Defense Centers for Public Health – Aberdeen
(formerly U.S. Army Public Health Center)

Technical Guide 254

Standard Practice for Wildlife Toxicity Reference Values

February 2023

Approved for public release; distribution is unlimited.

Acknowledgments

Key Technical Authors:

Mark A. Williams Ph.D., FAAAAI
Toxicology Directorate
Defense Centers for Public Health – Aberdeen (DCPH-A)

Glenn J. Leach Ph.D.
Toxicology Directorate
(DCPH-A)

Mark S. Johnson, Ph.D., DABT, ATS
Toxicology Directorate
(DCPH-A)

Michael J. Quinn, Ph.D.
Toxicology Directorate
(DCPH-A)

Point of Contact
For further information or assistance, please contact the Center at the following office:

Mark A. Williams Ph.D., FAAAAI
Defense Centers for Public Health – Aberdeen (DCPH-A)
Toxicology Directorate, Health Effects Division ATTN: MCHB-TS-TOX; Building E2100
8252 Blackhawk Road, Aberdeen Proving Ground MD 21010-5403
410-436-3980/DSN 584-3980
Email: usarmy.apg.medcom-phc.mbx.tox-info@health.mil

When referencing this document, please use the following citation:
Defense Centers for Public Health – Aberdeen (DCPH-A). 2023. Technical Guide (TG) 254, *Standard Practice for Wildlife Toxicity Reference Values,* Report No. HEF-0120190001, Aberdeen Proving Ground, Maryland.

The mention of any non-federal entity and/or its products is not to be construed or interpreted, in any manner, as federal endorsement of that non-federal entity or its products.

Preface

This technical guide (TG) provides the Center's Standard Practice for the development and documentation of toxicity reference values for wildlife, which will assist in the evaluation of hazards to mammals, birds, reptiles, and amphibians from exposures to military-related chemicals. Informed and defensible environmental health risk management is limited by the quality of the risk assessments used to support them. Therefore, design of this TG will improve the analyses that support and strengthen these risk management decisions. This TG provides the process for deriving these values.

This TG should not be construed as official Department of the Army policy unless so designated by other authorizing documents. This document provides guidance and technical reference material based on scientific information current at the time of publication. As available information and supporting data are continuously being advanced, users are cautioned to ascertain the existence of any updated information.

The Surgeon General is responsible for providing policy and technical expertise on human health and ecological aspects of pollution resulting from Army activities and operations (Army Regulation (AR) 200-1, *Environmental Protection and Enhancement* and AR 40-5, *Army Public Health Program*). The Surgeon General has delegated this responsibility through the U.S. Army Medical Command to the Defense Centers for Public Health – Aberdeen (DCPH-A; formerly U.S. Army Public Health Center). This guide was developed pursuant to this authority.

TABLE OF CONTENTS

SECTION 1: INTRODUCTION....1
1.1 Purpose....1
1.2 Limitations of Use/Exceptions....1
1.3 Audience....1
1.4 Application....2
1.5 Technical Guide Revisions....2
1.6 Background....2

SECTION 2: METHODOLOGY....2
2.1 Data Collection/Literature Search Procedures....3
2.2 Identification of Relevant Studies....14
2.3 The Toxicity Profile....18
2.4 TRV Derivation....20
2.5 Confidence Level Assignment....31
2.6 The TRV Report....32

APPENDICES
A. References....A-1
B. Terms....B-1
C. Quality Control Guidelines and Criteria for the WTA Literature Search Strategy and TRV Development....C-1
D. An Exemplar of Documenting the Literature Search Strategy....D-1
E. Example of Detailed Required for a Data Extraction Spreadsheet of a Chemical of Toxicological Interest Derived from the Literature Search Strategy....E-1

LIST OF TABLES
1. Descriptive List of the Key Recommended Resourced Primary, Secondary and Grey Literature Databases for the Literature Search Strategy....3
2. Descriptive PECO Statement and Conceptual Framework for all WTA reports….....7
3. Inclusion/Exclusion Criteria to Guide Strategic Distillation of Collated and Critically Analyzed Articles and Reports….....8
4. Uncertainty Factors Accounting for Differences in Response Due to Exposure Duration and Endpoint....28

LIST OF FIGURES
1. Flow Chart Describing the Iterative Strategy on Searching Primary and Secondary/Grey Literature Databases....10
2. Flow Chart Describing and Summarizing the Overall Process of the WTA Reporting Requirement....11
3. Example of a Scatter Plot to Summarize Multiple Studies and Toxicological Endpoints....19
4. Basic BMD Methodology....23

This technical guide supersedes USACHPPM TG 254, dated October 2000.

1. INTRODUCTION

1.1 Purpose

This Defense Centers for Public Health – Aberdeen (DCPH-A; formerly U.S. Army Public Health Center (APHC)) Technical Guide (TG) 254 outlines a Standard Practice that establishes a methodology for deriving scientifically-based wildlife[1] **toxicity reference values**[2] (TRVs) [1] for chemicals of interest in Army ecological risk assessment (ERA) programs. TRVs are toxicity-based criteria and represent environmental exposure concentrations (or pathway specific exposures or doses), which would have a minimal risk of adverse population level effects.

1.2 Limitations of Use/Exceptions

By definition, the procedures described herein result in toxicity benchmarks (i.e., TRVs) that are useful in environmental risk assessment applications—for screening and for decision making that are intended to be protective of adverse effects in individual organisms, which are assumed to be relevant to a population of organisms in the wild. This TG does not specifically address how the measures, or resulting risk estimates, relate to **demographic rates** (or outcomes) for any population of interest. These methods create both a biased risk estimate for use in screening-level evaluations (i.e., TRV-low) and one intended to be approximate to the threshold for adverse effects (i.e., TRV-high).

Assessing risk to populations involves using these methods and other lines of evidence[3] before any risk management action to protect populations can be recommended and often based upon site-specific scientific information [2].

Methodological exceptions to this Standard Practice may be warranted in some circumstances. These circumstances are when—

- The procedures are not consistent with promulgated Federal or State law.
- There are new or persuasive scientific evidence to bear on the specific issue in question.

1.3 Audience

Ecological risk assessors and toxicologists are the target audience for this Standard Practice. Army risk managers and staff responsible for coordination of ERA programs should ensure that their project teams consider this Standard Practice during project design and implementation (e.g., problem formulation).

[1] Use of the term “wildlife” specifically refers to *vertebrate* organisms other than fish that inhabit the wild.
[2] Definitions of terms in bold font are provided in Appendix B.
[3] For example: site-specific fieldwork, evaluations of reproductive success, demographic (population) modeling, and/or biological monitoring.

1.4 Application

This TG is primarily intended for use by DCPH-A to generate wildlife TRVs for military-related substances that are scientifically robust; thus, more defensible than those TRVs that are typically used in many U.S. Army risk assessments. If a TRV relevant to a particular ERA was generated by DCPH-A using the methodology described herein, then its use is expected unless an alternative can be reasonably defended. The DCPH-A will apply the methodology in a phased approach, focusing upon the highest priority chemicals first. Other U.S. Army and military entities are encouraged to use this Standard Practice within their ERA programs. The methods described herein are consistent with currently used evidence integration techniques as described in Lent et al. (2020) [3].

1.5 Technical Guide Revisions

This TG will be reviewed every 7 years for consistency with current practice and availability of new science. If the standard practice is determined to be inconsistent with current procedures and/or regulations, it will be revised and reissued.

1.6 Background

An integral component of a wildlife ERA is the development of a quantitative measure of a chemical's toxicity to the species (or receptors) of concern. In the past, toxicity benchmarks developed for this application have been repetitive, inconsistent, lacked scientific rigor, and in some cases, scientifically indefensible. The purpose of this guide is to provide a robust procedure, consistent with current and available scientific methods, to further the development of sound TRVs for four wildlife classes (i.e., terrestrial air-breathing mammals, birds, reptiles, and amphibians). These values are used in screening and remedial decision-making.

2. METHODOLOGY

In general, TRVs are needed to represent exposure concentrations that are associated with low risk levels for entire taxonomic classes (e.g., mammals) or for selected foraging **guilds** (e.g., carnivorous mammals). This TG focuses upon the development of chemical-specific TRVs for these receptor groups. The TG only addresses TRV development for mammals, birds, and herpetofauna (reptiles and amphibians); however, it is consistent with evidence integration techniques used in the development of toxicity benchmarks described in Lent et al. (2020) [3]. This TG excludes fish, as well as aquatic and terrestrial invertebrates. The methodology for generating wildlife TRVs and for document preparation to support such TRVs consists of two phases:

- **Phase 1 – Toxicity Profile.**
  - Search the literature and perform a comprehensive narrative review (NR) of the compound of interest.
  - Identify relevant studies and prepare a toxicity profile from synthesized evidence.

- **Phase 2 – TRV Report.**
  - Derive TRVs and document the rationale for TRV-aligned study selection.
  - Assign confidence levels to the derived TRVs by applying professional judgement of the available data and variability associated with exposure/effect endpoints.
  - Complete the TRV report.

The outcome from both phases is integrated into a comprehensive "Wildlife Toxicity Assessment (WTA)" report of the chemical(s) under review. Each WTA report shall contain a list of the primary author(s), contact information, and a report date.

2.1 Data Collection/Literature Search Procedures

The first step in developing the WTA is to perform a modified NR of primary, secondary, and the "grey" literature (i.e., information produced outside of traditional publishing and distribution channels, such as reports, government documents, white papers, etc.). An NR (also referred to as a comprehensive review or critical overview of the literature), serves to analyze and synthesize the literature on a specialized topic. An NR is a comprehensive literature search that follows a defined process, which qualitatively summarizes evidence on a specialized topic of interest using informed or subjective methods, to collect and interpret studies. The NR approach also permits inclusion of several study designs (e.g., experimental/non-experimental, and theoretical studies/empirical literature) and is the most common form of critically analyzing and synthesizing data from the scientific literature [3-5].

Table 1. Descriptive List of the Key Recommended Resourced Primary, Secondary, and Grey Literature Databases for the Literature Search Strategy

Primary Literature Resource	Secondary and the Grey Literature
PubMed: https://pubmed.ncbi.nlm.nih.gov/	U.S. Environmental Protection Agency (EPA) Environmental Topics and EPA ChemView: https://chemview.epa.gov/chemview/
PubChem: https://pubchem.ncbi.nlm.nih.gov/	Risk Assessment: IRIS at EPA Toxicology ITER: https://www.epa.gov/iris
ALTBIB – Alternatives to Animal Testing: https://ntp.niehs.nih.gov/go/altbib	EPA CompTox Chemicals Dashboard: https://comptox.epa.gov/dashboard/
Web of Science – Clarivate Analytics: https://clarivate.com/webofsciencegroup/solutions/web-of-science/	EPA Substance Registry Services – Includes ECOTOX: https://sor.epa.gov/sor_internet/registry/substreg/LandingPage.do
Scopus – Abstract and Citation Database: https://www.elsevier.com/solutions/scopus	EPA TOX21SL Chemical Screening Library and Dashboard: https://comptox.epa.gov/dashboard/chemical-lists/TOX21SL

Table 1. Descriptive List of the Key Recommended Resourced Primary, Secondary and Grey Literature Databases for the Literature Search Strategy (continued)

