデータベースを持つものと持たないものがある。ヒトでの重大な事例があり、特定の農薬の MOA／AOP や PK についてほとんど知られていない場合には、WOE 解析を用いて新たな研究分野を特定することができる。

疫学データとヒトでの事例データを組み込むための当該フレームワークにおける OPP のアプローチは、初めてのものもしくは新しいものではない。むしろ、このアプローチは、既存のツールや方法を合理的かつ論理的に拡張したものである。当該文書は、農薬リスク評価に使用する疫学データ及びヒトでの事例データを検討するための出発点として、既存のガイダンス文書及びフレームワーク（表 1）に依存している。ヒトの健康リスク評価における疫学データ及びヒトでの事例データの使用に関する当該フレームワークは、2009 年の「科学と政策調査」に関する報告書での NRC の提言や、特にリスク管理の必要性の文脈で、利用可能なものを調査、企画、検討するためのツールとして問題の定式化の使用を強調する点に関して、政府機関の最新のヒト健康リスク評価フレームワーク（USEPA、2014 年 a）と一致している。

同様に、OPP のフレームワークは、問題の定式化の重要性と生物学的体系の多種多様なレベルでの情報統合の必要性を強調した「世界保健機関／化学物質の安全性に関する国際プログラム MOA／ヒト関連(MOA／HR) フレームワーク」の最新版と一致している（Meek ら、2014 年）。MOA／HR フレームワークは、用量反応や時間的一致、ならびに生物学的な妥当性、整合性及び一貫性などの要素を考慮に入れて、Bradford Hill によって記述されたような原則を使用して、エビデンスの重み付けに基づいて確立された因果経路に沿った一連の重要な事象を特定することから始まる（Hill、1965 年）。この解析アプローチを用いることで、疫学的知見は、他のヒト情報（ヒトでの事例所見を含む）や実験的研究の文脈で評価することができ、不確実分野や今後の研究分野を特定することができる。しかし、完全に解明された MOA／AOP があることが、ヒト健康リスク評価に疫学研究を利用するための要件ではないことは注目に値する。政府機関が 2007 年及び 2009 年の NRC 報告書にある変化するアプローチの実施を引き続き進め、OPP が疫学情報とヒトでの事例情報の統合の経験を積む中で、OPP は当該フレームワークを適切に再評価し、更新していく予定である。

図 1. 有害性発現経路の概略図。Ankley ら (2010 年) より引用。

# 有害性発現経路

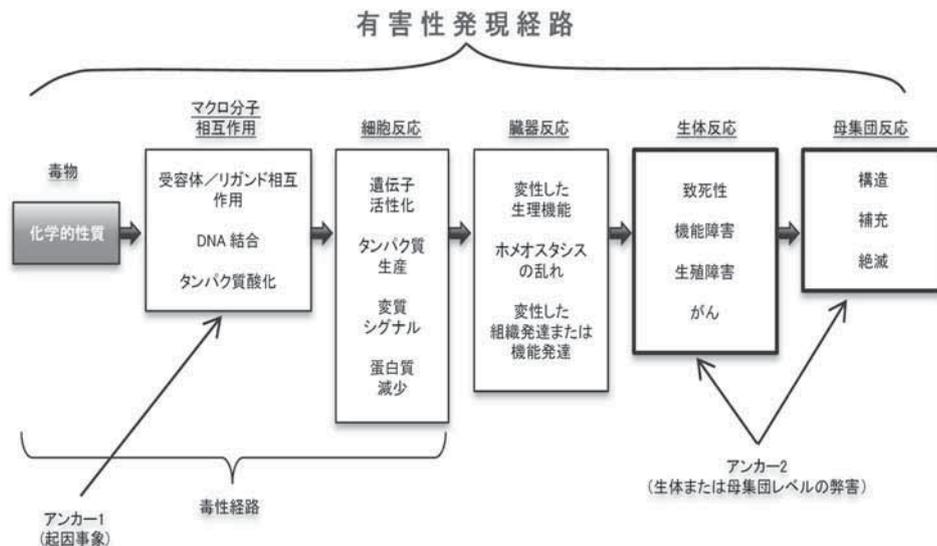


表 1. OPP が使用している主要なガイダンス文書とフレームワーク

NAS	1983: 連邦政府におけるリスクアセスメント：プロセスの管理
	1994: 科学と判断
	2007: 21 世紀の毒性試験
	2009: 科学と政策調査：リスクアセスメントの推進
	2011: ホルムアルデヒドに関する NAS 報告書
	2014: EPA の統合リスク情報システム (IRIS) プロセスの見直し
WHO/IPCS	2001～2007: 作用機序／ヒト関連のフレームワーク
	2005: 化学物質固有の補正係数 (CSAF)
	2014: 作用機序／種のコンコルダンス解析に関する WHO/IPCS フレームワークの進化と応用における新たな展開

環境保護庁	1991～2005: リスクアセスメントのためのリスクアセスメントフォーラム指針（発がん性、生殖毒性、発生毒性、神経毒性、生態毒性及びばく露評価のための指針、ベンチマーク用量モデリングのための指針、基準用量と基準濃度プロセスのレビューなど） <sup>4</sup>
	2000: リスクの判定に関する科学政策ハンドブック
	2006b. 生理的薬物動態（PBPK）モデルのリスク評価への応用のためのアプローチとその裏付けとなるデータ
	2014a. 政策調査に情報提供するためのヒト健康リスク評価のフレームワーク
	2014b. 異種間・同種間の推定のためのデータ由来の推定係数を開発するための定量的データの適用ガイダンス
農薬プログラム	2001: 総合的なリスク評価
	2001 と 2002 年: 累積リスク評価
経済協力開発機構	2013: 有害性発現経路の開発と評価に関する経済協力開発機構（Organisation for Economic Co-operation and Development）ガイダンス文書

ヒトの情報源は他にもあるが、このフレームワークの意図は、ヒトのリスク評価における**疫学データ**と**ヒトでの事例データ**の解釈と利用にあり、他にもあるヒトの情報源については当該文書で深く触れていない。特に、当該文書では、意図的なヒトを対象としたばく露に関する農薬の研究<sup>5</sup>や、農業従事者の活動に伴う経皮や吸入によるばく露量を測定するために行われた研究<sup>6,7</sup>については、広く論じていない。

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

<sup>5</sup> このような研究の実施及び OPP がこのような研究から得られたデータに依存することは、EPA の研究対象者保護規則（40 CFR Part 26）によって管理されている。その他に、この規則は妊婦、授乳婦、子供への意図的なばく露に関する研究を禁じ、EPA-OPP やヒト研究評価委員会（HSRB）による新規研究計画の事前評価、ならびに研究成果報告の評価を求めている。

<sup>6</sup> 過去数年間に OPP は、農業従事者のばく露評価に使用されるデータとアプローチを改善するために、農業従事者を対象とした既存の観察研究を広範囲に評価してきた。それらの評価は（[http://www.epa.gov/scipoly/sap/meetings/2007/010907\\_mtg.htm](http://www.epa.gov/scipoly/sap/meetings/2007/010907_mtg.htm)）で閲覧できる。

<sup>7</sup> OPP が農業従事者ばく露研究をどのように実施し、成果を利用するかについての追加情報は、PPP-48「農薬とヒト健康リスク評価：Policy, Processes, and Procedures」（<https://www.extension.purdue.edu/extmedia/PPP/PPP-48.pdf> で入手可能）を参照のこと。

### III. 農薬のリスク評価におけるシステマティックレビュー：疫学

近年、NRC は、規制上の政策調査に情報を提供するための化学物質固有のリスク評価をサポートする科学的文献レビューの明白性を高めるために、システマティックレビュープロセスに移行するよう政府機関に奨励している（NRC、2011 年、2014 年）。NRC はシステマティックレビューを「特定の疑問に焦点を当て、明確で事前に指定された科学的方法を用いて、類似しているが別個の研究の結果を特定、選択、評価、要約する科学的調査」と定義している（NRC、2014 年）。NRC の勧告に沿って、化学物質安全・公害防止局（OCSPP）は、政策調査を支える科学的データの収集、評価、統合のための標準的な方法に依存する、目的にかなったシステマティックレビューを採用している。

NRC によると、システマティックレビューには、「明白性があり明確に文書化された方法、関連するすべての文献の一貫した厳しい評価、エビデンスの強さを評価するための標準化されたアプローチの適用、明白で一貫性のある総括的用語（NRC、2014 年）」といういくつかの共通要素がある。近年、いくつかの集団（Rooney ら、2014 年； Woodruff and Sutton、2014 年； Hartung、2010 年）が、環境保健科学で使用するためのシステマティックレビューのアプローチを発表している。システマティックレビューでの OCSPP のアプローチは、Cochrane Handbook for Systematic Reviews of Interventions for evidence-based medicine に明記されている原則と、Grading of Recommendations Assessment, Development and Evaluation (GRADE) の原則と一致している。GRADE ガイドラインは、国立環境保健科学研究所（NIEHS）健康評価・翻訳局（OHAT）（Rooney ら、2014 年）とカリフォルニア大学サンディエゴ校（Woodruff and Sutton、2014 年）が開発した環境保健科学のためのシステマティックレビューアプローチで使用されているものである。コクラン・ハンドブックによると、システマティックレビューの主な特性は以下の通りである。

- 目的が明確に示されており、研究の適格性基準があらかじめ定義されていること。
- 明示的で再現性のある方法論。
- 適格性基準を満たすすべての研究を特定するための系統的検索。
- 特定された研究から得られた知見の妥当性の評価。
- 収録された研究の特性と知見を体系的に提示し、総合的にまとめたもの。

上記の各アプローチは共通のテーマとワークフローを共有しており、科学的な文脈（例えば、問題の定式化やプロトコル）の記述から始まり、明確な検索戦略の方法を用いた文献レビュー、研究の質の解析（しばしばバイアスのリスクと呼ばれる）、エビデンス全体の質の評価（例えば、統合）、そして最終的に結論に至るまでの流れを示している。それぞれのアプローチでは、科学文献の包含/除外、研究の質の評価、研究の質（例：高、中、低）の報告のための透明性のある事前に設定された基準を推奨している。各アプローチでは、結論の基礎となるデータ統合のための事前に定められたツールを推奨している。

これまでのところ、リスクアセスメントに関係する人々の間では、システマティックレビューのための単一の命名法は合意されておらず、OCSPP は、より広範な経験が積まれるにつれて用語が進化していくことを期待している。OCSPP は、システマティックレビューのプロセスとワークフローを、問題の定式化から始まり、データ収集、データ評価、データ統合、そして重要なデータギャップを特定した結果の要約と考えている。

科学的な解析は、新しい知識が得られるたびに、繰り返し行われることが多い。

## A. 問題の定式化

NRC の報告書「*Science and Decisions-Advancing Risk Assessment (科学と意思決定- リスク評価の促進)*」において、全米科学アカデミー (NAS) は EPA に対して、政策立案者に有用なリスク評価と関連する科学的解析を開発するよう勧告した。この目標達成のために NRC は、政府機関に対してリスク評価の開発において一層幅広く問題の定式化を用いるよう勧告した。NRC の勧告を受けて、USEPA は「ヒト健康リスク評価フレームワーク」(USEPA、2014 年) を公表し、問題の定式化の重要性を強調している。問題の定式化は、科学者とリスク管理者の間の初期対話を伴うものであり、科学的解析のための規制上の背景を提供し、解析の範囲を定義するのに役立つ。問題の定式化は、評価の規制、政策調査及び政策の背景から導き出され、評価の技術的なアプローチの情報を提供し、考慮すべき主要な要因を体系的に特定するものである。このように、各システマティックレビューの複雑さと範囲は、異なるリスク評価の文脈によってさまざまである。言い換えれば、OCSPP のシステマティックレビューは、問題設定から事前に設定された範囲と目的に基づいた「目的にかなったもの」(NRC、2009 年) として実施される。

問題の定式化には、利用可能な情報とともにデータや科学的情報の主要なギャップを考慮することが含まれる。OPP は、ばく露経路と潜在的な健康影響を特定するためのツールとして問題の定式化を用いており、適切な方法、情報源及び科学的解析のアプローチとともに使用している。欠損データが評価において重要な場合、情報 (例えば、必要とされる試験) を得るための最良の選択肢が議論される。ピアレビュープロセスが特定され、評価を完了するためのタイムラインが定義される。

システマティックレビューは、利用可能な情報を整理し、解析に対する規制目的に関する情報のギャップを特定するための明白なツールを提供する。このように、問題の設定では、科学的解析の規制上の文脈が説明され、それによって科学的文献の収集と評価の範囲と目的が定義される。問題の設定における考慮すべき事項は、集団またはライフステージ、ばく露経路 (例: 経路、期間、頻度) 及び/または *in vitro* または *in vivo* の実験室での研究、疫学またはヒトの事故による研究から特定された目的的健康結果、ならびに利用可能な資源及び規制の時間枠に関することであろう。疫学情報及びヒトでの事例情報を考慮する文脈では、試験の質、試験デザイン及び不確実性の初期評価が考慮される。

問題の設定で考慮されるハザード評価に関連する主な科学的課題は以下の通りである。ばく露に関連する影響とは何か？これらの影響に関連する MOA/AOPs はどのようなものか？影響の時間的側面は？影響を受けやすい集団があるのか、あるとすれば、その集団は誰であり、どのような要因によって影響を受けやすいのか。実験動物とヒトの間で PK や薬力学 (PD) に違いはあるか？また、ばく露情報は問題の設定においても評価される。問題の設定で考慮されるばく露評価に関連する主な科学的課題は以下の通りである。農薬はどのように使用されるのか？ばく露の関連する使用場所のすべてはどのようなものか？ヒトはどのような化学物質にばく露されるのか？ばく露の方法、期間、頻度はどのようなものか？誰がばく露される可能性があるか？ばく露は異なる集団（例えば、年齢や行動パターンのため）に異なるリスクをもたらすのか？疫学データの具体的なケースでは、このレビューでは、研究仮説、研究デザイン（すなわち、サンプルサイズ、十分な対照、測定方法の質など）、ばく露量/ばく露濃度、統計解析及び結論を含むが、これらに限定されない様々な要因を考慮して検討する。

## B. データ収集

システマティックレビューのデータ収集段階では、公開科学文献や農薬登録のために提出された研究論文など、様々な公開・未公開の情報源から利用可能な情報を収集する。OPP では、経済協力開発機構 (OECD) 試験ガイドライン、OCSPP 調和試験ガイドライン及び農薬ガイドライン (OPP ガイドライン) に基づいて収集されたデータをレビューしている。これらのガイドライン研究論文は、主に提出された研究の社内データベースから収集され、そのような社内データベースの検索によって発見される。

疫学の場合、ほとんどの研究は公開されている科学文献に掲載されていることが予想される。場合によっては補足的な解析や情報が得られることもあるが、研究者との対話によって、疫学研究を理解し解釈する上で、原著論文には掲載されていない追加の重要な情報が得られることもある。ヒトでの事例の情報源はセクション IV にまとめられている。

公開文献検索戦略では、特定の基準を用いて、公開されている科学文献や未発表の情報源から健康影響情報を検索する。最も適切な情報源/データベースを特定して選択し、分類コード、医学的主題の見出し及び/またはキーワードを利用した最もリソース効率の良い戦略を調査した後、文献の検索が行われる。科学的評価の複雑さに応じて、文献管理責任者のサポートが必要な場合もあれば、必要でない場合もある。ヒトの健康に関する文献検索の目標は、文献検索プロセスの適切な文書を提供することにより、信頼性が高く再現性のある文献検索を行うことである。有用な研究を検索するには、以下の手順で行う。

- 科学的解析の目的と包括基準が確立されている。
- 用語/キーワードや MeSH (Medical Subject Heading) 用語の組み合わせと、それらのブール考案の組み合わせ (AND; OR; NOT) を使用し、文書化する。

必要に応じて、著者名、タイトル、キーワード、主題見出しによる高度な検索やフィールド検索を行うこともできる。データベースの構造を知り、特定のデータベースに対して別の検索戦略を使用することは、有用な研究論文を検索する際に役立つ。最初の包括的な検索に加えて、文献リストを更新するために定期的な検索を行うこともある。

- すべてのデータベースのすべての検索が再現可能であることを保証するために、検索戦略は検索の日付を含めて文書化されている。
- 検索された論文のリファレンスリスト<sup>2</sup>は、追加の文脈を調べたり、最初の検索では発見されなかった論文を探したりするために調査されている。
- 異なるデータベースから検索された論文を組み合わせて重複を除去した後、利用可能なタイトルと要約を選別する。タイトルと要約から関連性が判断できなかった論文については、その論文を検索し、さらなる再検討を行う。
- 最初の選別の後、有用性のない論文（除外基準）-例えば、持論の論文、英語で記述されていない研究論文、要約のみで構成された論文は、除外される。追加の除外基準は個別に特定される。すべての除外基準は文書化されている。残りの論文は、有害な健康影響が認められなかったものであっても、検討と評価のために含まれる。

### C. データ評価

データ評価段階では、データの質が検討され、そのようなデータの有用性について結論が出される。研究の質は、報告された結果が正しいという全体的な信頼性を反映する（Balslem ら、2011年）。このように、研究の質には以下のようなものがある。

- 報告の質（研究の報告がどの程度に十分であるか、または完全であるか）。
- 試験のデザインと実施に基づいて、試験結果がどの程度に信頼性のあるものか。
- さらに、その研究が評価下の主題にどの程度に十分な対応しているか（Rooney ら、2014年）。

研究の質は、まず個々の研究基盤で検討され、その質が判断される。例えば、不完全に実施されたコホート研究よりも、適切に実施された症例対照研究の方が信頼性は高い場合がある。科学的知見の信頼性は、しばしばバイアスのリスクと呼ばれ、試験のデザインと実施に関連する特定の分野について、あらかじめ定められた基準を用いて評価される（表2参照）。

OPP は当初、「ヒトの健康リスク評価を支援するための公表されている科学文献の毒性研究の検討と使用に関するガイダンス」（Guidance for Considering and Using Open Literature Toxicity Studies to Support Human Health Risk Assessment）（USEPA、2012年）と呼ばれる公表されている科学文献の考えの使用に関するガイダンスを作成し、一般的にはこのガイダンスに従っている。しかし、リスク評価におけるシステムティックレビューの促進に伴い、この文献ガイダンスの一部の側面は将来的に更新が必要になる可能性がある。

データの質についての結論が出され、結論の記述またはカテゴリー（例えば、許容可／許容不可；低、中、高）で記述される。

農薬に関する疫学研究の評価に用いられる具体的な考えは、以下のセクション III.C に記載されている。データレビューの一環として、研究の簡潔なレビュー文書を作成する。このレビュー文書には、研究のデザイン、結果、結論と、研究の長所と短所が記載されている。疫学的ばく露評価の質は、疫学的データがリスク評価においてどのような役割を果たすかを調査する上で重要な要素である。そのため、疫学的ばく露評価で使用された仮定と、その仮定が疫学的知見の解釈と一般化の可能性にどの程度影響を与えるかを十分に特性評価することが重要である。疫学的ばく露評価には、過去及び現在のばく露パターン（例えば、ばく露集団、ばく露経路、ばく露方法及びばく露レベル）の検討が含まれ、また、使用パターンの大幅な変化（例えば、リスク軽減措置や新たな使用パターン）が含まれている場合がある。メタ分析の評価に関しては、Stroup ら（2000 年）によって報告された「疫学における観察研究のメタアナリシス」*Meta-analysis Of Observational Studies in Epidemiology (MOOSE)* のガイドラインが、メタ分析の質の評価や解釈に有用である。

#### D. データ統合：エビデンスの重み付け (WOE)

OPP のヒト健康影響評価では、ヒトでの調査／疫学、バイオモニタリングデータ、*in vitro* の毒性試験及び *in vivo* の毒性試験、MOA/AOP 情報、薬物動態試験、構造活性関係 (SAR) などを含む（但し、これらに限定されない）利用可能なすべての関連データを考慮する必要がある。さまざまな種類のハザードデータを収集し、関連する各試験の完全な評価を実施して記録した後、次のステップは複数のエビデンスを統合することである。

データ統合は、関連するハザードについて利用可能なネガティブデータとポジティブデータを総合的に判断するという原則に基づいている。OPP は、疫学データ及びヒトでの事例データの評価に WOE 解析を使用しているが、結論は、特定の研究に依拠するのではなく、情報の優越性に基づいて出される。OPP は、異なる情報源からの情報を整理して統合するためのツールとして、MOA/HR のような修正された Bradford Hill 基準を使用している (Hill, 1965 年; U.S. EPA, 1999 年、2005 年; Sonich-Mullin ら、2001 年; Meek ら、2003 年; Seed ら、2005 年; OECD AOP Wiki Users Handbook<sup>8</sup>)。Hill 基準は、チェックボックス方式ではなくて、エビデンスの総合的評価の検討を示唆していることに注目することが重要である。また、MOA/AOP が完全に解明されていることは、ヒト健康リスク評価に疫学研究を利用するための必須条件ではない。しかし、完全に解明された MOA/AOP がいない場合でも、メカニズムデータの収集と評価は、生物学的な妥当性の裏付けとなり、生体組織の感受性、生物種、性別、ライフステージ、またはその他の要因での違いを説明するのに役立つ可能性がある。MOA/HR は、厳密ではない情報を整理するための基盤を提供する柔軟なツールである。

---

<sup>8</sup> [https://aopwiki.org/wiki/index.php/Main\\_Page#OECD\\_User\\_Handbook](https://aopwiki.org/wiki/index.php/Main_Page#OECD_User_Handbook)

これは因果関係の評価、複数の科学的根拠に基づく情報の統合、データのギャップや今後の研究分野の特定などの様々な目的に有用なツールとなる柔軟性を備えている。この解析では、疫学的知見やヒトでの事例データは、生物学的な妥当性を評価するために、他のヒト情報や試験の文脈で評価され、不確実分野や今後の研究分野を特定することができる。Bradford Hill の側面が WOE 評価でどのように考慮されるかを説明するために、OPP は、EPA の全米環境大気質基準 (NAAQS) のレビューのための科学的基盤となる統合科学評価 (ISA) の EPA 前文に記載されている用語の定義を使用している (USEPA、2015 年)。

- **重要事象。** 特定の健康影響について MOA/AOP が確立されている場合には、MOA/AOP の基礎となる重要事象 (すなわち、測定可能なパラメータ) をそれぞれ明確に記述する。重要事象を示すデータは、*in vitro* または *in vivo* のデータソース (ヒトまたは動物) の組み合わせから得られる可能性がある。これらの重要事象は、PK 事象と PD 事象の組み合わせである。しかし、完全に解明された MOA/AOP を利用できることは、ヒト健康リスク評価に疫学研究を利用するための必須条件ではないことは注目に値する。
- **生物学的勾配/ばく露-反応/用量-反応の一致と関係。** ISAs の前文では、「疫学の文脈では、よく特徴づけられたばく露-反応関係 (例: ばく露量が多いほど影響が増加する) は、特にばく露期間 (例: ばく露時間が長くなると影響が増加する) についても因果関係を強く示唆する」 (USEPA, 2015) と指摘している。MOA/AOP が明らかである場合、各重要事象について用量反応関係が特定される。用量反応関係は、重要事象間で比較される。場合によっては、初期の重要事象の方が後期の重要事象よりも感度が高い場合がある。他の場合では、重要事象は類似した量反応曲線を共有しているかもしれない。
- **時間的関連付け。** 薬剤の投与と影響の出現との間に時間的な順序があるというエビデンスは、因果関係を支持するもう 1 つの根拠となる (USEPA、2015 年)。ISAs の前文では、「ばく露の変化が健康影響の発現または発生頻度の変化をもたらすことが判明した場合には、『自然実験』を通じて因果関係を示す強固なエビデンスを提供することができる」と指摘している。

この解析では、影響の整合性を確保するために急速に発現する重要事象 (例えば、ばく露後数分以内に発現する活性代謝物への代謝) と長い継続期間の後に発現する重要事象 (例えば、腫瘍の発生) を考慮している。疫学データを検討する際には、ばく露と健康影響の間の時間的関係が考慮される。

- **強固性、一致性、特異度。**

**一致性。**複数の独立した研究でリスクの上昇が観察された場合に、因果関係の推測は強固になる。結果の再現性は因果関係を示す最も強力な根拠の一つである。統計的な有意性は、効果の有無を判断する唯一の基準ではない。研究間で不一致な結果がある場合は、ばく露の違い、交絡因子、研究の検出力などの理由が考えられる（USEPA、2015年）。

複数の研究における結果の一貫性は、複数の独立した研究における効果または関連付けの繰り返しの観察から得られる。さらに、異なる状況下での異なる母集団における結果の再現性によっても支持される。しかし、独立した調査の間での結果の不一致は、調査方法の違い、ランダムエラー、ばく露、交絡因子、または研究の検出力の違いによって説明される可能性があり、因果関係を否定するために使用することはできない（USEPA、2015年）。

**観察された関連付けの強固性。**大規模で正確なリスクの発見は、関連付けが偶然、偏りまたはその他の要因によるものではない可能性が高いという確信を得られる。しかし、影響の推定が小さい場合は、母集団における実質的な影響を表す場合もあれば、そうでない場合もあることに注意が必要である（USEPA、2015年）。

**観察された関連付けの特異度。**特定の結果をばく露に結びつけるエビデンスは、因果関係を示す強い根拠となる。しかし、環境ばく露が結果の発生を必ず予測できることは稀であり、結果には複数の原因がある可能性を認識しなければならない（USEPA、2015年）。

- **生物学的な妥当性と整合性。**

**整合性。**1つのエビデンス（例えば、疫学的に管理されたヒトのばく露研究、毒性試験、または生態毒性試験）からの因果関係の推測は、関連付けの因果関係の解釈を支持する他のエビデンスによって強化されることがある。1つの科学的エビデンスグループの中で、さまざまな分野にわたるエビデンス、複数の試験デザインにわたるエビデンス、または健康影響に関連するエビデンスから、効果を示す一貫性があるかもしれない（USEPA、2015年）。

動物でのデータとヒトでのデータが、定量的にも定性的にも同様の毒性プロファイルを示している場合、ヒト健康リスク評価の信頼性は高い。一方、動物でのデータとヒトでのデータが定性的には類似した毒性プロファイルを示している場合、定量的な相違が観察される場合もある。例えば、ある化学物質が動物とヒトでは同じMOA/AOPを示している場合、用量反応特性に種差がある場合がある。このような用量反応性の相違は、生体組織薬量測定（すなわちPK）に起因するか、または異なる反応特性（すなわちPD）に起因する可能性がある。一方、動物でのデータとヒトでのデータは、場合によっては質的に異なる結果を示すことがある。このような状況から、種、性別及びライフステージの違いや潜在的な感受性などの要因を考慮するために、入手可能なすべての情報を透明性のある包括的な方法で十分かつ客観的に評価する必要性が明らかになった。

**生物学的妥当性。** 因果関係の推測は、生物学的に妥当なメカニズムを示す試験またはその他の情報源からの結果によって強化される。提案されたメカニズムは、実験的エビデンスに基づいており、ある物質へのばく露を所定の効果に結びつけるもので、因果関係を支持する重要な情報源である（USEPA、2015年）。

同様に、多くの構造的類似物の1つとしての化学物質のMOA/AOPに関する情報は、因果関係の可能性が高いかどうかの判断に役立てることができる。構造活性相関及び薬剤の構造的類似物に関する情報は、関連性が因果関係にあるかどうかについての洞察が提供される（USEPA、2015年）。

EPA のがんガイドライン（2005年）は以下の通りに指摘している。

「疫学研究で観察された関連付けの生物学的妥当性の評価は、潜在的な作用機序（MOA）の特定に関連する毒性学的エビデンスとばく露関連因子の両方を考慮したものである。同様に、疫学的文献で報告された健康影響の関連付けの整合性を考慮することは、毒性試験や疫学研究で評価された生物学的マーカーの特性に関する情報を幅広く考慮することを示している [p.39] 。

