

# CASE STUDY SUMMARY – Hypothesis-Based Weight of Evidence (Naphthalene as an Example)

---

## 1. Summary of Method Illustrated by Case Study

The Hypothesis-Based Weight of Evidence (HBWoE) framework is described in, and has evolved with, several of our recent publications (Rhombert *et al.*, 2010; 2011; Prueitt *et al.*, 2011; Bailey *et al.*, 2012). It is hypothesis-based in the sense that it emphasizes articulation of the proposed bases for the relevance of the data to the causal question at hand, specifying the logic and reasoning. The approach integrates all of the relevant data (epidemiology, animal toxicology, mechanistic, toxicokinetic, *etc.*), both positive and negative, in terms of quality and relevance to humans in a way that allows each data set to inform interpretation of the other. The approach further synthesizes all of the data to determine overall plausibility for causality in humans, considering uncertainties and inconsistencies in the data sets and ad hoc assumptions that may be required for some of the hypotheses put forth.

The hypothesized basis for inference about human risk from particular data should be seen not just as an extrapolation, but as a generalization. It is a proposal about something in common regarding the causal processes in the study situation and the human population of interest. As a generalization, it ought to apply to other situations as well, or at least have reasons why it does not, and one can evaluate the success of the hypothesis at being in accord with the whole suite of relevant observations at hand. If there are limits to the generalization (*e.g.*, it applies to one species but not another, to males but not females, at this dose but not that dose), then the plausibility of such exceptions in view of available evidence and broader knowledge becomes part of the evaluation of the hypothesis against available data.

Often the mode of action (MoA) is what provides the underlying commonality, or explains the lack of commonality, across species, and its consideration is key to weighing and integrating evidence from a large dataset in the HBWoE framework. This is particularly true if there are contrasting modes of action that have been put forth within the scientific community. If an MoA is yet to be established, however, the HBWoE approach can ask appropriate questions of the available data to inform future studies and a potential MoA hypothesis.

Although intended to be flexible in its application, the HBWoE approach generally consists of the following seven aspects:

1. Systematically review individual studies relevant to the causal question at hand, focusing on evaluation of study quality.

2. Within a given realm of evidence (e.g., epidemiology, animal toxicology, mechanistic, or toxicokinetic), systematically examine, organize, and present the data for particular endpoints.
3. Identify and articulate overarching hypotheses that bear on the available data and on establishing potential human risk.
4. Evaluate the logic of the proposed hypotheses with respect to each realm and line of evidence, considering plausibility, specificity, and consistency across studies.
5. Evaluate the logic of the proposed hypotheses with respect to all realms and lines of evidence so that all of the data are integrated and allowed to inform interpretation of one another.
6. Describe and compare the various accounts of the observations at hand, with a discussion of how well each overarching hypothesis is supported by all of the available data, the uncertainties and inconsistencies in the data set, and any *ad hoc* assumptions required to support each hypothesis.
7. Formulate conclusions and any proposed next steps (e.g., sharpening or reworking of proposed hypotheses already put forth; propose additional testing to clarify data gaps).

As discussed, our approach has evolved over several publications. Although the seven aspects described here formed the basic guide to our HBWoE evaluation for naphthalene (Rhomberg *et al.*, 2010), they are more explicitly presented in our later publications (Rhomberg *et al.*, 2011; Prueitt *et al.*, 2011), and more generally in Bailey *et al.* (2012). Although these steps should be generally adhered to, they are not intended to be a checklist, and may involve an approach that is not necessarily in the order presented.

Steps 4 and 5 describe the data integration portion of the evaluation. We find it useful, as part of these steps, to articulate specific questions regarding consistency and plausibility across studies that have become apparent while working through steps 1-3, and in answering these questions, to discuss how the data, as a whole, fit together, noting similarities across studies, strengths and limitations, and discordance. The answers to these questions provide a basis for judging the weight of evidence in support of a causal association.