Primary Literature Resource	Secondary and the Grey Literature
SciFinder Web – Chemical Abstracts: https://scifinder.cas.org	Defense Technical Information Center (DTIC): Technical-Reports: https://discover.dtic.mil/technical-reports/
American Psychological Association PsycINFO database and Collections: https://www.apa.org/pubs/databases/psycinfo	European Chemicals Agency (ECHA) and OECD eChemPortal: https://echa.europa.eu/information-on-chemicals
Embase – Biomedical Research: https://www.embase.com/	Toxicological Benchmarks for Wildlife (Oak Ridge National Laboratory) [10]: https://www.esd.ornl.gov/programs/ecorisk/benchmark_reports.html
The Cochrane Library Collections: https://www.cochranelibrary.com/	American Chemical Society – CAS Common Chemistry Portal: https://commonchemistry.cas.org/
Academic Search Ultimate – EBSCO: https://www.ebsco.com/products/research-databases/academic-search-ultimate	Royal Society of Chemistry – ChemSpider Chemical Database: http://www.chemspider.com/
Academic Search Ultimate – DB Finder: https://www.ebsco.com/products/research-databases/ultimate-databases	U.S. Federal Register – Electronic Code of Federal Regulations – Reports and Records of Chemicals and Toxicology: https://www.federalregister.gov/
U.S. Dept. of Agriculture National Agricultural Library (AGRICOLA): https://agricola.nal.usda.gov/	U.S. Federal Government Resources – Data: https://resources.data.gov/
ECOTOX Database: https://cfpub.epa.gov/ecotox/	U.S. FDA FEMA Flavor Ingredient Library – Safety Assessments, GRAS Reports: https://www.femaflavor.org/flavor-library
ATSDR Toxicological Profiles: https://www.atsdr.cdc.gov/toxprofiledocs/index.html	U.S. Government Publishing Office – Govinfo Portal: https://www.govinfo.gov/
	Comprehensive Toxicology Information: HSDB in PubChem: https://pubchem.ncbi.nlm.nih.gov/source/11933
	Agency for Toxic Substances and Disease Registry (ATSDR) Portal: https://www.atsdr.cdc.gov/
	International Toxicity Estimates for Risk (ITER) - Toxicology Excellence for Risk Assessment (TERA) Database: http://www.iter.tera.org/
	U.S. CDC NIOSH – Registry of Toxic Effects of Chemical Substances (RTECS): https://www.cdc.gov/niosh/rtecs/default.html
	U.S. Dept. of Labor – OSHA Chemical Database: https://www.osha.gov/chemicaldatabase
	The Merck Index Online: https://www.rsc.org/merck-index

Table 1. Descriptive List of the Key Recommended Resourced Primary, Secondary and Grey Literature Databases for the Literature Search Strategy (continued)

Primary Literature Resource	Secondary and the Grey Literature
	U.S. DHSS – National Toxicology Program Reports and Evaluations: https://ntp.niehs.nih.gov/
	The Joanna Briggs Institute Grey Literature Collection: https://jbi.global/#
	OpenGrey World Database: http://www.opengrey.eu/search/request?q=greynet http://www.greynet.org/opengreyrepository.html
	WorldWideScience Grey Literature: https://worldwidescience.org/
	Global Collection of Theses and Dissertations – Many Resources: https://www.worldcat.org/ https://oatd.org/ http://search.ndltd.org/ https://ethos.bl.uk/Home.do https://www.bac-lac.gc.ca/eng/services/theses/Pages/theses-canada.aspx

Since the objective of an NR is to provide a comprehensive summary or overview of a topic of interest, the next critical step in this process is to methodically search the literature and select the studies to be included in the assessment (see Table 1 above). A literature search strategy serves to ensure that all relevant literature on the test article of interest is methodologically searched, evaluated, and synthesized when deriving a TRV. Each step is documented in a transparent and logical format. All relevant sources are searched for by specific toxicological data aligned to classes of animals of interest, including mammals, birds, and herpetofauna.

The key components of the NR must include an Introduction and adoption of the employed Methodology in deriving the TRVs, including the TRV report. The methodology must further include details of the data collection and literature search strategies, the toxicity profile of the targeted test article of interest, and details of the mathematical and/or statistical approaches in deriving TRVs. The NR provides an overview of the extent of available literature and identifies possible data gaps for the compound of interest. The NR also provides a critical assessment of the quality and relevance of the test data and largely follows the NR process described by Lent et al. (2020) [3].

Please refer to Table 1 above for a list of primary and secondary/grey literature databases and platforms where website uniform resource locators (URLs) are provided for the key sources of information.

The literature search will provide the following:

- Qualitative synthesis of toxicological characteristics for the chemical(s) of interest.
- Careful consideration of a set of relevant studies in the development of TRVs.

Further, since data mining and researching of the available toxicological literature is a logic-driven repetitive process to generate an expected outcome, a conceptual and iterative framework is used to search relevant databases, as well as subsequent integration of the available literature. There is no single source that provides a comprehensive list of primary data sources for substances of concern; thus, this TG recommends the following six-step process for effective database literature searching of primary, secondary, and grey literature databases:

1. Specifically identify and record the topic area that is aligned to the scope of the NR; for example, “acute oral toxicity to xylene in mammals.” Record any keywords used.
2. Identify applicable technical resources or databases (see Figure 1A).
3. Create a list of controlled vocabulary terms, synonyms, and related terms that can be reproducibly used across animal classes of interest on exposure to a chemical of interest—this process assists in designing a consistent literature search strategy.
4. Conduct literature search; screen records for inclusion/exclusion (Figures 1A and 1B).
5. Critically evaluate the information (see Figures 1A and 1B).
6. Record the findings to satisfy the requirements of Phase 1 and Phase 2 of the WTA.

Consistent with the conceptual and iterative framework described earlier, is the development and inclusion of a standardized Populations, Exposure, Comparator, Outcome (PECO) (or control) statement in the problem formulation and iterative process (Figure 1 and Table 2). This generalized PECO will be applied to all WTA reports (with modifications when deemed necessary) and represents a key component driving the NR. The considerations and thoughts used to develop these initial statements will assist in structuring the actual literature search. The populations of interest for a WTA NR can be broader than for a typical human health assessment and include mammals as well as birds, amphibians, and reptiles.

Synthesis of an NR adopts an integrated approach and methodology (Figure 1; Tables 1-3) to generate new frameworks or perspectives on the topic., Subsequently, this generates evidence-based and defensible TRVs, which are determined after a comprehensive review of the available literature (Figure 1). Currently, no single source provides a comprehensive list of primary data for any compound of concern. Selected databases may include the following (see Table 1):

- PubMed
- Embase
- Web of Science
- BIOIS (Biological Abstracts)

Table 2. Descriptive PECO Statement and Conceptual Framework for all WTA reports

PECO Components	Criteria
Populations	Animals: Non-human mammalian species, birds, amphibians, and reptiles that were exposed to the test article in a laboratory setting. Animals exposed in a field setting may also be considered if appropriate exposure information is available; otherwise, these studies would be considered as supporting data.
Exposures	Exposures to test article of interest performed by the oral route. Dermal exposures may be relevant for amphibians. Non-physiologically relevant exposures (e.g., intra-venous, intra-peritoneal, and sub-cutaneous), post-natal exposures, and *in vitro* study designs may be included as supporting information, which is consistent with integrative approaches to narrative review reports [3]. Co-exposures are also considered only as supporting information. Stabilizers are not considered co-exposures and, therefore, are relevant for this search but are addressed during weight of evidence assessment.
Comparators or Controls	Vehicle-control treated groups. Studies lacking a control group will not be considered. The appropriateness of the control group will also be assessed (i.e., concurrent sham exposure to vehicle that is closely matched to the experimental group), and greater value will be placed on studies with adequate controls, well controlled dose-response vales, and/or modeling analyses.
Outcomes	All relevant adverse health outcomes will be considered. Highest consideration will be ascribed to those outcomes that are deemed biologically relevant (e.g., irreversible effects) and direct (e.g., physiological effects) where exposures can be quantified. Outcomes with an unclear relevance to disease development, progression, or prognosis may be considered as supporting information and are consistent with integrative approaches to narrative review reports [3].

The WTA document shall record a description of the search protocols, databases searched, dates that the databases were searched, and all search terms and keywords, and the CAS number (this will be documented under Appendix B in the WTA document). Documentation also includes all results, for example, number of titles retrieved, abstracts reviewed, and publications retrieved for the final analysis and report (see Figure 1).

To ensure all potentially relevant information is collected, the literature search should be inclusive of all intra-class foraging guilds (e.g., small mammalian herbivores and mammalian invertivores). After compiling a list of all references retrieved from the literature search, identify the references for inclusion in the synthesized NR and for the TRV derivations included in the final WTA report. This process will require developing a list of inclusion and exclusion criteria (Table 3). The PECO or problem formulation statement (Table 2) can also assist in identifying and refining such criteria.

All of the references identified in the literature search will be screened against these criteria. The inclusion and exclusion criteria used for studies included in the EPA Ecotox database are described in Table 3 below; these criteria will be adapted for documentation purposes to guide

inclusion and exclusion criteria employed in WTA reporting and TRV development (see also Figure 1). On screening primary and secondary/grey literature databases and platforms, the adopted approach must be comprehensive and must follow a logical process. Articles and reports will be screened by study/report title, and then the scientific content of the technical abstracts and report summaries will be screened against the inclusion and exclusion criteria described in Table 3 [see also 3-5].

These criteria should be based on requirements that were identified in the PECO statement (Table 2), and any other relevant criteria (review article vs original primary research, not an animal study, etc.). The inclusion and exclusion criteria and the results from the screening steps will be tracked and documented under Appendix B in the WTA document. Complementing the structured guidance provided by inclusion/exclusion criteria will be an approach that benefits from the professional judgement of the named investigators conducting the WTA and derivation of the TRVs. Iterative practices and workflows identified above (Figure 1, Tables 1 and 2) will help guide and instruct that process.

Table 3. Inclusion/Exclusion Criteria to Guide Strategic Distillation of Collated and Critically Analyzed Articles and Reports (Adapted from EPA's Ecotox Database)

PECO Components	Inclusion Criteria	Exclusion Criteria
Chemical	Exposure to single chemical of interest	Mixtures of chemicals in the study design; lack of analytical data or confirmation of substance purity including poor or inadequate characterization of the test compound; no information on preparation or storage of test compound.
Species	Environmentally relevant species Priority species are wildlife (test results for terrestrial, domestic and laboratory species are used to fill data gaps when needed). *In vitro* assays of relevance that infer on mechanism/mode of action.	Human, monkey, bacteria, viral and yeast. Also, species or strain, or sex or number of test animals per group or per study not reported.
Effect/Response	Biological effect on live, whole organisms (but see previous regarding *in vitro*) Modeling data, *in vitro* information and metabolism included as supporting information	Endpoint assessment insufficiently sensitive (e.g., only mortality assessed); endpoint assessment that differs between controls and treatment groups; no (or inappropriate) concurrent negative control group or data from controls that is not reported; positive control group not included when necessary or responses unacceptable; and biased allocation of animals to treatment groups.

Table 3. Inclusion/Exclusion Criteria to Guide Strategic Distillation of Collated and Critically Analyzed Articles and Reports (Adapted from EPA's Ecotox Database) (continued)

PECO Components	Inclusion Criteria	Exclusion Criteria
Exposure Route	Oral, Dermal, Inhalation	IV, IP, or SC routes of exposure in study design. Exposures that differ between treatment groups; exposure frequency and/or duration not reported or not appropriate for the study type; dose or concentration range inappropriate (e.g., all lethal or no effects observed). Note: Non-physiologically relevant exposures (e.g., intra-venous, intra-peritoneal and sub-cutaneous), post-natal exposures, and *in vitro* study designs may be included as supporting information, consistent with integrative approaches to narrative review reports, which requires professional judgement practices.
Exposure Duration	Repeated dose exposures. Acute exposure data is permitted since situations might present where only acute exposure data is available from certain test species that may be informative in the context of species extrapolation.	Lack of concurrent controls or reference data and fatal flaws as previously described [3] to include exposure frequency/duration not reported or not appropriate for the study type.
Publication/Data Format	Primary data source Full text English (some Non-English papers are with an English abstract can be included as supporting information).	Reviews Full text in a foreign language Abstract only format of article.