しかし、このガイドラインはさらに、「しかしながら、メカニズムデータの欠如は因果関係を否定する理由にはならない [p.41] 」と述べている。このように、確立されたMOA/AOPの欠如は、疫学データを使用する際に必要な知識ではなく、確立されたMOA/AOPがなくても疫学的関連付けは有効である可能性があり、また、潜在的なMOA/AOPについて洞察も得られる可能性がある。

- ・ **不確実性。** 不確実性は WOE で明白かつ客観的に述べられている。

#### **E. 全体的な結論、リスク評価の推奨事項、信頼性と不確実性の分野の陳述**

エビデンスの要約、エビデンスの重み付けに使用した手順または方法、WOE の結論または勧告の根拠、不確実性及び更なる研究分野を文書化することが重要である。また、リスク評価における疫学的データまたはヒト由来のデータの役割についても推奨されている。一般的に、OPP はヒトでの事例情報を定量的リスク評価には使用しないが、その代わりに、新たなリスク評価や新たなリスク管理措置の必要性を示すこと、リスク軽減措置が実施された後の成功度を評価すること及び可能性のある施行活動を対象とすることなど、リスク評価/リスク管理の活動を対象とする。

ヒトでの事例データの役割がより限定的であるのとは対照的に、疫学研究は様々な方法でリスク評価の複数の構成要素に情報を提供するのに役立つ可能性を持っている。正確なばく露評価を行った質の高い研究は、リスク指標を定量的に推定するために使用することができる。

また、疫学的研究で報告された結果は、*in vitro* 試験や動物試験で見られた結果と定性的に比較して動物試験の結果のヒトへの関連性を評価することもでき（Hertz-Picciotto、1995年）、疫学的結果の生物学的妥当性を評価することにも有用である。提案されている WOE 解析の最後の部分では、全体的な結論と信頼性と不確実性の分野の記述が行われる。このセクションでは、追加研究の分野も特定している。本セクションでは、規制値の情報源及び（必要に応じて）種間及びヒトの間での外挿法での推定のための適切なアプローチを推奨している。

#### IV. 農薬のリスク評価に用いるための疫学研究のレビュー

##### A. 緒言

疫学は、化学的、物理的、または生物学的要因へのばく露と集団の健康状態との関係を特定し、評価することを目指す科学である（Boyesら、2007年）。疫学とは、「集団における疾病の分布と、その分布に影響を与えたり、調査したりする要因の研究」と定義されている（Gordis、2009年）。より広義には、「特定の集団における健康に関連する事象、状態、プロセスの発生と分布の研究であり、そのようなプロセスに影響を与える調査要因の研究と、この知識に関連する健康問題の制御に応用することを含む」（Porta、2014年）と考えられている。多くの疫学研究の目的は、病気の発生に対する潜在的な原因の影響を有効かつ正確に推定することである。疫学の主な目的は、他の科学と同様に、原因と結果を明らかにすることであり、別の言い方をすれば、病気や健康影響の病因とそれに関連する可能性のあるリスク因子を特定することである。Calderon（2000年）は、このような研究の4つの主要な用途を説明している。1）集団の健康状態を記述し、疾病やばく露頻度の重要な時間的傾向を発見すること、2）特定の疾病や傾向に関連する因子を特定することで疾病の発生を説明すること、3）特定の集団における疾病の発生数や健康状態の分布を予測すること、4）環境やヒト健康に悪影響を及ぼす因子を特定することで集団の健康状態を改善すること、である。農薬の場合、疫学は、一般集団と、職業上ばく露される農業従事者や農薬散布者などの特定の小集団におけるばく露と健康への悪影響との関係に焦点を当てている。

疫学研究は、様々な方法でリスク評価の複数の要素に情報を提供するのに役立つ可能性がある。正確なばく露評価を行った質の高い研究は、リスクを定量的に推定するために、あるいはオッズ比やリスク比のような適切なリスク・サロゲート（代理）を用いることができる。しかし、農薬や農薬ばく露を扱う疫学研究の多くは、規模、範囲、ばく露評価、またはデータ解析になんらかの制限があり、それが定量的リスク評価への十分な利用を妨げている（Ntzaniら、2013年）。

米国での農薬使用は、ここ数十年で大きく変化した。使用が変化すると、農業従事者へのばく露も変化する。農薬使用の変化は、EPA によるリスク軽減措置、抵抗性管理活動、新しい化学物質の導入、遺伝子組換え作物の使用量の増加によって生じている。疫学研究を解釈し、最終的には定量的リスク評価にそのような研究を使用するかどうかを調査する際には、ばく露におけるこれらの大きな変化を考慮しなければならない。そうであっても、疫学研究は、実験的動物研究から得られたエビデンスと比較して、そのような値を導き出す際の仮定を特徴づけるために使用することが可能である。また、疫学研究で報告された結果を *in vitro* 試験や動物試験で得られた結果と定性的に比較して、動物実験で得られた結果の生物学的妥当性や人間との関連性を評価する場合もある (Hertz-Picciotto, 1995 年)。疫学研究で得られたようなヒトでの情報は、NRC (2007 年) が推奨する毒性試験の新しいビジョンにおいて重要な役割を果たす可能性があることと期待されている。具体的には、疫学研究は、ヒトの実社会での化学物質ばく露から生じる可能性のある健康影響についての知見を提供し、その結果、問題の設定やハザード/リスクの特性評価に貢献することができる。ヒトでの情報は、追加的な研究 (例えば、*in vitro* 試験や目標とする *in vivo* 試験で使用するための用量やエンドポイントの選択) の指針となり、今後の研究で調査される新たな健康影響や感受性を特定することができる。

動物実験から得られた実験データがハザードの特性評価のための主要な情報源となる場合、潜在的な不確実性の原因の一つとして、動物モデルのヒトへの関連性が挙げられる。これに反するデータがない場合には、動物の結果はヒトとの関連性があると想定される。さらに EPA はヒトは実験動物よりも感受性が高いと仮定しているが、それを裏付けるデータがない。実際には、ヒトは他の動物種よりも農薬に対して感受性が高い場合も低い場合もある。疫学データ及びヒトでの事例データは、科学的な情報を提供し、外挿法での推定に関連する不確実性を伝えるための裏付けとなり得る。母集団の変動性に関しては、動物試験よりも疫学研究の方が潜在的な変動性の特徴をよく表している。具体的には、疫学データは、ヒトの集団に固有の遺伝的多様性と変動性を含むため、実験動物よりも環境化学物質に対する実際の集団での反応をよりよく説明し、代表することができる (Calderon, 2000 年)。

用量反応の特性評価に関して、動物の毒性試験には、広範囲のばく露レベルをカバーするように試験デザインできるという利点がある。しかし、動物の毒性試験では、一般的に環境中でのばく露レベルよりもはるかに高い (時には桁違いの) ばく露レベルが用いられている。動物試験では、このような高いばく露レベルがあるため、高用量から低用量の外挿法での推定が必要である。この外挿法での推定はリスク評価に不確実性をもたらす。疫学研究データやヒトでの事例データは、ヒトの実社会でのばく露を含んでいるため、多くの場合、高用量の外挿法での推定は必要ないと考えられる。疫学研究は低用量から高用量までの広い範囲のばく露が最も有用である。

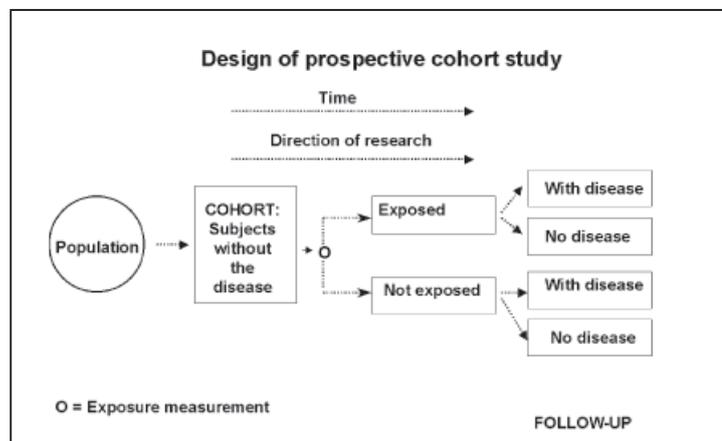
動物試験では、ヒトが経験するばく露の時間、程度、期間、ばく露経路及びばく露の変動性を再現するものではない (Calderon, 2000 年)。ヒトでのばく露は、食品、水、空気、屋内外の環境を含む多岐のばく露経路を介して行われることが多い。対照的に、制御された実験室での研究では、通常、単一のばく露経路である。

さらに、ほとんどの動物試験では単一の化学的ストレス因子を用いているのに対し、ヒトでは複数の化学物質及び／または非化学的ストレス因子へのばく露を同時に受けている可能性がある。一方で、疫学研究における複数の化学物質へのばく露は、疫学研究の結果が単一の農薬に帰属させようとする場合に問題となり得る。一方で、疫学研究は実社会のばく露を考慮しており、実験的アプローチと併せて考慮することで、実験環境では評価が困難な複数の化学物質ばく露に関連した問題に対処するのに役立つ可能性がある。

## B. 疫学研究の種類

観察疫学研究の主なタイプを、その長所と短所を考慮しながら以下に簡単に説明する (Lilienfeld and Lilienfeld, 1979 年; Mausner and Kramer, 1985 年; Kelsey ら、1996 年; Rothman and Greenland, 2012 年; Paddle and Harrington, 2000 年; USEPA, 2005 年; Purdue Pesticide Programs, PPP-43)。

**コホート研究**では、共通の特性を共有する人々の集団 (コホート) から始まり、病気の発生が記録されている長期の追跡期間にわたって健康状態を評価を行う (van den Brandt ら (2002 年) から引用の図を参照)。共通の特性は、しばしば「リスク因子」(ばく露など)<sup>9</sup>の有無である。このような研究では疾病発生率の違いを調べるために、「ばく露者」と「非ばく露者」の間の疾病発生率の違いを特定し、時間をかけて調査する<sup>10</sup>。次に、この疾病発生率の差を調査して、「ばく露群」と「非ばく露群」の間で疾病発生率が異なるかどうかを判断する。コホート研究は、複数の疾病発生を同時に評価する能力を持っている (これは症例対照研究では当てはまらない。症例対照研究では、一般的には後述するように単一の (事前に指定された) 疾病発生のみを評価することに限定される)。



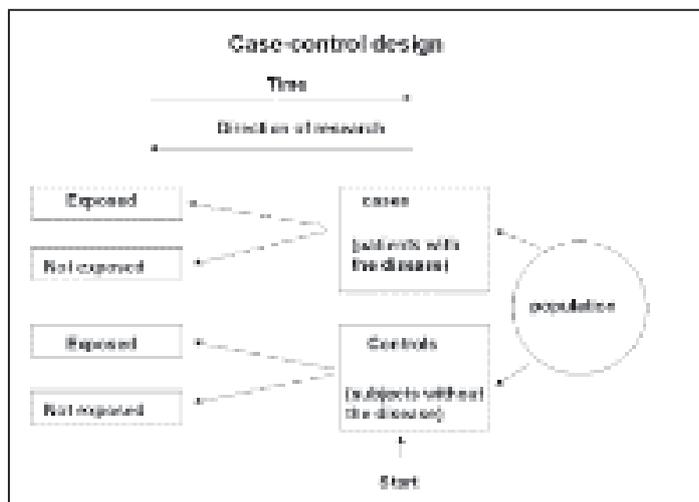
コホート研究は、Agricultural Health Study (AHS, <http://aghealth.nci.nih.gov/>) のように前向きで実施することもできるし、過去の記録から後ろ向きに実施することもできる。前向きコホート研究は、現在の時点から将来の時点までの人々の集団に焦点を当てている。後ろ向きコホート研究は、過去のある時点でばく露された集団に焦点を当て、ばく露が発生した後の疾病率を比較するものである (一般的には、一人 (個人) 単位で利用できる既存のばく露データベース (または記録) を利用)。

<sup>9</sup> コホート研究では、ばく露はしばしばばく露者と非ばく露者に二分されるが、ばく露は定量的尺度 (例えば、連続的尺度または定量値) で測定することも可能である。

<sup>10</sup> コホート研究では、一般的に疾病発生率の違いを研究しているが、これらの研究には、出生時の体重、知能、血圧など、関係のある他の限局的な結果も含まれる。

前向きコホート研究は、特に希少疾病の場合、実施に比較的時間がかかり、費用がかかることがあり、研究対象となる被験者の数が多いことが必要である。重要なことは、質の高いデータを収集するためには、長期間にわたって多大な資源と専門スタッフが必要であるということである。

**症例対照研究**は、特定の疾病を持つ（症例）個人と、そうでない（対照）個人の集団を特定し、ばく露（一般的には過去<sup>11</sup>のばく露）に関して比較することで、関係ある疾病を持つ人が、関係ある薬剤あるいは要因にばく露されていたか否かを判断する研究である。すなわち、症例対照研究の解析では、症例において関係ある薬剤または要因のばく露頻度と、対照群におけるばく露頻度を対比させて、これらが異なるかどうか、要するに関連があるかどうかを判断する。症例対照研究では、一般的に、疾病状態（すなわち、疾病を持つ症例と疾病を持たない対照群）を決定することが、通常、ばく露状態の調査に先立って行われる（van den Brandt ら、（2002年）から引用の図を参照）。この研究デザインは症例対照研究に選択された時点ですでに疾病が発生しているの、希な疾病や潜伏期間の長い疾病の研究で特に有用であり、多くのばく露と関係ある特定の（事前に指定された）疾病発生との関連性を評価するために利用することができる。また、症例対照研究は、その疾病に罹患している個人を対象としているので、コホート研究に比べて被験者数が少なく、比較的短い期間で終了することができる。症例対照研究の問題は、適切な対照群の選択と、疾病が診断されるずっと前に発生した可能性のあるばく露の評価である（Rothman, 2012年; Wacholder ら、1992年 a; Wacholder ら、1992年 b; Wacholder ら、1992年 c; Shultz and Grimes、2002年; Grimes and Schultz、2005年）。症例対照研究は、特に" 想起バイアス"の影響を受けやすい。これは、罹患した人は、対照あるいは健康人と比べてばく露や出来事を違った形で（一般的にはよく）覚えている可能性がある。



また、症例対照研究は、その疾病に罹患している個人を対象としているので、コホート研究に比べて被験者数が少なく、比較的短い期間で終了することができる。症例対照研究の問題は、適切な対照群の選択と、疾病が診断されるずっと前に発生した可能性のあるばく露の評価である（Rothman, 2012年; Wacholder ら、1992年 a; Wacholder ら、1992年 b; Wacholder ら、1992年 c; Shultz and Grimes、2002年; Grimes and Schultz、2005年）。症例対照研究は、特に" 想起バイアス"の影響を受けやすい。これは、罹患した人は、対照あるいは健康人と比べてばく露や出来事を違った形で（一般的にはよく）覚えている可能性がある。

**コホート内症例対照研究**は、ハイブリッドデザインの一例であり、コホート研究と症例対照研究の要素を含んでいる。この研究デザインは、農薬ばく露を調査するための解析コストが、高すぎて、コホート全体を調査することができない場合に有用である。

<sup>11</sup> 時間をかけて症例を研究に登録した状況下では研究開始後まで発症していない症例で前向きに症例対照研究を実施することは可能である。

例えば、現在進行中のコホート研究で疾病または健康影響を生じた症例は、解析時に関係ある疾病や影響を生じていない研究からの適切な対照群と組み合わせることが可能である。コホート内症例対照研究（標準的な症例対照研究とは対照的に）の利点として認識されているのは、選択バイアスと想起バイアスの問題が最小限に抑えられることである。

**横断研究**では、疾病（例えば、先天性異常、胎児発育遅延（SGA））、症状、生物学的／身体的及び生理学的反応の測定値（例えば、肺機能検査、血圧、胸部X線、臨床検査、肝臓及び腎臓のバイオマーカー）の有病率に焦点を当てている。このような研究の主な特徴は、サンプリング時またはサンプリング期間中に疾病の有無を判定し、有病率を頻度の尺度として重視した観察研究であるということである。有病率とは、ある集団の中で罹患している人の割合であり、「点有病率」または「期間有病率」として調査される<sup>12</sup>。有病率は比率ではなく、割合であるため、横断研究では追跡調査期間を必要としない。一般的に、ばく露状態（例えば、ばく露または非ばく露）、疾病の状態／発生、人口統計学的特性は、ある時点（またはその時点以後）で調査される。この研究デザインにおける主な比較は、ばく露集団における疾病発生の有病率と非ばく露集団における疾病発生の有病率の比較であり、リスク尺度は有病率リスク比またはオッズ比である。横断研究は、一般的に、長年にわたる疾病発生または地理的に異なる場所のパターンや傾向を明らかにするために使用され、迅速かつ比較的安価に実施することができる。しかし、これらの調査では、疾病の発生率（新規症例の発生率）と疾病の持続期間の両方に影響される疾病発生の有病率を測定しているため、これらの要因を十分に分離することは、どのような解析においても困難な場合がある。したがって、これらの研究は「生存者集団」を対象としており、病気になったために関係ある集団から離れてしまった人を測定、評価、考慮することはない。横断研究のもう一つの重要な限界は、ばく露が疾病に先行しているかどうかを判断できないことである。そのため、横断研究では疾病とばく露の時間的関係を確立することができず、横断研究で示唆された仮説に基づく因果関係を確認するために、追加の研究を必要とするのが一般的である。

**生態学的研究**では、グループまたは集団のレベルでのデータを反映した情報を用いてばく露や疾病のパターンを調べる。生態学的研究では、解析の単位は個人ではなく集団である<sup>13</sup>。ここでは、被験者の集団がサンプリングされ、ばく露、疾病及び潜在的な交絡因子を、この集団（またはクラスター）レベルで測定される。集団は一般的に地理的、行政的、組織的な単位（例えば、地区、町、郡、学校、職場など）で定義され、すべてのばく露、疾病、交絡因子の測定は、個人のレベルではなく、集団のレベルで行われたり、まとめられたりする。生態学的（集団ベース）研究は、個人レベルの研究とは対照的であり、前者では症例がばく露された実際の個人であるかどうかの情報がないのに対し、後者ではばく露情報が個人に結びついている。

<sup>12</sup> 前者は特定の場所及び／または特定の時間での測定であるのに対し、後者は特定の期間における症例の割合の調査である。

<sup>13</sup> 一部の研究では、ばく露または疾病発生のいずれかが集団のレベルで測定されるが、他の変数は個人のレベルで測定され、研究者は個人のレベルで推測を行うという「部分的に生態学的」研究もある。

例えば、水中の汚染物質レベルによる疾病率の研究は、ばく露の評価に関しては生態学的であるが、健康影響や疾病状態は個人で調査されている場合がある。このような場合、ばく露は集団レベルで決定されるが、結果は個人レベルで調査される場合に「半生態学的」という用語が使用されることがある。

この研究デザインでは、ばく露集団のすべてのメンバーが個別にばく露されているかどうか（または個人のばく露レベル）を知ることはできないし、グループレベルの影響から個人レベルの影響を推測することもできない。研究の目的が（個人ではなく）**集団**への直接的な推測を目的としている場合には、これは問題ではなく、特に個人レベルでの測定が制約されていたり、困難であったり、集団内のばく露が一般的に均質であったりする場合には、このような研究が適切であると考えられる。研究の意図が個人への直接的な推測である場合、この研究デザインは生態学的誤謬と呼ばれる問題を抱えていることになる。集合またはグループ化されたデータセットで観察された関係が、個人レベルでサンプリングが行われていたら観察されていたであろうことを反映するという仮定である。この生態学的誤謬の問題に加えて、交絡因子に関する情報が集団レベルでしか得られない場合、集団とは対照的に個人レベルで交絡変数を適切にコントロールできない結果、追加のバイアスが発生する。

ほとんどの場合、生態学的研究は仮説を生み出す研究と考えられており、今後の研究のための研究仮説を提案するのに最適であり、問題の提起に貢献する可能性がある。とはいえ、生態学的研究をデザインの質に基づいて評価することは重要であり、デザインが優れていれば、生態学的研究から有用な情報を得ることができる（FIFRA SAP、2010年）。一般的に、生態学的研究だけでは、因果関係を立証する能力はない。これらの研究を他の研究と併用した場合、特定の状況下では有用となる可能性があり、ハザード特性評価の際に留意すべきである。特に、安定した母集団、明確なばく露の対比、リスクの大きな差は、これらの研究の有用性を高める重要な要因となり得る。

### C. 農薬のリスク評価に用いるための疫学研究の評価

OPP は、化学物質の使用に関連した潜在的な急性及び慢性の健康影響に関する観察疫学研究を査読済み文献で検索する。文献検索の手順と戦略に関する詳細は、別の場所で提供されている。コホート研究、症例対照研究、または横断研究のデザインを利用した疫学研究は、潜在的なハザード、ばく露反応の特性、ばく露シナリオ、または評価方法、そして最終的にはリスクの特性評価に関する OPP の理解を深めるための情報を提供することができる（van den Brandt、2002年）。さらに、説得力のある症例報告や症例集積解析は、これまで明らかにされていなかった健康影響や作用機序を明らかにする可能性がある。

一般的には、OPP のリスク評価で用いられる公開文献からの疫学研究を評価する際には、疫学研究の質、研究の文書化の十分さ（研究デザインと結果）、リスク評価との関連性が考慮される。

これらの基準は、懸念されるエンドポイントに応じて様々な方法論の詳細が重要になったり、ならなかったりするため、エンドポイントを特定することが重要である。疾病発生の確認に影響を与える関連因子を理解することが重要である（例えば、影響の有無を示す検査やバイオマーカーが利用できるか、あるいは初期段階では非特異的な症状であり、進行した病状では医師の診断につながるかなど）。さらに、環境及び職域疫学研究では、ばく露評価の質が極めて重要である。検討中の問題に対して、ばく露と交絡要因の評価の側面を事前に考慮しなければならない。

個々の研究の質を考える際には、疫学研究のデザイン、実施、解析、解釈の様々な側面が重要である。これらには以下が含まれる。

1. 仮説を明確に明示することで、たとえその研究が本質的に仮説生成的なものであったとしても、その仮説を明確に示されていること。
2. 健康影響の関連する重大な時期、リスク評価対象集団の関係あるばく露範囲、試験から得られる用量／ばく露反応の傾向の入手可能性など、ばく露評価の資質の中で、適切なばく露評価が十分であること。
3. 合理的に有効で信頼性の高い結果の確認（研究集団における健康影響の有無を正しく識別されていること）。
4. 対象集団を代表するサンプル集団となり、系統的な偏りがない適切な組み入れ基準と除外基準。
5. 観察されたリスク推定値における複数の農薬ばく露、または混合物ばく露の役割の評価または考察を含む、潜在的な交絡変数の適切な評価及び解析。
6. 参加者の選択や情報収集における誤りを含む、研究における潜在的な系統的な偏りの全体的な特性。これには、提示されたリスク推定値に対する系統的誤差の潜在的な影響を調査するための感度分析の実施を含む（例：Greenland's 公式）。
7. 健康影響を観察するための研究の統計的検出力の評価と適切な考察及び／または説明。
8. 研究デザインと対象となる結果の性質を考慮した適切な統計的モデル化技術の使用。

他の連邦及び非連邦の機関が次のようなガイドを提供している（例えば、OHAT、ナビゲーションガイド、National Toxicology Program [NTP] Report on Carcinogens [ROC<sup>14</sup>]、IRIS、Cochrane ACROBAT-Non-Randomized Studies of Interventions）（Sterne ら、2015 年）、及び観察疫学研究に関する STROBE（Strengthening the Reporting of Observational Studies in Epidemiology）声明（[www.strobe-statement.org](http://www.strobe-statement.org), Vandenbroucke ら、2007 年; Von Elm、2014 年）を参照）。OPP が疫学研究をヒトの健康リスク評価に統合する経験を積むにつれ、評価アプローチに関連する調整が行われることになる。

---

<sup>14</sup> <http://ntp.niehs.nih.gov/pubhealth/roc/index.html>

独立した研究評価は、ハザードの特定及びばく露反応評価に有益である詳細な要約表を作成する前に実施され、文書化されている。表 2 は、評価された主な考慮事項と、各考慮事項に関連する重み（低、中、高）の構造を示している。表 2 は考慮事項の一般的な検討事項を示したものであり、チェックリストと考えるべきではない。特定の科学的解析に適した特定の科学的考慮事項は、ケース・バイ・ケースで調整される。

研究評価プロセスの最終段階では、研究固有の評価を通じて特定された各研究における系統的な誤りから生じる可能性のあるバイアスの性質についての専門家／エキスパートの意見を提供し、疫学データベースに対する全体的な信頼性の評価を行うことである。このようにして、データ統合（動物、ヒト、メカニズム、その他）は、環境ばく露や職業上ばく露のヒト健康影響を伝えるヒト疫学研究の信頼度に基づいて行われることになる。

表 2. 研究の質に関する考慮事項<sup>a</sup> (Munoz-Quezada ら、2013 年 ; LaKind ら、2014 年より引用)

パラメータ	高	中	低
ばく露評価	MOA/AOP に関連している外部ばく露、内部ばく露量、または目標ばく露量との正確で精密な定量関係。  アンケートを利用した場合は、化学物質固有のばく露についての調査対象者からのアンケート及び/またはインタビューの回答。	指定されたマトリックス内のバイオマーカーと外部ばく露、内部ばく露量、または目標ばく露量との間の関係についてのエビデンス。  調査対象者または代理人が回答した化学物質固有のばく露に関するアンケート及び/またはインタビュー。	わずかなサロゲート(代理)  質の低いアンケート及び/またはインタビュー ; 化学物質固有の情報ではなく、化学物質群について収集された情報 ; 化学物質固有のばく露情報が収集されていない ; 農薬の使用経験の有無に関する評価。
発現事象評価	調査対象集団で有効性が確認されている標準化ツール ; 訓練を受けたスタッフによるカルテのレビュー/診断の確認 ; 有病率/症例の発生率に関する適切な評価。	母集団で有効性が確認されていない標準化ツール、またはスクリーニングツール ; または、方法が明記されていないカルテのレビュー。	検査の選択された部分、または母親の報告、その他 ; または母親/父親の自己報告 ; 有病症例または事例症例が適切であるかどうか不明確/不評価。
交絡因子コントロール	科学的な疑問に関連する重要な交絡因子及び標準的な交絡因子を良好に制御。	科学的な疑問に関連する交絡因子、標準変数(すべての変数ではない)を中程度に良好に制御。	調整なしで多変数解析は実施していない ; 不層化、不制限、または不調和。
統計解析	研究の課題とデザインが適切であること、適切なサンプル規模に支えられていること、データを最大限に活用していること、適切に報告されていること(選択的ではない)。	容認可能な方法、疑わしい調査力(特にサブアナリシス)、情報が無く明確に報告されていない解析の選択。	統計解析への最小限の注意、比較の不実施、または不明記。
(その他の)バイアスのリスク(選択、差のある誤分類、効果の規模、その他)	その他の潜在的なバイアスの主な原因は、存在する可能性は低く、存在するがリスク推定の規模と方向性に影響を与える可能性は低く解析される。	その他のバイアスの原因は、存在しても研究では取り上げられていないが、推定の方向性ではなく、推定の規模に影響を与える可能性がある。	主な研究バイアスは、存在するが研究では取り上げられておらず、研究の結果について他の説明を排除することはできない。

<sup>a</sup> パラメータ全体の総合的な評価に基づく総合的な研究の質のランキング。

## 1. ばく露評価

ばく露評価は、「ある物質へのばく露の規模、頻度及び期間を、ばく露された集団の数と特性とともに推定または評価するプロセス」と定義することができる。理想的には、評価において原因、経路、方法及び不確実性を記述すると記載（Zartarianら、2005年）されている。環境疫学において、ばく露評価は、特に環境媒体中に低濃度で存在する毒性物質については、独特の課題となっている（NRC、1991年；NRC、1997年）。ばく露経路の複雑さを考慮して、研究者はばく露を評価するための多くの異なるアプローチを開発してきたが、その正確さ、精度及び必要な資源は様々である（Niewenhuijsen、2003年）。これらのアプローチのいくつかは、疫学研究に特化したものではないが、様々な科学的解析におけるばく露評価の情報を提供するために使用される可能性がある。これらのアプローチには、過去の記録、アンケート、環境モニタリングに基づく間接的な方法と、個人モニタリングやバイオモニタリングに基づく直接的な方法がある。それぞれの方法の簡単な説明とその長所と限界を以下に要約する。