A key aspect of the HBWoE framework is the importance of analysis of these lines of argument, or consideration of alternate “accounts” (or interpretations) of the available data and how each is supported by the available data. Hill (1965) makes explicit the importance of considering alternative “accounts” of the observations at hand in stating:

None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*. What they can do with greater or less strength, is to help us to make up our minds on **the fundamental question – is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?** (Hill, 1965) [emphasis added]

Therefore, a key outcome of the HBWoE framework is the evaluation and comparison of these alternative and contrasting accounts (Step 6). In the end, each account (that is, each tentative “story” as to why the facts are as they are) can be compared to other accounts. In this way, various competing overarching hypotheses can be weighed by comparing their relative success at explaining phenomena seen in the data, the relative reasonableness of ad hoc assumptions needed for each, and the relative naturalness and plausibility of the means whereby potentially refuting observations are reconciled with the account’s central hypothesis. Although it is hard to reduce this evaluative process into checklists, scores, or enumerations, the hope is that, by not simply conducting such evaluations of alternative accounts but also by writing them down to be scrutinized and debated, the relative explanatory success of each account, and the relative “epistemological baggage” associated with defending each alternative interpretation, will be evident. This can then serve as the basis for assigning the relative degree of credence that should be given to each account. For example, it provides the basis for comparing an account that asserts the existence of a causal role of the exposures of interest in the disease versus accounts that ascribe apparent patterns of association of exposure and disease to other noncausal factors. In addition, from this assessment, one can more clearly define hypotheses and propose areas of research needed to fill data gaps for each account or to put their hypotheses to the test.

As part of the comparison of accounts, the HBWoE approach considers all data relevant to the causal question at hand, even negative data and (particularly when they are the bases for a particular line of argument) data of questionable quality or from studies with significant design shortcomings. In this last case, it is important to demonstrate the analysis and logic of how poor quality data have been interpreted within an account, how critical they are to the account’s assertions, and the ad hoc assumptions required to fit these data to the proposed hypothesis. In the HBWoE framework, such questionable data are automatically down weighted by their poor ability to discriminate between accounts. This is because the face-value of interpretation of these data is not markedly more compelling than alternative explanations that ascribe the outcomes to those extraneous factors or alternative possible causes that better-designed studies would have eliminated. That is, the results are relatively easily and credibly explained away as artifacts.

As discussed in our recent HBWoE evaluations (Rhombert *et al.*, 2010, 2011; Prueitt *et al.*, 2011; and Bailey *et al.*, 2012), the explanations in each account need not be proven—what is important is that one set out the following questions to be considered throughout the evaluation:

- What is being proposed as causal and generalizable phenomena (i.e., what constitutes the basis for applying observations of biological perturbations or realized risks in other contexts to project potential risks to exposed humans)?
- In the case of observations that do not fit the hypothesized causal model, what is being proposed as the basis for these deviations (i.e., that would otherwise be counterexamples or refutations)?
- What assumptions are made that are ad hoc (i.e., to explain particulars, but for which the evidence consists of their plausibility and the observations they are adduced to explain)?
- What further auxiliary assumptions have to be made, and how reasonable are they in view of our wider knowledge and understanding?

- What is relegated to error, happenstance, or other causes not relevant to the question at hand?
- For those events or processes proposed as critical for a given account, what other observable manifestations should they have? Are these other manifestations indeed found?
- If either the operation or necessity of the proposed critical events for a given account were disproven, how else would one explain the array of outcomes?

Clearly, there may be many accounts, but the major contending accounts will be those that require the fewest ad hoc explanations for why certain observations do not fit with the data at hand. As an explicit process to the HBWoE framework, the scientific judgment (or logical rationale) required for each account needs to be illustrated and discussed in narrative text to describe how the data are being weighed, and what ad hoc assumptions are required to account for some of the problematic facts within the observations at hand. Different methods can be applied (e.g., organizational tables or figures), depending on the nature of the data, to organize and illustrate the consistencies and inconsistencies of the data as applied to various lines of evidence and various accounts. The point is to illustrate how one is tracing the logic through various competing accounts, and this will vary depending on the data set, likely requiring illustration as well as narrative text. As such, each HBWoE analysis can be constructed in a way that optimizes transparency and logic for the particular set of relevant data.