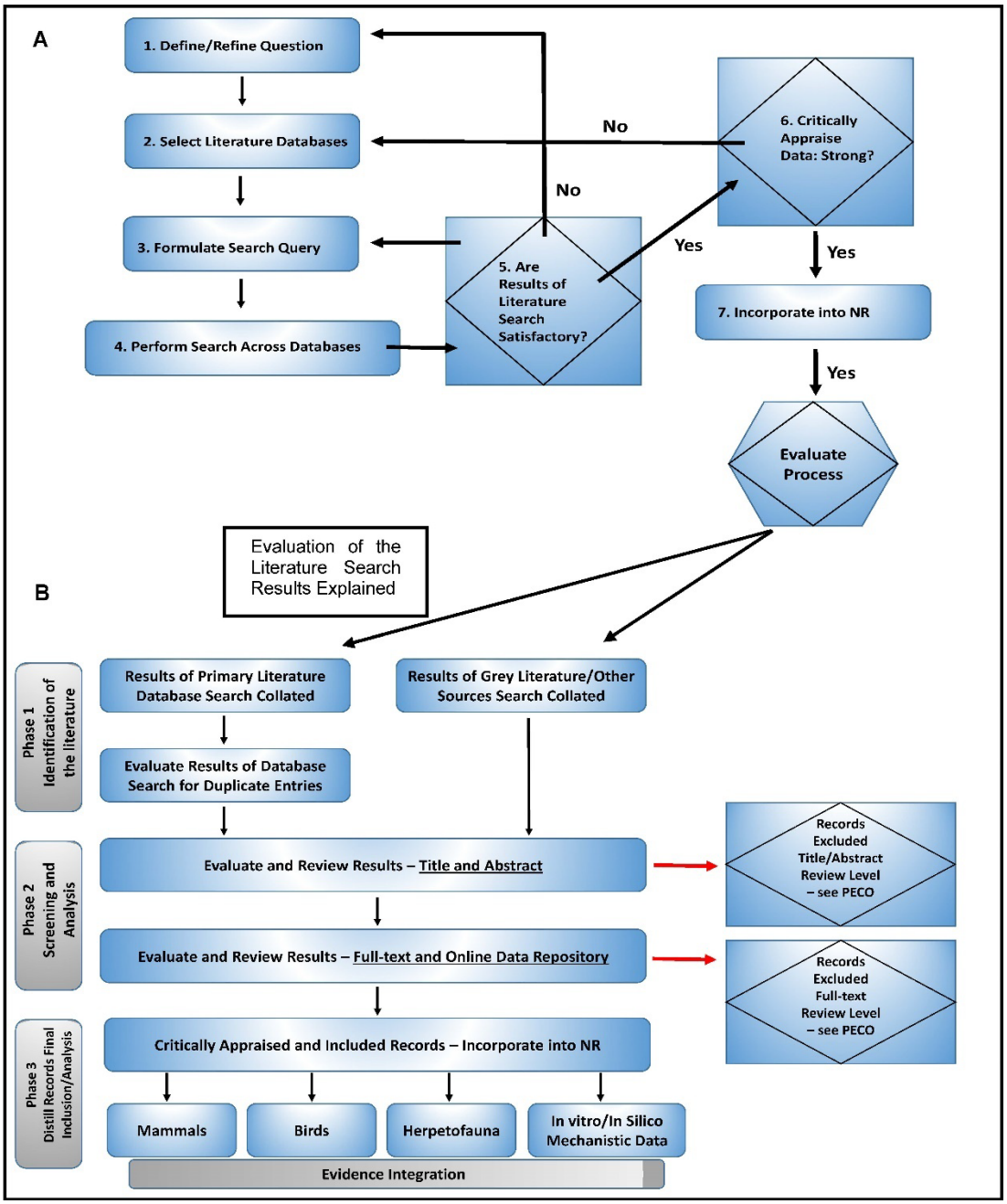

Figure 1. Flow Chart Describing the Iterative Strategy on Searching Primary and Secondary/Grey Literature Databases (A; Table 1) and problem formulation aligned to the PECO statement (see Table 2). Evaluation and appraisal frameworks (B) enable distillation and collation of primary and secondary/grey literature by screening, evaluation and appraisal of inclusion and exclusion criteria (B; and Table 3). This process requires prudent assessment of the search strategy and analyzed data through professional judgement. A WTA report using the NR approach outlined above (Figure 1) typically takes months to complete for any given compound of interest. Modified and adapted with revisions from Vandenberg et al. (2016) [6].

Following a careful review of the titles and abstracts, full text copies of all original articles, and other sources of data or technical reports that meet the inclusion criteria will be acquired and collated for subsequent critical analysis. Unpublished studies that are within the scope will be provided in the final WTA report.

A summary flowchart of the distilled literature search results and the framework used to exclude studies will be included in the toxicity profile. Figure 1B provides an example of a flow chart that was adapted with revisions, from Vandenberg et al. (2016) [6]. Figure 2 (below) provides an overall summary of the NR and WTA reporting process, which details the critical analysis and evaluation of the selected literature at the conclusion of the review process. This includes the PECO criteria, literature search strategy, report screening, the quality and relevance evaluation and data extraction that is required to determine TRVs from the selected studies, and as documented in the final WTA report. An exemplar for documenting the literature search strategy is found in Appendix D, and includes details on keywords used, databases searched, dates, number of hits, and other technical details.

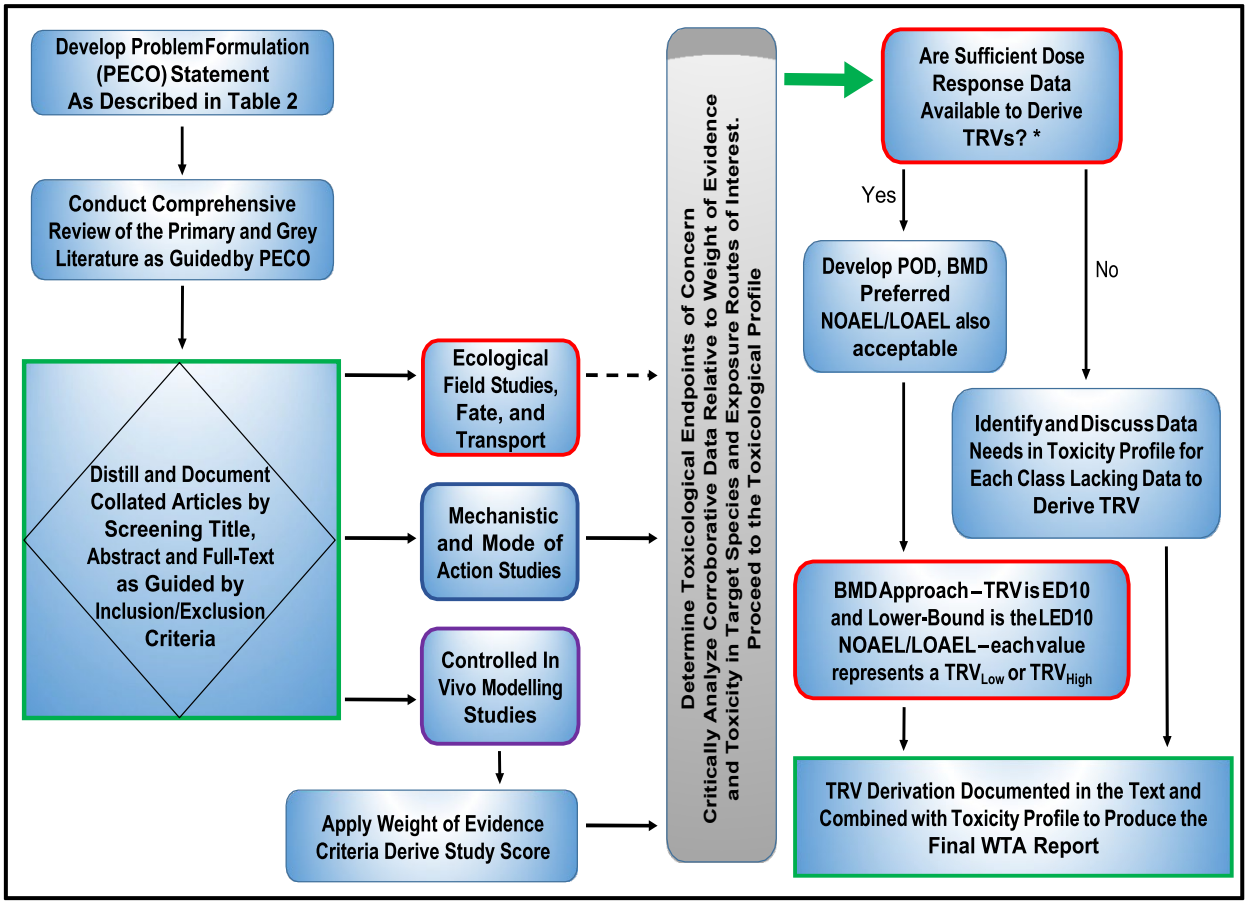

Figure 2. Flow Chart Describing and Summarizing the Overall Process of the WTA Reporting Requirement. The process includes the PECO criteria (as described in Table 2), the comprehensive literature of primary and secondary/grey sources of data, retrieved literature screening and an evaluation

of its relevance, and data screening.
Note: * indicates an ability to derive TRVs that are sufficiently determined by confidence in the database for each endpoint and set of dose-response data.

One of the exclusion criteria involved original primary literature sources as compared secondary sources. The TRVs and the data used in their derivation should be based on original primary literature to the extent possible. Environmental fate and transport and other related information not directly used in TRV derivation may be based on information from secondary/grey literature resources. Use of secondary/grey resources are also acceptable for corroborative comparisons and otherwise supportive information. In addition to details of *in vivo* toxicity, information relevant to environmental fate and transport (i.e., exposure criteria) should be collected to include:

- Physical and chemical properties of the compound of interest,
- The nature of fate and transport processes,
- *In vitro* data and absorption, and
- Details of metabolism and excretion.

Field studies often do not provide sufficient exposure information to be useful in deriving TRVs, but they may contain corroborating evidence of toxic effects and serve as supporting information.

In vitro data may provide mode of action or explain toxic mechanisms and aid in understanding species differences in toxic effects. Fate and transport studies can answer the questions about how the material was released to the environment, processes that change the material in the environment, and how it moves through the environment. This information may be relevant and should be included in the WTA; however, *in vivo* animal data are typically considered the most useful in developing points of departure (PODs) for TRV derivation.

Using the problem formulation/PECO statement and the iterative processes described in Figures 1 and 2 above, the selected resources from the literature searches can be categorized into four groups: field data, environmental (fate and transport), *in vitro*/mode of action/mechanistic data, and *in vivo* animal modeling studies.

Once the animal studies are identified, the next step in the NR process involves data extraction from the selected studies. Detailed information on the species, age, sex, study conditions, study duration, study endpoints, and results should be identified and tabulated. Further, to assist the tracking of the literature review process and to create a record of the collated resources and the data extracted following the literature search strategy, a Microsoft® Excel® datasheet may be used as described in Appendix E.

Regardless of the approach used, the collated data and study details will also be presented in tabular form. These data tables are incorporated into the final text of the WTA report. To derive TRVs, there is a requirement to analyze the toxicity data and provide an accompanying interpretation of that data in one or more species and vertebrate classes. When toxicity data are

unavailable for a class of animal (e.g., birds), data from other classes of animals will not be used to derive a quantitative toxicity measurement [7]. Physiological differences between taxonomic classes are assumed too great to make any extrapolation meaningful in predicting effects to another taxonomic class of animal (e.g., using mammalian data for extrapolation to birds). This science policy is based on the following three points:

1. As the taxonomic distance increases between any two groups of organisms, physiological differences that alter absorption, distribution, metabolism, excretion, and toxicodynamic processes also increase, as does the uncertainty associated with extrapolations [8, 9]. An appropriate exception to this policy is permitted when the toxicological mechanism is clearly known, and an understanding of the physiological differences permits extrapolation [7].
2. Extrapolations between two species may be more credible if factors to include similarities in food preferences, body mass, physiology, and seasonal behavior are considered [9, 10].
3. Extrapolation also requires context. Employing use of large (i.e., 3 - 4 orders of magnitude) uncertainty factors is unrealistic according to EPA guidance [2, 9, 10].

In these cases, the use of the following strategies can assist in the process of providing corroborative evidence in supporting an estimated TRV.

- Acknowledge the uncertainty due to the lack of appropriate data. Qualify the extent and direction in which inter-class physiological differences are expected to influence any toxicity estimate.
- Apply methods using Quantitative Structure-Activity Relationships (QSARs) or qualitative read-across to estimate the toxicity when there is information on a structurally similar organic compound that has a suspected similar mode of action. This is useful when assessments have historically used a chemical presumed to be the most toxic of a class of chemicals. For example, using the 2,3,7,8-tetrachlordibenzodioxin TRV for other similar dioxins, furans, and coplanar polychlorinated biphenyls when useful toxicity data are not available for other similarly structured compounds.
- Apply *in vitro* or mechanistic evidence where it can be demonstrated or reasonably presumed that such biological pathways are conserved across species.
- Use of field data or site-specific evidence of toxicity. Examples include the following:
    - Evidence of tissue chemical concentrations and observed toxicity that is associated and consistent with observations of adverse effect;
    - Forensic evidence;
    - Measurements of species diversity or abundance associated with exposure;
    - Tissue-based chemical concentrations and/or histopathological analysis of wildlife samples; and
    - Measurements of fitness and reproductive success at active contamination sites with evidence that suggests lack of an adverse effect (i.e., the use of negative data).

