June 11, 2013

Comments on the Draft Office of Health Assessment and Translation (OHAT) Approach For Systematic Review and Evidence Integration For Literature-Based Health Assessments (78 FR 37 published February 25, 2013)

Submitted by email to: andrew.rooney@nih.gov or
And, submitted online at: http://ntp.niehs.nih.gov/go/38673

These comments are submitted by in response to the NTP invitation for public comments on the Draft OHAT Approach – February 2013. The NTP included two protocols illustrate the application of this framework: (1) BPA exposure and obesity and (2) PFOA or PFOS exposure and immunotoxicity.

More information is here: http://ntp.niehs.nih.gov/?objectid=960B6F03-A712-90CB-8856221E90EDA46E

Thank you for the opportunity to provide comments on the OHAT draft approach to systematic review. Overall we are very pleased with the effort and results. The framework is organized into the following seven basic steps:

Step 1 – Prepare topic
Step 2 – Search for and select studies for inclusion
Step 3 – Extract data from studies
Step 4- Assess the quality of individual studies
Step 5- Rate the confidence in the body of evidence
Step 6 – Translate confidence ratings into level of evidence for health effect
Step 7 – Integrate evidence to develop hazard identification conclusions

The systematic review methodology that OHAT has proposed provides an elegant framework – simple, functional, and effective - for enhancing transparency and communication, promoting consistency, and facilitating reproducibility across literature-based evaluations of hazardous chemicals.

The OHAT draft represents the state of the science with regards to applying a systematic review methodology to environmental health assessments of hazardous chemicals – developed on the sturdy foundation of established and validated methods developed for other datasets, while pushing ahead

with cutting edge approaches to accommodate new types of data and new fields of information relevant to environmental health.

The key points of our comments are:

- The OHAT plan adapts a framework designed for evaluating treatments (vaccination strategies) to environmental health data, requiring NTP to be vigilant to ensure that the outcomes are health-protective.
- We support a comprehensive literature search in Step 2.
- We support the systematic and transparent approach to data extraction in Step 3.
- We support the framework approach of evaluating "believability" of study results, rather than internal validity, reporting quality, GLP compliance, or ToxRTool in Step 4.
- We support the exclusion criteria for low-quality studies, but suggest that the NTP consider an additional confidence rating to reflect high confidence studies in Step 5.
- Mechanistic data could raise the hazard identification in step 7, but should not be used to explain away hazard evidence. We support the framework approach for evaluating the consistency across studies to determine monotonic or non-monotonic dose-response patterns.
- We support the framework approach of excluding underpowered studies that fail to find an effect (null-association), but not studies that find an effect despite being underpowered.
- When information is missing or unreliable, the framework should use established defaults that will protect health, and set stringent criteria for when to depart from health-protective defaults.

The OHAT plan adapts a framework designed for evaluating treatments (vaccination strategies) to environmental health data, requiring NTP to be vigilant to ensure that the outcomes are health-protective

The OHAT draft builds on the GRADE framework - "Grading of Recommendations Assessment, Development and Evaluation". The GRADE framework is used by the Center for Disease Control and Prevention (CDC) to make recommendations regarding vaccinations for at-risk groups (for example, age groups or high-exposure groups), and for specific individuals in the context of a patient-clinician setting.[1] The GRADE framework considers the type and quality of evidence about a vaccine's expected health benefits and risks, the values and preferences of an affected person or group, and the health economic impacts. While the GRADE framework is a reasonable starting-place, it is designed for translating medical research into clinical strategies specifically addressing vaccination needs. For example, it is best suited to randomized human clinical trial data and human clinical observational studies, which are rarely available for environmental health endpoints. Conversely, the GRADE framework is fairly silent on how to assess data from animal studies, *in vitro* or *in silico* (computer simulation) studies, and non-clinical observational studies, which are the data that is often available for environmental health endpoints. The NTP will need to be flexible and vigilant to make sure that the draft OHAT framework is maximizing its ability to effectively use the datasets relevant to environmental health to support decision strategies that result in improved environmental and health protections.