Table 1 provides an example table that can be used to illustrate the comparison of accounts. The table should present the overall "big picture" assumptions, and should tell the story for how the data are used to support both hypotheses, focusing on how each addresses uncertainties and inconsistencies in the data. The content of the table should deal predominantly with the more uncertain and controversial issues within the data set; e.g. inconsistencies across species and tissues, human relevance, and threshold vs. non-threshold MoA. Ad hoc assumptions should be pointed out, and assumptions for which there is unlikely to be further support from additional data, based on what is already known from the current data set, should also be pointed out. There should also be text accompanying the table that clearly summarizes the basis for the reasoning and walks through the table. By this point in the text (should be presented in the conclusion of the HBWoE), however, these assumptions and categorizations should be very clear. There should be nothing new at this point. The table should be a point-by-point comparison of the reasoning for one account against the other. The point of the table is to be explicit about each time an assumption is considered ad hoc or that additional data will not support it, based on what is already known, and to clearly spell out the counter arguments so that the relative weights of the accounts can be assessed. The weaker account is the one with more ad hoc assumptions and/or where additional studies are unlikely to support assumption.

The HBWoE approach has some similarities to other frameworks, but also some important differences. Like the Mode-of-Action/Human-Relevance (MoA/HR) Framework, HBWoE uses an assessment of the understanding of mode of action and its component key events to probe the relevance of animal studies to human risk potential. The MoA/HR approach, however, is focused on assessing the human relevance of *particular* studies based on an assessment of whether the agent's ability to cause all of the key events in the MoA (to which the study's results

are attributed) are known to, or can be expected to, operate in parallel in humans. The MoA/HR starts by asking whether the animal MoA is known, and if not, it does not proceed, whereas the HBWoE can play through the evaluation of a hypothetical mode of action, assessing both the plausibility of the MoA in animals (based on weighing available evidence for and against) and the implications for that MoA in humans, if it were true. Indeed, several alternative MoA hypotheses, and their differing human risk consequences, can be evaluated to show how conclusions are contingent on accepting certain assumptions about MoA. A second difference is that, where MoA/HR focuses on the applicability of particular animal studies and endpoints to humans (a question of extrapolating how particular studies and outcomes relate to human risk potential), HBWoE has a broader focus on evaluating the whole base of available studies. HBWoE asks not only how each study relates to human risk potential, but also how those studies relate to one another in terms of consistency in outcomes, further evidence for or against the proposed MoA events and their roles, or insights into how well the proposed causal effects generalize across situations. Importantly, it tracks what further assumptions might be needed to reconcile apparent contradictory or inconsistent results among the set of studies, and the plausibility of these assumptions (and evidence for and against them) becomes part of the overall weight-of-evidence evaluation. In short, HBWoE is not just about assessing applicability of pieces of evidence, but about integrating interpretation of bodies of evidence. As such, it is naturally focused on important questions such as how to incorporate both animal and human data into evaluations, how to bring to bear *in vitro* information about metabolism, kinetics, gene expression, and so on. This integration is not just adding up bits of evidence, but rather using the whole array of information to aid in the interpretation of each part. It uses, for instance, animal study results to help in the interpretation of whether an epidemiology study's observed patterns of association are consistent with understanding of biology of the agent and its interaction with living systems.