The data collection/literature search effort will predominantly result in identifying relevant controlled toxicity studies; although it is recognized that other lines of evidence may support

typical assumptions used in extrapolation of these data to species of concern (e.g., in vitro, read-across, and field studies). As described by Lent et al. (2020) [3], and in Table 3 above, identified fatal flaws may preclude the use of sourced toxicological data in the derivation of TRVs; however, those data and information should be retained as they may provide corroborative support for other questions involved in OEL derivation.

Fatal flaws that would exclude animal studies for a quality dose-response assessment and TRV development, would consider many criteria (please see Table 3 above, and Lent et al. (2020) [3]). These criteria include:

- Poor or inadequate characterization of the test compound;
- No information on preparation or storage of test compound;
- Inadequate or poorly described methodological approaches used in inhalation studies to generate the test substance concentration and an appropriate dose metric;
- A dose or concentration range being inappropriate (e.g., all lethal or no effects observed);
- No (or inappropriate) concurrent negative control group or data from controls not reported;
- A positive control group not included when necessary or responses not acceptable;
- Biased allocation of animals to treatment groups;
- Exposures that differ between treatment groups;
- Exposure frequency and/or duration not reported or inappropriate for the study type;
- Species/strain/sex/number of test animals not reported;
- The variability in response (i.e., power of the statistical comparisons) must be assessed and confirmed to be relevant and of equivalent value to other studies being considered for a specific compound and class of vertebrates;
- The bioavailability of the substance in the field and the one used in the toxicity studies must be comparable; endpoint assessment insufficiently sensitive (e.g., only mortality assessed);
- An endpoint assessment that differs between controls and treatment groups;
- Repeatability of the study, wherein sufficient information is presented that subsequently permits a given study and its results to be repeated; and
- Corroboration with other similar data sets.

2.2 Identification of Relevant Studies

Following study screening and data extraction from the literature search results, relevant studies for developing TRVs that are applicable to wildlife are evaluated for study quality, relevance, and risk of bias. In general, studies describing a toxic effect with a dose response in a target species would be the preferred studies for scoring. If there are a large number of studies, the results from this process can be used to select the most relevant high-quality studies for deriving the TRV. If the chemical of interest produces multiple toxic effects, dose-response data for each endpoint should be included and evaluated separately.

Study quality considers how well the study was designed and executed, and hence, how dependable the results are for predicting adverse effects. Did the study use adequate numbers of test species, and did it include concurrent controls? Did the authors include appropriate statistical analysis of the data? Relevance includes how well the study addressed the overall requirements for TRV derivation. Consideration of the relevance must consider whether the route of exposure was appropriate and whether the exposure duration and test endpoints were applicable to the problem.

In addition to quality and relevance, consideration must be given to potential bias in the study and factored into the overall evaluation and decision-making on accepting or rejecting the study in TRV derivation [3]. Bias can occur in several places in the study. Selection bias may occur in the assignment of test animals to the different study groups. Performance bias may occur in the way the study was conducted (e.g., if study groups experienced different housing conditions, or the study personnel were not blinded to the different test groups). Bias could also exist in the way results were reported. All data should be included in the results, or a valid explanation given for excluding information. All of these examples can skew the results and conclusions of the study.

A number of qualitative and quantitative systems are available for assessing study quality and relevance [3]. The ToxRTool [11], the European Food Safety Authority (EFSA) [12], the Grading of Recommendations Assessment, Development and Evolution (GRADE) framework [13], the EPA Ecological Soil Screening Levels [14], and the SYRCLE risk of bias tool [15] are a few well-known evaluation systems available.

For cases where there is a paucity of relevant data, qualitative means to evaluate quality, relevance, and risk of bias are recommended, which is consistent with professional judgement of the study design.

The paragraphs below discuss the criteria used to select toxicity data relevant to TRV development. The available studies in the literature may not satisfy all of these criteria; therefore, those studies that satisfy as many of these criteria as possible are considered relevant. In most cases, it is expected that a small set of studies will be identified that are “nearly equivalent” in terms of their relevance.

The identified toxic effects are most clearly linked to factors that are suspected to greatly influence population sustainability (e.g., demographic rates: birth, death, and dispersal rates). Nonetheless, prior knowledge of factors most relevant in population-specific regulation is needed to select toxic endpoints up front. Thus, any adverse effect that could potentially harm an individual organism should be considered. Often times, information that is specific to the animal species of concern (i.e., the measurement endpoint) will not be available. Under such circumstances, a focus on selecting those endpoints that are protective of the other endpoints is recommended (i.e., when considering sensitive endpoints). Toxicological endpoints should be evaluated regarding their relevance to the health and ecology of the whole organism(s).

The following includes several examples of endpoints:

(1) Mortality
(2) Reproduction
(3) Development
(4) Growth
(5) Behaviors relevant to reproduction, feeding, and predation avoidance
(6) Decreased resistance to disease (stress)

Other indirectly acting endpoints might also be important. Examples include factors that influence energy allocation, which could indirectly influence reproductive performance and success. In the absence of sound ecological knowledge for the species of concern, the aforementioned endpoints can be considered as nearly equivalent.

The exposure duration in each study should be clearly identified. Typically, repetitive chronic or sub-chronic exposures are most logically connected and relevant to risk assessment and TRV derivation. Occasionally, acute exposures are also important in the identification of short-term exposures (e.g., in spatially explicit exposure models) or for understanding differences between species wherein only acute data are available. All exposure periods should be considered because of the differences in species responsiveness, methods, observed effects, dispersal characteristics and habitat use in the field, as well as all potential toxicological endpoints.

The following guidelines are used to determine the exposure duration of a toxicity study:

- Chronic exposures are those equal to or greater than 10% of the test organism's lifespan. A notable exception to this criterion is found on exposure of a test organism during a sensitive life stage (e.g., during gestation). Classifying such tests as "chronic" is considered reasonable for endpoints that are specific to that life stage (e.g., embryonic development and clutch size).
- Sub-chronic exposures are those repetitive exposures that are less than 10% of the test organism's lifespan, yet greater than 14 days.
- Acute exposures are those of a single or repetitive exposure with a duration of less than 14 days or 10% of the test organism's lifespan.

The EPA [16, 17] developed these exposure duration definitions primarily from their regulations concerning regulatory toxicity testing under the Toxic Substances Control Act (TSCA) [16] and the Risk Assessment Guidance for Superfunds. Also considered were references provided in the EPA Great Lakes Water Quality Initiative Technical Support Document for Wildlife Criteria [18] and the work of Sample et al. (1996) [10].

For mammalian studies, defining tests that exceed 10% of the test organism's lifespan as chronic is consistent with EPA regulations for conducting toxicity studies under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) and TSCA (please refer to the following URL: https://www.epa.gov/test-guidelines-pesticides-and-toxic-substances/final-test-guidelines-pesticides-and-toxic). Exposure during a sensitive life stage (e.g., gestation and embryo development) is considered a reasonable criterion to classify a test as a chronic exposure

because of the potential for impaired reproduction and development. This is consistent with the method of Sample et al. (1996) [10]. For sub-chronic mammalian tests, the EPA defines a 90-day exposure duration as standard practice for mice and rats and describes those exposures as approximately 10% of the lifespan of the animal [16, 19]. Tests that are single exposures of extremely short duration (<14 days) are considered acute.

The effect levels in the study should be those most clearly associated with no-to-low adverse effects. The type of effect levels that satisfy this criterion are as follows:

- (Bayesian or Frequentist/Linear) Benchmark Dose (BMD or BMDL)
- No-observable-adverse-effect-level (NOAEL)
- Lowest-observable-adverse-effect- level (LOAEL)
- Effect Dose (EDx), where x is less than 50

The effect levels most useful for an environmental risk assessment (ERA) are those at the low end of the dose-response function.

The exposure pathway in the study will closely match the pathway that contributes the most to the exposure in the field. This will be a professional judgment determination. For example, for oral exposures, a feeding study may be preferred to a gavage study if the dose in food was well characterized and more applicable to the exposure route and matrix in the field.

The overall validity of the study design (e.g., exposure conditions and chemical form) relative to the appropriate exposure pathways in the environment will ensure the best possible toxicological risk estimate.

The quality of the study must be assessed and determined to meet general minimal requirements appropriate for inclusion. Criteria that must be considered include those described above (see Table 3) and by Lent et al. (2020) [3].

Important: Before completing the tabulation or drafting of a scatter plot of the data, a statement will be provided that describes the quality of all included relevant studies (or minimal criteria) in the toxicity profile, which is done once the toxicological and/or physiological effects have been fully characterized.

The final step during relevant study identification is to determine whether the relevant studies collected from the literature review provide the necessary data to meet the minimum data set requirement. The minimum data requirements are as follows:

- Data exist from at least three studies of sufficient quality to be deemed relevant (using the above criteria), which collectively provide data for three or more species within the taxonomic class.
- Data exist for at least two different taxonomic orders.
- At least two sub-chronic or chronic LOAELs and at least one sub-chronic or chronic NOAEL are available.

These minimum data requirements for test organisms are consistent with the number of species required for the certification of substances for Food and Drug Administration (FDA) approval for human applications [20]. Given the current state of the toxicological database and the general variation in toxicological response between species within a class, these requirements are considered reasonable. The minimum requirement for endpoint selection is based on professional judgment and experience with the literature.

Section 2.4.4 discusses procedures for addressing those cases where the minimum data set requirements are not met.

2.3 The Toxicity Profile

The toxicity profile is the written documentation of collected information and data aligned to the toxicological characteristics of the chemical(s) or compound(s) of interest before selecting or developing TRVs. The toxicity profile must be designed to provide all the necessary and required documentation for a clear and transparent final TRV report in the context of defending risk management decisions.

A toxicity profile consists of the following two components:

1. Documentation of the literature search, and how the relevant studies were selected.
2. Presentation of the data relevant to the development of TRVs, which includes a table identifying specific adverse effects and a scatter plot.

The toxicity profile should summarize the basic physicochemical characteristics of the chemical(s) and basic environmental fate and transport information. Such information is useful for the understanding of the potential exposure and toxicity of the chemical(s).

The main portion of the profile should present the available toxicity data. The extent of the discussion should provide all known information regarding the nature of the exposure and toxicity that is necessary for a risk assessor to understand the general characteristics of the chemical(s). The discussion should also be sufficiently limited in scope. For example, it should identify the major target organs and endpoints to include methodological details of the exposure, but not necessarily the effects seen at higher exposures to non-target tissues. Major sources of information and data should be cited.

Major section headings should be organized first by class (e.g., mammals), then by route of exposure (e.g., oral, inhalation, or dermal), and then by exposure duration (i.e., acute, sub-chronic, and chronic). Exceptions for appropriate mesocosm/microcosm or field studies are permitted.

When sufficient information exists, the profile should conclude with a hazard assessment, where each toxic endpoint is assessed across all three lines of evidence. The hazard assessment should consist of a table and summarize, for each toxic endpoint, whether there is “sufficient” or “insufficient” evidence for each toxic effect. A subsequent table should then summarize across all three lines of evidence (i.e., controlled laboratory animal, *in vitro*/mode of action, field data)

where sufficient evidence exists across all three. Points of departure (PODs) are then only developed for those toxic endpoints where sufficient evidence exists.

The profile may include a scatter plot that presents the quantitative data of the relevant studies specific to each taxonomic class presented in the table that was previously described. The scatter plot will contain all reliable data regarding a specific exposure route (e.g., oral), and categorized based on the observed endpoint (e.g., mortality, reproductive, developmental, systemic, and behavioral). Each data point presented in the scatter plot will also be presented in tabular format to include toxicological endpoint, species studied, concentration, and specific reference citation.