表 3. 間接的ばく露評価方法と直接的ばく露評価方法のまとめ

アプローチ	方法／ツール	例	ばく露の推定
間接的	過去の記録	住所情報を利用した農地への近接推定	二分法または順序のばく露
	アンケート	農薬使用の回答に基づくばく露の可能性の調査	二分法または順序のばく露
	環境モニタリング	地域の水道における農薬レベルの測定	モデル化を用いてばく露を推定することができるが、二分法または順序のばく露
直接的	個人モニタリング	吸入と経皮でばく露された農薬の測定	定量化されたばく露
	バイオモニタリング	血中・尿中の農薬の測定	定量化された内部ばく露量

過去の記録やアンケートは、化学物質へのばく露に関連する可能性のある主要な特性を評価するために使用される。疫学研究で使用される場合、過去の記録やアンケートは通常、定量的なばく露レベルを予測するためには使用されない。むしろ、過去の記録情報やアンケートの回答は、ばく露のカテゴリーレベルを調査するために使用される。ばく露レベルを調査するために使用できる過去の記録情報の例としては、農地に近接した住所や、職位や職歴に関する雇用履歴情報などが挙げられる。同様に、アンケートは、個人が農薬を使用したことを思い出すかどうかを調査するために、またはばく露の可能性を高める特定の職務を行っている個人を特定するために使用される。過去の記録やアンケートは、潜在的なばく露に関するデータの費用対効果の高い情報源になり得るが、限界がある。過去の記録やアンケートから収集されたデータは、ばく露のサロゲート（代理）に過ぎない。

その結果、これらの情報源はばく露の単純化しすぎているため、個人のばく露の可能性を正確に並べていない可能性がある。

**環境モニタリング**は、大気、水、土壌、食品、家庭環境や職場環境などの環境媒体中の汚染物質のレベルを特性評価するために使用される。多くの州及び連邦政府のプログラムでは、疫学研究に有用な環境モニタリングデータを収集している。環境モニタリングは、大気汚染や飲料水など、地理的な境界によって定義されるばく露に対して特に有用である。そのため、多くの疫学研究では、大気汚染と飲料水汚染へのばく露を特性評価するために、大気モニタリングデータと地域の水道のデータをそれぞれ利用している。環境モニタリングデータは、地理的境界によって定義されるばく露量を推定するのに有用であるが、個人レベルのばく露量を調査する目的では、特に個人が多く異なる場所で生活し、仕事をし、時間を過ごす場合には、信頼性が低くなる可能性がある。

**個人モニタリング**は、体の境界の接触点でのばく露を特性評価するために使用される。個人モニタリングの例としては、農薬との経皮接触を評価するための線量計の使用、吸入ばく露を評価するための個人用空気サンプリング装置、食品中の農薬レベルを調査するための対の食品サンプルの収集などがある。個人モニタリングの利点は、間接的な方法よりも個人レベルのばく露量をより正確に推定できる可能性が高いことである。また、個人モニタリングは、異なる経路でのばく露の関連性の優先順位付けに有用なばく露レベルの定量化を可能にする。さらに、個人モニタリングは、時間をかけて測定を繰り返すことで、縦断的なばく露量を評価するためにも使用される。個人的なモニタリングは間接的なアプローチに比べて多くの利点があるが、労力と資源が必要とする傾向がある（Niewenhuijsen、2003年）。その結果、個人モニタリングを用いてばく露を評価する大規模な疫学研究を実施することは一般的には不可能である。さらに、個人的なモニタリングは、サンプルを取得するために使用される測定技術や解析ツールに大きく依存しており、関連する期間（通常は過去）のばく露を特性評価する情報が利用可能になる可能性は低い。さらに、対象となる期間のばく露の全容が捕捉される可能性は低く、ピークと変動を捕捉するのに十分な期間でのサンプリングがではないかもしれない。そのため、疫学研究で使用される個人モニタリング手法の科学的厳密さと信頼性を考慮することが非常に重要であり、そのようなモニタリングは他のモニタリング（例えば、環境データ、生物学的データ及び／またはインタビュー／アンケートのデータ）で補完する必要があるかもしれない。

**バイオモニタリング**は、血液、尿、唾液、乳、脂肪、その他の生体組織などの生物学的試料中の化学物質、その代謝物、または反応生成物を測定することであり、ばく露の特性を明らかにするために使用される（Needhamら、2007年）。Zartarianら（2005年）は、「バイオマーカー／生物学的マーカーは、「生物学的システムにおける変化や事象の指標」として定義されてきた」と述べている。「ばく露の生物学的マーカーとは、組織、細胞、体液などの生物学的媒体から得られる細胞や生化学的、分析的、または分子的な測定値を指し、ある物質へのばく露指標である」と述べている。したがって、バイオマーカーは、ばく露を評価に用いられ、また健康影響の指標として使用される（LaKindら、2014年）。表4は、疫学研究から収集されたバイオモニタリングデータの質と関連性を評価するための科学的考察を示している。

解析ツールがより費用対効果の高いものになり、より多くのバイオマーカーが同定されるようになったため、バイオモニタリングを用いたばく露評価は急速に拡大してきた。自己申告のアンケートやインタビューデータと比較して、バイオモニタリングはばく露の誤分類を減らし、リスク推定の精度を高めることができるかもしれない。同様に、バイオモニタリングは、異なる経路からのばく露を統合し、体内に吸収されるばく露量を調査するために使用されている（Checkoway ら、2004 年）。さらに、疾病の履歴におけるバイオマーカーの役割が知られるようになってきている場合もあり、バイオマーカーはばく露の時間的な問題を解決するのに役立つかもしれない。

バイオモニタリングは他のばく露評価方法に比べて多くの利点があるが、それ自体にも限界がある。多くの研究では、生物学的サンプルはある時点での一点からしか採取されず、特にばく露が大きく変動する場合には、経時的なパターンを正確に反映していない可能性がある。さらに、バイオマーカーの評価には、環境とヒトの両方での化学物質の分解と代謝についての理解も必要である。そのため、ばく露のバイオマーカーはばく露以外の理由で個人間で異なる場合がある。代謝の違い、尿の測定値に関連した腎臓病などの併発疾患やバイオマーカーが有効成分のばく露を測定しているのか、環境中分解物のばく露を測定しているのかといった不確実性が、個人間や想定される比較集団間におけるばく露のバイオマーカーの見かけ上の違いの原因となっているかもしれない。

表 4. 環境疫学研究におけるバイオモニタリングデータの考察 (LaKind ら、2014 年より引用)

バイオマーカーの考察	ティア 1	ティア 2	ティア 3
ばく露バイオマーカー	バイオマーカーは、外部ばく露、内部ばく露量、目標ばく露量と正確かつ精密な定量的関係にある。	バイオマーカーは、外部ばく露、内部ばく露量、目標ばく露量との定量的関係が不明であるか、ばく露/用量に関してわずかなサロゲート [代理] (正確さ・精度が低い) である。	NA
効果バイオマーカー	MOA/AOP では key events のバイオインジケーター。	健康影響との関係に対する効果のバイオマーカーは明らかである。	バイオマーカーは結論をもたらさない (例えば、バイオマーカーは健康影響を明確にしない)。
特異度	バイオマーカーは、1 つの親化合物へのばく露に由来する。	バイオマーカーは、類似の毒性を持つ複数の親化合物に由来する。	バイオマーカーは、様々な有害エンドポイントがある複数の親化合物に由来する。
手法の感度	検出限界は、研究課題に取り組むのに十分な割合のサンプルから化学物質を検出するのに十分に低い。	研究仮説に取り組むには、検出頻度が低すぎる。	NA
バイオマーカーの安定性	既知の履歴と文書化された安定性データを持つサンプル。	サンプルは保存中の損失が知られているが、低量ばく露と高量ばく露の違いを定性的に評価することができる	履歴が不明なサンプル及び/または目的の分析物の安定性データがないサンプル。

バイオマーカーの考察	ティア 1	ティア 2	ティア 3
サンプルの汚染	サンプルは、収集時から測定時まで汚染されていない（例えば、分析物の無いことが保証された収集用品及び参照物質の使用及びフィールドとラボの両方でのブランクの適切な使用による）。研究には、研究データの信頼性を保証するために必要な手順の文書化が含まれる。	これらの手順を使用しない／文書化しない研究。	汚染の問題が知られているが、問題を取り上げた文書はない。
手法の要件	バイオマーカーの明確な同定と定量を必要な感度で提供する装置（例：GC-HRMS、GC-MS/MS、LC-MS/MS）。	高い信頼性と必要な感度でバイオマーカーの同定を可能にする装置（例：GC-MS、GC-ECD）。	バイオマーカーの定量のみが可能であるが、その方法には既知の干渉物質がある装置（例：GC-FID、分光法）
マトリックス調整	研究には、調整濃度と非調整濃度の結果が含まれている。	研究では、1つの方法（マトリックス調整済みかどうか）を用いた結果のみを提供している。	NA

FP = 偽陽性、FN = 偽陰性、GC-HRMS = ガスクロマトグラフィー／高分解能質量分析、GC-MS = ガスクロマトグラフィー／質量分析、GC-ECD = ガスクロマトグラフィー-電子補足検出器；GC-FID = ガスクロマトグラフィー-水素炎イオン化型検出器]、ICC = クラス内相関係数；NA = 該当しない；PFP = 偽陽性の確率

**間接的ばく露評価**方法は、後ろ向き研究では一般的であり、化学物質のばく露のサロゲート（代理）となる因子に基づいている。上述したように、間接的なばく露データは、追加のモデル化を行わない限り、一般的には定量的なばく露レベルを推定するためには使用できない。例えば、ある個人が農薬を使用したことがあるかどうかを調査するためにアンケートを使用することはできるが、その個人のばく露量を計算するために必要なすべての環境的・行動的な要因に関するデータを確実に集めることは困難である。そのため、間接的ばく露データは、二分法のばく露変数（すなわち、ばく露／非ばく露）または順序ばく露尺度を用いてばく露を分類するために使用されることが多い。一方、直接的ばく露評価方法は、個人のモニタリングやバイオモニタリングによる実際の個人レベルのばく露に関するデータに基づいている。したがって、直接法は、個人のばく露または内部ばく露量レベルを推定するために使用することができる。直接法は前向き研究で用いられることが多いが、十分に定義された母集団から既存の生物学的サンプルが入手可能な場合には、後ろ向き研究でも使用される。

**定量化された個人評価**には、個人モニタリングやバイオモニタリングがあり、一般的に実際のばく露レベルを推定するための最良のデータソースと考えられている（NRC、1991年；NRC、1997年）。その一方で、間接的な方法による正確なばく露の質的測定（例えば、二分式や順序式のばく露測定）も、疫学の目的では同じように正確であると考えられる。さらに、間接法は解釈が容易であることが多く、ばく露評価における有用性を示すための追加の研究開発が少なく済む可能性がある。

どのようなアプローチであっても、ばく露評価方法は信頼性と妥当性のあるばく露推定値を提供できなければならない。疫学の文脈では、一般的に**信頼性**とは結果を再現する能力を意味し、**妥当性**とは、一般的にばく露推定値が真のばく露レベルを反映しているかどうかを意味している（Checkowayら、2004年）。特定のばく露評価の信頼性と妥当性を評価する際には、研究目的の文脈でばく露評価の長所と短所を考慮することが重要である。あまり洗練されていないばく露評価は、探索的研究に適しているかもしれない。これは、探索的研究は潜在的なハザードについての認識を高めるのに役立つため、より焦点を絞った研究への専心を促すことができるからである。逆に、より焦点を絞った仮説を持つ研究は、より洗練されたばく露評価法を用いることで大幅に強固となる。したがって、間接的ばく露評価法と直接ばく露評価法は補完的なツールであり、ばく露と疾病の関係を調査する際に研究の様々な段階で使用することができる。

## 2. 交絡因子

交絡因子は、ばく露と疾病との関係がある程度で第二のリスク因子（交絡因子）の影響に起因している場合に発生する。これは、この第二のリスク因子（すなわち、交絡因子）が、疾病の独立した因果関係のあるリスク因子でありながら、解析対象のばく露と因果関係（因果関係または非因果関係）があり、ばく露と関係のある疾病との間の因果経路の中間変数としての役割を果たしていない場合に起こりうる。交絡因子は、適切に測定され、考慮されなければ、ばく露と健康影響との間の推定された関連性の大きさ（そして場合によっては方向性）を変化させる可能性がある。

これは、2つのリスク因子の影響が適切に分離、あるいは「区別」されていないため、ばく露と疾病の関係が過大または過小に評価される結果となりうる。例えば、ある研究では、ある農薬が肺がんに関連しているかもしれないが、これは農場のトラクターのディーゼル排気ガスの交絡効果によるものかもしれない。このように疾病発生（ここでは肺がん）との因果関係が農薬ばく露とも関連する場合、第二の因子の「農場のトラクターのディーゼル排気ガス」は交絡因子である。交絡因子には、喫煙、食事因子（例えば、高エネルギー／高カロリーの食事）、身体活動（例えば、身体活動の欠如）、遺伝子、併存症、薬物使用、アルコール摂取などのあまり直感的でないライフスタイルでのばく露が含まれることがあり、これらはすべて健康に悪影響を及ぼし、農薬使用と統計的に関連している可能性がある。疫学的解析では、交絡因子は研究サンプルで測定され、通常、研究のデザイン段階または解析段階のいずれかで最終的なリスク推定値に「調整」される。前者に関しては、疫学研究者は、研究者がコントロールしたい特性を共有する個人に研究集団を「限定」することができる。これにより、その特性（現在はコントロールされている）に起因する交絡の潜在的な影響を取り除くことができる。2番目に利用可能な方法は、研究のデザイン段階にも適用可能な方法で、研究者が交絡変数に基づいて個人を「マッチング」することで交絡をコントロールすることである。これにより、交絡変数が2つの比較群の間で均等に分布し、交絡変数を効果的にコントロールすることができる。主効果のリスク推定値に影響を与えるためには、交絡因子とばく露または疾病との関係が統計的に有意である必要はないことに注目することが重要である<sup>15</sup>。

解析の段階で、交絡因子をコントロールできる方法の1つは、層化がある。この方法では、交絡変数（の可能性はある）ごとに関連性を測定し、適切であると判断された場合には、個別の推定値を統計的に「まとめる」ことで、各層で評価された推定値に重みをつける Mantel-Haenszel アプローチを使用して共通のオッズ比または他の効果を算出する。層化は同時にコントロールする必要のある潜在的な交絡因子が複数存在する場合には困難である。このような場合、交絡因子は通常、統計的モデル化によって対処される。（例：ロジスティック回帰）。

疫学研究が現場で開始される前に交絡因子を慎重に考慮することが重要であり、どのような研究でも交絡因子をどのようにして考慮したかを適切に記述することが重要である。疫学研究は、潜在的な交絡因子を無視したり、あるいは十分な注意を払っていないと批評されることか多い。このため、感度分析は、欠落している交絡因子や考慮されていない交絡因子が観察された効果量に与える潜在的な影響を示すのに役立つ（Gustafson and McCandless, 2010年を参照）。

---

<sup>15</sup> これが、一般的に、交絡因子を調整する必要があるかどうかを判断するために交絡因子を「統計的に検定」することが不適切であると考えられている理由である。その代わりに、潜在的な交絡因子を調整した（交絡因子を含む）後の効果量の変化が10%以上であれば、交絡因子を分析に組み込むのに十分なエビデンスであると考えられる人もいる。

測定されていない交絡因子が結果に悪影響を及ぼすと考えられる場合、研究者は感度分析を実施して、影響の範囲とその結果として得られる調整後の効果測定値の範囲を推定すべきである。このような感度分析は、一般的にほとんどの公表されている疫学研究では一様には行われていないが、既知ではあるが評価されていないリスク因子によるバイアスや潜在的な交絡因子の影響を推定するために、利用可能な場合には使用することができる。

特定のばく露-疾病関連付けに応じて、ある因子はコントロールするために必要な交絡因子である場合とそうでない場合がある。交絡因子によって推定された効果量に大きな歪みが生じるためには、交絡因子は、関係ある疾病が比較的強いリスク因子でなければならない<sup>16</sup>し、関係あるばく露とも強固に関連付けられていなければならない。潜在的な交絡因子の評価は研究ごとに行われ、測定されていない交絡因子が結果に悪影響を与えると考えられる場合には、研究者は感度分析を行い、影響の範囲とその結果として得られる調整された効果評価の範囲を推定すべきである。観察疫学研究の質を評価する際、OPP は、関連する交絡因子が適切に同定、記述、評価、解析されているかどうかを検討し、研究対象となっている特定の関連性を偏りのない推定が可能かどうかを検討する。可能な場合は、交絡因子が推定された効果量の推定値に影響を及ぼす範囲の判断を助けるツールとして感度分析を考慮する。交絡因子は、疾患の比較的強いリスク因子でなければならないが、リスク推定値に大きな歪みをもたらすためには、対象となるばく露と強く関連していなければならないということを強調すべきである。このような場合には、交絡因子の可能性を提起するだけでは十分ではなく、なぜリスク因子が交絡因子となりうるのか、その影響がどのようなものなのか、そしてその影響が結果の解釈にとってどの程度重要なものなのかについて、説得力のある議論が必要である。(p.23-25、FIFRA SAP レポート、2010年4月22日)

最後に、共変量の交絡、効果修正、相乗効果、その他の媒介効果を区別することが重要である。交絡因子とは、疾病と因果関係のある変数及び対象となるばく露と因果関係があるか、否かにかかわらず、関連する変数をコントロールしないことによって生じるバイアスである。疫学研究者はこのバイアスを最小化しようと努める。一方、効果修飾因子とは、層化（例えば、年齢、人種／民族、SES の状態、遺伝的多型）によって、効果量に異なる影響を与える変数である。効果修飾因子は、交絡因子である場合もあればそうでない場合もある。一般的に、効果修飾因子は多変量モデルに交互作用項を導入するか、効果修飾因子のレベルによってデータを層化した後、層化に効果量を評価することによってモデル化される。対象となる各集団層において十分なサンプルサイズを確保するためには、効果修飾因子を評価するための特別な研究デザインが必要となることが多い。疫学研究者は、母集団の層化にわたるリスクの違いの評価、ばく露と関係のある効果の関連付けの評価、影響を受けやすい亜集団の特定が重要であることから、効果修飾因子を理解することを目指す（交絡因子のように最小化するのではなく、効果修飾因子の理解に努めている）。

---

<sup>16</sup> 交絡因子がそれ自体が本当にリスク因子であることを確認するだけでなく、関係のあるばく露との関連性だけではないことを確認することも考慮する必要がある。関係のある疾病との関連性を持つ因子の全部または一部を、関係のあるばく露との関連性のために調整することは、ばく露と疾病の関係が本当に存在する場合には、ばく露と疾病の関係を減衰させることにつながる。

効果修飾因子は交絡因子である場合もあればそうでない場合もある。例えば、タバコによく使用される殺虫剤と肺がんを関連付ける研究では、喫煙は交絡因子になるかもしれないが、この殺虫剤へのばく露のリスクが非喫煙者よりも喫煙者の方が高い場合には、効果修飾因子になるかもしれない。相乗効果は、疫学的な概念ではなく、生物学的または薬理的／毒物学的な概念として紹介されることが多く、2つの化学物質が一緒になって共同で作用することで、各化学物質の効果の（数学的な）合計を考慮した場合の効果を超えて、効果を拡大したり、増強したりする能力に関係している。疫学や統計学の用語では、これは効果の修飾や相互作用として表現されることが多い。

### 3. 統計解析

疫学研究は、特定のばく露と疾病との関連付けを評価することを目的としている。農業疫学研究の質を評価する際、OPPは使用された統計的手法も考慮する。具体的には、研究に記載されている分析方法が研究課題に適切であるかどうか、使用された統計的方法の記述の完全性、ばく露と疾病の関係における潜在的な交絡変数の特定、評価、調整のための方法の適切性、実施された可能性のある小集団解析の説明、範囲、提示（多重比較のための統計的補正が行われているかどうかを含む）が考慮される。

疫学研究では通常、リスクを推定するために統計的モデル（例えば、ロジスティック回帰（オッズ比）やポアソン回帰（カウントデータ）のような一般化された線形モデル）を利用する。そのためには、研究者は関連する主要なばく露変数と転帰変数だけでなく、関連する交絡因子を考慮しなければならない。また、研究における関連付けがこれらの因子、すなわち、効果の修飾や相互作用のレベルによって異なるかどうかを考慮しなければならない（Szkloら、2004年）。交絡の可能性のある変数（調査対象となる関連性の大きさや方向性を実質的に変化させる変数）が同定された場合、回帰モデルによる調整を行うことで、目的のリスク推定値、すなわち調査対象となる関連性を分離することができる。さらに、OPPは、研究における潜在的な効果修飾因子のレベルまたは統計的相互作用の評価による関連付けの層化を評価する。関係のある関連付けの規模と方向性が第三の変数のレベルによって大きく異なる場合、層化された結果は一次的なものとなさるべきである。

発現事象がまれな場合やサンプル規模が比較的小さい場合に統計的モデリングを行う場合、モデルにおける共変量が多すぎないように注意することが重要である。結果として得られる効果量の推定が高すぎたり、低すぎたりすることがあり、効果の真の推定を反映しているとは考えにくい。稀な事象やサンプル規模が小さいことによるこのような問題は、条件付き手法（例えば、デザインに研究における比較群の組み合わせを含む場合の条件付きロジスティック回帰）を使用する場合にも起こり得る。すなわち不一致ペア（または不一致セット）の観察数が少なすぎると、推定された効果量もまた信頼性が低くなる可能性がある。このように、交絡因子及びその他の共変量をコントロールすることは重要であるが、信頼性の高い検定を行うために過度にコントロールしたり、自由度が少なくしすぎたりしないように評価者は注意しなければならない。

このような場合には、有効なサンプルサイズを十分に確保するため、最も影響力のある交絡因子や他の共変量をより少ない数だけ調整して、より解析的なモデルを求めることがより重要である。

最後に、どのような統計モデリング演習においても、結果の臨床的／生物学的／科学的な意義の文脈で統計的有意性を考慮することが重要である。統計的には有意だが、臨床／生物学的／科学的な文脈では重要ではないということもある。逆に、統計的には有意ではないが、臨床的／生物学的／科学的な文脈では重要な結果である場合もある。前者の場合はサンプルサイズが必要以上に大きいことを示唆し、後者の場合は必要以上に小さなサンプル規模を示唆しているかもしれない。後者の場合は、公衆衛生の観点から重要であり、（不正確であるにもかかわらず）特に関連付けが強固である場合には、さらなる調査が必要である。

#### 4. 観察研究における潜在的なバイアス

バイアスとは、（観察されていない）真の状況とは系統的に異なる研究結果をもたらす研究のデザインまたは実施における系統的な誤差のことである。これは、効果量の評価におけるサンプリングの変動や精度（または同義で信頼限界）に関するランダム誤差とは対照的であるが、結果をある特定の方向（例えば、帰無仮説に向かって、または帰無仮説から離れて）に「追い込む（drive）」または「押し込む（push）」ようなことはない。

バイアスとは、研究のデザインや実施における方法論的な不完全性の反映であり、研究者は解析の一部として、それに対処したり、議論したりすべきである。研究にバイアスが導入される方法はいくつかある。研究は、参加者を研究に選択する方法（選択バイアス）や、ばく露や疾病に関する情報を収集する方法（情報バイアス、先に説明した症例対照研究の想起バイアスを含む）にバイアスがかかっている可能性がある。一般的な職業選択バイアスの例としては、「健康な労働者効果」があり、職域疫学研究において重要なバイアスを生じさせ、帰無値に向かってバイアスを生じさせ、さらにはそれ以下のバイアス（ばく露が「保護的」であるという解釈を生じさせる）につながる。バイアスの全くない研究はなく、公表した研究の著者は研究における潜在的バイアスどの程度説明したか、研究におけるバイアスを（もしあれば）どのように対処し、研究の特徴を明らかにしたかを検討すべきである。バイアスは、差別的（differential）または非差別的（non-differential）な誤分類から生じる可能性がある（Greenland、1998年）。差別的誤分類（バイアス）とは、他の変数の値に依存する方法で誤分類が発生したことを意味し、非差別的誤分類（バイアス）とは、他の変数の値に依存しない誤分類を意味する。誤分類バイアス（差別的または非差別的のいずれか）は、そのようなばく露を分類するために使用される試験方法の感度と特異度に依存し、特定の（限定された）条件の下でバイアスの方向性を予測可能な影響を及ぼすことがある。研究の方法と解析の知識に基づいてバイアスの方向を特定できるということは、規制当局の意思決定者にとって有用である。なぜなら、意思決定者は、検討されている疫学的効果の大きさ（例：OR、RR）が、真の効果の大きさの過小評価または過大評価である可能性がどの程度あるかを判断することができるからである<sup>17</sup>。

誤分類の程度が 10~20%の範囲にあることは異例ではなく、このような誤分類の程度（それが差別的か非差別的か）と、そのような誤分類の大きさと方向性の両方に関してオッズ比または相対リスクに影響を与える可能性の程度を評価する感度（または「もしも」）解析の形態を検討することは、疫学研究をレビューする際に有用であると考えられる<sup>17</sup>。（p.25、FIFRA EPA SAP 報告書、2010年4月22日）。交絡因子に関して前述したように、このような定量的感度分析は、公表されている疫学研究ではほとんど実施されておらず、バイアスの潜在的な規模や疫学的効果量推定への影響の定量的な評価は行われずに、むしろ一般的には解説文で評価されている（Jurek at al., 2006）。これは次のような理由によるものである。その理由の一部は、疫学者がこのような解析のための計算ツールを一般的に利用できないことや、そのようなツールに慣れていないことにある。このようなツールは次第に利用可能になりつつあり、潜在的なバイアスを評価するためのより厳密な定量的手法を開発する上で有用であると考えられる。

## 5. 帰無研究の解釈

「帰無」研究、すなわち、農薬へのばく露と有害な健康影響との間に関連付けがないと報告するよく行われている研究は、ヒト健康リスク評価に有用な可能性があるかどうか慎重に評価される。研究が無効という結果になるのは、調査された関連付けが実際には存在しないか、あるいは、実施された研究が所定の有意水準で関連付けを検出することができなかったためである。後者の結果（関連性を検出できなかった）は、必ずしも関連付けが存在しないことを意味するものではなく、特定の研究で関連付けが検出されなかったと解釈すべきである<sup>18,19</sup>。「帰無」研究をレビューする際に、これら 2 つの条件のどちらが正しいかを評価するためには、同一または類似の研究課題について報告された他の研究、ばく露と発現事象の評価方法、ばく露の誤分類が研究を帰無に偏らせた可能性の程度、関連付けの交絡変数の同定と分析を含む使用された統計的方法、影響が発生する、あるいは影響が検出される閾値を下回っているばく露の程度、研究の検出力と実質的な関係のある効果量を検出する能力を考慮する必要がある。統計的検出力とは、実際に真の違い（または関連付け）がある場合、研究者が 2 つの比較群間に差があること、すなわち、ばく露と疾病の間に関連付けがあることを正しく識別できる確率を意味する。

---

<sup>17</sup> このような感度分析は、多くの場合、交絡因子よりも効果量推定に実質的な影響を及ぼすと予想されるばく露の誤分類バイアスに対して特に推奨されるかもしれない。

<sup>18</sup> 「エビデンスがないからといって、エビデンスがないと解釈すべきではない→エビデンスがないことを、ないことのエビデンスと解釈してはならない」という古い格言は、ここでは真実である。

