Updates and Clarification to the
OHAT Approach for Systematic Review and Evidence Integration

The Office of Health Assessment and Translation (OHAT) Handbook (NTP 2015) outlines the standard operating procedures for systematic review and evidence integration for conducting OHAT literature-based assessments (herein referred to as the "OHAT approach"). The purpose of these assessments is to perform an evidence evaluation and communicate the resulting findings (e.g., state of the science or hazard conclusions). As outlined in the OHAT handbook, the procedures are a living document with the expectation that approaches will be updated as methodological practices are refined and strategies identified that improve the ease and efficiency of conducting a systematic review.

This document clarifies and updates the OHAT Handbook (NTP 2019) to address two topics that have been identified during the conduct of evidence evaluations: 1) the process for reaching hazard conclusions from human health data alone (i.e., in the absence of animal data or when there is low confidence in the available animal data); and 2) the process for developing confidence conclusions in the overall body of evidence across multiple outcomes, study types, or exposures.
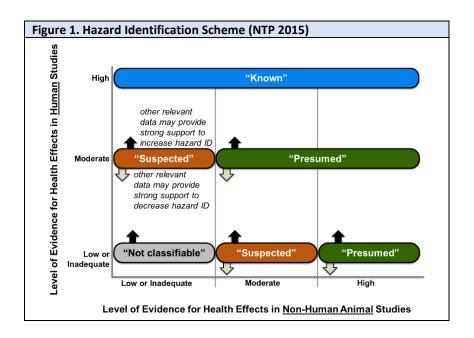
Clarifications and Updates

1) **Hazard conclusions based on human data alone**

   The OHAT approach (NTP 2015):
   The hazard identification scheme has four categories: known, presumed, suspected, and not classifiable. Three of the four hazard categories (see below) can result from evidence integration when there is human evidence and "low or inadequate" confidence in the non-human animal evidence. "Presumed to be a hazard to humans" is not an option (Figure 1).

   - o High confidence in human data (with no animal data/low confidence in available animal data) results in "known to be a hazard to humans".

   - o Moderate confidence in the human data (with no animal data/low confidence in available animal data) results and in "suspected to be hazard to humans".

   - o Low confidence or inadequate confidence in the human data (with no animal data/low confidence in available animal data) results in "not classifiable".

Figure 1. Hazard Identification Scheme (NTP 2015)

Note: the up arrows indicate hazard conclusions for which other relevant data may provide strong support to increase the hazard conclusion. Similarly, the down arrows indicate hazard conclusions for which other relevant data may provide strong support to decrease the hazard conclusion.

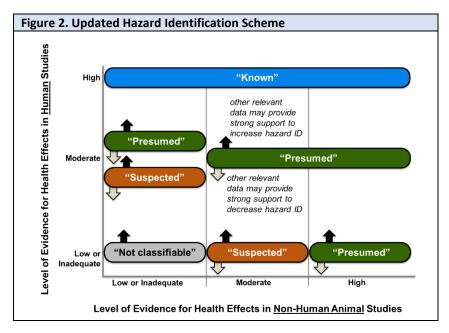**The updated OHAT approach (NTP 2019):**
The updated evidence integration approach (Figure 2) outlines how all four hazard categories could be reached when there is human evidence and "low or inadequate" confidence in the non-human animal evidence. Characteristics of a body of evidence can differ such that moderate confidence in a body of evidence for human data alone may support a hazard conclusion of suspected in some cases and presumed in other cases. The justification for the final hazard conclusion will be based on transparent evaluation criteria appropriate for the body of evidence and scientific judgement

- o NO CHANGE - Low confidence or inadequate confidence in the human data (with no animal data/low confidence in available animal data) results in "not classifiable"

- o NEW[1] - Moderate confidence in the human data (with no animal data/low confidence in available animal data) will result in either a conclusion of "suspected to be hazard to humans" or "presumed to be a hazard to humans" based on scientific judgement as to the robustness of the body of evidence that supports moderate confidence and consideration of the potential impact of additional studies.

  - ▪ The hazard rating reflects the likelihood that additional studies could impact the conclusions. For "suspected", there is a reasonable expectation that the data from new studies would impact the hazard conclusion and result in a change in the hazard rating. For "presumed", there is a low expectation that new studies would impact the hazard conclusion.