All test species will be identified, as well as the effect levels (e.g., BMD, NOAELs, and LOAELs). The scatter plot approach is an optimal approach at summarizing the data that are relevant to TRV development. In this type of graphical representation, patterns of variability among species, endpoints, and exposures are clearly visible and evaluated.

Figure 3 shows a sample scatter plot for multiple mammalian studies with multiple toxicological endpoints for TNT.

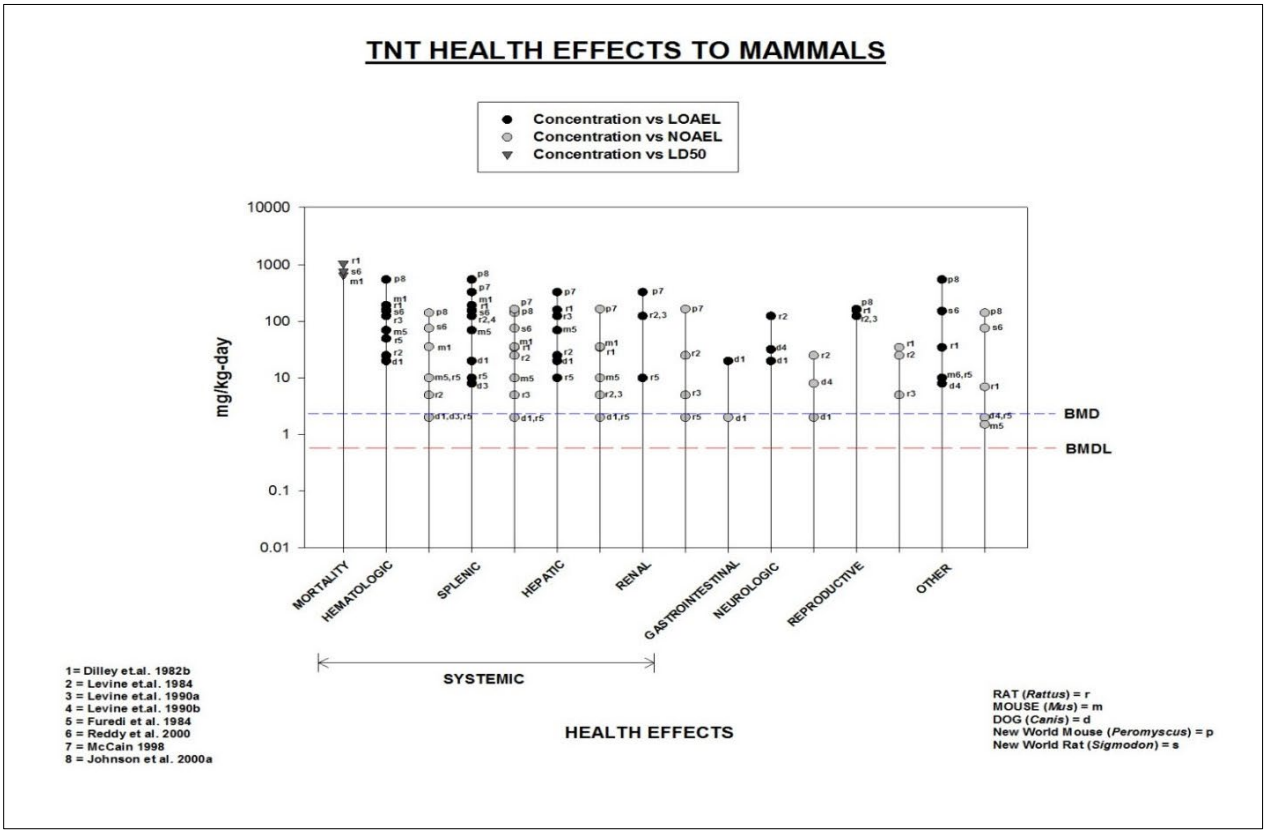

Figure 3. Example of a Scatter Plot to Summarize Multiple Studies and Toxicological Endpoints. Toxicity endpoints are shown along the X axis and concentration on the Y axis. For each endpoint, there

are two plots, one each for the NOAEL and LOAEL. The BMD and BMDL values are shown as dotted lines.

2.4 TRV Derivation

At this point in the process, the toxicity profile is completed and all available data within a taxonomic class that are relevant to TRV development have been summarized. The toxicity profile will provide data for developing TRVs that are protective of the entire taxonomic class and, in some cases, TRVs that are more specific to a lower taxonomic category (i.e., order and family).

The DCPH-A Wildlife TRV report will develop two levels of TRVs for each taxonomic class, where sufficient data exists. The TRV-low is intended to be used as a screening-level benchmark, while the TRV-high is intended for decision-making. To proceed through the ERA process under conditions of limited resources, the screening-level approach provides a feasible and an efficient means to evaluate the potential hazards of many substances [21-23]. This approach helps to reduce the generally long list of potential chemicals of concern at many sites to a more manageable list. The screening-level approach is also biased towards protection. It is

intended to support prioritization of resources towards understanding the probability of adverse effects from exposures to substances that pose significant risks.

When more specific TRVs are needed for a particular project (i.e., TRVs for a guild association or particular species), the data provided in the toxicity profile section of the WTA TRV report can be used to develop a specific TRV for the intended measurement endpoint, but only if appropriate data are available. Depending upon the available resources, each WTA TRV report produced by DCPH-A might provide one or more guild association TRVs, in addition to the class-specific TRVs. Standard Practice found in this TG does not result in species-specific TRVs that may be needed for some environmental assessments (see Sample and Arenal (1999) [24] for an approach that is based upon allometric scaling).

2.4.1 TRV Development Approaches

The available data (as documented in the toxicity profile) will determine which of the following three procedures will be used: 1) BMD approach; 2) NOAEL/LOAEL approach; and 3) the approximation approach.

Regardless of the procedure, two TRVs are developed for use: a low and a high. A bracketed range provides the risk assessor with a level of confidence between which no observed adverse effects may occur and where the threshold for sub-lethal adverse effects may occur. A range can be used to discriminate the relative importance of exposures that exceed the low TRV (e.g., when the HQ > 1). Although procedurally different, this concept is based on the collaborative work of the U.S. Navy, EPA Region 9, California EPA, and others [25] and is consistent with current EPA guidance [23].

- Benchmark dose approach. Data that show a clear dose/response relationship in a unimodal design are best used to derive two TRVs, wherein one is based on a BMD value (TRV_{Low}) and one based on the BMD value of (TRV_{High}).
- NOAEL/LOAEL approach. Data that do not have clear dose-response relationships within well-designed and conducted parameters should be used to derive two TRVs, one based on a NOAEL (TRV_{Low}) and one based on a LOAEL (TRV_{High}).
- Approximation approach. Where data are scarce and cannot be used for the aforementioned procedures, then the second approach will be approximated with the use of uncertainty factors (UFs) to derive TRVs that estimate a NOAEL and/or a LOAEL.

Each of these approaches describes development of specific toxicity values that can be used to evaluate an exposure pathway consistent with the one of interest. For some organisms (e.g., terrestrial amphibians or pulse-feeding reptiles), a pathway-specific exposure TRV of daily oral exposure may not be appropriate since total exposure to the media may best describe exposure and would most likely be represented in the literature (e.g., a soil concentration rather than a mg/kg-d oral intake value). In these cases, media concentrations (i.e., in soil) can be derived using the same logic presented in each of the above procedures.

2.4.2 Benchmark Dose Approach

The BMD approach is the preferred method for identifying points of departure for use in risk assessments. Advantages of the BMD are that it uses the entire dose-response curve rather than a single point on that curve and provides a probability of a response at a given exposure [26]. Typically, BMD uses the best fit dose-response curve to select the mean (50%) dose that corresponds to a 10% response (the ED10 or benchmark dose) and a dose that corresponds to the lower bound on the ED10 (the LED10; based on the lower 95% confidence limit). These two doses (the ED10 and the LED10) are selected as the TRV-high and TRV-low, respectively.

One of the clear advantages of applying BMD methodology is that it is not limited to certain types of endpoints. The BMD represents the dose level that is associated with the effect level of concern. Since the precise shape of the dose/response relationship is critical at low estimates [27], a 10% benchmark response (BMR) is recommended as the “threshold for adverse sublethal effects” [21, 23] for the measurement endpoint. This infers that there is a 95% chance that 10% or fewer animals will exhibit effects at this exposure. This value often corresponds well to NOAELs that are reported in the literature.

The point at the 10% response level (or ± 1 standard deviation for continuous variable data) on the best-fit BMD curve represents the level defined as an effective dose (e.g., ED10). Exceeding exposures beyond this value is considered the level where adverse changes in the assessment endpoint will begin to become unacceptable. In this procedure, a study is chosen from those determined relevant, based on endpoint, design, model, and overall quality. The endpoint selection should be one that is either suggestive of a population-relevant endpoint (see Section 2.2) or, when that is not known, is protective of the other endpoints.

The use of this approach is expected if available toxicological data can support it (i.e., if the data from the relevant studies identified in the toxicity profile can be used to develop a reasonable

dose-response curve). The EPA states that the “advantages of curve-fitting approaches include using all of the available experimental data and the ability to interpolate to values other than the data points measured” [23]. These curves are more defensible and more useful in predicting and communicating risk. The shape of the dose-response curve can be used to determine the presence or absence of an effects threshold, to evaluate incremental risks, and used as input for effects models (e.g., demographic models) [23].

Recently, BMD analysis has been refined to include Bayesian statistics to the BMD models. Bayesian statistics incorporates knowledge of prior events to improve the reliability of the dose-response modeling.

In Bayesian Benchmark Dose (BBMD) analyses, dichotomous data (e.g., right/wrong answers or viable/non-viable; see Figure 4, illustrating BMD analysis for dichotomous data, and Box 1) can be analyzed as binary “success” (1) or “failure” (0) variables to describe the status of studied subjects. BBMD can also analyze individual dichotomous data, where two quantities are required, e.g., dose and status (1 or 0), or as summary dichotomous data that requires three quantities to be analyzed (e.g., dose, total number of studied subjects, and number of subjects displaying an observed effect).

The statistical method and framework for an analysis of dichotomous data is defined as:

– $\Pr(Y = 1|d, n, \boldsymbol{\theta}) = binomial\,[n, f(d|\boldsymbol{\theta})]$

Where the Log-likelihood function as defined as:

$$\log[p(d, n, y|\boldsymbol{\theta})] = \sum_{i=1}^{G} \left\{ \log \binom{n_i}{y_i} + y_i \log[f(d_i|\boldsymbol{\theta})] + (n_i - y_i)\log[1 - f(d_i|\boldsymbol{\theta})] \right\}$$

For dichotomous data, the BMD (logistical; See Figure 4) is defined as:

$$\mathrm{BMD}_{\mathrm{Logistic}} = \log\left[\frac{\mathrm{BMR} + \exp(a)}{1 - \mathrm{BMR}}\right] - a / b$$