<sup>19</sup> 米国統計協会の統計的意義と P 値に関する声明 <https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf> も参照のこと。

「低検出力」の研究は、関連付けが実際に存在しているのに、関連付けがないと誤って結論を下すことがある<sup>20</sup>。

最後に、帰無研究の解釈に関しては、文献のシステマティックレビューにおける出版バイアスの影響を考慮することが重要である。出版バイアスとは、利用可能な公表文献が、そのような帰無研究を不釣り合いに除外する傾向のことである。疾病または健康影響の間にこのような「帰無」関連付けを示す研究は、研究課題が疫学的レビュープロセスの一部として確立された質の基準を満たしていれば、その研究と同様に有益な情報を得ることができる。これらには、研究デザイン、優先的仮説と探索的解析の比較、関係のある効果量を検出するためのサンプルサイズと統計的検出力、感度と特異度に対する影響の適切な確認、ばく露評価の質と差別的・非差別的な誤分類の可能性、主要な潜在的交絡因子とその他のバイアス（情報、選択等）の評価の適切性、効果修飾因子の評価（優先的仮説による裏付けがない仮説、生物学的妥当性、またはその他の裏付け情報についての多重比較の検討とその補正を含む適切な統計解析）などの要因が含まれる。

## 6. 外的妥当性（一般化可能性）

上述したように、妥当性とは、一般的にはばく露推定が真のばく露レベルを反映している程度を指す（Checkoway ら）。外的妥当性または一般化可能性とは、母集団のサンプル（例：農薬散布者）から得られた疫学的研究結果を他の母集団（例：すべての農業従事者）に拡張する能力のことである。外的妥当性を評価するために、サンプルの特性とより大きな集団（既知の場合）との比較を行うことができる。このような評価には、人口統計学的要因だけでなく、ばく露（例えば、用量、時期、期間）が類似しているかどうか及び重要な影響修飾因子（例えば、脆弱者集団の感受性）が考慮されているかどうかも含まれるべきである。一般化可能性は特に重要であり、個々の研究結果が、規制当局のリスク評価において、より大きな集団または対象となる亜集団に適用できるかどうか、またどのように適用できるかを理解することが重要である。例えば、AHS は、いくつかの農薬について、いくつかのがんとがん以外の健康影響との間に統計的な関連付けがあることを報告している。OPP は、報告された知見が、ノースカロライナ州とアイオワ州以外の州の農薬散布者や、主に散布後の作業を行う農業従事者にどの程度適用できるかを評価することに関心が持たれている。

---

<sup>20</sup> 検出力は低くても統計的に有意な効果が見られる研究は、効果量の拡大という現象である可能性があり、これを調査することが重要である（Ioannidis、2008年）。

## V. ヒトでの事例調査データ

一般的に言えば、上記のような農薬に関する疫学研究は、急性の臨床症状をもたらす可能性が小さい低量ばく露（より長い期間にわたって）に焦点を当てている。OPP はまた、より高く、より短い間隔（多くの場合、急性の"一回限り"で）のばく露に関心を持っている。この「ヒトでの事例」または中毒データは、問題となっているに起因すると考えられる短期的で高濃度の高量ばく露シナリオを評価するのに有用である。

OPP はこのような「ヒトでの事例情報」をいくつかの目的で使用している。最も広い意味では、OPP はリスク評価／リスク管理の活動に情報を提供するためにヒトでの事例データを使用している。これは、農薬登録改善法（PRIA）の責任のもと、登録審査活動の重要な一部となっている。この目的のために、OPP は人の事例データを評価して経時的な傾向を調べ、さまざまな農薬ばく露の重大さの程度と頻度のパターンを調べている。場合によっては、事例情報は、追加の情報や追加のリスク管理措置の必要性を示すことができる。また、事例情報は、リスク軽減措置が実施された後の達成の評価にも役立ち、事例情報は、OPP がヒト健康と環境を保護するために行ってきたリスク管理措置の有効性を保証するための OPP のパフォーマンス・アカウンタビリティ・システムの重要な部分である。最後に、事例情報は、使用方法に関する実際の使用情報を提供する上で有用であり、また、必要に応じて、施行や教育の活動の対象となる可能性がある。

OPP はこの情報を様々な情報源から入手している。ヒトでの事例データの情報源には、医学的及び毒物学的文献に掲載されている（ヒトの）**医学的症例報告**や急性及び慢性のハザードの特定を促進し、リスク評価プロセスに不可欠な共同で検討されている急性農薬中毒の様々な国の**中毒サーベイランス活動**からの情報が含まれている<sup>21</sup>。

**医学的症例報告**（医師によって書かれた直接の証言）や医学的症例集積（共通の原因や症状を持つ個人の症例報告をまとめたもの）は、健康影響に関する利用可能なすべてのエビデンスを解析するための貴重なツールであり、動物試験や疫学研究の知見を補完するものでもある。さらに、特定の農薬の使用に関連すると思われる有害な健康影響の稀な発生や新規の発生を特定することができ、中毒警戒の行政機関に「事前通知」を提供できる。公表されている農薬の症例報告は、通常、非定型的な（高量ばく露／大量投与、違法、適用外）急性または短期のばく露による影響を記述している。報告書はしばしば個々の事例に基づいており、その性質上、選択性が高い。しかし、これらの報告書は、ヒトにおいてこれまでに確認されていなかった毒性影響を特定し、高量ばく露後の影響、健康影響及び医学的後遺症を知る上で、特に有用である。これらの研究には、より詳細な医学的情報（後遺症を含む）や詳細な追跡調査が行われていることが多く、一般的に用量に関する高い質及び／または定量的な情報が得られている。

---

<sup>21</sup> OPP は、IPSC がリスク評価においてヒトでの事例データを考慮する成果を認識している。  
[http://www.who.int/ipcs/publications/methods/human\\_data/en/index.html](http://www.who.int/ipcs/publications/methods/human_data/en/index.html)

複数の医療事例研究で類似点が見られたり、症状、ばく露シナリオ、使用方法などにパターンが見られた場合には、リスク評価プロセスと結果の裏付けに有用な情報を提供することができる。医学的症例研究や定量的なばく露情報集積は、必要に応じて、典型的な使用から予想されるばく露マージンを特性評価することで（ラベルに明示された適用率や代替ばく露情報に基づく）リスク評価のばく露推定値と比較できる。

医学的症例報告や医学的症例集積を評価する際には、以下の点を考慮する。

- ばく露の詳細な履歴（いつ、どのように、どのくらい）、健康影響の発現時期、患者の徴候や症状が報告されている。
- 商品名、農薬表示、登録番号など、商品・化学品・農薬に関する情報。
- 患者情報（例：年齢、人種、性別）；基本的な健康状態及び同様の徴候や症状をもたらす可能性のある医薬品の使用；関連する病歴及びリスク因子の有無。
- 疾病の説明と診断に至った経緯。
- 患者の管理・治療、農薬や化学物質の血中濃度などの検査データ（治療前、治療中、治療後）。
- 医学的報告は信頼性があるか、妥当であるか。他の研究、レビュー、ガイドラインを含む最新の知見と一致しているかどうか。
- 疾病の臨床経過及び患者の転帰（例：患者の回復及び退院；退院後の患者の状態、農薬または化学物質へのばく露に関連した慢性的な健康影響または早期死亡）。

実際のばく露情報及び毒物学的情報の情報源として医学的症例の報告／集積を利用することに加えて、OPP は様々な農薬中毒事例データベースを利用した中毒サーベイランス活動も行っている。具体的には、OPP は以下の 5 つのヒトでの事例データソースにアクセスできる。OPP *Incident Data System* (IDS)、American Association of Poison Control Centers (PCC) の *National Poison Data System* (NPDS) からの要約報告書、現在オレゴン州立大学にある EPA が資金提供している *National Pesticide Information Center* (NPIC) からのデータ、疾病管理予防センター（Centers for Disease Control and Prevention）/National Institute for Occupational Safety and Health for Occupational Risk - Pesticides (NIOSH SENSOR-Pesticides) 及び *California Pesticide Illness Surveillance Program* (PISP) のデータ。以下、それぞれについて順番に説明していく。

- **OPP Incident Data System (IDS)** は OPP が管理しており、FIFRA 第 6 条 (a)(2)<sup>22</sup>項に基づき登録者が提出したデータ及び EPA に直接報告されたその他の事例を含んでいる。

---

<sup>22</sup> FIFRA 6(a)(2)に基づき、農薬登録者は、「農薬の環境への不適切な悪影響に関する事実情報」を知った場合には、EPA に通知するよう求められている。

OPPは1992年以来、IDSに農薬関連の事例報告をまとめている。IDSには、登録者、他の連邦・州の保健・環境機関及び個人消費者からの強制的なFIFRA第6章(a)(2)報告を含む、様々な情報源から主張されたヒトでの健康影響事例の報告が含まれている。IDSには、ヒト、植物、野生動物、家畜への悪影響を主張している事例に関する情報が含まれている。IDS報告の大部分は、政府機関が紙形式で受け取っている。IDSの項目は、OPPの情報サービスセンターにあるマイクロフィルムやスキャンした画像に保存されているオリジナルの報告書のコピーへの「ポインタ」として機能する。

IDSには職業上及び非職業上の両方の事例が含まれているが、報告された事例の大部分は非職業／住居のシナリオに関するものである。全米から報告された事例のほとんどは製品情報（EPA登録番号など）の記載に関連している。IDSはほとんどがFIFRA 6(a)(2)報告要件に基づき農薬登録者から提供された情報で構成されているため、政府機関は特定の製品を識別することに対して比較的高い信頼性を持っている。重大さの程度の順位付けは、各事例について（CFR §159.184で規定されている）記載されている。症状情報は、事例の解説部分に含まれることがあるが、この情報は通常、医療専門家による検証／確認を受けていない。IDSには、ばく露シナリオやハザード情報の解説情報も含まれている。多くの企業は、標準化された、業界で開発された自主的な事例報告フォームを使用している。

OPPはIDSからデータを収集・評価し、農薬ばく露による健康影響の程度や重大さの程度に関して潜在的なパターンを識別している。IDSの報告書は範囲が広く、場合によっては詳細な情報を含むが、システムは必ずしも職業上ばく露状況や医学的転帰などの発現事象に関する詳細な情報を一貫して把握しているわけではない。

また、IDSに入る症例データのほとんどは、中毒管理センターへの電話からの収集もあるが、確認・照合されない。とはいえ、事例情報は、製品の初期登録後、製造販売後の重要なフィードバックループを政府機関に提供できる。重大な性質のIDS事例や、それほど重大ではない事例の間で示唆されたパターンや傾向は、特定の化学物質や製品をさらに調査するように政府機関に求めることができる。IDSはこのように広範囲をカバーしているため、一時的な傾向情報を提供し、リスク軽減によって潜在的な農薬ばく露が減少し、IDSに報告された潜在的な事例の数が減少したかどうかを判断するのに役立つ可能性がある。全体的に見ると、IDSは農薬の全国的な傾向と事例の頻度に関する良好な情報を提供し、農薬のハザード及び／またはばく露の可能性について有用な予測を提供することができる。

- 全米中毒データシステム（NPDS）（以前は毒性影響監視システム（TESS）と呼ばれていた）は、米国中毒管理センター協会（AAPCC）によって維持されており、いくつかの連邦政府機関からの資金援助を受けている。NPDS はコンピュータ化された情報システムで、地理的に特定されたほぼリアルタイムの報告が可能である。中毒管理センターの主な任務は、緊急事態に対応する通報者を支援することであるが、NPDS のデータは、化学製品の安全性に関する新たな問題を特定するのに役立つ。全国にある 61 の PCC のホットラインは、365 日 24 時間年中無休で開設されており、特別な訓練を受けた看護師、薬剤師及びその他の臨床医療専門家が中毒情報を提供している。コンピュータによるデータ入力、標準化されたプロトコル、厳格なデータ入力基準を使用して、現地の通報者が事例を報告する。これらの報告された事例は現地で保存され、AAPCC によって維持されている全国データベースに要約された形で更新される。情報通報は個別に集計され、事例としてはカウントされない。PCC システムは、ほぼすべての米国とその領土をカバーしており、2001 年以来、コンピュータの大幅な強化が行われてきた。

NPDS には主に非職業上の事例が含まれている。NPDS には解説情報は含まれておらず、製品情報は完全ではない。NPDS は、訓練を受けた専門家が指定／記録した重大さの程度の順位と症状情報を提供しており、政府機関はこれらの情報に対して比較的高い信頼性を持っている。NPDS はまた、報告されたばく露による健康影響の可能性に関する情報も提供している。全体として、NPDS は全国的な傾向、農薬の事例発生頻度、特定の農薬の潜在的ハザードについての良好な情報を提供している。しかし、情報源の制限により、毎年発行される AAPCC のサマリーレポート（例：<http://www.aapcc.org/annual-reports/>を参照）にしかアクセスできず、これらのレポートは他の情報源を補完する役割を果たしているため、NPDS はより完全なアクセスが可能となっている。

#### □ National Pesticide Information Center (NPIC)

<http://npic.orst.edu/index.html>は、EPA の資金提供を受けており、客観的で科学的根拠に基づいた農薬情報を提供し、問い合わせに対応したり、事例に対応したりすることを目的としている。NPIC は平日の営業時間内に全国的に機能しており、オレゴン州立大学（現在）と EPA との間で協力している。中毒管理センターと同様に、NPIC の主な目的は事例データの収集ではなく（NPIC の年間の問い合わせの約 10%は「事例」に関連していると考えられている）、むしろ幅広い農薬に関する情報を問い合わせ者に提供し、農薬での事例の調査や緊急処置のための他の情報源に誘導することである。とはいえ、NPIC では問い合わせ者からの事例情報（年間約 4000 件）を収集し、データベースに記録している。NPIC は全国の事例の情報源であるが、一般的に IDS に比べて報告件数は少ない。

特定の農薬や製品が IDS で高頻度に観察された場合でも、NPIC は全国のデータセット間の整合性を判断する上で有用な情報源を提供している。

IDS や PCC と同様に、NPIC の事例は主に非職業上のものである。NPIC の事例には、通報者が情報を提供する際に解説情報や製品情報が含まれている。対象範囲は全国であるが、NPIC に報告される事例の数は NPDS や IDS に比べてかなり少ない。一方、提供される情報量はかなり多く、必要に応じて政府機関がカスタムレポートを要求することも可能である。ハザード情報には、重大さの程度の順位付け、ばく露経路、症状などが含まれており、これらは訓練を受けた担当者によって記録される。また、ばく露と有害影響の関連性がどの程度であるかについての情報も提供している（これを確実性指数と呼んでいる）。また、NPIC は年次報告や解析結果を公開しており、有用な情報源となっている。

- 疾病管理予防国立産業衛生研究所（CDC/NIOSH）は、**SENSOR（Sentinel Event Notification System for Occupational Risk） - Pesticides** と題した農薬監視プログラムとデータベースを管理している<sup>23</sup>。このデータベースには、1998 年から 2013 年までの 12 の州における農薬による疾病症例報告が含まれている。参加している州は、カリフォルニア州、フロリダ州、アイオワ州、ルイジアナ州、ミシガン州、ネブラスカ州、ニューメキシコ州、ニューヨーク州、ノースカロライナ州、オレゴン州、テキサス州、ワシントン州である。ある年の参加州は、州と連邦政府が農薬監視のために資金を提供しているかどうかによって異なる。

SENSOR-農薬データベース内の農薬関連の疾病症例は、様々な情報源（現地の中毒管理センターからの報告、医師によって報告された際に農業従事者の補償請求の州労働局からの報告、州農業局からの報告及び州保健局への医師の報告）から把握されている。職業上及び非職業上の事例の両方がデータベースに含まれているが、SENSOR 責任者は、主に職業上の農業従事者事例に関する追跡症例調査の成果に焦点を当てている。州保健局の SENSOR 責任者は、症例を追跡し、ばく露シナリオ、症状、重大さの程度、健康影響を確認するために、カルテを取得する。標準化されたプロトコルと症例の定義を使用して、州保健局の SENSOR 責任者は、症例、カルテ、医師と患者によって提供された事例インタビュー解説を SENSOR データシステムに入力する。

すべての SENSOR-Pesticides の症例は、EPA が事例の解析に使用する集合データベースに含まれるために、最低 2 つの健康影響を報告しなければならない。

---

<sup>23</sup> SENSOR-Pesticides のホームページ : <http://www.cdc.gov/niosh/topics/pesticides/overview.html>

各症例のエビデンスは、NIOSH の症例分類マトリックスに基づいて、ばく露と疾病との因果関係について評価される。SENSOR-Pesticides の症例の 98%は、確定、可能性高い、または可能性ありに分類され、2%は疑わしい症例に分類されている。可能性の低い症例、無症候性の症例、無関係の症例及び情報が不十分な症例は、SENSOR-Pesticides のデータベースには含まれない。

全体的には、職業上及び非職業上の事例について非常に有用な情報を提供しており、時には農薬のハザードやばく露の可能性についての有用な予想を提供している。また、SENSOR-Pesticides は独自のデータの解析を行い、Morbidity and Mortality Weekly に公表している。前述のデータベースとは異なり、非職業上/住居上と職業上の両方の事例が含まれているが、SENSOR'S は伝統的に農業従事者での事例に焦点を当てており、その情報を提供する上で特に価値のあるものとなっている。1998 年から 2011 年までの SENSOR-Pesticides のデータはオンラインで入手可能である：<http://wwwn.cdc.gov/Niosh-whc/Home/Pesticides>。

- **カリフォルニア州農薬疾病サーベイランスプログラム (PISP)** は、カリフォルニア州によって管理されている。このデータベースは、農薬に関連した疾病や傷害を記録している。症例報告は、医師からの報告と農業従事者補償記録を介して行われる。現地の郡農業委員会がばく露の状況を調査する。カルテと調査結果は、カリフォルニア州農薬規制局 (DPR) の技術専門家によって評価され、疾病の登録簿に登録される。カリフォルニア州の PISP プログラムで報告されたすべての農薬による疾病は、郡の農業委員会によって調査され、DPR は報告書の評価してデータベースにまとめ、農業従事者やその他の人々を農薬ばく露の悪影響から保護するための州のプログラムを改善するために使用される (<http://apps.cdpr.ca.gov/calpiq/>)。

現在、OPP は化学物質別にヒトの事例データを評価している。各データベースからの事例は、ハザードポテンシャル (死亡者数、より重大な事例の頻度、報告された症状のパターン/傾向) とばく露ポテンシャル (事例の頻度/経時的傾向、ばく露シナリオのパターン/傾向、ばく露に影響を与える要因のパターン/傾向、または製品のパターン/傾向) について解析されている。上記のデータベースからのヒトでの事例データを評価する際、OPP はいくつかの一般的な基準を考慮している。OPP は症状の相対的な重大さの程度と頻度を考慮している。さらに、OPP は一般的に、時間的な関連付けが確認できる報告を、あるいは少なくとも妥当と思われる報告を、より信頼性の高いものとしている。最後に、ヒトの事例データを評価するために使用される他の要因には、ばく露反応の関連付けのエビデンス、報告された健康影響の一貫性、報告された健康影響の生物学的妥当性、医薬品の使用などの健康影響の代替原因の排除及び観察された症状や健康影響の特異度が含まれる。さらに、より重大な事例の解説は、ばく露の時間と報告された影響の間に時間的な関連付けがあるかどうかを評価し、関連付けが状況によって裏付けられているかどうかを判断することが多い。

例えば、屋内での農薬散布後 3 か月に発生した高齢者の心臓発作は、因果関係があるとは判断されない。一方で、農薬種類に応じた既知の症状と一致する症状を伴うばく露時またはばく露後間もなく発生した重大な事例は、事前の病歴がなく発生した場合、因果関係の推測がより正当化されることを示唆している。

以上のことから、医学的文献やヒトの毒性学的文献に掲載されている医学的症例報告／症例シリーズと、EPA が維持、資金提供またはアクセスしているデータベースから得られる中毒サーベイランスデータの両方からなる事例データは、実社会での農薬のリスクを評価する際に OPP を支援する有用で補完的な情報を提供することが可能である。

## VI. 要約と結論

このフレームワークでは、疫学とヒトでの事例データをレビューする際の重要な要素を説明し、そのようなデータを農薬のヒト健康リスク評価に組み込むために提案されている WOE 解析について説明している。OPP では、複数のエビデンスと、*in vitro*、*in vivo* 及び *in silico* の情報源から得られた入手可能な最善のデータを使用する。OPP では、複数の情報源からのデータを統合する際に、系統化されたツールとして修正された Bradford Hill 基準を用いて、品質、一貫性、関連性、整合性、生物学的妥当性を考慮するため、WOE アプローチを使用している。WOE 解析の適用は、リスク評価ガイドラインに概説されている科学的解析に従って EPA が日常的に使用している統合的かつ解釈的なプロセスである。WOE 解析はまた、結合されたデータセットの質も評価し、特定の科学的評価に適切である成果と複雑性のレベルと一致している（U.S.EPA、2002 年）。OPP は、21 世紀の毒性試験の新たなビジョンの実施に向けて、現在、毒性学とリスク評価が変革期にあることを認識している。これらの変革に伴い、OPP はこのアプローチを適宜更新して行くであろう。

## VII. 参考文献

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. Lippincott-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 P1, 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 MT1, 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](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](http://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](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](http://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](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<sup>3/4</sup> 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](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, Pocock 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 TJ1 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

**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**



農薬のリスク評価において  
ヒトでの疫学データと事例データを導入するための  
農薬プログラム局のフレームワーク（基本概念）

2016年12月28日

農薬プログラム局  
米国環境保護庁

## Table of Contents

<b>I. Purpose &amp; Scope</b>	3
<b>II. Introduction</b>	4
<b>III. Systematic Review in Pesticide Risk Assessment: Epidemiology</b>	8
A. Problem Formulation	9
B. Data Collection	10
C. Data Evaluation	11
D. Data Integration: Weight of Evidence (WOE)	12
E. Overall conclusions, recommendations for risk assessment, statement of areas of confidence and uncertainty	15
<b>IV. Reviewing Epidemiology Studies for Use in Pesticide Risk Assessment</b>	16
A. Introduction	16
B. Types of Epidemiology Studies	18
C. Evaluating epidemiology studies for use in pesticide risk assessment	21
1. Exposure Assessment	25
2. Confounding Factors	30
3. Statistical Analysis	33
4. Potential Bias in Observational Research	34
5. Interpretation of Null studies	35
6. External Validity (Generalizability)	36
<b>V. Human Incident Surveillance Data</b>	37
<b>VI. Summary &amp; Conclusions</b>	43
<b>VII. References</b>	44

<b>I. 目的と範囲</b>	<b>3</b>
<b>II. 序章</b>	<b>4</b>
<b>III. 農薬のリスク評価におけるシステマティックレビュー：疫学</b>	<b>8</b>
A. 問題の定式化	9
B. データ収集	10
C. データ評価	11
D. データ統合：エビデンスの重み付け (WOE)	12
E. 全体的な結論、リスク評価の推奨事項、信頼性と不確実性の分野の陳述	15
<b>IV. 農薬のリスク評価に用いるための疫学研究のレビュー</b>	<b>16</b>
A. 緒言	16
B. 疫学研究の種類	18
C. 農薬のリスク評価に用いるための疫学研究の評価	21
1. ばく露評価	25
2. 交絡因子	30
3. 統計解析	33
4. 観察研究における潜在的なバイアス	34
5. 帰無研究の解釈	35
6. 外的妥当性（一般化可能性）	36
<b>V. ヒトでの事例調査データ</b>	<b>37</b>
<b>VI. 要約と結論</b>	<b>43</b>
<b>VII. 参考文献</b>	<b>44</b>

## 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<sup>1</sup> and from the National Institute of Environmental Health Sciences (NIEHS)/EPA Children's Centers<sup>2</sup>, 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

<sup>1</sup> <https://aghealth.nih.gov/>

<sup>2</sup> <https://www.epa.gov/research-grants/niehsepa-childrens-environmental-health-and-disease-prevention-research-centers>

## I. 目的と範囲

米国環境保護庁 (EPA) の農薬プログラム (OPP) は、連邦殺虫剤・殺菌剤・殺鼠剤法 (FIFRA) 及び連邦食品・医薬品・化粧品法 (FFDCA) に基づき、米国内の農薬を規制する認可プログラムである。このプログラムの一環として、OPP は毒性とばく露の多くの主要データを評価し、農薬がヒトの健康及び環境に及ぼす影響を評価している。ヒト健康影響を評価する際、EPA は、まず疫学データを含めたヒトへの影響の可能性を直接評価する情報に注目する。しかし、これまでは農薬の潜在的な毒性を報告した疫学研究はほとんどない。したがって、OPP はハザードの可能性を評価し、ヒトの健康リスクを推定するため、これまでは実験動物を用いた毒性試験に主に依存してきた。Agricultural Health Study<sup>1</sup> や National Institute of Environmental Health Sciences (NIEHS) /EPA Children's Centers<sup>2</sup> などから数多くの論文が発表されたことで、米国で実施された農薬へのばく露に関する疫学研究の有用性が高まっている。とはいえ、質の高い疫学データが存在する、あるいは充実しつつある農薬の数は、今後短期的には比較的少ないので、ほとんどの農薬の定量的リスク評価に用いるためのデータの主な情報源は試験データであることがおそらく継続するだろう。

OPP の目標は、そのような情報が入手可能な場合には、科学的に正確で明白な方法で利用することである。これを達成するために OPP は、この文書に記載されているように総合的な疫学的フレームワークを開発した。このフレームワークでは、そのような研究や科学的情報を農薬のリスク評価にどのようにして一層完全に統合することができるかを評価する際に OPP が検討する科学的に考慮すべき要点を述べている。現在の文書は、拘束力のある規則ではなく、疫学データを評価または使用するためのレビュアー向けのガイドやマニュアル、標準作業手順書であることを意図したものでも、その役割を果たすことを意図したものでもない。また、より最新で進歩した疫学的方法の完全な専門書となることも、この種の研究を解釈するために重要なニュアンスや複雑さを適切に伝えることも意図していない。したがって、因果推論及び因果関係図 (Rothman ら、2009 年; Greenland and O'Rourke, 2012 年)、交絡因子の特定・評価・制御に対する最新の研究法、メタ分析と異質性の評価 (Borenstein ら、2009 年; Greenland and O'Rourke, 2012 年)、または疫学データの感度/定量的バイアス解析 (Lash ら、2009 年; Lash ら、2014 年; Ioannidis, 2008 年; Greenland and Lash, 2012 年; Jurek ら、2007 年) といった重要な疫学的话题は (詳細に) 論じていない。これらの話題、概念及び問題点のすべては、農薬に関する疫学研究に応用できるが、この OPP フレームワーク文書では取り上げていない。代わりに、この文書では、OPP の FIFRA 及び FFDCA の機能支援のためのヒト健康リスク評価の文脈において農薬の疫学研究の評価と使用に関する全体的な概念を提供している。

<sup>1</sup> <https://aghealth.nih.gov/>

<sup>2</sup> <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 21<sup>st</sup> century toxicity testing advocates a shift away from the current focus of using apical toxicity endpoints to using toxicity pathways<sup>3</sup> 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 21<sup>st</sup> 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.

<sup>3</sup> Toxicity pathways are cellular response pathways that, when sufficiently perturbed, are expected to result in adverse health effects.