- For example, bodies of evidence that would lead to a conclusion of suspected to be a hazard include, but are not limited to: 1) a single well-designed and conducted study including multiple populations with small group sizes and/or a small magnitude of effect; 2) a few well-designed and conducted studies with small study populations or group sizes and/or small magnitude of effect; or 3) a larger number of studies with some inconsistencies in outcomes but an overall small magnitude of effect across the body of evidence.

- For example, bodies of evidence that would lead to a conclusion of presumed to be a hazard include, but are not limited to: 1) a few well-designed and conducted studies with large magnitude of effect; 2) a few well-designed and conducted studies with large study populations or group sizes with a small magnitude of effect; or 3) a larger number of studies showing a consistent pattern of a small magnitude of effect across the body of evidence.

o NO CHANGE - High confidence in human data (with no animal data/low confidence in available animal data) results in "known to be hazard to humans"

o NO CHANGE - High confidence in human data with moderate or high confidence in animal data will result in "known to be hazard"

Figure 2. Updated Hazard Identification Scheme

Note: the up arrows indicate hazard conclusions for which other relevant data may provide strong support to increase the hazard conclusion. Similarly, the down arrows indicate hazard conclusions for which other relevant data may provide strong support to decrease the hazard conclusion.

3

2) **Process for developing confidence conclusions in an overall body of evidence**

   The OHAT approach (NTP 2015):
   The 2015 Handbook addresses developing confidence conclusions in an overall body of evidence across multiple outcomes or across multiple study types but does not address considerations across multiple exposure metrics.

   o   Biologically related outcomes may inform confidence in the overall body of evidence for a health effect. In such situations, biologically related outcomes are evaluated separately for Step 5 (Synthesize the evidence and rate confidence in the body of evidence) and then considered together and re-evaluated for properties that relate to downgrading and upgrading the body of evidence to reach a confidence determination.

   o   Separate bodies of evidence by study type may inform confidence in the overall body of evidence across study types. Similar to evaluating biologically related outcomes, bodies of evidence are evaluated separately then considered together and re-evaluated for properties that relate to downgrading and upgrading the overall body of evidence to reach a confidence determination.

   The updated OHAT approach (NTP 2019):
   The updated approach describes the development of confidence conclusions for an overall body of evidence across multiple exposure metrics in Step 5.

   o   Different exposure metrics can individually provide information about the association between an exposure and a health outcome and may collectively inform confidence in the body of evidence for that association.   This update allows for evidence from each exposure metric to be evaluated separately and then considered together to inform confidence in the overall body of evidence analogous to the stepwise process for evaluating biologically related outcomes to inform the overall body of evidence.

   o   Different exposure metrics are evaluated separately for Step 5 (Synthesize the evidence and rate confidence in the body of evidence) and then considered together and re-evaluated collectively for properties that relate to downgrading and upgrading the body of evidence to reach a confidence determination.

Background

1) Selected text below is copied from the OHAT approach (NTP 2015) as additional background.

Integration of Human and Animal Evidence

Hazard identification conclusions are initially reached by integrating the highest level-of-evidence conclusion for a health effect(s) from the human and the animal evidence streams. On an outcome basis, this approach applies to whether the data support a health effect conclusion or provide evidence of no health effect. Hazard identification conclusions may be reached on individual outcomes (health effects) or groups of biologically related outcomes, as appropriate, based on the evaluation's objectives and the available data. The five hazard identification conclusion categories are as follows:

- • Known to be a hazard to humans
- • Presumed to be a hazard to humans
- • Suspected to be a hazard to humans
- • Not classifiable as a hazard to humans
- • Not identified as a hazard to humans

When the data support a health effect, the level-of-evidence conclusion for human data from Step 6 is considered together with the level of evidence for non-human animal data to reach one of four hazard identification conclusions [*link to the Hazard Identification Scheme (2015) or Figure 1 in this 2019 update*]. If one evidence stream (either human or animal) is characterized as "Inadequate Evidence," then conclusions are based on the remaining evidence stream alone (which is equivalent to treating the missing evidence stream as "Low" in Step 7).