Finally, it is noteworthy that, although the HBWoE approach is not explicitly about dose-response uncertainty, it has important contributions to make to this question. Dose-response analysis has "statistical" uncertainties about curve fits and measurement errors, but the larger uncertainties are more qualitative – which endpoints are reliably concluded to be caused by an agent, which data sets best represent those endpoints, which models (with which low-dose extrapolations) should be used, what interactions with other agents or background processes might contribute to risk levels, what basis for variation in human sensitivity might exist, *etc.* These are not readily treated as quantitative measures of uncertain extrapolations, but the uncertainty in dose-response evaluation can be better characterized by doing a dose-response analysis for each viable choice, and the relative defensibility among the various alternatives assessed by noting the judgments about their relative appropriateness and plausibility as drawn from the HBWoE analysis. That is, HBWoE provides a route for using the insights into the basis for (and uncertainties about) human risk inference that are developed during the Hazard Characterization process and bringing them to bear on the understanding of uncertainty in quantitative risk of the dose-response relationships for those hazards.

**Example Table 1. Comparative Reasoning for Accounts**

<b>Account for Hypothesis #1</b>	<i>Ad hoc</i> <b>explanation?</b>	<b>Plausibility that additional data will support explanation</b>	<b>Account for Hypothesis #2</b>	<i>Ad hoc</i> <b>explanation?</b>	<b>Plausibility that additional data will support explanation</b>
<b>Animal Data</b>					
explanation and reasoning for key observation	yes	plausible	explanation and reasoning for key observation - may be counter to hypothesis #1		plausible
explanation and reasoning for key observation		plausible	explanation and reasoning for key observation - may be counter to hypothesis #1		plausible
explanation and reasoning for key observation		plausible	explanation and reasoning for key observation - may be counter to hypothesis #1		plausibility can reasonably be excluded
<b>Epidemiology Data</b>					
explanation and reasoning for key observation		plausible	explanation and reasoning for key observation - may be counter to hypothesis #1		plausibility can reasonably be excluded
<b>Mechanistic Data</b>					
explanation and reasoning for key observation		plausible	explanation and reasoning for key observation - may be counter to hypothesis #1		plausibility can reasonably be excluded
<b>Human Relevance</b>					
explanation and reasoning for key observation	yes	plausible	explanation and reasoning for key observation - may be counter to hypothesis #1		plausibility can reasonably be excluded
explanation and reasoning for key observation		plausible	explanation and reasoning for key observation - may be counter to hypothesis #1	yes	plausibility can reasonably be excluded
<b>Relative weight of evidence for accounts</b>	stronger			weaker	

Shaded cells are *ad hoc* assumptions and/or where additional data are unlikely to support explanation.

Accounts with the fewest *ad hoc* assumptions and/or assumptions where additional data are unlikely to support explanation are considered stronger.

## 2. Problem Addressed by the Method

The National Academy of Sciences (NAS) Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde (NRC, 2011) proposed a "roadmap" for reform and improvement of the Agency's risk assessment process. Specifically, it called on the Agency to undertake a program to develop a transparent and defensible methodology for weight-of-evidence (WoE) assessments. In a broad sense, all chemical risk assessment is a form of WoE evaluation, in that it involves the interpretation of a body of scientific studies to discern what they can tell us about estimation of potential exposure impacts on populations and/or systems of interest. This means that risk assessment needs a way to clearly illustrate the methodology and logic applied to the WoE processes; *i.e.*, sorting out the potential interpretations with respect to the methods of science, acknowledging data limitations, combining and integrating evidence, providing a basis for applying sound judgment, and identifying the most supportable conclusions.

The HBWoE approach provides a framework for weighing evidence for large and complex datasets in a way that can be clearly and logically applied to the risk assessment process, and therefore addresses the ultimate goal expressed in the NAS review. HBWoE provides a practical and transparent approach that takes the reader through the logic of the WoE analysis – how and why available data are considered to inform the judgments about the existence and nature of causal processes. On a case-specific basis, problem formulation within the HBWoE approach occurs as one of the first steps in the evaluation, and is aimed at articulating the various, and often competing, overarching hypotheses, and articulating the goal of creating a thorough characterization of current scientific knowledge to judge the potential toxicological hazards or risks of a particular agent, in a manner that illustrates sufficiently the tracing of the logic through the various competing accounts of the data so that the various interpretations can be compared.