[1] MMWR May 11, 2012. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6118a3.htm

We support a comprehensive literature search in Step 2

For selecting studies for inclusion (Step 2) we support the draft OHAT framework approach of being as comprehensive as reasonably possible, by gathering the published, unpublished, and "grey" literature (publicly available government reports, etc.) as part of the literature search. It is appropriate to be as comprehensive as possible in gathering all the available data as a first step. These data can then be peer-reviewed internally within the OHAT and NIEHS established review mechanisms.

We support the systematic and transparent approach to data extraction in Step 3

For extracting data from studies (Step 3) we support the framework approach of using template forms customized for the type of study (animal, human, in vitro) and the specific needs of the evaluation. Quality control is built into the process during this step, and all data extraction files will be made publicly available. This level of transparency will increase meaningful public participation and communication.

We support the framework approach of evaluating "believability" of study results, rather than internal validity, reporting quality, GLP compliance, or ToxRTool in Step 4

For assessing the quality of individual studies (Step 4), the OHAT framework evaluates the risk of outcome-specific bias using five domains: selection bias, performance bias, attrition bias, detection bias, and reporting bias. We support this as reflecting the state of the science for approaches to identifying bias in a systematic and transparent manner. We support the OHAT framework approach of considering more than reporting quality for a measure of study quality for animal and human studies. Instead, the OHAT framework wisely uses an approach that is a much better measurement of the internal validity of a study, which really gets to the heart of the "believability" of the study results, rather than simply the articulation or reporting quality. Reporting quality is not the same as internal validity, which is a better measure of study quality.

If the systematic framework were to only rely on reporting quality as a measure of study quality, it would favor/bias towards GLP-compliant (Good Laboratory Practice) studies, when, ironically, the GLP-compliant studies may actually be the most likely to be insensitive to health endpoints being measured. "Good Laboratory Practices" is a standard for animal care and data collection required for industry laboratories in response to fraudulent practices documented in the 1970s. Industry-funded studies are required by EPA and FDA to follow so-called Good Laboratory Practices (GLP) standards, which include specified approaches to recordkeeping to facilitate audits and reduce fraud (54 Fed. Reg. 34034 (August 17, 1989). GLP requirements are not necessarily associated with higher quality research, proper study design or correct statistical analysis.[2] In most cases, GLP studies have not even undergone scientific peer-review and publication. GLP studies are most often designed to identify major toxic effects (apical effects) like cancer. The problem is that major (apical) endpoints will not be predictive or indicate early-warnings of potential toxicity leading to "major" adverse health outcomes. GLP studies don't necessarily use modern methods for evaluating chemicals and aren't designed to grapple with the problems of low-

[2] Myers, J. P., F. S. vom Saal, et al. (2009). "Why public health agencies cannot depend on good laboratory practices as a criterion for selecting data: the case of bisphenol A." Environ Health Perspect 117 (3): 309-15.

dose exposures, endocrine or hormonal effects, behavioral or learning effects, immunotoxicity, cardiotoxicity, or upstream effects like reduced sperm count or reduced anogenital distance which are predictors of infertility. The draft OHAT framework represents the state of the science for approaches to evaluating a study's internal validity in a systematic and transparent manner.

The ToxRTool was developed to assess the reporting quality of a study, and is not an appropriate measure of the internal validity of *in vitro* studies or of risk of bias or overall study quality. NTP says that the reliability categories utilized in the ToxRTool are the same as the Klimisch codes of reliability (Klimisch et al. 1997) (See BPA assessment, page 25, footnote 7).[3] Since the Klimisch codes favor GLP compliant studies, than using ToxRTool would be subject to the same criticism as using either Klimisch codes or GLP as measures of study quality.

We support the exclusion criteria for low-quality studies, but suggest that the NTP consider an additional confidence rating to reflect high confidence studies in Step 5

We support the draft OHAT framework approach for identifying and excluding very low-quality studies (Step 5). To do otherwise would be to skew the outcome of the assessment. We also support the exclusion or downgrading of null association studies that are underpowered, if they are inconsistent with the whole body of literature (BPA case study p. 45). The OHAT framework identifies some study aspects that would lead to downgrading the confidence rating in that study include: risk of bias, unexplained inconsistency, indirectness in the relationship between a measured outcome and a health effect, imprecision, and publication bias serious enough to significantly decrease confidence in the body of evidence. Ultimately, the NTP conclusions will be based on the whole body of literature, excluding the really low-confidence studies. We support this as reflecting the state of the science for approaches to evaluating study quality in a systematic and transparent manner.