本文書の以前のバージョンは、2010年2月にFIFRA科学諮問パネル(SAP)によって好意的にレビューされた(USEPA, 2010年, FIFRA SAP, 2010年)。この文書は、SAPが推奨する改善点、パブリックコメント及び2010年以降に疫学データが利用可能であった農薬の評価から得られた知識が盛り込まれており、進捗及び必要に応じて随時更新される文書と考えるべきである。

## II. 序章

全米科学アカデミー(NAS)の米国研究評議会(NRC)による2つの報告書、「21世紀の毒性試験: ビジョンと戦略(2007年)」と「科学と政策調査(2009年)」は、共に毒性学とリスク評価の新たな方向性を示している。これら2つのNRC報告書は、毒性試験の実施方法、データの解釈方法、そして最終的に規制上の政策調査の方法に大きな変更を提唱している。具体的には、21世紀の毒性試験に関する2007年の報告書では、毒性試験、リスク評価、そして最終的な政策調査に情報を提供するために、現在の先端毒性エンドポイントの使用に焦点を当てたものから、毒性発現経路<sup>3</sup>を使用するようにシフトすることを提唱している。このアプローチは、遺伝子、タンパク質、低分子がどのように相互作用してヒト細胞の機能を維持する分子経路を形成しているかという急速に進化する科学的理解に基づいている。新しい毒性試験のパラダイムの目標は、環境因子へのばく露がこれらの経路をどのように乱すかを明らかにし、それによって有害な健康影響につながる重要な事象を連鎖的に引き起こす原因を明らかにすることである。この新しいアプローチでは、疫学研究、ヒトでの事例データベース、バイオモニタリング研究に見られるようなヒトの情報と実験毒性学的情報が重要な役割を果たすことが期待されている。具体的には、このようなヒト情報は、実際の化学物質ばく露による影響についての見識を提供し、問題の定式化やハザード/リスクの特性評価に貢献することができる。さらに、疫学データやヒトでの事例データは、追加の解析やデータ生成(例えば、*in vitro*試験や目標とする*in vivo*試験で使用するための用量やエンドポイントの選択)の指針となり、影響を受けやすい集団を特定したり、新たな健康影響を特定したり、既存の毒性学的結果を確認したりすることができる。

この毒性試験とリスク評価の新しいビジョンには、生物学的体系の多種多様なレベル(分子レベルから化学物質の影響をその発生源から最終的な健康影響や集団への影響まで考慮することを目的とした集団を基にした調査レベルまで)のデータが含まれる。このようなデータには、*in vitro*試験や*in vivo*試験から得られるデータのほかに、*in silico*モデルデータがある。疫学データと事例データを取り入れるOPPのフレームワークは、21世紀の毒性試験に関する2007年のNRCの報告書と概念的に一致しており、どちらも多種多様なデータソースから得られる最良の情報の用途はエビデンスの重み付け(WOE)解析にまとめて使用することを強調している。

<sup>3</sup> 毒性発現経路とは、十分に攪乱された場合に有害な健康影響を及ぼすことが予想される細胞応答経路のことである。

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.

一般的な原則として、産業環境疫学研究は広く使用されている農薬のみを対象に実施されており、これらの農薬も科学的文献で十分に研究されている必要がある。したがって、疫学データが利用可能な多くの場合、毒性、ばく露、薬物動態 (PK)、作用機序/有害性発現経路 (MOA/AOP) の情報に関する主要な文献データも利用可能であると OPP は予想している。ヒトでの事例データは、より広範囲の化学物質について入手可能であるが、中には正確なデータベースを持つものと持たないものがある。ヒトでの重大な事例があり、特定の農薬の MOA/AOP や PK についてほとんど知られていない場合には、WOE 解析を用いて新たな研究分野を特定することができる。

疫学データとヒトでの事例データを組み込むための当該フレームワークにおける OPP のアプローチは、初めてのものもしくは新しいものではない。むしろ、このアプローチは、既存のツールや方法を合理的かつ論理的に拡張したものである。当該文書は、農薬リスク評価に使用する疫学データ及びヒトでの事例データを検討するための出発点として、既存のガイダンス文書及びフレームワーク (表 1) に依存している。ヒトの健康リスク評価における疫学データ及びヒトでの事例データの使用に関する当該フレームワークは、2009 年の「科学と政策調査」に関する報告書での NRC の提言や、特にリスク管理の必要性の文脈で、利用可能なものを調査、企画、検討するためのツールとして問題の定式化の使用を強調する点に関して、政府機関の最新のヒト健康リスク評価フレームワーク (USEPA, 2014 年 a) と一致している。

同様に、OPP のフレームワークは、問題の定式化の重要性と生物学的体系の多種多様なレベルでの情報統合の必要性を強調した「世界保健機関/化学物質の安全性に関する国際プログラム MOA/ヒト関連(MOA/HR) フレームワーク」の最新版と一致している (Meek ら、2014 年)。MOA/HR フレームワークは、用量反応や時間的一致、ならびに生物学的な妥当性、整合性及び一貫性などの要素を考慮に入れて、Bradford Hill によって記述されたような原則を使用して、エビデンスの重み付けに基づいて確立された因果経路に沿った一連の重要な事象を特定することから始まる (Hill, 1965 年)。この解析アプローチを用いることで、疫学的知見は、他のヒト情報 (ヒトでの事例所見を含む) や実験的研究の文脈で評価することができ、不確実分野や今後の研究分野を特定することができる。しかし、完全に解明された MOA/AOP があることが、ヒト健康リスク評価に疫学研究を利用するための要件ではないことは注目に値する。政府機関が 2007 年及び 2009 年の NRC 報告書にある変化するアプローチの実施を引き続き進め、OPP が疫学情報とヒトでの事例情報の統合の経験を積む中で、OPP は当該フレームワークを適切に再評価し、更新していく予定である。

図 1. 有害性発現経路の概略図。Ankley ら (2010 年) より引用。

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

### Adverse Outcome Pathway



### 有害性発現経路

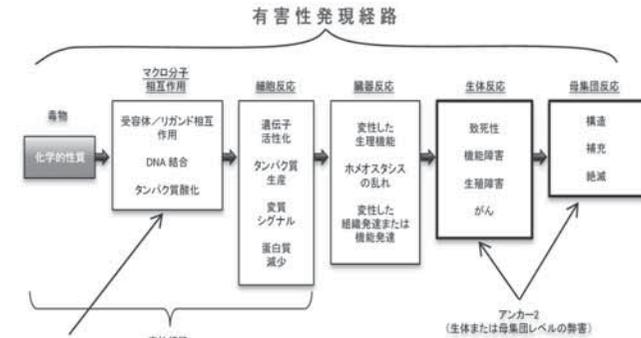


Table 1. Key guidance documents and frameworks used by OPP

<b>NAS</b>	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
	2001-2007: Mode of Action/Human Relevance Framework
<b>WHO/IPCS</b>	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.

表 1. OPP が使用している主要なガイダンス文書とフレームワーク

<b>NAS</b>	1983: 連邦政府におけるリスクアセスメント：プロセスの管理
	1994: 科学と判断
	2007: 21 世紀の毒性試験
	2009: 科学と政策調査：リスクアセスメントの推進
	2011: ホルムアルデヒドに関する NAS 報告書
	2014: EPA の統合リスク情報システム (IRIS) プロセスの見直し
<b>WHO/IPCS</b>	2001~2007: 作用機序/ヒト関連のフレームワーク
	2005: 化学物質固有の補正係数 (CSAF)
	2014: 作用機序/種のコンコルダンス解析に関する WHO/IPCS フレームワークの進化と応用における新たな展開

<b>EPA</b>	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) <sup>4</sup>
	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
<b>OPP</b>	2001: Aggregate risk assessment
	2001 and 2002: Cumulative risk assessment
<b>OECD</b>	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<sup>5</sup> or on studies done to measure dermal or inhalation exposures in agricultural workers as they perform their activities<sup>6,7</sup>.

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

<sup>5</sup> 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.

<sup>6</sup> 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](http://www.epa.gov/scipoly/sap/meetings/2007/010907_mtg.htm))

<sup>7</sup> 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>.

<b>環境保護庁</b>	1991～2005: リスクアセスメントのためのリスクアセスメントフォーラム指針 (発がん性、生殖毒性、発生毒性、神経毒性、生態毒性及びばく露評価のための指針、ベンチマーク用量モデリングのための指針、基準用量と基準濃度プロセスのレビューなど) <sup>4</sup>
	2000: リスクの判定に関する科学政策ハンドブック
	2006b. 生理的薬物動態 (PBPK) モデルのリスク評価への応用のためのアプローチとその裏付けとなるデータ
	2014a. 政策調査に情報提供するためのヒト健康リスク評価のフレームワーク
	2014b. 異種間・同種間の推定のためのデータ由来の推定係数を開発するための定量的データの適用ガイダンス
<b>農業プログラム</b>	2001: 総合的なリスク評価
	2001 と 2002 年: 累積リスク評価
<b>経済協力開発機構</b>	2013: 有害性発現経路の開発と評価に関する経済協力開発機構 (Organisation for Economic Co-operation and Development) ガイダンス文書

ヒトの情報源は他にもあるが、このフレームワークの意図は、ヒトのリスク評価における疫学データとヒトでの事例データの解釈と利用にあり、他にもあるヒトの情報源については当該文書で深く触れていない。特に、当該文書では、意図的なヒトを対象としたばく露に関する農業の研究<sup>5</sup>や、農業従事者の活動に伴う経皮や吸入によるばく露量を測定するために行われた研究<sup>6,7</sup>については、広く論じていない。

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

<sup>5</sup> このような研究の実施及び OPP がこのような研究から得られたデータに依存することは、EPA の研究対象者保護規則 (40 CFR Part 26) によって管理されている。その他に、この規則は妊婦、授乳婦、子供への意図的なばく露に関する研究を禁じ、EPA-OPP やヒト研究評価委員会 (HSRB) による新規研究計画の事前評価、ならびに研究成果報告の評価を求めている。

<sup>6</sup> 過去数年間に OPP は、農業従事者のばく露評価に使用されるデータとアプローチを改善するために、農業従事者を対象とした既存の観察研究を広範囲に評価してきた。それらの評価は ([http://www.epa.gov/scipoly/sap/meetings/2007/010907\\_mtg.htm](http://www.epa.gov/scipoly/sap/meetings/2007/010907_mtg.htm)) で閲覧できる。

<sup>7</sup> OPP が農業従事者ばく露研究をどのように実施し、成果を利用するかについての追加情報は、PPP-48 「農業とヒト健康リスク評価: Policy, Processes, and Procedures」 (<https://www.extension.purdue.edu/extmedia/PPP/PPP-48.pdf> で入手可能) を参照のこと。

### 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).

### III. 農薬のリスク評価におけるシステマティックレビュー：疫学

近年、NRC は、規制上の政策調査に情報を提供するための化学物質固有のリスク評価をサポートする科学的文献レビューの明白性を高めるために、システマティックレビュープロセスに移行するよう政府機関に奨励している (NRC、2011 年、2014 年)。NRC はシステマティックレビューを「特定の疑問に焦点を当て、明確で事前に指定された科学的方法を用いて、類似しているが別個の研究の結果を特定、選択、評価、要約する科学的調査」と定義している (NRC、2014 年)。NRC の勧告に沿って、化学物質安全・公害防止局 (OCSPP) は、政策調査を支える科学的データの収集、評価、統合のための標準的な方法に依存する、目的がなかったシステマティックレビューを採用している。

NRC によると、システマティックレビューには、「明白性があり明確に文書化された方法、関連するすべての文献の一貫した厳しい評価、エビデンスの強さを評価するための標準化されたアプローチの適用、明白で一貫性のある総括的用語 (NRC、2014 年)」といういくつかの共通要素がある。近年、いくつかの集団 (Rooney ら、2014 年； Woodruff and Sutton、2014 年； Hartung、2010 年) が、環境保健科学で使用するためのシステマティックレビューのアプローチを発表している。システマティックレビューでの OCSPP のアプローチは、Cochrane Handbook for Systematic Reviews of Interventions for evidence-based medicine に明記されている原則と、Grading of Recommendations Assessment, Development and Evaluation (GRADE) の原則と一致している。GRADE ガイドラインは、国立環境保健科学研究所 (NIEHS) 健康評価・翻訳局 (OHAT) (Rooney ら、2014 年) とカリフォルニア大学サンディエゴ校 (Woodruff and Sutton、2014 年) が開発した環境保健科学のためのシステマティックレビューアプローチで使用されているものである。コクラン・ハンドブックによると、システマティックレビューの主な特性は以下の通りである。

- 目的が明確に示されており、研究の適格性基準があらかじめ定義されていること。
- 明示的で再現性のある方法論。
- 適格性基準を満たすすべての研究を特定するための系統的検索。
- 特定された研究から得られた知見の妥当性の評価。
- 収録された研究の特性と知見を体系的に提示し、総合的にまとめたもの。

上記の各アプローチは共通のテーマとワークフローを共有しており、科学的な文脈 (例えば、問題の定式化やプロトコル) の記述から始まり、明確な検索戦略の方法を用いた文献レビュー、研究の質の解析 (しばしばバイアスのリスクと呼ばれる)、エビデンス全体の質の評価 (例えば、統合)、そして最終的に結論に至るまでの流れを示している。それぞれのアプローチでは、科学文献の包含/除外、研究の質の評価、研究の質 (例: 高、中、低) の報告のための透明性のある事前に設定された基準を推奨している。各アプローチでは、結論の基礎となるデータ統合のための事前に定められたツールを推奨している。

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.

これまでのところ、リスクアセスメントに関係する人々の間では、システマティックレビューのための単一の命名法は合意されておらず、OCSPP は、より広範な経験が積まれるにつれて用語が進化していくことを期待している。OCSPP は、システマティックレビューのプロセスとワークフローを、問題の定式化から始まり、データ収集、データ評価、データ統合、そして重要なデータギャップを特定した結果の要約と考えている。

科学的な解析は、新しい知識が得られるたびに、繰り返し行われることが多い。

### A. 問題の定式化

NRC の報告書「*Science and Decisions-Advancing Risk Assessment* (科学と意思決定- リスク評価の促進)」において、全米科学アカデミー (NAS) は EPA に対して、政策立案者に有用なリスク評価と関連する科学的解析を開発するよう勧告した。この目標達成のために NRC は、政府機関に対してリスク評価の開発において一層幅広く問題の定式化を用いるよう勧告した。NRC の勧告を受けて、USEPA は「ヒト健康リスク評価フレームワーク」(USEPA, 2014年) を公表し、問題の定式化の重要性を強調している。問題の定式化は、科学者とリスク管理者の間の初期対話を伴うものであり、科学的解析のための規制上の背景を提供し、解析の範囲を定義するのに役立つ。問題の定式化は、評価の規制、政策調査及び政策の背景から導き出され、評価の技術的なアプローチの情報を提供し、考慮すべき主要な要因を体系的に特定するものである。このように、各システマティックレビューの複雑さと範囲は、異なるリスク評価の文脈によってさまざまである。言い換えれば、OCSPP のシステマティックレビューは、問題設定から事前に設定された範囲と目的に基づいた「目的にかなったもの」(NRC, 2009年)として実施される。

問題の定式化には、利用可能な情報とともにデータや科学的情報の主要なギャップを考慮することが含まれる。OPP は、ばく露経路と潜在的な健康影響を特定するためのツールとして問題の定式化を用いており、適切な方法、情報源及び科学的解析のアプローチとともに使用している。欠損データが評価において重要な場合、情報 (例えば、必要とされる試験) を得るための最良の選択肢が議論される。ピアレビュープロセスが特定され、評価を完了するためのタイムラインが定義される。

システマティックレビューは、利用可能な情報を整理し、解析に対する規制目的に関する情報のギャップを特定するための明白なツールを提供する。このように、問題の設定では、科学的解析の規制上の文脈が説明され、それによって科学的文献の収集と評価の範囲と目的が定義される。問題の設定における考慮すべき事項は、集団またはライフステージ、ばく露経路 (例: 経路、期間、頻度) 及び/または *in vitro* または *in vivo* の実験室での研究、疫学またはヒトの事故による研究から特定された目的的健康結果、ならびに利用可能な資源及び規制の時間枠に関することであろう。疫学情報及びヒトでの事例情報を考慮する文脈では、試験の質、試験デザイン及び不確実性の初期評価が考慮される。

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.

問題の設定で考慮されるハザード評価に関連する主な科学的課題は以下の通りである。ばく露に関連する影響とは何か？これらの影響に関連する MOA/AOPs はどのようなものか？影響の時間的側面は？影響を受けやすい集団があるのか、あるとすれば、その集団は誰であり、どのような要因によって影響を受けやすいのか。実験動物とヒトの間で PK や薬力学 (PD) に違いはあるか？また、ばく露情報は問題の設定においても評価される。問題の設定で考慮されるばく露評価に関連する主な科学的課題は以下の通りである。農薬はどのように使用されるのか？ばく露の関連する使用場所のすべてはどのようなものか？ヒトはどのような化学物質にばく露されるのか？ばく露の方法、期間、頻度はどのようなものか？誰がばく露される可能性があるか？ばく露は異なる集団（例えば、年齢や行動パターンのため）に異なるリスクをもたらすのか？疫学データの具体的なケースでは、このレビューでは、研究仮説、研究デザイン（すなわち、サンプルサイズ、十分な対照、測定方法の質など）、ばく露量/ばく露濃度、統計解析及び結論を含むが、これらに限定されない様々な要因を考慮して検討する。

## B. データ収集

システマティックレビューのデータ収集段階では、公開科学文献や農薬登録のために提出された研究論文など、様々な公開・未公開の情報源から利用可能な情報を収集する。OPP では、経済協力開発機構 (OECD) 試験ガイドライン、OCSPP 調和試験ガイドライン及び農薬ガイドライン (OPP ガイドライン) に基づいて収集されたデータをレビューしている。これらのガイドライン研究論文は、主に提出された研究の社内データベースから収集され、そのような社内データベースの検索によって発見される。

疫学の場合、ほとんどの研究は公開されている科学文献に掲載されていることが予想される。場合によっては補足的な解析や情報が得られることもあるが、研究者との対話によって、疫学研究を理解し解釈する上で、原著論文には掲載されていない追加の重要な情報が得られることもある。ヒトでの事例の情報源はセクション IV にまとめられている。

公開文献検索戦略では、特定の基準を用いて、公開されている科学文献や未発表の情報源から健康影響情報を検索する。最も適切な情報源/データベースを特定して選択し、分類コード、医学的テーマの見出し及び/またはキーワードを利用した最もリソース効率の良い戦略を調査した後、文献の検索が行われる。科学的評価の複雑さに応じて、文献管理責任者のサポートが必要な場合もあれば、必要でない場合もある。ヒトの健康に関する文献検索の目標は、文献検索プロセスの適切な文書を提供することにより、信頼性が高く再現性のある文献検索を行うことである。有用な研究を検索するには、以下の手順で行う。

- 科学的解析の目的と包括基準が確立されている。
- 用語/キーワードや MeSH (Medical Subject Heading) 用語の組み合わせと、それらのブール考案の組み合わせ (AND; OR; NOT) を使用し、文書化する。

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<sup>2</sup> 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 (Balslem 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.

必要に応じて、著者名、タイトル、キーワード、主題見出しによる高度な検索やフィールド検索を行うこともできる。データベースの構造を知り、特定のデータベースに対して別の検索戦略を使用することは、有用な研究論文を検索する際に役立つ。最初の包括的な検索に加えて、文献リストを更新するために定期的な検索を行うこともある。

- すべてのデータベースのすべての検索が再現可能であることを保証するために、検索戦略は検索の日付を含めて文書化されている。
- 検索された論文のリファレンスリスト<sup>2</sup>は、追加の文脈を調べたり、最初の検索では発見されなかった論文を探したりするために調査されている。
- 異なるデータベースから検索された論文を組み合わせることで重複を除去した後、利用可能なタイトルと要約を選別する。タイトルと要約から関連性が判断できなかった論文については、その論文を検索し、さらなる再検討を行う。
- 最初の選別の後、有用性のない論文（除外基準）-例えば、持論の論文、英語で記述されていない研究論文、要約のみで構成された論文は、除外される。追加の除外基準は個別的に特定される。すべての除外基準は文書化されている。残りの論文は、有害な健康影響が認められなかったものであっても、検討と評価のために含まれる。

### C. データ評価

データ評価段階では、データの質が検討され、そのようなデータの有用性について結論が出される。研究の質は、報告された結果が正しいという全体的な信頼性を反映する（Balslem ら、2011年）。このように、研究の質には以下のようなものがある。

- 報告の質（研究の報告がどの程度に十分であるか、または完全であるか）。
- 試験のデザインと実施に基づいて、試験結果がどの程度に信頼性のあるものか。
- さらに、その研究が評価下の主題にどの程度に十分な対応しているか（Rooney ら、2014年）。

研究の質は、まず個々の研究基盤で検討され、その質が判断される。例えば、不完全に実施されたコホート研究よりも、適切に実施された症例対照研究の方が信頼性は高い場合がある。科学的知見の信頼性は、しばしばバイアスのリスクと呼ばれ、試験のデザインと実施に関連する特定の分野について、あらかじめ定められた基準を用いて評価される（表 2 参照）。

OPP は当初、「ヒトの健康リスク評価を支援するための公表されている科学文献の毒性研究の検討と使用に関するガイダンス」（Guidance for Considering and Using Open Literature Toxicity Studies to Support Human Health Risk Assessment）（USEPA、2012年）と呼ばれる公表されている科学文献の考えの使用に関するガイダンスを作成し、一般的にはこのガイダンスに従っている。しかし、リスク評価におけるシステムティックレビューの促進に伴い、この文献ガイダンスの一部の側面は将来的に更新が必要になる可能性がある。

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<sup>8</sup>). 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

<sup>8</sup> [https://aopwiki.org/wiki/index.php/Main\\_Page#OECD\\_User\\_Handbook](https://aopwiki.org/wiki/index.php/Main_Page#OECD_User_Handbook)

データの質についての結論が出され、結論の記述またはカテゴリー（例えば、許容可/許容不可；低、中、高）で記述される。

農薬に関する疫学研究の評価に用いられる具体的な考えは、以下のセクション III.C に記載されている。データレビューの一環として、研究の簡潔なレビュー文書を作成する。このレビュー文書には、研究のデザイン、結果、結論と、研究の長所と短所が記載されている。疫学的ばく露評価の質は、疫学的データがリスク評価においてどのような役割を果たすかを調査する上で重要な要素である。そのため、疫学的ばく露評価で使用された仮定と、その仮定が疫学的知見の解釈と一般化の可能性にどの程度影響を与えるかを十分に特性評価することが重要である。疫学的ばく露評価には、過去及び現在のばく露パターン（例えば、ばく露集団、ばく露経路、ばく露方法及びばく露レベル）の検討が含まれ、また、使用パターンの大幅な変化（例えば、リスク軽減措置や新たな使用パターン）が含まれている場合がある。メタ分析の評価に関しては、Stroup ら（2000 年）によって報告された「疫学における観察研究のメタアナリシス」Meta-analysis Of Observational Studies in Epidemiology (MOOSE) のガイドラインが、メタ分析の質の評価や解釈に有用である。

#### D. データ統合：エビデンスの重み付け (WOE)

OPP のヒト健康影響評価では、ヒトでの調査/疫学、バイオモニタリングデータ、*in vitro* の毒性試験及び *in vivo* の毒性試験、MOA/AOP 情報、薬物動態試験、構造活性関係 (SAR) などを含む（但し、これらに限定されない）利用可能なすべての関連データを考慮する必要がある。さまざまな種類のハザードデータを収集し、関連する各試験の完全な評価を実施して記録した後、次のステップは複数のエビデンスを統合することである。

データ統合は、関連するハザードについて利用可能なネガティブデータとポジティブデータを総合的に判断するという原則に基づいている。OPP は、疫学データ及びヒトでの事例データの評価に WOE 解析を使用しているが、結論は、特定の研究に依拠するのではなく、情報の優越性に基づいて出される。OPP は、異なる情報源からの情報を整理して統合するためのツールとして、MOA/HR のような修正された Bradford Hill 基準を使用している（Hill, 1965 年; U.S. EPA, 1999 年, 2005 年; Sonich-Mullin ら, 2001 年; Meek ら, 2003 年; Seed ら, 2005 年; OECD AOP Wiki Users Handbook<sup>8</sup>）。Hill 基準は、チェックボックス方式ではなくて、エビデンスの総合評価の検討を示唆していることに注目することが重要である。また、MOA/AOP が完全に解明されていることは、ヒト健康リスク評価に疫学研究を利用するための必須条件ではない。しかし、完全に解明された MOA/AOP がなくても、メカニズムデータの収集と評価は、生物学的な妥当性の裏付けとなり、生体組織の感受性、生物種、性別、ライフステージ、またはその他の要因での違いを説明するのに役立つ可能性がある。MOA/HR は、厳密ではない情報を整理するための基盤を提供する柔軟なツールである。

<sup>8</sup> [https://aopwiki.org/wiki/index.php/Main\\_Page#OECD\\_User\\_Handbook](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.