The approach does not eliminate the need for scientific judgment, and often may not lead to a definitive choice of one interpretation over the other, but it will clearly lay out the logic for how one weighs the evidence for and against each interpretation. Only in this way is it possible to have constructive scientific debate about potential causality that is focused on an organized, logical "weighing" of the evidence.

## 3. General Applicability of the Method

The HBWoE method is generally applicable to all chemical risk assessment. The approach is particularly useful when the data sets are large and complex and contain conflicting results that are difficult to interpret. Although weighing evidence in a clear and transparent manner is also necessary for small data sets, a full HBWoE evaluation is not required for data sets that are more clearly consistent across studies, or when the data sets are such that the accounting of the data and how they bear on the conclusions of the assessment are not so varied within the scientific community. That is, although it is necessary to integrate the data from various realms of evidence and to take the reader through the logic of how the available data support the conclusions of the assessment (even for small, less complicated data sets), a full comparison of accounts for these types of data sets is not likely to be necessary.

Another aspect to consider is whether potential modes of action for a given agent have been clearly articulated. Depending on how rich the data base is for a given chemical, there may or may not be enough data to bear on a potential mode of action. Since the proposed modes of action often form the basis for overarching and competing hypotheses, a thin data set may not lend itself to a full weighing of the evidence and comparison of accounts based on proposed modes of action. In this case, the approach can be applied to lay out the key pieces of information that are available and ask appropriate questions of the data to guide future studies aimed at obtaining a full data set (*i.e.*, epidemiology, animal toxicology, mechanistic, and toxicokinetic) that can be integrated in a way that will allow the different realms of evidence to inform interpretation of each other, with the ultimate goal of proposing a robust, biologically plausible mode of action that can be used to guide risk assessment processes.

#### 4. Overall Strengths and Weaknesses of the Method

The strengths of the HBWoE method are that it:

- emphasizes tracing the logic of how the available data support (or refute) the conclusions of the assessment;
- compares the tracing of logic for alternate accounts of the available data and how each is supported (or refuted) by the available data;
- emphasizes integration of all realms of evidence (*i.e.*, epidemiology, animal toxicology, mechanistic, and toxicokinetic) so that different realms of evidence are allowed to inform interpretation of each other;
- emphasizes the need to clearly convey (through use of illustrations, tables, *etc.*) the comparison of accounts and overall WoE analysis results so that the analysis can guide constructive discourse with the scientific community and future risk management decisions; and
- is flexible in specific application, yet systematic in the overall goal as guided by the seven aspects of the framework.

One may consider a weakness of the method to be that, by necessity, it is flexible and therefore may not always be applied adequately; *i.e.*, one cannot assume that use of the method will simply provide the best analysis. The approach requires judgment that needs to be checked by those interpreting the results, and may lead to disagreements and stimulate further scientific debate and discussion, and possible refinement of proposed modes of action or other overarching hypotheses. This is, in fact, part of the method. As such, the HBWoE method should be viewed as being iterative, interactive, and flexible. And, therefore, the approach will often feel unstructured and complex. The challenge is keeping the ultimate goal in mind – integration of all relevant data logically and transparently so that biological plausibility and human relevance will guide future risk assessment processes.



## 5. Minimum Data Requirements and Types of Data Sets Needed for Method

There are no minimum data requirements for the HBWoE method. As discussed in #3 above, however, depending on how rich the data base is for a given chemical, there may or may not be enough data to bear on potential modes of action. Or the data may not contain significant uncertainties and inconsistencies from which various interpretations have been made that require a full comparison of the different accounts. Although rich, complex data sets are not a requirement for application of the method, these types of data sets benefit most effectively from the HBWoE framework. The framework can be useful for smaller, less complicated data sets as a guide for proposing new biologically plausible modes of action, or for developing a more complete data set that can further inform the risk assessment processes.