We support the "confidence ratings" approach (Step 5, Box 1), but believe that the gap is too great between the highest and second highest levels of confidence, which are characterized as "highly likely" and "may be" reflected by the apparent relationship. We suggest that you consider that there should be some version of "probably reflected" between those two.

We suggest that NTP reconsider its evaluation of observations studies (lacking controlled exposures), to rank them as higher confidence where they have measurements over a useful range of exposure in Step 5

Why should an observational study that does not establish controlled exposures not be able to get a "high confidence" rating (Step 5, page 4), if the exposures were measured, and span a useful range? The framework provides as an example that a prospective cohort study – considered a very strong study design – would get a rating of "moderate confidence" even if it had all three remaining features (exposures occur prior to development of outcome, outcome is assessed on an individual level, and a comparison group is used within the study), simply because it lacked a controlled exposure. This seems an unrealistic bar that would unfairly reduce the confidence ranking of human observational studies that are highly regarded in the scientific community and provide important real-world human dose-response

[3] http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/BPAProtocolDraft.pdf

information. We suggest that the OHAT framework reconsider this unscientific approach that would artificially diminish some of the most valuable data available to the NTP.

Mechanistic data could raise the hazard identification in step 7, but should not be used to explain away hazard evidence.

We support the draft OHAT framework approach of considering – but, not requiring – mechanistic data as part of the overall evaluation, but not as any more or less valuable than other evidentiary data (Step 7). Mechanistic (or mode of action, MOA) data should be treated as a parallel stream of data, and can be very helpful in interpreting human and animal data. However, mechanistic data should not be seen as either necessary or sufficient for interpreting or evaluating other data. Numerous hazardous materials share the same mechanism of toxicity, and a single material can have numerous mechanisms of toxicity. Because of these complexities, a committee of the National Academies recommended that cumulative risk assessments group chemicals together that cause common adverse outcomes rather than focusing exclusively on structural similarity or on similar mechanisms of action." (National Academies report on Phthalates and Cumulative Risk Assessment, p. 9) Further, OHAT/NTP should dismiss arguments about MOA that are really arguments about the potency or degree of risk, and are therefore not relevant to the listing decision. However, strong evidence from mechanistic/MOA studies could support a conclusion and raise it to a level of increased concern.

We support the framework approach for evaluating the consistency across studies to determine monotonic or non-monotonic dose-response patterns

We support the draft OHAT framework regarding evaluating monotonic and non-monotonic dose response patterns (BPA case study p. 53). The framework requires looking across all the relevant studies to see a consistency across studies in the dose-gradient. If the observed consistent pattern is non-monotonic, than the framework would upgrade for that pattern, same as would be done for a consistent monotonic dose-gradient. The framework approach would lead to an upgrade for a consistent observed pattern across studies in which there is an acceptable level of confidence, even without a prior knowledge. In other words, there should be the same data requirement for making a determination of either a monotonic and or a non-monotonic dose gradient. This represents the state of the science.

We support the framework approach of excluding underpowered studies that fail to find an effect (null-association), but not studies that find an effect despite being underpowered

For a continuous endpoint, the draft OHAT framework requires an assessment to determine if the study was adequately powered, and if the results are consistent across studies. The framework proposes that if there is inconsistency across studies, and it's the underpowered studies are showing null association, than they will be excluded from consideration. In other words, the framework proposes to eliminate null-association studies that are inconsistent and underpowered, but not to eliminate studies that may be underpowered if they do find and effect. This is because an underpowered study that fails to find an effect cannot be interpreted, but an underpowered study that does find an effect indicates that the effect is real. As an analogy, if you reach into a haystack a few times (an underpowered study) and don't find a needle (a null study), you cannot conclude whether or not there may be needles in the haystack, whereas if you do find a needle (an underpowered study that finds an effect), than there is at least one

needle, and probably more, in the haystack (the effect is real). The OHAT framework approach represents the state of the science.