これは因果関係の評価、複数の科学的根拠に基づく情報の統合、データのギャップや今後の研究分野の特定などの様々な目的に有用なツールとなる柔軟性を備えている。この解析では、疫学的知見やヒトでの事例データは、生物学的な妥当性を評価するために、他のヒト情報や試験の文脈で評価され、不確定分野や今後の研究分野を特定することができる。Bradford Hill の側面が WOE 評価でどのように考慮されるかを説明するために、OPP は、EPA の全米環境大気質基準 (NAAQS) のレビューのための科学的基盤となる統合科学評価 (ISA) の EPA 前文に記載されている用語の定義を使用している (USEPA, 2015 年)。

- **重要事象。** 特定の健康影響について MOA/AOP が確立されている場合には、MOA/AOP の基礎となる重要事象 (すなわち、測定可能なパラメータ) をそれぞれ明確に記述する。重要事象を示すデータは、*in vitro* または *in vivo* のデータソース (ヒトまたは動物) の組み合わせから得られる可能性がある。これらの重要事象は、PK 事象と PD 事象の組み合わせである。しかし、完全に解明された MOA/AOP を利用できることは、ヒト健康リスク評価に疫学研究を利用するための必須条件ではないことは注目に値する。
- **生物学的勾配/ばく露-反応/用量-反応の一致と関係。** ISAs の前文では、「疫学の文脈では、よく特徴づけられたばく露-反応関係 (例: ばく露量が多いほど影響が増加する) は、特にばく露期間 (例: ばく露時間が長くなると影響が増加する) についても因果関係を強く示唆する」 (USEPA, 2015) と指摘している。MOA/AOP が明らかである場合、各重要事象について用量反応関係が特定される。用量反応関係は、重要事象間で比較される。場合によっては、初期の重要事象の方が後期の重要事象よりも感度が高い場合がある。他の場合では、重要事象は類似した量反応曲線を共有しているかもしれない。
- **時間的関連付け。** 薬剤の投与と影響の出現との間に時間的な順序があるというエビデンスは、因果関係を支持するもう 1 つの根拠となる (USEPA, 2015 年)。ISAs の前文では、「ばく露の変化が健康影響の発現または発現頻度の変化をもたらすことが判明した場合には、『自然実験』を通じて因果関係を示す強固なエビデンスを提供することができる」と指摘している。

この解析では、影響の整合性を確保するために急速に発現する重要事象 (例えば、ばく露後数分以内に発現する活性代謝物への代謝) と長い継続期間の後に発現する重要事象 (例えば、腫瘍の発生) を考慮している。疫学データを検討する際には、ばく露と健康影響の間の時間的關係が考慮される。

- **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., epidemiologic, controlled 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

- **強固性、一致性、特異度。**

**一致性。**複数の独立した研究でリスクの上昇が観察された場合に、因果関係の推測は強固になる。結果の再現性は因果関係を示す最も強力な根拠の一つである。統計的な有意性は、効果の有無を判断する唯一の基準ではない。研究間で不一致な結果がある場合は、ばく露の違い、交絡因子、研究の検出力などの理由が考えられる (USEPA, 2015年)。

複数の研究における結果の一貫性は、複数の独立した研究における効果または関連付けの繰り返しの観察から得られる。さらに、異なる状況下での異なる母集団における結果の再現性によっても支持される。しかし、独立した調査の間での結果の不一致は、調査方法の違い、ランダムエラー、ばく露、交絡因子、または研究の検出力の違いによって説明される可能性があり、因果関係を否定するために使用することはできない (USEPA, 2015年)。

**観察された関連付けの強固性。**大規模で正確なリスクの発見は、関連付けが偶然、偏りまたはその他の要因によるものではない可能性が高いという確信を得られる。しかし、影響の推定が小さい場合は、母集団における実質的な影響を表す場合もあれば、そうでない場合もあることに注意が必要である (USEPA, 2015年)。

**観察された関連付けの特異度。**特定の結果をばく露に結びつけるエビデンスは、因果関係を示す強い根拠となる。しかし、環境ばく露が結果の発生を必ず予測できることは稀であり、結果には複数の原因がある可能性を認識しなければならない (USEPA, 2015年)。

- **生物学的な妥当性と整合性。**

**整合性。**1つのエビデンス (例えば、疫学的に管理されたヒトのばく露研究、毒性試験、または生態毒性試験) からの因果関係の推測は、関連付けの因果関係の解釈を支持する他のエビデンスによって強化されることがある。1つの科学的エビデンスグループの中で、さまざまな分野にわたるエビデンス、複数の試験デザインにわたるエビデンス、または健康影響に関連するエビデンスから、効果を示す一貫性があるかもしれない (USEPA, 2015年)。

動物でのデータとヒトでのデータが、定量的にも定性的にも同様の毒性プロファイルを示している場合、ヒト健康リスク評価の信頼性は高い。一方、動物でのデータとヒトでのデータが定性的には類似した毒性プロファイルを示していても、定量的な相違が観察される場合もある。例えば、ある化学物質が動物とヒトでは同じMOA/AOPを示していても、用量反応特性に種差がある場合がある。このような用量反応性の相違は、生体組織薬量測定 (すなわちPK) に起因するか、または異なる反応特性 (すなわちPD) に起因する可能性がある。一方、動物でのデータとヒトでのデータは、場合によっては質的に異なる結果を示すことがある。このような状況から、種、性別及びライフステージの違いや潜在的な感受性などの要因を考慮するために、入手可能なすべての情報を透明性のある包括的な方法で十分かつ客観的に評価する必要性が明らかになった。

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

**生物学的妥当性。** 因果関係の推測は、生物学的に妥当なメカニズムを示す試験またはその他の情報源からの結果によって強化される。提案されたメカニズムは、実験的エビデンスに基づいており、ある物質へのばく露を所定の効果に結びつけるもので、因果関係を支持する重要な情報源である (USEPA, 2015 年)。

同様に、多くの構造的類似物の 1 つとしての化学物質の MOA/AOP に関する情報は、因果関係の可能性が高いかどうかの判断に役立てることができる。構造活性相関及び薬剤の構造的類似物に関する情報は、関連性が因果関係にあるかどうかについての洞察が提供される (USEPA, 2015 年)。

EPA のがんガイドライン (2005 年) は以下の通りに指摘している。

「疫学研究で観察された関連付けの生物学的妥当性の評価は、潜在的な作用機序 (MOA) の特定に関連する毒性学的エビデンスとばく露関連因子の両方を考慮したものである。同様に、毒性学的文献で報告された健康影響の関連付けの整合性を考慮することは、毒性試験や疫学研究で評価された生物学的マーカーの特性に関する情報を幅広く考慮することを示している [p.39]。」

しかし、このガイドラインはさらに、「しかしながら、メカニズムデータの欠如は因果関係を否定する理由にはならない [p.41] 」と述べている。このように、確立された MOA/AOP の欠如は、疫学データを使用する際に必要な知識ではなく、確立された MOA/AOP がなくても疫学的関連付けは有効である可能性があり、また、潜在的な MOA/AOP について洞察も得られる可能性がある。

- **不確実性。** 不確実性は WOE で明白かつ客観的に述べられている。

#### **E. 全体的な結論、リスク評価の推奨事項、信頼性と不確実性の分野の陳述**

エビデンスの要約、エビデンスの重み付けに使用した手順または方法、WOE の結論または勧告の根拠、不確実性及び更なる研究分野を文書化することが重要である。また、リスク評価における疫学的データまたはヒト由来のデータの役割についても推奨されている。一般的に、OPP はヒトでの事例情報を定量的リスク評価には使用しないが、その代わりに、新たなリスク評価や新たなリスク管理措置の必要性を示すこと、リスク軽減措置が実施された後の成功度を評価すること及び可能性のある施行活動を対象とすることなど、リスク評価/リスク管理の活動を対象とする。

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

ヒトでの事例データの役割がより限定的であるのとは対照的に、疫学研究は様々な方法でリスク評価の複数の構成要素に情報を提供するのに役立つ可能性を持っている。正確なばく露評価を行った質の高い研究は、リスク指標を定量的に推定するために使用することができる。

また、疫学的研究で報告された結果は、*in vitro* 試験や動物試験で見られた結果と定性的に比較して動物試験の結果のヒトへの関連性を評価することもでき (Hertz-Picciotto, 1995 年)、疫学的結果の生物学的妥当性を評価することにも有用である。提案されている WOE 解析の最後の部分では、全体的な結論と信頼性と不確実性の分野の記述が行われる。このセクションでは、追加研究の分野も特定している。本セクションでは、規制値の情報源及び (必要に応じて) 種間及びヒトの間での外挿法での推定のための適切なアプローチを推奨している。

#### IV. 農薬のリスク評価に用いるための疫学研究のレビュー

##### A. 緒言

疫学は、化学的、物理的、または生物学的要因へのばく露と集団の健康状態との関係を特定し、評価することを目指す科学である (Boyes ら、2007 年)。疫学とは、「集団における疾病の分布と、その分布に影響を与えたり、調査したりする要因の研究」と定義されている (Gordis, 2009 年)。より広義には、「特定の集団における健康に関連する事象、状態、プロセスの発生と分布の研究であり、そのようなプロセスに影響を与える調査要因の研究と、この知識を関連する健康問題の制御に応用することを含む」 (Porta, 2014 年) と考えられている。多くの疫学研究の目的は、病気の発生に対する潜在的な原因の影響を有効かつ正確に推定することである。疫学の主な目的は、他の科学と同様に、原因と結果を明らかにすることであり、別の言い方をすれば、病気や健康影響の病因とそれに関連する可能性のあるリスク因子を特定することである。Calderon (2000 年) は、このような研究の 4 つの主要な用途を説明している。1) 集団の健康状態を記述し、疾病やばく露程度の重要な時間的傾向を発見すること、2) 特定の疾病や傾向に関連する因子を特定することで疾病の発生を説明すること、3) 特定の集団における疾病の発生数や健康状態の分布を予測すること、4) 環境やヒト健康に悪影響を及ぼす因子を特定することで集団の健康状態を改善すること、である。農薬の場合、疫学は、一般集団と、職業上ばく露される農業従事者や農薬散布者などの特定の小集団におけるばく露と健康への悪影響との関係に焦点を当てている。

疫学研究は、様々な方法でリスク評価の複数の要素に情報を提供するのに役立つ可能性がある。正確なばく露評価を行った質の高い研究は、リスクを定量的に推定するために、あるいはオッズ比やリスク比のような適切なリスク・サロゲート (代理) を用いることができる。しかし、農薬や農薬ばく露を扱う疫学研究の多くは、規模、範囲、ばく露評価、またはデータ解析になんらかの制限があり、それが定量的リスク評価への十分な利用を妨げている (Ntzani ら、2013 年)。

(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

米国での農薬使用は、ここ数十年で大きく変化しました。使用が変化すると、農業従事者へのばく露も変化します。農薬使用の変化は、EPA によるリスク軽減措置、抵抗性管理活動、新しい化学物質の導入、遺伝子組換え作物の使用量の増加によって生じています。疫学研究を解釈し、最終的には定量的リスク評価にそのような研究を使用するかどうかを調査する際には、ばく露におけるこれらの大きな変化を考慮しなければならない。そうであっても、疫学研究は、実験的動物研究から得られたエビデンスと比較して、そのような値を導き出す際の仮定を特徴づけるために使用することが可能である。また、疫学研究で報告された結果を *in vitro* 試験や動物試験で得られた結果と定性的に比較して、動物実験で得られた結果の生物学的妥当性や人間との関連性を評価する場合もある (Hertz-Picciotto, 1995 年)。疫学研究で得られたようなヒトでの情報は、NRC (2007 年) が推奨する毒性試験の新しいビジョンにおいて重要な役割を果たす可能性があるとして期待されている。具体的には、疫学研究は、ヒトの実社会での化学物質ばく露から生じる可能性のある健康影響についての知見を提供し、その結果、問題の設定やハザード/リスクの特性評価に貢献することができる。ヒトでの情報は、追加的な研究 (例えば、*in vitro* 試験や目標とする *in vivo* 試験で使用するための用量やエンドポイントの選択) の指針となり、今後の研究で調査される新たな健康影響や感受性を特定することができる。

動物実験から得られた実験データがハザードの特性評価のための主要な情報源となる場合、潜在的な不確実性の原因の一つとして、動物モデルのヒトへの関連性が挙げられる。これに反するデータがない場合には、動物の結果はヒトとの関連性があると想定される。さらに EPA はヒトは実験動物よりも感受性が高いと仮定しているが、それを裏付けるデータがない。実際には、ヒトは他の動物種よりも農薬に対して感受性が高い場合も低い場合もある。疫学データ及びヒトでの事例データは、科学的な情報を提供し、外挿法での推定に関連する不確実性を伝えるための裏付けとなり得る。母集団の変動性に関しては、動物試験よりも疫学研究の方が潜在的な変動性をの特徴をよく表している。具体的には、疫学データは、ヒトの集団に固有の遺伝的多様性と変動性を含むため、実験動物よりも環境化学物質に対する実際の集団での反応をよりよく説明し、代表することができる (Calderon, 2000 年)。

用量反応の特性評価に関して、動物の毒性試験には、広範囲のばく露レベルをカバーするように試験デザインできるという利点がある。しかし、動物の毒性試験では、一般的に環境中でのばく露レベルよりもはるかに高い (時には桁違いの) ばく露レベルが用いられている。動物試験では、このような高いばく露レベルがあるため、高用量から低用量の外挿法での推定が必要である。この外挿法での推定はリスク評価に不確実性をもたらす。疫学研究データやヒトでの事例データは、ヒトの実社会でのばく露を含んでいるため、多くの場合、高用量の外挿法での推定は必要ないと考えられる。疫学研究は低用量から高用量までの広い範囲のばく露が最も有用である。

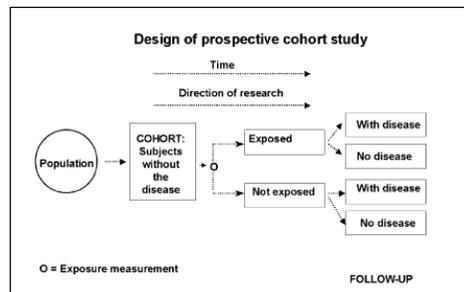
動物試験では、ヒトが経験するばく露の時間、程度、期間、ばく露経路及びばく露の変動性を再現するものではない (Calderon, 2000 年)。ヒトでのばく露は、食品、水、空気、屋内外の環境を含む多岐のばく露経路を介して行われることが多い。対照的に、制御された実験室での研究では、通常、単一のばく露経路である。

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; Matusner 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)<sup>9</sup>. 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<sup>10</sup>. 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



<sup>9</sup> 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)

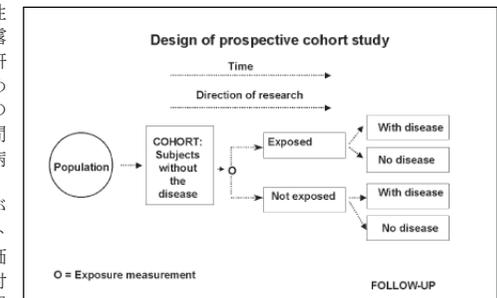
<sup>10</sup> Cohort studies commonly study differences in rates of disease, but these can also include other focal outcomes of interest such as birth weight, mental abilities, blood pressure, etc.

さらに、ほとんどの動物試験では単一の化学的ストレス因子を用いているのに対し、ヒトでは複数の化学物質及び/または非化学的ストレス因子へのばく露を同時に受けている可能性がある。一方で、疫学研究における複数の化学物質へのばく露は、疫学研究の結果が単一の農業に帰属させようとする場合に問題となり得る。一方で、疫学研究は実社会のばく露を考慮しており、実験的アプローチと併せて考慮することで、実験環境では評価が困難な複数の化学物質ばく露に関連した問題に対処するのに役立つ可能性がある。

## B. 疫学研究の種類

観察疫学研究の主なタイプを、その長所と短所を考慮しながら以下に簡単に説明する (Lilienfeld and Lilienfeld, 1979 年; Matusner and Kramer, 1985 年; Kelsey ら、1996 年; Rothman and Greenland, 2012 年; Paddle and Harrington, 2000 年; USEPA, 2005 年; Purdue Pesticide Programs, PPP-43)。

**コホート研究**では、共通の特性を共有する人々の集団 (コホート) から始まり、病気の発生が記録されている長期の追跡期間にわたって健康状態を評価を行う (van den Brandt ら (2002 年) から引用の図を参照)。共通の特性は、しばしば「リスク因子」(ばく露など)<sup>9</sup>の有無である。このような研究では疾病発生率の違いを調べるために、「ばく露者」と「非ばく露者」の間の疾病発生率の違いを特定し、時間をかけて調査する<sup>10</sup>。次に、この疾病発生率の差を調査して、「ばく露群」と「非ばく露群」の間で疾病発生率が異なるかどうかを判断する。コホート研究は、複数の疾病発生を同時に評価する能力を持っている (これは症例対照研究では当てはまらない。症例対照研究では、一般的には後述するように単一の (事前に指定された) 疾病発生のみを評価することに限定される)。



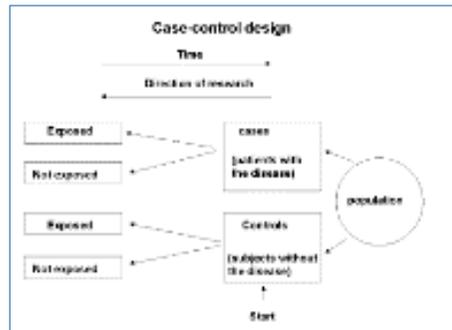
コホート研究は、Agricultural Health Study (AHS, <http://aghealth.nci.nih.gov/>) のように前向きで実施することもできるし、過去の記録から後ろ向きに実施することもできる。前向きコホート研究は、現在の時点から将来の時点までの人々の集団に焦点を当てている。後ろ向きコホート研究は、過去のある時点でばく露された集団に焦点を当て、ばく露が発生した後の疾病率を比較するものである (一般的には、一人 (個人) 単位で利用できる既存のばく露データベース (または記録) を利用)。

<sup>9</sup> コホート研究では、ばく露はしばしばばく露者と非ばく露者に二分されるが、ばく露は定量的尺度 (例えば、連続的尺度または定量値) で測定することも可能である。

<sup>10</sup> コホート研究では、一般的に疾病発生率の違いを研究しているが、これらの研究には、出生時の体重、知能、血圧など、関係のある他の限局的な結果も含まれる。

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<sup>11</sup>) 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.

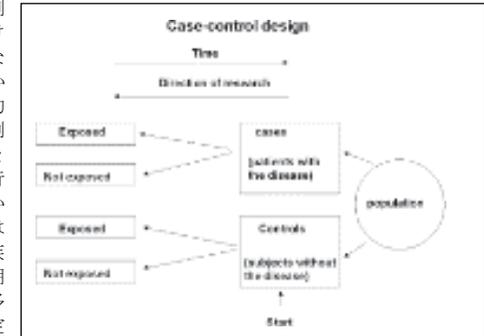


**Nested case-control studies** are an example of a hybrid design and contain the elements of a cohort and a case-control study. These designs can be useful when the analytical costs for determining pesticide exposure are too high for the entire cohort to be studied. For example, a cases that that have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

<sup>11</sup> 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.

前向きコホート研究は、特に希少疾病の場合、実施に比較的時間がかかり、費用がかかることがあり、研究対象となる被験者の数が多いことが必要である。重要なことは、質の高いデータを収集するためには、長期間にわたって多大な資源と専門スタッフが必要であるということである。

**症例対照研究**は、特定の疾病を持つ（症例）個人と、そうでない（対照）個人の集団を特定し、ばく露（一般的には過去<sup>11</sup>のばく露）に関して比較することで、関係ある疾病を持つ人が、関係ある薬剤あるいは要因にばく露されていたか否かを判断する研究である。すなわち、症例対照研究の解析では、症例において関係ある薬剤または要因のばく露頻度と、対照群におけるばく露頻度を対比させて、これらが異なるかどうか、要するに関連があるかどうかを判断する。症例対照研究では、一般的に、疾病状態（すなわち、疾病を持つ症例と疾病を持たない対照群）を決定することが、通常、ばく露状態の調査に先立って行われる（van den Brandt ら、(2002年)から引用の図を参照）。この研究デザインは症例対照研究に選択された時点ですでに疾病が発生しているので、希な疾病や潜伏期間の長い疾病の研究で特に有用であり、多くのばく露と関係ある特定の（事前に指定された）疾病発生との関連性を評価するために利用することができる。また、症例対照研究は、その疾病に罹患している個人を対象としているので、コホート研究に比べて被験者数が少なく、比較的短い期間で終了することができる。症例対照研究の問題は、適切な対照群の選択と、疾病が診断されるずっと前に発生した可能性のあるばく露の評価である（Rothman, 2012年; Wacholder ら、1992年 a; Wacholder ら、1992年 b; Wacholder ら、1992年 c; Shultz and Grimes、2002年; Grimes and Schultz、2005年）。症例対照研究は、特に“想起バイアス”の影響を受けやすい。これは、罹患した人は、対照あるいは健康人と比べてばく露や出来事を違った形で（一般的にはよく）覚えている可能性がある。



**コホート内症例対照研究**は、ハイブリッドデザインの一例であり、コホート研究と症例対照研究の要素を含んでいる。この研究デザインは、農薬ばく露を調査するための解析コストが、高すぎて、コホート全体を調査することができない場合に有用である。

<sup>11</sup> 時間をかけて症例を研究に登録した状況下では研究開始後まで発症していない症例で前向きに症例対照研究を実施することは可能である。

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”.<sup>12</sup> 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<sup>13</sup>. 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

<sup>12</sup> 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.

<sup>13</sup> 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 measured at an individual level with the researcher making inferences to the individual level.

例えば、現在進行中のコホート研究で疾病または健康影響を生じた症例は、解析時に関係ある疾病や影響を生じていない研究からの適切な対照群と組み合わせることが可能である。コホート内症例対照研究（標準的な症例対照研究とは対照的に）の利点として認識されているのは、選択バイアスと想起バイアスの問題が最小限に抑えられることである。

**横断研究**では、疾病（例えば、先天性異常、胎児発育遅延（SGA））、症状、生物学的/身体的及び生理学的反応の測定値（例えば、肺機能検査、血圧、胸部X線、臨床検査、肝臓及び腎臓のバイオマーカー）の有病率に焦点を当てている。このような研究の主な特徴は、サンプリング時またはサンプリング期間中に疾病の有無を判定し、有病率を頻度の尺度として重視した観察研究であるということである。有病率とは、ある集団の中で罹患している人の割合であり、「点有病率」または「期間有病率」として調査される<sup>12</sup>。有病率は比率ではなく、割合であるため、横断研究では追跡調査期間を必要としない。一般的に、ばく露状態（例えば、ばく露または非ばく露）、疾病の状態/発生、人口統計学的特性は、ある時点（またはその時点以後）で調査される。この研究デザインにおける主な比較は、ばく露集団における疾病発生の有病率と非ばく露集団における疾病発生の有病率の比較であり、リスク尺度は有病率リスク比またはオッズ比である。横断研究は、一般的に、長年にわたる疾病発生または地理的に異なる場所のパターンや傾向を明らかにするために使用され、迅速かつ比較的安価に実施することができる。しかし、これらの調査では、疾病の発生率（新規症例の発生率）と疾病の持続期間の両方に影響される疾病発生の有病率を測定しているため、これらの要因を十分に分離することは、どのような解析においても困難な場合がある。したがって、これらの研究は「生存者集団」を対象としており、病気になったために関係ある集団から離れてしまった人を測定、評価、考慮することはない。横断研究のもう一つの重要な限界は、ばく露が疾病に先行しているかどうかを判断できないことである。そのため、横断研究では疾病とばく露の時間的関係を確立することができず、横断研究で唆された仮説に基づく因果関係を確認するために、追加の研究を必要とするのが一般的である。

**生態学的研究**では、グループまたは集団のレベルでのデータを反映した情報を用いてばく露や疾病のパターンを調べる。生態学的研究では、解析の単位は個人ではなく集団である<sup>13</sup>。ここでは、被験者の集団がサンプリングされ、ばく露、疾病及び潜在的な交絡因子を、この集団（またはクラスター）レベルで測定される。集団は一般的に地理的、行政的、組織的な単位（例えば、地区、町、郡、学校、職場など）で定義され、すべてのばく露、疾病、交絡因子の測定は、個人のレベルではなく、集団のレベルで行われたり、まとめられたりする。生態学的（集団ベース）研究は、個人レベルの研究とは対照的であり、前者では症例がばく露された実際の個人であるかどうかの情報がないのに対し、後者ではばく露情報が個人に結びついている。

<sup>12</sup> 前者は特定の場所及び/または特定の時間での測定であるのに対し、後者は特定の期間における症例の割合の調査である。

<sup>13</sup> 一部の研究では、ばく露または疾病発生のいずれかが集団のレベルで測定されるが、他の変数は個人のレベルで測定され、研究者は個人のレベルで推測を行うという「部分的に生態学的」研究もある。

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 been 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 as 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

例えば、水中の汚染物質レベルによる疾病率の研究は、ばく露の評価に関しては生態学的であるが、健康影響や疾病状態は個人で調査されている場合がある。このような場合、ばく露は集団レベルで決定されるが、結果は個人レベルで調査される場合に「半生態学的」という用語が使用されることがある。

この研究デザインでは、ばく露集団のすべてのメンバーが個別にばく露されているかどうか（または個人のばく露レベル）を知ることはできないし、グループレベルの影響から個人レベルの影響を推測することもできない。研究の目的が（個人ではなく）*集団*への直接的な推測を目的としている場合には、これは問題ではなく、特に個人レベルでの測定が制約されていたり、困難であったり、集団内のばく露が一般的に均質であったりする場合には、このような研究が適切であると考えられる。研究の意図が個人への直接的な推測である場合、この研究デザインは生態学的誤謬と呼ばれる問題を抱えていることになる。集合またはグループ化されたデータセットで観察された関係が、個人レベルでサンプリングが行われていたら観察されていたであろうことを反映するという仮定である。この生態学的誤謬の問題に加えて、交絡因子に関する情報が集団レベルでしか得られない場合、集団とは対照的に個人レベルで交絡変数を適切にコントロールできない結果、追加のバイアスが発生する。

ほとんどの場合、生態学的研究は仮説を生み出す研究と考えられており、今後の研究のための研究仮説を提案するのに最適であり、問題の提起に貢献する可能性がある。とはいえ、生態学的研究をデザインの質に基づいて評価することは重要であり、デザインが優れていれば、生態学的研究から有用な情報を得ることができる（FIFRA SAP, 2010年）。一般的に、生態学的研究だけでは、因果関係を立証する能力はない。これらの研究を他の研究と併用した場合、特定の状況下では有用となる可能性があり、ハザード特性評価の際に留意すべきである。特に、安定した母集団、明確なばく露の対比、リスクの大きな差は、これらの研究の有用性を高める重要な要因となり得る。

### C. 農業のリスク評価に用いるための疫学研究の評価

OPP は、化学物質の使用に関連した潜在的な急性及び慢性の健康影響に関する観察疫学研究を査読済み文献で検索する。文献検索の手順と戦略に関する詳細は、別の場所で提供されている。コホート研究、症例対照研究、または横断研究のデザインを利用した疫学研究は、潜在的なハザード、ばく露反応の特性、ばく露シナリオ、または評価方法、そして最終的にはリスクの特性評価に関する OPP の理解を深めるための情報を提供することができる（van den Brandt, 2002年）。さらに、説得力のある症例報告や症例集積解析は、これまで明らかにされていなかった健康影響や作用機序を明らかにする可能性がある。

一般的には、OPP のリスク評価で用いられる公開文献からの疫学研究を評価する際には、疫学研究の質、研究の文書化の十分さ（研究デザインと結果）、リスク評価との関連性が考慮される。

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<sup>14</sup>], 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](http://www.strobe-statement.org) and Vandenberg 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.

<sup>14</sup> <http://ntp.niehs.nih.gov/pubhealth/roc/index.html>

これらの基準は、懸念されるエンドポイントに応じて様々な方法論の詳細が重要になったり、ならなかったりするため、エンドポイントを特定することが重要である。疾病発生の確認に影響を与える関連因子を理解することが重要である（例えば、影響の有無を示す検査やバイオマーカーが利用できるか、あるいは初期段階では非特異的な症状であり、進行した病状では医師の診断につながるかなど）。さらに、環境及び職域疫学研究では、ばく露評価の質が極めて重要である。検討中の問題に対して、ばく露と交絡要因の評価の側面を事前に考慮しなければならない。

個々の研究の質を考える際には、疫学研究のデザイン、実施、解析、解釈の様々な側面が重要である。これらには以下が含まれる。

1. 仮説を明確に明示することで、たとえその研究が本質的に仮説生成的なものであったとしても、その仮説を明確に示されていること。
2. 健康影響の関連する重大な時期、リスク評価対象集団の関係あるばく露範囲、試験から得られる用量/ばく露反応の傾向の入手可能性など、ばく露評価の資質の中で、適切なばく露評価が十分であること。
3. 合理的に有効で信頼性の高い結果の確認（研究集団における健康影響の有無を正しく識別されていること）。
4. 対象集団を代表するサンプル集団となり、系統的な偏りがない適切な組み入れ基準と除外基準。
5. 観察されたリスク推定値における複数の農薬ばく露、または混合物ばく露の役割の評価または考察を含む、潜在的な交絡変数の適切な評価及び解析。
6. 参加者の選択や情報収集における誤りを含む、研究における潜在的な系統的な偏りの全体的な特性。これには、提示されたリスク推定値に対する系統的誤差の潜在的な影響を調査するための感度分析の実施を含む（例：Greenland's 公式）。
7. 健康影響を観察するための研究の統計的検出力の評価と適切な考察及び/または説明。
8. 研究デザインと対象となる結果の性質を考慮した適切な統計的モデル化技術の使用。

他の連邦及び非連邦の機関が次のようなガイドを提供している（例えば、OHAT、ナビゲーションガイド、National Toxicology Program [NTP] Report on Carcinogens [ROC<sup>14</sup>]、IRIS、Cochrane ACROBAT-Non-Randomized Studies of Interventions）（Sterne ら、2015 年）、及び観察疫学研究に関する STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) 声明 ([www.strobe-statement.org](http://www.strobe-statement.org), Vandenberg ら、2007 年; Von Elm、2014 年) を参照)。OPP が疫学研究をヒトの健康リスク評価に統合する経験を積みにつれ、評価アプローチに関連する調整が行われることになる。

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.

独立した研究評価は、ハザードの特定及びばく露反応評価に有益である詳細な要約表を作成する前に実施され、文書化されている。表 2 は、評価された主な考慮事項と、各考慮事項に関連する重み（低、中、高）の構造を示している。表 2 は考慮事項の一般的な検討事項を示したものであり、チェックリストと考えるべきではない。特定の科学的解析に適した特定の科学的考慮事項は、ケース・バイ・ケースで調整される。

研究評価プロセスの最終段階では、研究固有の評価を通じて特定された各研究における系統的な誤りから生じる可能性のあるバイアスの性質についての専門家／エキスパートの意見を提供し、疫学データベースに対する全体的な信頼性の評価を行うことである。このようにして、データ統合（動物、ヒト、メカニズム、その他）は、環境ばく露や職業上ばく露のヒト健康影響を伝えるヒト疫学研究の信頼度に基づいて行われることになる。

**Table 2. Study Quality Considerations <sup>a</sup> (Adapted from Munoz-Quezada et al., 2013; LaKind et al., 2014)**

Parameter	High	Moderate	Low
<b>Exposure assessment</b>	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
<b>Outcome Assessment</b>	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
<b>Confounder control</b>	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
<b>Statistical Analysis</b>	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
<b>Risk of (other) bias (selection, differential misclassification, effect size magnification, other)</b>	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

<sup>a</sup> Overall study quality ranking based on comprehensive assessment across the parameters.