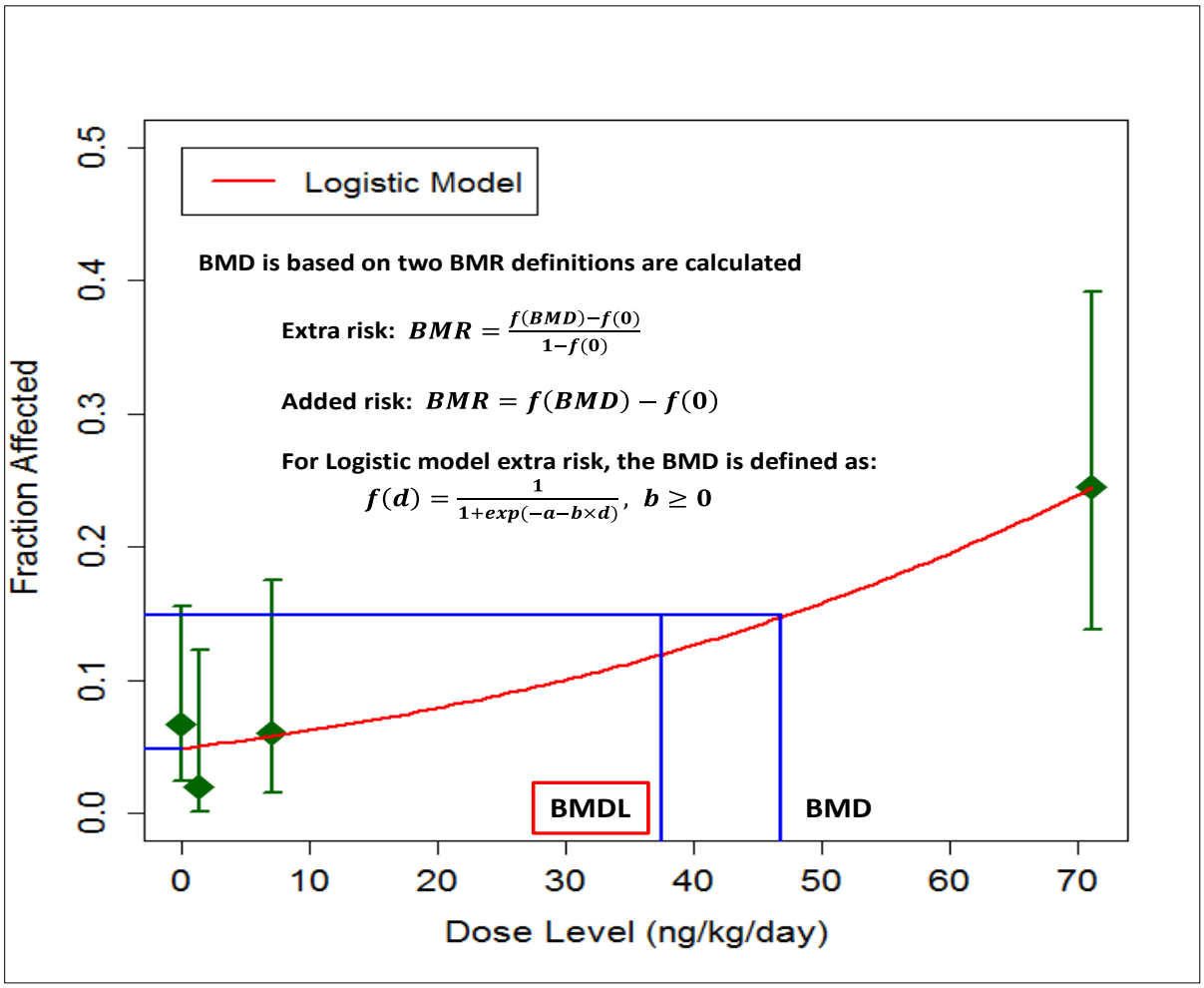

Figure 4. Basic BMD Methodology – dichotomous data**.** Basic steps for determining a BMD are shown here, using typical dichotomous data sets as an example. First, fit a dose-response model to the data, i.e., estimate the parameters in a dose-response model. For a frequentist method, this would include the maximum likelihood estimation (MLE) approach or better the Bayesian Markov Chain Monte Carlo (MCMC method). Once the BMD is determined, it permits calculation of the benchmark response or BMR, which can be determined for dichotomous, categorical, and continuous datasets. BMD uses the best fit dose-response curve to select the mean (50%) dose that corresponds to a 10% response (the ED10 or BMD) and a dose that corresponds to the lower bound on the ED10 or BMDL (the LED10; based on the lower 95% confidence limit).

Quantal-linear model: $f(d) = a + (1 - a) \times [1 - exp(-b \times d)],\ 0 \le a \le 1,\ b \ge 0$

- **Probit model:** $f(d) = \Phi(a + b \times d),\ b \ge 0$
- **Logistic model:** $f(d) = \frac{1}{1+exp(-a-b\times d)},\ b \ge 0$
- **Weibull model:** $f(d) = a + (1 - a) \times [1 - exp(-b \times d^{g})],\ 0 \le a \le 1,\ b \ge 0,\ g \ge restriction$
- **Multistage (2nd degree) model:** $f(d) = a + (1 - a) \times [1 - exp(-b \times d - c \times d^{2})],\ 0 \le a \le 1,\ b \ge 0,\ c \ge 0$
- **LogLogistic model:** $f(d) = a + \frac{1-a}{1+exp[-b-g\times log(d)]},\ 0 \le a \le 1,\ g \ge restriction$
- **LogProbit model:** $f(d) = a + (1 - a) \times \Phi[b + g \times log(d)],\ 0 \le a \le 1,\ g \ge restriction$
- **Dichotomous Hill model:** $f(d) = a \times b + \frac{a-a\times b}{1+exp[-c-g\times log(d)]},\ 0 < a \le 1,\ 0 < b < 1,\ g \ge restriction$

BOX 1: Illustrating the eight typical Dichotomous Dose-Response Models in BBMD Analysis. Shown here are the mathematical equations that are the basis of the dose-response modeling, which can be collated and analyzed to provide the best fit. Determination of a single estimate of the BMD/BMDL can thus be derived.

BBMD can analyze continuous data (see Box 2); i.e., data recorded on a continuous scale (e.g., body weight, relative liver weight). BBMD analysis finds that continuous data displays a lognormal distribution. Several models are run to analyze continuous dose-response data (Box 2). BBMD has the added advantage of automatically averaging the models, which provide the best fit, from where a single estimate of BMD/BMDL is derived (see Box 2 for a list of the continuous dose-response models in BBMD analysis).

The statistical method and framework for an analysis of continuous data is defined as:

- $\Pr(y|d, \boldsymbol{\theta}) = Lognormal[f(d_i|\boldsymbol{\theta}), \gamma]$

Where the Log-likelihood function as defined as:

$$\log[p(Data|\theta)] = -\frac{N}{2}\log(2\pi) - \sum_{i=1}^{G}\left\{\frac{n_i}{2}\log(\gamma^2) + \frac{n_i \times \{\bar{y}_i' - \log[f(d_i|\theta')]\}^2 + (n_i - 1) \times s_i'^2}{2\gamma^2}\right\}^2$$

- **Linear model:** $f(d) = a + b \times d, \quad a > 0$
- **Power model:** $f(d) = a + b \times d^g, \quad a > 0, \quad g \geq restriction$
- **Michaelis-Menten model:** $f(d) = a + \frac{b \times d}{c + d}, \quad a > 0, \quad c > 0$
- **Hill model:** $f(d) = a + \frac{b \times d^g}{c^g + d^g}, \quad a > 0, \quad c > 0, \quad g \geq restriction$
- **Exponential 2 model:** $f(d) = a \times exp(b \times d), \quad a > 0$
- **Exponential 3 model:** $f(d) = a \times exp(b \times d^g), \quad a > 0, \quad g \geq restriction$
- **Exponential 4 model:** $f(d) = a \times [c - (c - 1) \times exp(-b \times d)], \quad a > 0, \quad b > 0, \quad c > 0$
- **Exponential 5 model:** $f(d) = a \times [c - (c - 1) \times exp(-(b \times d)^g)], \quad a > 0, \quad b > 0, \quad c > 0, \quad g \geq restriction$

BOX 2: Illustrating the eight typical Continuous Dose-Response Models in BBMD Analysis. Shown here are the mathematical equations that are the basis of the dose-response modeling, which can be collated and analyzed to provide the best fit, and thus determination of a single estimate of the BMD/BMDL can be derived.

In BBMD analyses, it is generally accepted that there are two ways to define BMR for continuous data that include 1) an approach that is based on changes to the central tendency and 2) an approach that is based on changes in the proportion of distribution on the tail of the dose-response curve, i.e., a hybrid approach (as discussed by Crump (1995) [28]).

Shao and Shapiro (2018) [29] provide a detailed description of the BBMD method. The new BBMD modeling tool is found at https://www.benchmarkdose.com and is designed to help investigators in the field of probabilistic risk assessment. It is capable of many types of analyses in dose-response modeling to include central tendency and the hybrid approach to continuous data analyses previously described. The online BBMD data analysis tool can analyze single

data sets and can allow batch processing for BMD analysis. This analysis is designed to provide an efficient but somewhat simplified approach to analyze many datasets simultaneously for continuous, dichotomous, and categorical data. Though it is not yet incorporated to the WTA reporting methodology, the BBMD software tool also permits BMD modeling, analysis, and estimation of genomic datasets.

Finally, the BBMD software has an option to determine probabilistic reference dose (RfD). This analysis is used to convert conventionally estimated BMD/BMDL or NOAEL/LOAEL derivations to a probabilistic dose, given some additional study design information that is easily inputted to the system. Some examples include the experimental model (i.e., rat, mouse, or guinea pig, etc.), the use of uncertainty factors, the geometric mean, and the standard deviation of the inputted data that can be a default setting of the provided program or user defined. In addition, it should be noted that the EPA BMD V3 offers a Bayesian option, though this approach is criticized since errors have been identified in the design of the computational modeling tool (Dr. Shan Kao – personal communication).

In the analysis of a single dataset, the BBMD online tool is designed to facilitate detailed analyses for critical effects of a targeted chemical toxicant of interest. Inputted data can be continuous, dichotomous, and categorical. Markov-Chain Monte Carlo (MCMC) settings are preset by this model and remain unaltered. However, other parametrized model settings can be adjusted as needed (e.g., non-informative or informative priors; and selection of a standard model or individual model), including the BMR settings (i.e., central tendency and tails). The BMD analysis comprises a feature for setting uncertainty factors that are distributional, default set, or user specified—please see below section on uncertainty factors. Following entry of specific dose-response data and parameters, the BMD and BMDL values are determined by the BBMD software program [29]. This software also has the capacity to determine probabilistic RfD values that combine default and user-defined settings requiring the input of key data sets to include BMD/BMDL values, the experimental model (e.g., mouse, rat, amphibian), Bayesian uncertainty factors, geometric mean, and standard deviation values of the data.

The disadvantages of using dose-response curves are that the number of data points needed to complete the analysis are often unavailable, the process is time intensive, it is not always practical for toxicants that have a complex dose-response relationship, and when models cannot adequately be fitted to the data [23]. If sufficient and appropriate data exists, however, the EPA guidance supports the use of this approach [23].

2.4.3 NOAEL/LOAEL Approach

The NOAEL/LOAEL approach also produces two TRVs for the wildlife group of interest; one based on a NOAEL (TRV_{Low}) and one based on a LOAEL (TRV_{High})—see Section 2.4.1. These TRVs will be selected from the scatter diagram provided in the toxicity profile.

When the minimum dataset requirements are met (see Section 2.2), then the TRVs are developed from the studies identified as relevant in the toxicity profile using the below procedures. Note that selections should be made or reviewed by a toxicologist familiar with the literature.

- Choose the LOAEL-based TRV by selecting the lowest documented LOAEL that is suggestive of a population-relevant endpoint (Section 2.2), or when that is unknown, the LOAEL that is clearly adverse and potentially protective of the other endpoints is selected.
- Choose the NOAEL-based TRV by selecting the highest NOAEL (i.e., lower than the selected LOAEL) within the same endpoint as the selected LOAEL. If a NOAEL from the same endpoint is unavailable, then the highest NOAEL (i.e., less than the selected LOAEL) within all relevant endpoints should be selected.

Use of the NOAEL in screening-level assessments is consistent with EPA guidance [21]. Selecting the highest NOAEL that is less than the lowest LOAEL is permitted under conditions where both toxic endpoints are relevant, is consistent with EPA guidance [21], and is consistent with unnecessary over-protection (i.e., where the lowest possible NOAEL is selected).

Toxic effects identified from repeated sub-chronic and chronic exposures are most useful in developing TRVs for risk assessments intended for wildlife; however, an acute or repeated sub-acute exposure period may include important toxicological endpoints for some species and may allow for an evaluation of inter-specific differences in sensitivity. If the exposure duration of concern in an ERA is other than chronic, then the chosen exposure duration for selecting the TRV should be determined according to the toxicologist's professional judgment. Deviations from this procedure are acceptable if the reported toxicity data are inconsistent with other data (e.g., outlier data, potentially false positive data) or if the endpoints are of questionable toxicological and biological relevance (e.g., enzyme induction).