The critical issue with the OHAT guidelines is confidence in results of studies of people, especially confidence that null results predict absence of risk. Studies in people depend on exposure assessments, but confidence in exposure assessments are limited by the range and duration of exposure in the studied population. No matter how elaborate and extensive an exposure assessment, the assessment can't make up for a cohort too small, exposed at too low a level, or for too short a time. For hazard identification purposes, a null study should be assessed taking into account the upper confidence limit of risk rate, and the lower confidence limit of exposure levels to estimate the risk possibly ruled out.

When information is missing or unreliable, the framework should use established defaults that will protect health, and set stringent criteria for when to depart from health-protective defaults

The framework is noticeably silent on the issue of health-protective default assumptions. In practice, in the absence of compelling data to the contrary, the framework will appropriately interpret animal data and other available data according to well-established principles. The use of mechanistic/MOA data must be interpreted relative to the plausibility of the default, and not as if the alternative to the proposed mechanism/MOA were no proposed mechanism/MOA. When information is missing or unreliable, the framework should be clear and consistent that its approach is to use scientifically-based default assumptions that will protect health to improve the timeliness of the chemical assessment and decision-making process, and set clear scientifically-based criteria for when to depart from these assumptions.[4] In the landmark "Science and Decisions" report (NAS, 2009), the NAS committee concluded that, "established defaults need to be maintained for the steps in the risk assessment that require inferences."[5] The NAS committee recommended that EPA and other agencies – presumably including the NTP - update default factors and assumptions based on the best current science, identify where unstated or implicit assumptions are used, and replace these with explicit assumptions wherever possible. These recommendations push Agencies to, "continue and expand use of the best, most current science to support or revise its default assumptions,"[6] making the assumptions stronger, rather than reducing reliance on them. In fact, the committee specifically recommended that EPA develop "clear standards for departures from defaults."[7] The committee also noted that establishing, "clear criteria for departure from defaults can provide incentives for third parties to produce research" that can reduce uncertainty and, over time, result in more accurate assessments. Importantly, by using the established defaults more often, the OHAT framework could avoid "the delay entailed by having to re-examine generic information with every new risk assessment."[8] The OHAT framework should also evaluate and quantify, when possible, the impact of the uncertainty associated with a default assumption, including a description of how using a default versus the chosen alternative assumption affects the decisions that protect the environment and public health.

---

[4] NRDC Issue paper. Strengthening toxic chemical risk assessments to protect human health. S Janssen, J Sass, T Schettler, G Solomon. February, 2012.
http://switchboard.nrdc.org/blogs/jsass/nrdc_issue_paper_better_risk_a.html

[5] Science and Decisions: Advancing Risk Assessment. National Research Council of the National Academies. (2009), p. 7.

[6] Science and Decisions, p. 207.

[7] Science and Decisions, p. 199.

[8] Science and Decisions, p. 191.

Thank you for the opportunity to provide comments.

Respectfully,

Jennifer Sass, Ph.D.
Senior Scientist, Natural Resources Defense Council (NRDC)**
Professorial Lecturer, George Washington University*

Adam M. Finkel, Sc.D., CIH
Senior Fellow and Executive Director, Penn Program on Regulation and
Professor of Environmental and Occupational Health, UMDNJ School of Public Health*
Voting member, NTP Executive Committee (1995-2000)

Franklin E. Mirer, PhD, CIH
Professor of Environmental and Occupational Health Sciences
CUNY School of Public Health*
Former member, NTP Board of Scientific Counselors

Charlotte Brody, RN
Vice President for Health Initiatives
BlueGreen Alliance*

*Affiliations listed for purposes of identification only

**The Natural Resources Defense Council (NRDC) is a national, non-profit environmental organization of lawyers, scientists, and other professionals. NRDC presents these comments on behalf of our 1.3 million members and online activists. NRDC does not have any financial interest in the topic of these comments.