**表 2. 研究の質に関する考慮事項 <sup>a</sup> (Munoz-Quezada ら、2013 年; LaKind ら、2014 年より引用)**

パラメータ	高	中	低
<b>ばく露評価</b>	MOA/AOP に関連している外部ばく露、内部ばく露量、または目標ばく露量との正確で精密な定量関係。  アンケートを利用した場合は、化学物質固有のばく露についての調査対象者からのアンケート及び/またはインタビューの回答。	指定されたマトリックス内のバイオマーカーと外部ばく露、内部ばく露量、または目標ばく露量との間の関係についてのエビデンス。  調査対象者または代理人が回答した化学物質固有のばく露に関するアンケート及び/またはインタビュー。	わずかなサロゲート (代理)  質の低いアンケート及び/またはインタビュー; 化学物質固有の情報ではなく、化学物質群について収集された情報; 化学物質固有のばく露情報が収集されていない; 農薬の使用経験の有無に関する評価。
<b>発現事象評価</b>	調査対象集団で有効性が確認されている標準化ツール; 訓練を受けたスタッフによるカルテのレビュー/診断の確認; 有病率/症例の発生率に関する適切な評価。	母集団で有効性が確認されていない標準化ツール、またはスクリーニングツール; または、方法が明記されていないカルテのレビュー。	検査の選択された部分、または母親の報告、その他; または母親/父親の自己報告; 有病症例または事例症例が適切であるかどうかが不明確/不評価。
<b>交絡因子コントロール</b>	科学的な疑問に関連する重要な交絡因子及び標準的な交絡因子を良好に制御。	科学的な疑問に関連する交絡因子、標準変数 (すべての変数ではない) を中程度に良好に制御。	調整なしで多変数解析は実施していない; 不層化、不制限、または不調和。
<b>統計解析</b>	研究の課題とデザインが適切であること、適切なサンプル規模に支えられていること、データを最大限に活用していること (選択的ではない)。	容認可能な方法、疑わしい調査力 (特にサブアナリシス)、情報が無く明確に報告されていない解析の選択。	統計解析への最小限の注意、比較の不実施、または不明記。
<b>(その他の) バイアスのリスク (選択、差のある誤分類、効果の規模、その他)</b>	その他の潜在的なバイアスの主な原因は、存在する可能性は低く、存在するがリスク推定の規模と方向性に影響を与える可能性は低く解析される。	その他のバイアスの原因は、存在しても研究では取り上げられていないが、推定の方向性ではなく、推定の規模に影響を与える可能性がある。	主な研究バイアスは、存在するが研究では取り上げられておらず、研究の結果について他の説明を排除することはできない。

<sup>a</sup> パラメータ全体の総合的な評価に基づく総合的な研究の質のランキング。

## 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 (Niewenhuisen, 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

## 1. ばく露評価

ばく露評価は、「ある物質へのばく露の規模、頻度及び期間を、ばく露された集団の数と特性とともに推定または評価するプロセス」と定義することができる。理想的には、評価において原因、経路、方法及び不確実性を記述すると記載 (Zartarian ら、2005 年) されている。環境疫学において、ばく露評価は、特に環境媒体中に低濃度で存在する毒性物質については、独特の課題となっている (NRC、1991 年; NRC、1997 年)。ばく露経路の複雑さを考慮して、研究者はばく露を評価するための多くの異なるアプローチを開発してきたが、その正確さ、精度及び必要な資源は様々である (Niewenhuisen、2003 年)。これらのアプローチのいくつかは、疫学研究に特化したものではないが、様々な科学的解析におけるばく露評価の情報を提供するために使用される可能性がある。これらのアプローチには、過去の記録、アンケート、環境モニタリングに基づく間接的な方法と、個人モニタリングやバイオモニタリングに基づく直接的な方法がある。それぞれの方法の簡単な説明とその長所と限界を以下に要約する。

**表 3. 間接的ばく露評価方法と直接的ばく露評価方法のまとめ**

アプローチ	方法/ツール	例	ばく露の推定
間接的	過去の記録	住所情報を利用した農地への近接推定	二分法または順序のばく露
	アンケート	農薬使用の回答に基づくばく露の可能性の調査	二分法または順序のばく露
	環境モニタリング	地域の水道における農薬レベルの測定	モデル化を用いてばく露を推定することができるが、二分法または順序のばく露
直接的	個人モニタリング	吸入と経皮でばく露された農薬の測定	定量化されたばく露
	バイオモニタリング	血中・尿中の農薬の測定	定量化された内部ばく露量

**過去の記録やアンケート**は、化学物質へのばく露に関連する可能性のある主要な特性を評価するために使用される。疫学研究で使用される場合、過去の記録やアンケートは通常、定量的なばく露レベルを予測するためには使用されない。むしろ、過去の記録情報やアンケートの回答は、ばく露のカテゴリーレベルを調査するために使用される。ばく露レベルを調査するために使用できる過去の記録情報の例としては、農地に近接した住所や、職位や職歴に関する雇用履歴情報などが挙げられる。同様に、アンケートは、個人が農薬を使用したことを思い出すかどうかを調査するために、またはばく露の可能性を高める特定の職務を行っている個人を特定するために使用される。過去の記録やアンケートは、潜在的なばく露に関するデータの費用対効果の高い情報源になり得るが、限界がある。過去の記録やアンケートから収集されたデータは、ばく露のサロゲート (代理) に過ぎない。

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 (Niewenhuisen, 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

その結果、これらの情報源はばく露の単純化しすぎているため、個人のばく露の可能性を正確に並べていない可能性がある。

**環境モニタリング**は、大気、水、土壌、食品、家庭環境や職場環境などの環境媒体中の汚染物質のレベルを特性評価するために使用される。多くの州及び連邦政府のプログラムでは、疫学研究に有用な環境モニタリングデータを収集している。環境モニタリングは、大気汚染や飲料水など、地理的な境界によって定義されるばく露に対して特に有用である。そのため、多くの疫学研究では、大気汚染と飲料水汚染へのばく露を特性評価するために、大気モニタリングデータと地域の水道のデータをそれぞれ利用している。環境モニタリングデータは、地理的境界によって定義されるばく露量を推定するのに有用であるが、個人レベルのばく露量を調査する目的では、特に個人が多くの異なる場所で生活し、仕事をし、時間を過ごす場合には、信頼性が低くなる可能性がある。

**個人モニタリング**は、体の境界の接触点でのばく露を特性評価するために使用される。個人モニタリングの例としては、農薬との経皮接触を評価するための線量計の使用、吸入ばく露を評価するための個人用空気サンプリング装置、食品中の農薬レベルを調査するための対の食品サンプルの収集などがある。個人モニタリングの利点は、間接的な方法よりも個人レベルのばく露量をより正確に推定できる可能性が高いことである。また、個人モニタリングは、異なる経路でのばく露の関連性の優先順位付けに有用なばく露レベルの定量化を可能にする。さらに、個人モニタリングは、時間をかけて測定を繰り返すことで、縦断的なばく露量を評価するためにも使用される。個人的なモニタリングは間接的なアプローチに比べて多くの利点があるが、労力と資源が必要とする傾向がある (Niewenhuisen, 2003 年)。その結果、個人モニタリングを用いてばく露を評価する大規模な疫学研究を実施することは一般的には不可能である。さらに、個人的なモニタリングは、サンプルを取得するために使用される測定技術や解析ツールに大きく依存しており、関連する期間 (通常は過去) のばく露を特性評価する情報が利用可能になる可能性は低い。さらに、対象となる期間のばく露の全容が捕捉される可能性は低く、ピークと変動を捕捉するのに十分な期間でのサンプリングがではないかもしれない。そのため、疫学研究で使用される個人モニタリング手法の科学的厳密さと信頼性を考慮することが非常に重要であり、そのようなモニタリングは他のモニタリング (例えば、環境データ、生物学的データ及び/またはインタビュー/アンケートのデータ) で補完する必要があるかもしれない。

**バイオモニタリング**は、血液、尿、唾液、乳、脂肪、その他の生体組織などの生物学的試料中の化学物質、その代謝物、または反応生成物を測定することであり、ばく露の特性を明らかにするために使用される (Needham ら、2007 年)。Zartarian ら (2005 年) は、「バイオマーカー/生物学的マーカーは、「生物学的システムにおける変化や事象の指標」として定義されてきた」と述べている。「ばく露の生物学的マーカーとは、組織、細胞、体液などの生物学的媒体から得られる細胞や生化学的、分析的、または分子的な測定値を指し、ある物質へのばく露指標である」と述べている。したがって、バイオマーカーは、ばく露を評価に用いられ、また健康影響の指標として使用される (LaKind ら、2014 年)。表 4 は、疫学研究から収集されたバイオモニタリングデータの質と関連性を評価するための科学的考察を示している。

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.

解析ツールがより費用対効果の高いものになり、より多くのバイオマーカーが同定されるようになったため、バイオモニタリングを用いたばく露評価は急速に拡大してきた。自己申告のアンケートやインタビューデータと比較して、バイオモニタリングはばく露の誤分類を減らし、リスク推定の精度を高めることができるかもしれない。同様に、バイオモニタリングは、異なる経路からのばく露を統合し、体内に吸収されるばく露量を調査するために使用されている (Checkoway ら、2004 年)。さらに、疾病の履歴におけるバイオマーカーの役割が知られるようになってきている場合もあり、バイオマーカーはばく露の時間的な問題を解決するのに役立つかもしれない。

バイオモニタリングは他のばく露評価方法に比べて多くの利点があるが、それ自体にも限界がある。多くの研究では、生物学的サンプルはある時点での一点からしか採取されず、特にばく露が大きく変動する場合には、経時的なパターンを正確に反映していない可能性がある。さらに、バイオマーカーの評価には、環境とヒトの両方での化学物質の分解と代謝についての理解も必要である。そのため、ばく露のバイオマーカーはばく露以外の理由で個人間で異なる場合がある。代謝の違い、尿の測定値に関連した腎臓病などの併発疾患やバイオマーカーが有効成分のばく露を測定しているのか、環境中分解物のばく露を測定しているのかといった不確実性が、個人間や想定される比較集団間におけるばく露のバイオマーカーの見かけ上の違いの原因となっているかもしれない。

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

Biomarker Consideration	Tier 1	Tier 2	Tier 3
<b>Exposure biomarker</b>	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
<b>Effect biomarker</b>	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).
<b>Specificity</b>	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.
<b>Method sensitivity</b>	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
<b>Biomarker stability</b>	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.
<b>Sample contamination</b>	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

**表 4. 環境疫学研究におけるバイオモニタリングデータの考察 (LaKind ら、2014 年より引用)**

バイオマーカの考察	ティア 1	ティア 2	ティア 3
<b>ばく露バイオマーカ</b>	バイオマーカは、外部ばく露、内部ばく露量、目標ばく露量と正確かつ精密な定量的関係にある。	バイオマーカは、外部ばく露、内部ばく露量、目標ばく露量との定量的関係が不明であるか、ばく露/用量に関してわずかなサロゲート〔代理〕（正確さ・精度が低い）である。	NA
<b>効果バイオマーカ</b>	MOA/AOP では key events のバイオインジケータ。	健康影響との関係に対する効果のバイオマーカは明らかである。	バイオマーカは結論をもたらさない（例えば、バイオマーカは健康影響を明確にしない）。
<b>特異度</b>	バイオマーカは、1 つの親化合物へのばく露に由来する。	バイオマーカは、類似の毒性を持つ複数の親化合物に由来する。	バイオマーカは、様々な有害エンドポイントがある複数の親化合物に由来する。
<b>手法の感度</b>	検出限界は、研究課題に取り組むのに十分な割合のサンプルから化学物質を検出するのに十分に低い。	研究仮説に取り組むには、検出頻度が低すぎる。	NA
<b>バイオマーカの安定性</b>	既知の履歴と文書化された安定性データを持つサンプル。	サンプルは保存中の損失が知られているが、低量ばく露と高量ばく露の違いを定性的に評価することができる	履歴が不明なサンプル及び/または目的的分析物の安定性データがないサンプル。

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.		
<b>Method requirements</b>	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 interferents (e.g., GC-FID, spectroscopy)
<b>Matrix adjustment</b>	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; PPP = probability of false positive

バイオマーカーの考察	ティア 1	ティア 2	ティア 3
<b>サンプルの汚染</b>	サンプルは、収集時から測定時まで汚染されていない（例えば、分析物の無いことが保証された収集用品及び参照物質の使用及びフィールドとラボの両方でのブランクの適切な使用による）。研究には、研究データの信頼性を保証するために必要な手順の文書化が含まれる。	これらの手順を使用しない/文書化しない研究。	汚染の問題が知られているが、問題を取り上げた文書はない。
<b>手法の要件</b>	バイオマーカーの明確な同定と定量を必要としない装置（例：GC-HRMS、GC-MS/MS、LC-MS/MS）。	高い信頼性と必要な感度でバイオマーカーの同定を可能にする装置（例：GC-MS、GC-ECD）。	バイオマーカーの定量のみが可能であるが、その方法には既知の干渉物質がある装置（例：GC-FID、分光法）
<b>マトリックス調整</b>	研究には、調整濃度と非調整濃度の結果が含まれている。	研究では、1つの方法（マトリックス調整済みかどうか）を用いた結果のみを提供している。	NA

FP = 偽陽性、FN = 偽陰性、GC-HRMS = ガスクロマトグラフィー/高分解能質量分析、GC-MS = ガスクロマトグラフィー/質量分析、GC-ECD = ガスクロマトグラフィー-電子捕獲検出器、GC-FID = ガスクロマトグラフィー-水素炎イオン化型検出器、ICC = クラス内相関係数、NA = 該当しない、PPP = 偽陽性の確率

**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

**間接的ばく露評価**方法は、後ろ向き研究では一般的であり、化学物質のばく露のサロゲート（代理）となる因子に基づいている。上述したように、間接的なばく露データは、追加のモデル化を行わない限り、一般的には定量的なばく露レベルを推定するためには使用できない。例えば、ある個人が農薬を使用したことがあるかどうかを調査するためにアンケートを使用することはできるが、その個人のばく露量を計算するために必要なすべての環境的・行動的要因に関するデータを確実に集めることは困難である。そのため、間接的ばく露データは、二分法のばく露変数（すなわち、ばく露／非ばく露）または順序ばく露尺度を用いてばく露を分類するために使用されることが多い。一方、直接的ばく露評価方法は、個人のモニタリングやバイオモニタリングによる実際の個人レベルのばく露に関するデータに基づいている。したがって、直接法は、個人のばく露または内部ばく露量レベルを推定するために使用することができる。直接法は前向き研究で用いられることが多いが、十分に定義された母集団から既存の生物学的サンプルが入手可能な場合には、後ろ向き研究でも使用される。

**定量化された個人評価**には、個人モニタリングやバイオモニタリングがあり、一般的に実際のばく露レベルを推定するための最良のデータソースと考えられている（NRC, 1991年; NRC, 1997年）。その一方で、間接的な方法による正確なばく露の質的測定（例えば、二分式や順序式のばく露測定）も、疫学の目的では同じように正確であると考えられる。さらに、間接法は解釈が容易であることが多く、ばく露評価における有用性を示すための追加の研究開発が少なく済む可能性がある。

どのようなアプローチであっても、ばく露評価方法は信頼性と妥当性のあるばく露推定値を提供できなければならない。疫学の文脈では、一般的に**信頼性**とは結果を再現する能力を意味し、**妥当性**とは、一般的にばく露推定値が真のばく露レベルを反映しているかどうかを意味している（Checkoway ら, 2004年）。特定のばく露評価の信頼性と妥当性を評価する際には、研究目的の文脈でばく露評価の長所と短所を考慮することが重要である。あまり洗練されていないばく露評価は、探索的研究に適しているかもしれない。これは、探索的研究は潜在的なハザードについての認識を高めるのに役立つため、より焦点を絞った研究への専心を促すことができるからである。逆に、より焦点を絞った仮説を持つ研究は、より洗練されたばく露評価法を用いることで大幅に強固となる。したがって、間接的ばく露評価法と直接ばく露評価法は補完的なツールであり、ばく露と疾病の関係を調査する際に研究の様々な段階で使用することができる。

## 2. 交絡因子

交絡因子は、ばく露と疾病との関係がある程度で第二のリスク因子（交絡因子）の影響に起因している場合に発生する。これは、この第二のリスク因子（すなわち、交絡因子）が、疾病の独立した因果関係のあるリスク因子でありながら、解析対象のばく露と因果関係（因果関係または非因果関係）があり、ばく露と関係のある疾病との間の因果経路の中間変数としての役割を果たしていない場合に起こりうる。交絡因子は、適切に測定され、考慮されなければ、ばく露と健康影響との間の推定された関連性の大きさ（そして場合によっては方向性）を変化させる可能性がある。

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<sup>15</sup>.

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

<sup>15</sup> 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.

これは、2つのリスク因子の影響が適切に分離、あるいは「区別」されていないため、ばく露と疾病の関係が過大または過小に評価される結果となりうる。例えば、ある研究では、ある農薬が肺がんと関連しているかもしれないが、これは農場のトラクターのディーゼル排気ガスの交絡効果によるものかもしれない。このように疾病発生（ここでは肺がん）との因果関係が農薬ばく露とも関連する場合、第二の因子の「農場のトラクターのディーゼル排気ガス」は交絡因子である。交絡因子には、喫煙、食事因子（例えば、高エネルギー/高カロリーの食事）、身体活動（例えば、身体活動の欠如）、遺伝子、併存症、薬物使用、アルコール摂取などのあまり直感的でないライフスタイルでのばく露が含まれることがあり、これらはすべて健康に悪影響を及ぼし、農薬使用と統計的に関連している可能性がある。疫学的解析では、交絡因子は研究サンプルで測定され、通常、研究のデザイン段階または解析段階のいずれかで最終的なリスク推定値に「調整」される。前者に関しては、疫学研究者は、研究者がコントロールしたい特性を共有する個人に研究集団を「限定」することができる。これにより、その特性（現在はコントロールされている）に起因する交絡の潜在的な影響を取り除くことができる。2番目に利用可能な方法は、研究のデザイン段階にも適用可能な方法で、研究者が交絡変数に基づいて個人を「マッチング」することで交絡をコントロールすることである。これにより、交絡変数が2つの比較群の間で均等に分布し、交絡変数を効果的にコントロールすることができる。主効果のリスク推定値に影響を与えるためには、交絡因子とばく露または疾病との関係が統計的に有意である必要はないことに注目することが重要である<sup>15</sup>。

解析の段階で、交絡因子をコントロールできる方法の1つは、層化がある。この方法では、交絡変数（の可能性はある）ごとに関連性を測定し、適切であると判断された場合には、個別の推定値を統計的に「まとめる」ことで、各層で評価された推定値に重みをつける Mantel-Haenszel アプローチを使用して共通のオッズ比または他の効果を算出する。層化は同時にコントロールする必要のある潜在的な交絡因子が複数存在する場合には困難である。このような場合、交絡因子は通常、統計的モデル化によって対処される。（例：ロジスティック回帰）。

疫学研究が現場で開始される前に交絡因子を慎重に考慮することが重要であり、どのような研究でも交絡因子をどのようにして考慮したかを適切に記述することが重要である。疫学研究は、潜在的な交絡因子を無視したり、あるいは十分な注意を払っていないと批評されることが多い。このため、感度分析は、欠落している交絡因子や考慮されていない交絡因子が観察された効果量に与える潜在的な影響を示すのに役立つ（Gustafson and McCandless, 2010年を参照）。

<sup>15</sup> これが、一般的に、交絡因子を調整する必要があるかどうかを判断するために交絡因子を「統計的に検定」することが不適切であると考えられている理由である。その代わりに、潜在的な交絡因子を調整した（交絡因子を含む）後の効果量の変化が10%以上であれば、交絡因子を分析に組み込むのに十分なエビデンスであると考えられる人もいる。

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<sup>16</sup>, 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

<sup>16</sup> 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.

測定されていない交絡因子が結果に悪影響を及ぼすと考えられる場合、研究者は感度分析を実施して、影響の範囲とその結果として得られる調整後の効果測定値の範囲を推定すべきである。このような感度分析は、一般的にほとんどの公表されている疫学研究では一様には行われていないが、既知ではあるが評価されていないリスク因子によるバイアスや潜在的な交絡因子の影響を推定するために、利用可能な場合には使用することができる。

特定のばく露・疾病関連付けに応じて、ある因子はコントロールするために必要な交絡因子である場合とそうでない場合がある。交絡因子によって推定された効果量に大きな歪みが生じるためには、交絡因子は、関係ある疾病が比較的強いリスク因子でなければならない<sup>16</sup>し、関係あるばく露とも強固に関連付けられていなければならない。潜在的な交絡因子の評価は研究ごとに行われ、測定されていない交絡因子が結果に悪影響を与えると考えられる場合には、研究者は感度分析を行い、影響の範囲とその結果として得られる調整された効果評価の範囲を推定すべきである。観察疫学研究の質を評価する際、OPP は、関連する交絡因子が適切に同定、記述、評価、解析されているかどうかを検討し、研究対象となっている特定の関連性を偏りのない推定が可能であるかどうかを検討する。可能な場合は、交絡因子が推定された効果量の推定値に影響を及ぼす範囲の判断を助けるツールとして感度分析を考慮する。交絡因子は、疾患の比較的強いリスク因子でなければならず、リスク推定値に大きな歪みをもたらすためには、対象となるばく露と強く関連していなければならないということを強調すべきである。このような場合には、交絡因子の可能性を提起するだけでは十分ではなく、なぜリスク因子が交絡因子となりうるのか、その影響がどのようなものなのか、そしてその影響が結果の解釈にとってどの程度重要なものなのかについて、説得力のある議論が必要である。(p.23-25, FIFRA SAP レポート、2010年4月22日)

最後に、共変量の交絡、効果修正、相乗効果、その他の媒介効果を区別することが重要である。交絡因子とは、疾病と因果関係のある変数及び対象となるばく露と因果関係があるか、否かにかかわらず、関連する変数をコントロールしないことによって生じるバイアスである。疫学研究者はこのバイアスを最小化しようと努める。一方、効果修飾因子とは、層化（例えば、年齢、人種／民族、SES の状態、遺伝的多型）によって、効果量に異なる影響を与える変数である。効果修飾因子は、交絡因子である場合もあればそうでない場合もある。一般的に、効果修飾因子は多変量モデルに交互作用項を導入するか、効果修飾因子のレベルによってデータを層化した後、層化に効果量を評価することによってモデル化される。対象となる各集団層において十分なサンプルサイズを確保するためには、効果修飾因子を評価するための特別な研究デザインが必要となることが多い。疫学研究者は、母集団の層化にわたるリスクの違いの評価、ばく露と関係のある効果の関連付けの評価、影響を受けやすい亜集団の特定が重要であることから、効果修飾因子を理解することを目指す（交絡因子のように最小化するのではなく、効果修飾因子の理解に努めている）。

<sup>16</sup> 交絡因子がそれ自体が本当にリスク因子であることを確認するだけでなく、関係のあるばく露との関連性だけではないことを確認することも考慮する必要がある。関係のある疾病との関連性を持つ因子の全部または一部を、関係のあるばく露との関連性のために調整することは、ばく露と疾病の関係が本当に存在する場合には、ばく露と疾病の関係を減衰させることにつながる。

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

効果修飾因子は交絡因子である場合もあればそうでない場合もある。例えば、タバコによく使用される殺虫剤と肺がんを関連付ける研究では、喫煙は交絡因子になるかもしれないが、この殺虫剤へのばく露のリスクが非喫煙者よりも喫煙者の方が高い場合には、効果修飾因子になるかもしれない。相乗効果は、疫学的な概念ではなく、生物学的または薬理的/毒物学的な概念として紹介されることが多く、2つの化学物質が一緒になって共同で作用することで、各化学物質の効果の(数学的な)合計を考慮した場合の効果を超えて、効果を拡大したり、増強したりする能力に関係している。疫学や統計学の用語では、これは効果の修飾や相互作用として表現されることが多い。

### 3. 統計解析

疫学研究は、特定のばく露と疾病との関連付けを評価することを目的としている。農薬疫学研究の質を評価する際、OPPは使用された統計的手法も考慮する。具体的には、研究に記載されている分析方法が研究課題に適切であるかどうか、使用された統計的方法の記述の完全性、ばく露と疾病の関係における潜在的な交絡変数の特定、評価、調整のための方法の適切性、実施された可能性のある小集団解析の説明、範囲、提示(多重比較のための統計的補正が行われているかどうかを含む)が考慮される。

疫学研究では通常、リスクを推定するために統計的モデル(例えば、ロジスティック回帰(オッズ比)やポアソン回帰(カウントデータ)のような一般化された線形モデル)を利用する。そのためには、研究者は関連する主要なばく露変数と転帰変数だけでなく、関連する交絡因子を考慮しなければならず、また、研究における関連付けがこれらの因子、すなわち、効果の修飾や相互作用のレベルによって異なるかどうかを考慮しなければならない(Szkloら、2004年)。交絡の可能性のある変数(調査対象となる関連性の大きさや方向性を実質的に変化させる変数)が同定された場合、回帰モデルによる調整を行うことで、目的のリスク推定値、すなわち調査対象となる関連性を分離することができる。さらに、OPPは、研究における潜在的な効果修飾因子のレベルまたは統計的相互作用の評価による関連付けの層化を評価する。関係のある関連付けの規模と方向性が第三の変数のレベルによって大きく異なる場合、層化された結果は一次的なもののみなされるべきである。

発現事象がまれな場合やサンプル規模が比較的小さい場合に統計的モデリングを行う場合、モデルにおける共変数が多すぎないように注意することが重要である。結果として得られる効果量の推定が高すぎたり、低すぎたりすることがあり、効果の真の推定を反映しているとは考えにくい。稀な事象やサンプル規模が小さいことによるこのような問題は、条件付き手法(例えば、デザインに研究における比較群の組み合わせを含む場合の条件付きロジスティック回帰)を使用する場合にも起こり得る。すなわち不一致ペア(または不一致セット)の観察数が少なすぎると、推定された効果量もまた信頼性が低くなる可能性がある。このように、交絡因子及びその他の共変数をコントロールすることは重要であるが、信頼性の高い検定を行うために過度にコントロールしたり、自由度が少なくすぎたりしないように評価者は注意しなければならない。

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 that 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

このような場合には、有効なサンプルサイズを十分に確保するため、最も影響力のある交絡因子や他の共変量をより少ない数だけ調整して、より解析的なモデルを求めることがより重要である。

最後に、どのような統計モデリング演習においても、結果の臨床的／生物学的／科学的な意義の文脈で統計的有意性を考慮することが重要である。統計的には有意だが、臨床／生物学的／科学的な文脈では重要ではないということもある。逆に、統計的には有意ではないが、臨床的／生物学的／科学的な文脈では重要な結果である場合もある。前者の場合はサンプルサイズが必要以上に大きいことを示唆し、後者の場合は必要以上に小さなサンプル規模を示唆しているかもしれない。後者の場合は、公衆衛生の観点から重要であり、（不正確であるにもかかわらず）特に関連付けが強固である場合には、さらなる調査が必要である。

#### 4. 観察研究における潜在的なバイアス

バイアスとは、（観察されていない）真の状況とは系統的に異なる研究結果をもたらす研究のデザインまたは実施における系統的な誤差のことである。これは、効果量の評価におけるサンプリングの変動や精度（または同義で信頼限界）に関係するランダム誤差とは対照的であるが、結果をある特定の方向（例えば、帰無仮説に向かって、または帰無仮説から離れて）に「追い込む（drive）」または「押し込む（push）」ようなことはない。

バイアスとは、研究のデザインや実施における方法論的な不完全性の反映であり、研究者は解析の一部として、それに対処したり、議論したりすべきである。研究にバイアスが導入される方法はいくつもある。研究は、参加者を研究に選択する方法（選択バイアス）や、ばく露や疾病に関する情報を収集する方法（情報バイアス、先に説明した症例対照研究の想起バイアスを含む）にバイアスがかかっている可能性がある。一般的な職業選択バイアスの例としては、「健康な労働者効果」があり、職域疫学研究において重要なバイアスを生じさせ、帰無値に向かってバイアスを生じさせ、さらにはそれ以下のバイアス（ばく露が「保護的」であるという解釈を生じさせる）につながる。バイアスの全くない研究はなく、公表した研究の著者は研究における潜在的バイアスどの程度説明したか、研究におけるバイアスを（もしあれば）どのように対処し、研究の特徴を明らかにしたかを検討すべきである。バイアスは、差別的（differential）または非差別的（non-differential）な誤分類から生じる可能性がある（Greenland, 1998 年）。差別的誤分類（バイアス）とは、他の変数の値に依存する方法で誤分類が発生したことを意味し、非差別的誤分類（バイアス）とは、他の変数の値に依存しない誤分類を意味する。誤分類バイアス（差別的または非差別的のいずれか）は、そのようなばく露を分類するために使用される試験方法の感度と特異度に依存し、特定の（限定された）条件の下でバイアスの方向性を予測可能な影響を及ぼすことがある。研究の方法と解析の知識に基づいてバイアスの方向性を特定できるということは、規制当局の意思決定者にとって有用である。なぜなら、意思決定者は、検討されている疫学的効果の大きさ（例：OR、RR）が、真の効果の大きさの過小評価または過大評価である可能性がどの程度あるかを判断することができるからである<sup>17</sup>。

size<sup>17</sup>. 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<sup>18</sup>. (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<sup>19,20</sup>. 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

<sup>17</sup> 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).