Importantly, no matter where one is in the processes of weighing evidence, the ultimate goal should be to ask the appropriate, logical questions of the available data (no matter how much) to either guide further studies or the natural tracing of the logic within the evidence at hand to reach conclusions that are scientifically sound, supported, and biologically plausible.

## 6. Does the Case Study:

### A. Describe the dose-response relationship in the dose range relevant to human exposure?

As described in the case study, and guided by the results of our HBWoE evaluation, our approach is to consider the applicability and limits on the animal responses to serve as a basis for estimation of potential human risk. We do this by considering the potential mode of action underlying the effects seen in animal bioassays, including evaluation of the metabolic activation and detoxification, as well as the nature, tissue locations, and dependence on tissue-dose of key precursor responses. Species differences in tissue dosimetry (*via* application of PBPK models, if available) are used to evaluate whether parallel tissues in humans will be subject to tissue doses that could prompt the key events of the apparent mode of action. Our approach further considers whether the tissue doses required to prompt a particular mode of action are achievable with typical human exposures; therefore, describing the dose-response relationship in the dose range relevant to humans.

Further, HBWoE does not treat different quantitative modeling approaches for given endpoints as quantitative measures of uncertain extrapolations. Rather, the uncertainty in dose-response evaluation can be better characterized by conducting dose-response analyses for each viable choice. The various analyses can then be compared and the relative defensibility among the alternatives assessed by noting the judgments about their relative appropriateness and plausibility as drawn from the HBWoE analysis.

### B. Address human variability and sensitive populations?

The HBWoE approach accommodates an evaluation of human variability and sensitivity, if necessary, for a given chemical agent, and can be a key part of the evaluation if the data support

it. Questions about variability and sensitivity in the human population should be asked upfront as part of the initial organizing of the relevant studies, and these studies should be integrated into and given appropriate weight in the evaluation similarly to how all other data are considered. Studies of human variability and sensitivity constitute lines of evidence within the epidemiology or mechanistic (*e.g.*, polymorphisms of genes known to be involved in the mechanism of action) data that need to be considered within and across all realms of evidence so that the data can be fully considered and integrated as part of the WoE evaluation. The practical outcome of working these data through the WoE evaluation may be realized through application of, or proposals to gain more insight into, appropriate uncertainty factors that are based on a more scientifically sound understanding of the actual variability within a human population and that can be used in place of default values.

### **C. Address background exposures or responses?**

The results of the HBWoE approach, particularly if applied to quantitative risk assessment for a given chemical, should be presented in a clear and transparent manner allowing for practical application of toxicity values to risk management decisions. As such, the ultimate goal of the HBWoE evaluation is to present a biologically plausible MoA that is most strongly supported by the WoE (comparing and contrasting to other proposed modes of action), the associated exposure concentration that would be necessary to lead to that MoA and associated adverse effect in humans, and how that exposure concentration compares to background and typical human exposure concentrations.

Consideration of background levels of cancer incidence in the human population may also be important in the HBWoE evaluation. For example, nasal cancer in the human population is very low. Therefore, occurrence of nasal cancer in naphthalene-exposed individuals should have been notable had it occurred (Rhomberg *et al.* 2010). This observation is an important part of the data integration phase of HBWoE evaluation.

Further, consideration of background levels of biomarkers of exposure or effect are important in the HBWoE evaluation. For example, interpretation of studies that evaluate formaldehyde DNA adduct or blood levels requires an understanding of how these levels compare to endogenous levels (Rhomberg *et al.* 2011), and is a key part of the data integration phase of HBWoE evaluation.

### **D. Address incorporation of existing biological understanding of the likely mode of action?**

Yes. As discussed above, the mode of action is what provides the underlying commonality across species and its understanding is key to weighing and integrating evidence from a large dataset in the HBWoE framework, particularly if there are contrasting modes of action that have been put forth within the scientific community. HBWoE can be used to evaluate data within the current data set or to guide future studies aimed at obtaining a full data set (*i.e.*, epidemiology, animal toxicology, mechanistic, and toxicokinetic) that can be integrated in a way that will allow the different realms of evidence to inform interpretation of each other, with the ultimate goal of

proposing a robust, biologically plausible mode of action that can be used to guide risk assessment processes.