When the minimum data needs are met, then the toxicity profile and the scatter plots represent all the available data within a class of animals (including sensitive species). Under these conditions, no UFs are needed to modify the values (NOAEL/LOAEL or BMD/BMDL) in deriving the TRVs. All relevant class-specific data for each substance (including sensitive species) would be included in the toxicity profile (e.g., all mammalian toxicity data for the chemicals of interest). This format accommodates the use of variability in the data to determine the taxonomic differences in toxicity instead of ambiguous UFs. This approach is consistent with the guiding principles of toxicity data extrapolation [2]. If the minimum data requirements are not met for the wildlife group, then the approximation approach should be used for TRV derivation.

2.4.4 Approximation Approach

When the data set requirements are not satisfied, the available toxicity data are insufficient to characterize toxicity for a class of animals with the desired degree of certainty and confidence. Therefore, it becomes necessary to use UFs in the development of TRVs until more toxicity data are available.

In this approach, the most relevant study that is identified in the toxicity profile and is identified as the most reliable in terms of quality and applicability, should be used for TRV derivation that approximates the NOAEL and LOAEL-based TRVs described previously. These TRVs are developed by dividing the effect level of interest by appropriate UFs (see Table 4). Extrapolation from a single study or from data that are unreliable, given an understanding of the

design, may not be appropriate. Professional judgment by a toxicologist is recommended to determine whether the development of TRV approximations from limited data are justified.

The UFs used for TRV derivation need to account for potential differences in responses between species and differences in response due to exposure duration (e.g., acute vs. chronic) and endpoint (e.g., lethality vs. sub-lethal NOAEL). A general UF of 10 to protect against potential interspecies differences should be used for screening-level assessments.

The UFs described in Table 4 should be used to account for differences in exposure duration and endpoint for chronic exposures protective of sublethal effects. Most of these factors are based on the work of Ford et al. (1992) [30] and are presented in the tri-Service guidelines [22]. The factor for the chronic LOAEL to chronic NOAEL conversion is 10, whereas Ford et al. (1992) [30] would apply a factor of 5. The EPA identifies an approach that would apply a factor of 10 [21] based on an evaluation by Dourson and Stara (1983) [7]. Note that where Ford et al. (1992) [30] uses a combined UF of 16 to account for interspecies variability, this procedure uses an UF of 10 (see paragraph above). The rationale behind this change is that Chapman et al. (1998) [2] recommends that any factor used in extrapolation should be limited to an order of magnitude.

Acute TRVs may be required when using spatially explicit exposure models or other applications where it is reasonable to assume individuals of some species have single day exposure (e.g., when characterizing a hot spot, or understanding the threshold for mortality). Therefore, UFs are not needed when sufficient species are represented that describe a lowest lethal dose. When slopes are available, LD01, LD05, or LD10 may be useful. Median lethal doses should not be used; however, it is generally considered that using an UF of 20 or 100 approximates a chronic LOAEL and NOAEL, respectively.

Table 4. Uncertainty Factors Accounting for Differences in Response Due to Exposure Duration and Endpoint

Type of Available Data	Uncertainty Factors to Approximate a TRV	
	NOAEL-Based[a]	LOAEL-Based[b]
Chronic NOAEL	1	N/A
Chronic LOAEL	10	1
Sub-chronic NOAEL	10	N/A
Sub-chronic LOAEL	20	4
Acute NOAEL	30	N/A
Acute LOAEL	50	10
LD50	100	20

Notes:
[a] Ford et al. (1992) [30], except for the chronic LOAEL
[b] The factors for approximating a LOAEL-based TRV are derived using the other factors, assuming the chronic LOAEL is 5 times the chronic NOAEL.
Legend:
N/A=not appropriate

Uncertainty factors may be updated as new, class, or when chemical-specific information becomes available.

2.4.5 Use of Probabilistic (Bayesian) Uncertainty Factors

There are generally two steps to derive an RfD or reference concentration (RfC) in a non-cancer risk assessment, i.e., POD (point of departure) derivation and low-dose extrapolation. Traditionally, and as described above, the NOAEL is chosen as the POD, but the statistical lower bound of an estimated benchmark dose (i.e., the BMDL) has emerged as the default choice for the POD to replace the NOAEL. In the second step, an RfD/RfC is calculated by dividing the POD by multiple uncertainty factors (UFs). Typically, these UFs are expressed as a single value to account for the differences between the critical study on which the POD is based and the target population intended to protect; thus, the resulting RfD/RfC is a point estimate.

One important criticism on the current approach of deriving RfD/RfC is that using single value uncertainty factors can only adjust differences between the subjects in the critical study and target population of protection; however, it does not take the uncertainty associated with these differences and adjustments into account. Therefore, as suggested by a few organizations and committees [31, 32], uncertainty factors should be expressed as distributions to consider both adjustment and uncertainty, and log-normal distribution became a convenient choice for UFs in both reports. Once the UFs are expressed as distributions, the resulting RfD/RfC will become a distribution and will indicate a safe dose or concentration level and its associated uncertainty (useful information for decision making). Although log-normal distributions are used to replace the classical single value UFs, the general framework to derive the RfD/RfC remains the same, i.e., first deriving a POD and then extrapolating it to an RfD.

In the new framework, a series of assumptions has been made to eventually estimate the distribution of RfD/RfC (i.e., the TRV). Simon et al. (2016) [33] demonstrated how to build distributional estimation into the traditional RfD/RfC equation, i.e., equation 1 (Eq.1) as described below:

$$RfD = \frac{POD}{UF} = \frac{POD}{UF_H \times UF_A \times UF_S \times \cdots} \qquad \text{Eq.1}$$

There are important assumptions to keep in mind when using probabilistic (Bayesian) UFs:

Assumption 1: The POD follows a log-normal distribution. When the BMD method is used as an example, then if only the BMD and BMDL are provided (as is the case with the EPA's BMD software program) and based on the ln(BMD) (the mean value on a log scale) and the ln(BMDL) (the 5th percentile on log scale), the μ_{POD} and σ_{POD} (i.e., the two parameters of the log-normal distribution of the BMD) can be estimated.

Assumption 2: All UFs follow a log-normal distribution. For UF_A, which is the uncertainty factor for interspecies uncertainty, the two parameters of the log-normal distribution are μ_{UFa} and σ_{UFa}, (i.e., the mean and the standard deviation on a log scale). For UF_H, which is the uncertainty factor for intraspecies variability, the two parameters of the lognormal distribution are μ_{UFh} and σ_{UFh} (i.e., the mean and standard deviation on log scale). All the remaining

uncertainty factors, such as a sub-chronic to chronic extrapolation or UF_S, and a database uncertainty factor or UF_D (as defined by the assessment professional), will follow the same pattern that have two parameters of a log-normal distribution.

By taking log on both sides of Eq 1 above, the result is the following equation 2 (Eq.2):

$$\ln(RfD) = \frac{\ln(POD)}{\ln(UF)} = \ln(POD) - \ln(UF_H) - \ln(UF_A) - \ln(UF_S) - \cdots \quad \text{Eq.2}$$

Since the POD and all associated UFs are assumed to follow a log-normal distribution, then $\ln(POD)$, $\ln(UF_H)$, etc., are all normally distributed. As a sum of multiple random variables of the normal distribution, the mean and standard deviation of $\ln(RfD)$ can be expressed as:

$$\mu_{RfD} = \mu_{POD} - \mu_{UFh} - \mu_{UFa} - \cdots = \ln(BMD) - \sum \mu_{UF} \quad \text{Eq.3}$$

$$\sigma_{RfD} = \sqrt{\sigma_{UFh}^2 + \sigma_{UFa}^2 + \sigma_{UFs}^2 + \cdots} \quad \text{Eq.4}$$

Therefore, to calculate the 5th percentile of the $\ln(RfD)$, see the following equation 5 (Eq.5):

$$5\text{th percentile of } \ln(RfD) = \ln(BMD) - \sum \mu_{UF} - Z_a \sqrt{\sigma_{UFh}^2 + \sigma_{UFa}^2 + \sigma_{UFs}^2 + \cdots} \quad \text{Eq.5}$$

Wherein Zα is the absolute Z-score, which is 1.645 for the 5th percentile that is associated with the BMDL. It is important to note that the value calculated in Eq.5 is on a log scale, and it will need to be transferred to a regular (linear) scale; thus exponential “exp()” should be taken.

Based on the steps illustrated above, the traditional RfD (i.e., the TRV) is derived by dividing the BMDL (i.e., 5th percentile of BMD) by multiple uncertainty factors (everything is on a regular scale). However, the Simon et al. (2016) approach [33] calculates the 5th percentile of the distribution of RfD (on the log normal scale) and then transfers the value back to regular scale.

Chiu and Slob (2015) [34] proposed a human dose (a new RfD) based on the World Health Organization-International Programme on Chemical Safety (WHO-IPCS) (2014) report [32] to take both magnitude (i.e., the seriousness of the endpoint) and protection level (e.g., 5th percentile to protect 95% of the population, or 1st percentile to protect the 99% of the population). It is a more complex concept, but the basic algorithm to calculate the human dose from BMD is similar to the process above, i.e., multiple lognormal distributions are combined together.

The Bayesian BMD software program and modeling (BBMD) system [29] provides a more reliable way to derive the distribution of the RfD or human dose. As shown above, one important assumption applied in the Simon et al. (2016) [33] approach is that the estimated BMD follow a log-normal distribution (then, the ln(BMD) and ln(BMDL) can be used to estimate the parameters of the log-normal distribution), which is not always true or plausible. By contrast, the

BBMD directly generates the posterior sample of the BMD distribution, which can be more smoothly integrated with the distributional uncertainty factors using Monte Carlo simulation to generate the distribution of RfD (which is not necessarily log-normally distributed as well). Theoretically, the distributions of these uncertainty factors can be more flexible and are not necessarily to be log-normal.

2.4.6 Other Approaches

Other approaches for deriving TRVs from dose-response information include the development of species sensitivity distributions (SSDs), site-specific values focused on meta-population regulating mechanisms, and straight regression fitted lines of dose-response information to calculate a sub-lethal effective concentration (ECx) or dose (EDx). Development of SSDs assume the variation in dose-response relationships are exclusively influenced by species differences; however, unless differences in study deign are considered (e.g., feeding versus gavage/bolus designs), variation in response that is attributable to species may be confounded. Predominantly, toxicological criteria measured or assessed in wildlife models mimic (and often use) the same for human health assessment and include clinical chemistries, histopathological examinations, and measurements of organ and cellular level responses. Therefore, unless study methods are directly comparable, SSDs for terrestrial wildlife species are not recommended. Further, since BMD approaches use the same dose-response data to best fit a function, straight-line methods to calculate an ECx would be inferior to a BMD or BMDL.

Use of site-specific criteria for site-specific applications promise to provide more accurate end-point selection and measurements and as such are typically preferred over other methods. The user is encouraged to use toxicity profile information to help corroborate site-specific observations and measurements to develop site-specific TRVs; however, this is beyond the scope of this TG.

2.5 Confidence Level Assignment

All measures of effect contain some degree of uncertainty. The data available to develop TRVs are usually limited and not equal in their ability to describe a threshold for toxicity for all species within a vertebrate class. An assigned level of confidence should be used to communicate this fact, as it can be helpful to risk assessors and risk managers in—

- Determining the accuracy of the risk estimate.
- Judging overall uncertainty.
- Deciding where to focus additional resources to increase certainty.

The purpose of this step is to ensure that a qualifying estimate of the reliability for each TRV is documented and available. The confidence levels should be qualitative (high, medium, and low) estimates of accuracy in the toxicity estimates. They should be based on professional judgment reflecting the confidence that the toxicologist that selects the TRV will be accurate in predicting benchmarks of toxicity. Factors considered may include the range of interspecific variation in response, completeness of the database, and overall quality of the experiments from which the conclusions were based [3].

This step is consistent with the methods used by the EPA in RfD derivation in human health risk assessment applications [23].

2.6 The TRV Report

The wildlife TRV report for a chemical shall describe the derivation of the TRV that, at a minimum, shall consist of the following components:

- Discussion of how the data were used to generate the TRVs.
- Documentation of the rationale behind all decisions made in TRV development.
- Documentation of the confidence associated with each measure (i.e., low, medium, high) that are based on professional judgement and the weight of evidence described in the report.