<sup>18</sup> 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.

<sup>19</sup> The old adage that “the absence of evidence does should not be interpreted as the evidence of absence” is true here.

<sup>20</sup> 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>

誤分類の程度が 10~20%の範囲にあることは異例ではなく、このような誤分類の程度（それが差別的か非差別的か）と、そのような誤分類の大きさと方向性の両方に関してオッズ比または相対リスクに影響を与える可能性の程度を評価する感度（または「もしも」）解析の形態を検討することは、疫学研究をレビューする際に有用であると考えられる<sup>17</sup>。（p.25、FIFRA EPA SAP 報告書、2010 年 4 月 22 日）。交絡因子に関して前述したように、このような定量的感度分析は、公表されている疫学研究ではほとんど実施されておらず、バイアスの潜在的な規模や疫学的効果量推定への影響の定量的な評価は行われずに、むしろ一般的には解説文で評価されている（Jurek at al., 2006）。これは次のような理由によるものである。その理由の一部は、疫学者がこのような解析のための計算ツールを一般的に利用できないことや、そのようなツールに慣れていないことにある。このようなツールは次第に利用可能になりつつあり、潜在的なバイアスを評価するためのより厳密な定量的手法を開発する上で有用であると考えられる。

### 5. 帰無研究の解釈

「帰無」研究、すなわち、農薬へのばく露と有害な健康影響との間に関連付けがないと報告するよく行われている研究は、ヒト健康リスク評価に有用な可能性があるかどうか慎重に評価される。研究が無効という結果になるのは、調査された関連付けが実際には存在しないか、あるいは、実施された研究が所定の有意水準で関連付けを検出することができなかったためである。後者の結果（関連性を検出できなかった）は、必ずしも関連付けが存在しないことを意味するものではなく、特定の研究で関連付けが検出されなかったと解釈すべきである<sup>18,19</sup>。「帰無」研究をレビューする際に、これら 2 つの条件のどちらが正しいかを評価するためには、同一または類似の研究課題について報告された他の研究、ばく露と発現事象の評価方法、ばく露の誤分類が研究を帰無に偏らせた可能性の程度、関連付けの交絡変数の同定と分析を含む使用された統計的方法、影響が発生する、あるいは影響が検出される閾値を下回っているばく露の程度、研究の検出力と実質的な関係のある効果量を検出する能力を考慮する必要がある。統計的検出力とは、実際に真の違い（または関連付け）がある場合、研究者が 2 つの比較群間に差があること、すなわち、ばく露と疾病の間に関連付けがあることを正しく識別できる確率を意味する。

<sup>17</sup> このような感度分析は、多くの場合、交絡因子よりも効果量推定に実質的な影響を及ぼすと予想されるばく露の誤分類バイアスに対して特に推奨されるかもしれない。

<sup>18</sup> 「エビデンスがないからといって、エビデンスがないと解釈すべきではない→エビデンスがないことを、ないことのエビデンスと解釈してはならない」という古い格言は、ここでは真実である。

<sup>19</sup> 米国統計協会の統計的意義と P 値に関する声明 <https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf> も参照のこと。

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<sup>21</sup>.

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.

<sup>21</sup> 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).

「低検出力」の研究は、関連付けが実際に存在しているのに、関連付けがないと誤って結論を下すことがある<sup>20</sup>。

最後に、帰無研究の解釈に関しては、文献のシステマティックレビューにおける出版バイアスの影響を考慮することが重要である。出版バイアスとは、利用可能な公表文献が、そのような帰無研究を不釣り合いに除外する傾向のことである。疾病または健康影響の間にこのような「帰無」関連付けを示す研究は、研究課題が疫学的レビュープロセスの一部として確立された質の基準を満たしていれば、その研究と同様に有益な情報を得ることができる。これらには、研究デザイン、優先的仮説と探索的解析の比較、関係のある効果量を検出するためのサンプルサイズと統計的検出力、感度と特異度に対する影響の適切な確認、ばく露評価の質と差別的・非差別的な誤分類の可能性、主要な潜在的交絡因子とその他のバイアス（情報、選択等）の評価の適切性、効果修飾因子の評価（優先的仮説による裏付けがない仮説、生物学的妥当性、またはその他の裏付け情報についての多重比較の検討とその補正を含む適切な統計解析）などの要因が含まれる。

### 6. 外的妥当性（一般化可能性）

上述したように、妥当性とは、一般的にはばく露推定が真のばく露レベルを反映している程度を指す（Checkoway ら）。外的妥当性または一般化可能性とは、母集団のサンプル（例：農薬散布者）から得られた疫学的研究結果を他の母集団（例：すべての農業従事者）に拡張する能力のことである。外的妥当性を評価するために、サンプルの特性とより大きな集団（既知の場合）との比較を行うことができる。このような評価には、人口統計学的要因だけでなく、ばく露（例えば、用量、時期、期間）が類似しているかどうか及び重要な影響修飾因子（例えば、脆弱者集団の感受性）が考慮されているのかも含まれるべきである。一般化可能性は特に重要であり、個々の研究結果が、規制当局のリスク評価において、より大きな集団または対象となる亜集団に適用できるかどうか、またどのように適用できるかを理解することが重要である。例えば、AHS は、いくつかの農薬について、いくつかのがんとがん以外の健康影響との間に統計的な関連付けがあることを報告している。OPP は、報告された知見が、ノースカロライナ州とアイオワ州以外の州の農薬散布者や、主に散布後の作業を行う農業従事者にどの程度適用できるかを評価することに関心が持たれている。

<sup>20</sup> 検出力は低くても統計的に有意な効果が見られる研究は、効果量の拡大という現象である可能性があり、これを調査することが重要である（Ioannidis, 2008年）。

## 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.<sup>22</sup>

**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

<sup>22</sup> 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](http://www.who.int/ipcs/publications/methods/human_data/en/index.html)

## V. ヒトでの事例調査データ

一般的に言えば、上記のような農薬に関する疫学研究は、急性の臨床症状をもたらす可能性が小さい低量ばく露（より長い期間にわたって）に焦点を当てている。OPP はまた、より高く、より短い間隔（多くの場合、急性の"一回限り"で）のばく露に関心を持っている。この「ヒトでの事例」または中毒データは、問題となっているに起因すると考えられる短期的で高濃度の高量ばく露シナリオを評価するのに有用である。

OPP はこのような「ヒトでの事例情報」をいくつかの目的で使用している。最も広い意味では、OPP はリスク評価/リスク管理の活動に情報を提供するためにヒトでの事例データを使用している。これは、農薬登録改善法 (PRIA) の責任のもと、登録審査活動の重要な一部となっている。この目的のために、OPP は人の事例データを評価して経時的な傾向を調べ、さまざまな農薬ばく露の重大さの程度と頻度のパターンを調べている。場合によっては、事例情報は、追加の情報や追加のリスク管理措置の必要性を示すことができる。また、事例情報は、リスク軽減措置が実施された後の達成の評価にも役立ち、事例情報は、OPP がヒト健康と環境を保護するためにやってきたリスク管理措置の有効性を保証するための OPP のパフォーマンス・アカウントビリティ・システムの重要な部分である。最後に、事例情報は、使用方法に関する実際の使用情報を提供する上で有用であり、また、必要に応じて、施行や教育の活動の対象となる可能性がある。

OPP はこの情報を様々な情報源から入手している。ヒトでの事例データの情報源には、医学的及び毒物学的文献に掲載されている（ヒトの）**医学的症例報告**や急性及び慢性のハザードの特定を促進し、リスク評価プロセスに不可欠な共同で検討されている急性農薬中毒の様々な国の**中毒サーベイランス活動**からの情報が含まれている<sup>21</sup>。

**医学的症例報告**（医師によって書かれた直接の証言）や医学的症例集積（共通の原因や症状を持つ個人の症例報告をまとめたもの）は、健康影響に関する利用可能なすべてのエビデンスを解析するための貴重なツールであり、動物試験や疫学研究の知見を補完するものでもある。さらに、特定の農薬の使用に関連すると思われる有害な健康影響の稀な発生や新規の発生を特定することができ、中毒警戒の行政機関に「事前通知」を提供できる。公表されている農薬の症例報告は、通常、非定型的な（高量ばく露/大量投与、違法、適用外）急性または短期のばく露による影響を記述している。報告書はしばしば個々の事例に基づいており、その性質上、選択性が高い。しかし、これらの報告書は、ヒトにおいてこれまでに確認されていなかった毒性影響を特定し、高量ばく露後の影響、健康影響及び医学的後遺症を知る上で、特に有用である。これらの研究には、より詳細な医学的情報（後遺症を含む）や詳細な追跡調査が行われていることが多く、一般的に用量に関する高い質及び/または定量的な情報が得られている。

<sup>21</sup> OPP は、IPSC がリスク評価においてヒトでの事例データを考慮する成果を認識している。 [http://www.who.int/ipcs/publications/methods/human\\_data/en/index.html](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)<sup>23</sup>, as well as other incidents reported directly to EPA. OPP has compiled the pesticide related

<sup>23</sup> 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."

複数の医療事例研究で類似点が見られたり、症状、ばく露シナリオ、使用方法などにパターンが見られた場合には、リスク評価プロセスと結果の裏付けに有用な情報を提供することができる。医学的症例研究や定量的なばく露情報集積は、必要に応じて、典型的な使用から予想されるばく露マージンを特性評価することで（ラベルに明示された適用率や代替ばく露情報に基づく）リスク評価のばく露推定値と比較できる。

医学的症例報告や医学的症例集積を評価するには、以下の点を考慮する。

- ばく露の詳細な履歴（いつ、どのように、どのくらい）、健康影響の発現時期、患者の徴候や症状が報告されている。
- 商品名、農薬表示、登録番号など、商品・化学品・農薬に関する情報。
- 患者情報（例：年齢、人種、性別）；基本的な健康状態及び同様の徴候や症状をもたらす可能性のある医薬品の使用；関連する病歴及びリスク因子の有無。
- 疾病の説明と診断に至った経緯。
- 患者の管理・治療、農薬や化学物質の血中濃度などの検査データ（治療前、治療中、治療後）。
- 医学的報告は信頼性があるか、妥当であるか。他の研究、レビュー、ガイドラインを含む最新の知見と一致しているかどうか。
- 疾病の臨床経過及び患者の転帰（例：患者の回復及び退院；退院後の患者の状態、農薬または化学物質へのばく露に関連した慢性的な健康影響または早期死亡）。

実際のばく露情報及び毒物学的情報の情報源として医学的症例の報告/集積を利用することに加えて、OPP は様々な農薬中毒事例データベースを利用した中毒サーベイランス活動も行っている。具体的には、OPP は以下の 5 つのヒトでの事例データソースにアクセスできる。*OPP Incident Data System* (IDS)、American Association of Poison Control Centers (PCC) の *National Poison Data System* (NPDS) からの要約報告書、現在オレゴン州立大学にある EPA が資金提供している *National Pesticide Information Center* (NPIC) からのデータ、疾病管理予防センター (Centers for Disease Control and Prevention) /National Institute for Occupational Safety and Health for Occupational Risk - Pesticides (NIOSH SENSOR-Pesticides) 及び *California Pesticide Illness Surveillance Program* (PISP) のデータ。以下、それぞれについて順番に説明していく。

- **OPP Incident Data System (IDS)** は OPP が管理しており、FIFRA 第 6 条 (a)(2)<sup>22</sup>項に基づき登録者が提出したデータ及び EPA に直接報告されたその他の事例を含んでいる。

<sup>22</sup> FIFRA 6(a)(2)に基づき、農薬登録者は、「農薬の環境への不適切な悪影響に関する事実情報」を知った場合には、EPA に通知するよう求められている。

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.

OPP は 1992 年以来、IDS に農薬関連の事例報告をまとめている。IDS には、登録者、他の連邦・州の保健・環境機関及び個人消費者からの強制的な FIFRA 第 6 章(a)(2)報告を含む、様々な情報源から主張されたヒトでの健康影響事例の報告が含まれている。IDS には、ヒト、植物、野生動物、家畜への悪影響を主張している事例に関する情報が含まれている。IDS 報告の大部分は、政府機関が紙形式で受け取っている。IDS の項目は、OPP の情報サービスセンターにあるマイクロフィルムやスキャンした画像に保存されているオリジナルの報告書のコピーへの「ポインタ」として機能する。

IDS には職業上及び非職業上の両方の事例が含まれているが、報告された事例の大部分は非職業/住居のシナリオに関するものである。全米から報告された事例のほとんどは製品情報 (EPA 登録番号など) の記載に関連している。IDS はほとんどが FIFRA 6(a)(2)報告要件に基づき農薬登録者から提供された情報で構成されているため、政府機関は特定の製品を識別することに対して比較的高い信頼性を持っている。重大さの程度の順位付けは、各事例について (CFR §159.184 で規定されている) 記載されている。症状情報は、事例の解説部分に含まれることがあるが、この情報は通常、医療専門家による検証/確認を受けていない。IDS には、ばく露シナリオやハザード情報の解説情報も含まれている。多くの企業は、標準化された、業界で開発された自主的な事例報告フォームを使用している。

OPP は IDS からデータを収集・評価し、農薬ばく露による健康影響の程度や重大さの程度に関して潜在的なパターンを識別している。IDS の報告書は範囲が広く、場合によっては詳細な情報を含むが、システムは必ずしも職業上ばく露状況や医学的転帰などの発現事象に関する詳細な情報を一貫して把握しているわけではない。

また、IDS に入る症例データのほとんどは、中毒管理センターへの電話からの収集もあるが、確認・照合されない。とはいえ、事例情報は、製品の初期登録後、製造販売後の重要なフィードバックループを政府機関に提供できる。重大な性質の IDS 事例や、それほど重大ではない事例の間で示唆されたパターンや傾向は、特定の化学物質や製品をさらに調査するように政府機関に求めることができる。IDS はこのように広範囲をカバーしているため、一時的な傾向情報を提供し、リスク軽減によって潜在的な農薬ばく露が減少し、IDS に報告された潜在的な事例の数が減少したかどうかを判断するのに役立つ可能性がある。全体的に見ると、IDS は農薬の全国的な傾向と事例の頻度に関する良好な情報を提供し、農薬のハザード及び/またはばく露の可能性について有用な予測を提供することができる。

- ❑ **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

- ❑ 全米中毒データシステム (NPDS) (以前は毒性影響監視システム (TESS) と呼ばれていた) は、米国中毒管理センター協会 (AAPCC) によって維持されており、いくつかの連邦政府機関からの資金援助を受けている。NPDS はコンピュータ化された情報システムで、地理的に特定されたほぼリアルタイムの報告が可能である。中毒管理センターの主な任務は、緊急事態に対応する通報者を支援することであるが、NPDS のデータは、化学製品の安全性に関する新たな問題を特定するのに役立つ。全国にある 61 の PCC のホットラインは、365 日 24 時間年中無休で開設されており、特別な訓練を受けた看護師、薬剤師及びその他の臨床医療専門家が中毒情報を提供している。コンピュータによるデータ入力、標準化されたプロトコル、厳格なデータ入力基準を使用して、現地の通報者が事例を報告する。これらの報告された事例は現地で保存され、AAPCC によって維持されている全国データベースに要約された形で更新される。情報通報は個別に集計され、事例としてはカウントされない。PCC システムは、ほぼすべての米国とその領土をカバーしており、2001 年以来、コンピュータの大幅な強化が行われてきた。

NPDS には主に非職業上の事例が含まれている。NPDS には解説情報は含まれておらず、製品情報は完全ではない。NPDS は、訓練を受けた専門家が指定/記録した重大さの程度の順位と症状情報を提供しており、政府機関はこれらの情報に対して比較的高い信頼性を持っている。NPDS はまた、報告されたばく露による健康影響の可能性に関する情報も提供している。全体として、NPDS は全国的な傾向、農薬の事例発生頻度、特定の農薬の潜在的ハザードについての良好な情報を提供している。しかし、情報源の制限により、毎年発行される AAPCC のサマリーレポート (例: <http://www.aapcc.org/annual-reports/> を参照) にしかアクセスできず、これらのレポートは他の情報源を補完する役割を果たしているため、NPDS はより完全なアクセスが可能となっている。

- ❑ **National Pesticide Information Center (NPIC)** (<http://npic.orst.edu/index.html>) は、EPA の資金提供を受けており、客観的で科学的根拠に基づいた農薬情報を提供し、問い合わせに対応したり、事例に対応したりすることを目的としている。NPIC は平日の営業時間内に全国的に機能しており、オレゴン州立大学 (現在) と EPA との間で協力している。中毒管理センターと同様に、NPIC の主な目的は事例データの収集ではなく (NPIC の年間の問い合わせの約 10% は「事例」に関連していると考えられている)、むしろ幅広い農薬に関する情報を問い合わせ者に提供し、農薬での事例の調査や緊急処置のための他の情報源に誘導することである。とはいえ、NPIC では問い合わせ者からの事例情報 (年間約 4000 件) を収集し、データベースに記録している。NPIC は全国の事例の情報源であるが、一般的に IDS に比べて報告件数は少ない。

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**.<sup>24</sup> 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

特定の農薬や製品が IDS で高頻度に観察された場合でも、NPIC は全国のデータセット間の整合性を判断する上で有用な情報源を提供している。

IDS や PCC と同様に、NPIC の事例は主に非職業上のものである。NPIC の事例には、通報者が情報を提供する際に解説情報や製品情報が含まれている。対象範囲は全国であるが、NPIC に報告される事例の数は NPDS や IDS に比べてかなり少ない。一方、提供される情報量はかなり多く、必要に応じて政府機関がカスタムレポートを要求することも可能である。ハザード情報には、重大さの程度の順位付け、ばく露経路、症状などが含まれており、これらは訓練を受けた担当者によって記録される。また、ばく露と有害影響の関連性がどの程度であるかについての情報も提供している（これを確実性指数と呼んでいる）。また、NPIC は年次報告や解析結果を公開しており、有用な情報源となっている。

- 疾病管理予防国立産業衛生研究所（CDC/NIOSH）は、**SENSOR (Sentinel Event Notification System for Occupational Risk) - Pesticides** と題した農薬監視プログラムとデータベースを管理している<sup>23</sup>。このデータベースには、1998 年から 2013 年までの 12 の州における農薬による疾病症例報告が含まれている。参加している州は、カリフォルニア州、フロリダ州、アイオワ州、ルイジアナ州、ミシガン州、ネブラスカ州、ニューメキシコ州、ニューヨーク州、ノースカロライナ州、オレゴン州、テキサス州、ワシントン州である。ある年の参加州は、州と連邦政府が農薬監視のために資金を提供しているかどうかによって異なる。

SENSOR-農薬データベース内の農薬関連の疾病症例は、様々な情報源（現地の中毒管理センターからの報告、医師によって報告された際に農業従事者の補償請求の州労働局からの報告、州農業局からの報告及び州保健局への医師の報告）から把握されている。職業上及び非職業上の事例の両方がデータベースに含まれているが、SENSOR 責任者は、主に職業上の農業従事者事例に関する追跡症例調査の成果に焦点を当てている。州保健局の SENSOR 責任者は、症例を追跡し、ばく露シナリオ、症状、重大さの程度、健康影響を確認するために、カルテを取得する。標準化されたプロトコルと症例の定義を使用して、州保健局の SENSOR 責任者は、症例、カルテ、医師と患者によって提供された事例インタビュー解説を SENSOR データシステムに入力する。

すべての SENSOR-Pesticides の症例は、EPA が事例の解析に使用する集合データベースに含まれるために、最低 2 つの健康影響を報告しなければならない。

<sup>24</sup> SENSOR-Pesticides webpage: <http://www.cdc.gov/niosh/topics/pesticides/overview.html>

<sup>23</sup> SENSOR-Pesticidesのホームページ: <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

各症例のエビデンスは、NIOSH の症例分類マトリックスに基づいて、ばく露と疾病との因果関係について評価される。SENSOR-Pesticides の症例の 98%は、確定、可能性高い、または可能性ありに分類され、2%は疑わしい症例に分類されている。可能性の低い症例、無症候性の症例、無関係の症例及び情報が不十分な症例は、SENSOR-Pesticides のデータベースには含まれない。

全体的には、職業上及び非職業上の事例について非常に有用な情報を提供しており、時には農薬のハザードやばく露の可能性についての有用な予想を提供している。また、SENSOR-Pesticides は独自のデータの解析を行い、Morbidity and Mortality Weekly に公表している。前述のデータベースとは異なり、非職業上/住居上と職業上の両方の事例が含まれているが、SENSOR'S は伝統的に農業従事者での事例に焦点を当てており、その情報を提供する上で特に価値のあるものとなっている。1998年から2011年までのSENSOR-Pesticides のデータはオンラインで入手可能である：<http://wwwn.cdc.gov/Niosh-whc/Home/Pesticides>。

- **カリフォルニア州農薬疾病サーベイランスプログラム (PISP)** は、カリフォルニア州によって管理されている。このデータベースは、農薬に関連した疾病や傷害を記録している。症例報告は、医師からの報告と農業従事者補償記録を介して行われる。現地の郡農業委員会がばく露の状況を調査する。カルテと調査結果は、カリフォルニア州農薬規制局 (DPR) の技術専門家によって評価され、疾病の登録簿に登録される。カリフォルニア州の PISP プログラムで報告されたすべての農薬による疾病は、郡の農業委員会によって調査され、DPR は報告書を評価してデータベースにまとめ、農業従事者やその他の人々を農薬ばく露の悪影響から保護するための州のプログラムを改善するために使用される (<http://apps.cdpr.ca.gov/calpiq/>)。

現在、OPP は化学物質別にヒトの事例データを評価している。各データベースからの事例は、ハザードポテンシャル (死亡者数、より重大な事例の頻度、報告された症状のパターン/傾向) とばく露ポテンシャル (事例の頻度/経時的傾向、ばく露シナリオのパターン/傾向、ばく露に影響を与える要因のパターン/傾向、または製品のパターン/傾向) について解析されている。上記のデータベースからのヒトでの事例データを評価する際、OPP はいくつかの一般的な基準を考慮している。OPP は症状の相対的な重大さの程度と頻度を考慮している。さらに、OPP は一般的に、時間的な関連付けが確認できる報告を、あるいは少なくとも妥当と思われる報告を、より信頼性の高いものとしている。最後に、ヒトの事例データを評価するために使用される他の要因には、ばく露反応の関連付けのエビデンス、報告された健康影響の一貫性、報告された健康影響の生物学的妥当性、医薬品の使用などの健康影響の代替原因の排除及び観察された症状や健康影響の特異度が含まれる。さらに、より重大な事例の解説は、ばく露の時間と報告された影響の間に時間的な関連付けがあるかどうかを評価し、関連付けが状況によって裏付けられているかどうかを判断することが多い。

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 in 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 21<sup>st</sup> century toxicity testing. As these transformation changes occur, OPP will update this approach as appropriate.

例えば、屋内での農薬散布後 3 か月に発生した高齢者の心臓発作は、因果関係があるとは判断されない。一方で、農薬種類に応じた既知の症状と一致する症状を伴うばく露時またはばく露後間もなく発生した重大な事例は、事前の病歴がなく発生した場合、因果関係の推測がより正当化されることを示唆している。

以上のことから、医学的文献やヒトの毒性学的文献に掲載されている医学的症例報告／症例シリーズと、EPA が維持、資金提供またはアクセスしているデータベースから得られる中毒サーベイランスデータの両方からなる事例データは、実社会での農薬のリスクを評価する際に OPP を支援する有用で補完的な情報を提供することが可能である。

## VI. 要約と結論

このフレームワークでは、疫学とヒトでの事例データをレビューする際の重要な要素を説明し、そのようなデータを農薬のヒト健康リスク評価に組み込むために提案されている WOE 解析について説明している。OPP では、複数のエビデンスと、*in vitro*、*in vivo* 及び *in silico* の情報源から得られた入手可能な最善のデータを使用する。OPP では、複数の情報源からのデータを統合する際に、系統化されたツールとして修正された Bradford Hill 基準を用いて、品質、一貫性、関連性、整合性、生物学的妥当性を考慮するため、WOE アプローチを使用している。WOE 解析の適用は、リスク評価ガイドラインに概説されている科学的解析に従って EPA が日常的に使用している統合的かつ解釈的なプロセスである。WOE 解析はまた、結合されたデータセットの質も評価し、特定の科学的評価に適切である成果と複雑性のレベルと一致している (U.S. EPA、2002 年)。OPP は、21 世紀の毒性試験の新たなビジョンの実施に向けて、現在、毒性学とリスク評価が変革期にあることを認識している。これらの変革に伴い、OPP はこのアプローチを適宜更新して行くであろう。

## 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.

[Balslem 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, 2<sup>nd</sup> 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.

## VII. 参考文献

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.

[Balslem 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, 2<sup>nd</sup> 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. Lippincott-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<sup>1</sup>](#), 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, Yucl 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.

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. Lippincott-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 P1, 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, Yucl 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*. 2<sup>nd</sup> 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*, 2<sup>nd</sup> ed. Oxford University Press, New York.

Mausner JS and Kramer S. 1985. *Epidemiology*, 2<sup>nd</sup> 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<sup>1</sup>](#), [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.

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*. 2<sup>nd</sup> 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*, 2<sup>nd</sup> ed. Oxford University Press, New York.

Mausner JS and Kramer S. 1985. *Epidemiology*, 2<sup>nd</sup> 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<sup>1</sup>](#), [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](http://www.nap.edu/catalog.php?record_id=18764)

Ntzani EE, Chondrogiorgi 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](http://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](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. 6<sup>th</sup> 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.

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](http://www.nap.edu/catalog.php?record_id=18764)

Ntzani EE, Chondrogiorgi 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](http://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](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*. 3<sup>rd</sup> 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.

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*. 3<sup>rd</sup> 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](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<sup>3/4</sup> 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). 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](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<sup>3/4</sup> 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/0/33E1AD305287588F85257D20006BE8C/\\$File/ISA\\_PREAMBLE\\_FINAL2015.PDF](https://yosemite.epa.gov/sab/sabproduct.nsf/0/33E1AD305287588F85257D20006BE8C/$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, Pocock SJ, Pool C, Schlesselman 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](#)<sup>1</sup> 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

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](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, Pocock SJ, Pool C, Schlesselman 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<sup>1</sup> 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