Confidence Ratings: Assessment of Body of Evidence

The confidence rating for a given health outcome is developed by considering the strengths and weaknesses in a collection of human and animal studies that constitute the body of evidence. The confidence rating reflects confidence that the study findings accurately reflect the true association between exposure to a substance and an effect. The confidence rating approach described below (Rooney et al. 2014); *Figure 6 from the OHAT Approach (2015) Assessing Confidence in the Body of Evidence*) is based primarily on guidance from the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group (Balshem et al. 2011, Guyatt et al. 2011a). The GRADE framework is applied most often to evaluate the quality of evidence and the strength of recommendations for health care interventions based on human studies (typically randomized clinical trials). The appeal of the GRADE framework is that (1) it is widely used (Guyatt et al. 2011f), (2) it is conceptually similar to the approach used by the Agency for Healthcare Research and Quality for grading the strength of a body of evidence of human studies (AHRQ 2012a), (3) the Cochrane Collaboration has adopted the principles of the GRADE system for evaluating the quality of evidence for outcomes reported in systematic reviews (Higgins and Green 2011), and (4) the GRADE Working Group is committed to method development/validation and has recently established subgroups to focus on application of GRADE to environmental health and animal studies. Embedded within the GRADE approach is consideration of principles that are consistent with causation as discussed by Sir Austin Bradford Hill (Hill 1965, Schünemann et al. 2011).

The OHAT framework includes several refinements to GRADE that were necessary to accommodate the need to integrate data from multiple evidence streams (human, animal, in vitro) and focus on observational human studies rather than randomized clinical trials. This is important because ethical considerations virtually preclude use of human controlled intervention studies to test the hazards of substances in order to address environmental health questions. The human studies generally available for environmental health assessments are observational studies of cross-sectional, case-control, cohort, or case reports/series design. However, the experience with GRADE in the environmental health context is as yet limited, and empirical evaluations of using GRADE in this context are also limited.

Four descriptors are used to indicate the level of confidence in the body of evidence for human and animal studies:

- **High Confidence (++++)** in the association between exposure to the substance and the outcome. The true effect is highly likely to be reflected in the apparent relationship.

- **Moderate Confidence (+++)** in the association between exposure to the substance and the outcome. The true effect may be reflected in the apparent relationship.

- **Low Confidence (++)** in the association between exposure to the substance and the outcome. The true effect may be different from the apparent relationship.

- **Very Low Confidence (+)** in the association between exposure to the substance and the outcome. The true effect is highly likely to be different from the apparent relationship.

In the context of identifying research needs, a conclusion of "High Confidence" indicates that further research is very unlikely to change confidence in the apparent relationship between exposure to the substance and the outcome. Conversely, a conclusion of "Very Low Confidence" suggests that further research is very likely to have an impact on confidence in the apparent relationship. It is possible that a single well-conducted study may provide sufficient evidence of toxicity or health effect. This is consistent with the US EPA's minimum evidence necessary to determine if a potential hazard exists: data demonstrating an adverse reproductive (or developmental) effect in a single appropriate, well-executed study in a single test species (EPA (US Environmental Protection Agency) 1991, 1996).

Combine Confidence Conclusions for All Study Types and Multiple Outcomes

Conclusions are primarily based on the evidence with the highest confidence when considering evidence across study types and multiple outcomes. Confidence ratings are initially set based on key design features of the available studies for a given outcome (e.g., for experimental studies separately from observational studies). The studies with the highest confidence rating form the basis for the confidence conclusion for each evidence stream.

Combined consideration across study types:

Consistent results across studies with different design features can increase confidence in the combined body of evidence and result in an upgraded confidence rating moving forward to Step 6.

Combined consideration across multiple related outcomes:

When outcomes are biologically related, they may inform overall confidence in the health effect, and confidence conclusions can be developed in two steps. Each outcome is first considered separately. Then, the related outcomes are considered together and re-evaluated for properties

that relate to downgrading and upgrading confidence in the body of evidence. This approach is especially helpful in circumstances where the combined body of evidence covers issues that may be uncertain or missing when outcomes are considered separately. For example, if the body of evidence for one outcome has studies that more clearly establishes that exposure proceeded the outcome, it may strengthen the overall body of evidence.

2) The text below is NEW text added to provide additional background on the development of confidence conclusions for an overall body of evidence across multiple exposure metrics.

NEW: Combined consideration across exposures from a common source:

When individual chemical or physical agent exposures are components of a broader relevant exposure derived from a common source, collectively they may inform overall confidence in the association of that broader exposure with the health effect. Confidence conclusions can be developed in three steps. Each individual exposure is first considered separately and a confidence rating in the body of evidence is reached. Then mechanistic data or other relevant considerations should be used to determine 1) if the individual exposures could independently affect the health outcome and 2) if there is evidence of an exposure-dependent relationship between the exposure and the health effect. 3) If both scenarios are true for the exposure, then evidence from the individual exposures is considered together and re-evaluated for properties that relate to downgrading or upgrading confidence in the body of evidence.