**E. Address other extrapolations, if relevant – insufficient data, including duration extrapolations, interspecies extrapolation?**

The HBWoE method focuses on integration of all relevant data (*i.e.*, epidemiology, animal toxicology, mechanistic, and toxicokinetic), including consideration of data quality and sufficiency, null and negative as well as positive studies, data from all species, tissues, and exposure durations. Therefore, this integration allows for a form of extrapolation that is more of a "generalization" across realms of evidence, guided by the need to identify something in common regarding the causal processes in the study situation and the human population of interest. This generalization is a form of extrapolation that involves initial qualitative evaluation and integration of the data that can then be used to guide a more quantitative extrapolation of the data for derivation of human toxicity values.

**F. Address uncertainty?**

The HBWoE approach is especially suited to deal with complex and conflicting data sets with large numbers of uncertainties. Thus, the HBWoE approach aids in addressing uncertainty in a qualitative manner, although it does not provide a quantitative uncertainty analysis.

**G. Allow the calculation of risk (probability of response for the endpoint of interest) in the exposed human population?**

As discussed above in A and C, the ultimate goal of the HBWoE evaluation is to present a biologically plausible MoA that is most strongly supported by the WoE (comparing and contrasting to other proposed modes of action), the associated exposure concentration that would be necessary to lead to that MoA and associated adverse effect in humans (derivation of toxicity values), and how that exposure concentration compares to typical human exposure concentrations. Thus, it aids in the evaluation of risk, although no novel methods are presented for calculating probability of response.

**H. Work practically? If the method still requires development, how close is it to practical implementation?**

Yes, the method is considered to work practically and has been applied to several chemicals (Rhombert *et al.*, 2010, 2011; Prueitt *et al.*, 2011; Bailey *et al.*, 2012), evolving with each application. Refinement of the method is likely through further application. It should be noted, however, that improvement of the method is likely to take the form of demonstration of its flexibility across different chemicals and data sets (thereby providing examples of different approaches) rather than becoming more structured. That is, the goal of the HBWoE is to integrate data across realms of evidence in a way that is transparent and logical. Therefore, the WoE practitioner will need to apply different approaches for different chemicals depending on

the specific nature and quality of the data, overarching hypotheses, and uncertainties/inconsistencies within the dataset.

## References

Bailey, LA; Prueitt, RL; Rhomberg, LR. 2012. "Hypothesis-based weight-of-evidence evaluation of methanol as a human carcinogen." *Regul. Toxicol. Pharmacol.* 62 :278-291.

Hill, AB. 1965. "The environment and disease: Association or causation?" *Proc. R. Soc. Med.* 58 :295-300.

National Research Council. April 2011. "Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde." National Academies Press (Washington, DC) at <http://www.nap.edu/catalog/13142.html> 194p.

Prueitt, RL; Goodman, JE; Bailey, LA; Rhomberg, LR. 2011. "Hypothesis-based weight-of-evidence evaluation of the neurodevelopmental effects of chlorpyrifos." *Crit. Rev. Toxicol.* 41 (10) :822-903.

Rhomberg, LR; Bailey, LA; Goodman, JE. 2010. "Hypothesis-based weight of evidence: A tool for evaluating and communicating uncertainties and inconsistencies in the large body of evidence in proposing a carcinogenic mode of action - Naphthalene as an example." *Crit. Rev. Toxicol.* 40 :671-696.

Rhomberg, LR; Bailey, LA; Goodman, JE; Hamade, A; Mayfield, D. 2011. "Is exposure to formaldehyde in air causally associated with leukemia? - A hypothesis-based weight-of-evidence analysis." *Crit. Rev. Toxicol.* 41 (7) :555-621.