In documenting the Toxicity Profile and TRV report, it is recognized that there is a risk of bias at all stages in the development of a TRV. The logical structure of this TG and the workflow described herein, provides a framework to significantly reduce risk of bias.

Finally, it is a requirement that for any TRV report aligned to a chemical of military interest that the information described in that report, as well as the TRVs contained therein, should be subjected to comprehensive external peer-review to mitigate against introducing subjective decision-making. This should mitigate against introducing analyses that do not follow a logical argument and, ultimately, strengthen confidence in the reported TRV derivations.

Appendix A

References

1. Department of the Army. 2020. Regulation 40-5*, Army Public Health Program.*

2. Chapman PM, A Fairbrother, and D Brown. 1998. "A critical evaluation of safety (uncertainty) factors for ecological risk assessment." *Environ Toxicol Chem* 17(1):99-108. https://doi.org/10.1002/etc.5620170112

3. Lent EM, TE Sussan, GJ Leach, and MJ Johnson. 2020. "Using Evidence Integration Techniques in the Development of Health-Based Occupational Exposure Levels." *Int J Toxicol* 40(2):178-195. doi: 10.1177/1091581820970494

4. Leenaars M, et al. 2012. "A step-by-step guide to systematically identify all relevant animal studies." *Lab Anim* 46(1):24-31. doi: 10.1258/la.2011.011087

5. National Toxicology Program. 2019. *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration.* Office of Health Assessment and Translation (OHAT), Division of the National Toxicology Program, National Institute of Environmental Health Sciences. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf

6. Vandenberg LN, et al. 2016. "A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals." *Environ Health* 15(1):74. doi: 10.1186/s12940-016-0156-6

7. Dourson ML and JF Stara. 1983. "Regulatory history and experimental support of uncertainty (safety) factors." *Regul Toxicol Pharmacol* 3(3):224-238. doi: 10.1016/0273-2300(83)90030-2

8. Suter II G and LW Barnhouse eds. 1993. *Ecological Risk Assessment*. Boca Raton, Florida: Lewis Publishers.

9. EPA. 1998. EPA/630/R-95/002F, *Guidelines for Ecological Risk Assessment.* https://www.epa.gov/sites/default/files/2014-11/documents/eco_risk_assessment1998.pdf

10. U.S. Department of Energy. 1996. ES/ER/TM-86/R3, *Toxicological Benchmarks for Wildlife: 1996 Revision,* by B Sample, DM Opresko, and GW Suter II, Lockheed Martin Energy Systems. https://rais.ornl.gov/documents/tm86r3.pdf

11. Klimisch HJ, M Andreae, and U Tillmann. 1997. "A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data." *Regul Toxicol Pharmacol* 25(1):1-5. doi: 10.1006/rtph.1996.1076

12. EFSA. 2006. “Opinion of the scientific panel on food additives, flavourings, processing aids and materials in contact with food (AFC) on a request from the commission related to a new long-term carcinogenicity study on aspartame.” *EFSA J* 4(5):1-44. https://doi.org/10.2903/j.efsa.2006.356

13. Morgan RL, et al. 2016. “GRADE: Assessing the quality of evidence in environmental and occupational health.” *Environ Int* 92-93:611-6. doi: 10.1016/j.envint.2016.01.004

14. EPA. 2005. OSWER Directive 9285.7-55, *Guidance for Developing Ecological Soil Screening Levels*. EPA Office of Solid Waste and Emergency Response. https://www.epa.gov/chemical-research/guidance-developing-ecological-soil-screening-levels

15. Hooijmans CR, et al. 2014. “SYRCLE’s risk of bias tool for animal studies.” *BMC Med Res Methodol* 14(1):43. doi: 10.1186/1471-2288-14-43

16. EPA. 1998. *Proposed Health Effects Test Standards for Toxic Substances Control Act Test Rules. 40 CFR Part 772. Standard for Development of Test Data. Subpart D;* Federal Register Vol. 44, pp. 27350-27362. Office of Testing and Evaluation. July 1998.

17. EPA. 1989. EPA/540/1-89/002, *Risk Assessment Guidance for Superfund. Volume 1, Human Health Evaluation Manual (Part A) Interim Final*.

18. EPA. 1995. EPA-820-B-95005, *OW. Great Lakes water quality initiative technical support document for the procedure to determine bioaccumulation factors.*

19. EPA. 1998. EPA 712-C-98-199, *Health Effects Testing Guidelines: OPPTS 870.3100 90-Day Oral Toxicity in Rodents*.

20. U.S. Food and Drug Administration. 1966. Guidelines for Reproductive Studies for Safety Evaluation of drugs for Human Use. Rockville, Maryland.

21. EPA. 1997. EPA 540/R-97/006, *Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments, Interim Final.*

22. Edgewood Research, Development and Engineering Center (ERDEC). 1996. Report number AD-A-314323/7/XAB, *Tri-service procedural guidelines for ecological risk assessments, Volume 1. Final report, January-May 1996, by* Wentsel RS, et al. Aberdeen Proving Ground, Maryland.

23. *Federal Register,* Vol. 63, No. 26846. 1998. “Guidelines for Ecological Risk Assessment.”

24. Sample BE and CA Arenal. 1999. “Allometric models for interspecies extrapolation of wildlife toxicity data.” *Bull Environ Contam Toxicol* 62(6):653-63. doi: 10.1007/s001289900924

25. EFA West (Engineering Field Activity, West). 1998. *Development of Toxicity Reference Values Part of a Regional Approach for Conducting Ecological Risk Assessments at Naval Facilities in California.* Interim Final. EFA West, Naval Facilities Engineering Comman. San Bruno, California

26. Haber LT, et al. 2018. “Benchmark dose (BMD) modeling: current practice, issues, and challenges.” *Crit Rev Toxicol* 48(5):387-415. doi: 10.1080/10408444.2018.1430121

27. Moore DR and PY Caux. 1997. “Estimating low toxic effects.” *Environ Toxicol Chem* 16(4):794-801. doi: 10.1002/etc.5620160425

28. Crump KS. 1995. “Calculation of Benchmark Doses from Continuous Data.” *Risk Analysis* 15(1):79-89. https://doi.org/10.1111/j.1539-6924.1995.tb00095.x

29. Shao K and AJ Shapiro. 2018. “A Web-Based System for Bayesian Benchmark Dose Estimation.” *Environ Health Perspect* 126(1):017002. doi: 10.1289/EHP1289

30. Ford K, F Applehans, and R Ober. 1992. *Development of toxicity reference values for terrestrial wildlife.* Proceeding of the Hazardous Materials Conference, Superfund 92. pp. 803-803. Hazardous Materials Control Research Institute.

31. National Research Council (NRC). 2014. *Review of EPA’s Integrated Risk Information System (IRIS) Process.* Washington, DC: The National Academies Press.

32. WHO-IPCS. 2014. *Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization*. Geneva, Switzerland.

33. Simon TW, Y Zhu, ML Dourson, and NB Beck. 2016. “Bayesian methods for uncertainty factor application for derivation of reference values.” *Regul Toxicol Pharmacol* 80:9-24. doi: 10.1016/j.yrtph.2016.05.018

34. Chiu WA and W Slob. 2015. “A unified probabilistic framework for dose response assessment of human health effects.” *Environ Health Perspect* 123(12):1241-54. doi: 10.1289/ehp.1409385

35. *Federal Register,* Vol 63, No. 26845-26924. 1998. “Guidelines for Ecological Risk Assessment.”

36. Simberloff D and T Dayan. 1991. “The guild concept and the structure of ecological communities.” *Annu Rev Ecol Sys* 22(1):115-43. doi:10.1146/annurev.es.22.110191.000555

Appendix B

Terms

Allometry — The EPA [35] (p. 26880) provides the following discussion: “Allometry is the study of change in the proportions of various parts of an organism as a consequence of growth and development. Processes that influence toxicokinetics (e.g., renal clearance, basal metabolic rate, and food consumption) tend to vary across species according to allometric scaling factors that can be expressed as a nonlinear function of body weight.”

Demographic Rates — Demographic rates refer to survival rate, birth rate, death rate, dispersal rate (i.e., immigration and emigration), and recruitment rate.

EDx Values — An effective dose (ED) is one that elicits a response in a percentage (x) of animals tested. For example, consider a test where 10 out of 100 animals experience reduced growth after they are exposed to chemical X at a concentration approximately equal to 25 units per day for their lifetime. This result, lifetime exposure of 25 units per day of chemical X, can be expressed as the ED10 for growth effects.

Endpoints — Adverse effects that are likely to occur in a terrestrial vertebrate as a result of exposure to a contaminant. These effects need to be considered in an ecological context where effects likely to alter reproductive performance (e.g., courtship, nest defense, etc.), subsequent reproductive success (e.g., mortality), or other factors (e.g., interspecific competition, dispersal) are important in the life history of the species, the population, or the community.

Guild or Guild Association — In a general sense, a guild (or guild association) is a group of species with similar functional roles within a community [36]. In this document, guild refers more specifically to a group of species that have similar foraging (i.e., feeding) behavior and are related taxonomically (currently defined as within the same class). The implicit assumptions are: (1) species with similar foraging behavior are likely to be exposed to chemicals in similar ways, and (2) the more taxonomically related species are, the more similar they are in terms of sensitivity to a toxicant. Guild associates are the individual species within a particular guild.

NOAEL and LOAEL — These are acronyms for two toxicological endpoints. The NOAEL (no-observed-adverse-effect-level) is a concentration associated with no observed adverse effects in the tested organisms. The LOAEL (lowest-observed-adverse-effect-level) is a concentration associated with the lowest observed level of adverse effects in the tested organism.

Reference Dose (RfD) — “An estimate (with uncertainty spanning perhaps an order of magnitude) of a daily oral exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime” [17]. It can be derived from a NOAEL, LOAEL, or benchmark dose, with UFs generally applied to reflect limitations of the data used. Generally used in EPA's non-cancer health assessments.

Taxonomy and Taxon — Taxonomy is the science of classification as applied to organisms. A taxon is any group of organisms to which any rank of taxonomic classification is applied.

Taxonomic nomenclatures are based on a hierarchy of phylogeny (or similarity) of groups. Examples include species, genus, family, order, class, and phylum.

Toxicological Data Extrapolation — The procedure that estimates dose-response relationships for organisms that have not or cannot be tested themselves. It entails the process of inferring toxicity characteristics from a set of empirical toxicity data for an organism or taxon to other organisms or taxonomic groups.

Toxicity Reference Value (TRV) — A chemical concentration expressed as an administered dose (e.g., oral, inhalation or dermal dose) or as a media concentration for terrestrial amphibians that is used in conjunction with an exposure prediction to estimate health hazard or ecological risk.

Uncertainty Factor (UF) — A numerical value used to adjust an estimate of toxicity or risk. It is an approach for dealing with uncertainty related to assessing chemical risks.

Appendix C

Quality Control Guidelines and Criteria for the WTA Literature Search Strategy and TRV Development

Table C-1. Strength of Methods

1. Test Material Characterization: Confidence that Study Results Due to Compound of Interest
1.1 Clear identification of the test material identity, form, and source.
1.2 Test material composition and purity confirmed analytically (e.g., ≥98% pure); OR purity is <98%, but impurities were characterized and deemed unlikely to contribute to the evaluated toxicological endpoint. Test material composition and purity can be stated in the text or determined on provision of sufficient source and catalog information.
2. Exposure Characterization: Confidence in the Accuracy of the Administered Dose
2.1 Concentration, homogeneity, and stability in selected vehicle are analytically determined throughout the study. For inhalation studies, monitoring of chamber concentration confirmed frequently/continuously during each exposure.
2.2 Concentrations were consistent throughout the study.
2.3 Analytical methods were clearly described.
3. Selection/Allocation: Are any Differences among Groups Due to Group Assignment of Animals
3.1 Randomization method is used to assign animals to treatment groups.
3.2 The number of subjects per group appropriate for the species and study (e.g., 10-20 rodents/sex/group for sub-chronic and chronic studies, 2-3 primates or dogs/group).
3.3 Treatment groups were similar at baseline (e.g., age, weight, sex, strain).
4. Exposure Methods: Confidence in Exposure Methods
4.1 Exposure route/method is appropriate for the test substance.
4.2 Frequency/timing/duration of exposures is specified and is appropriate for the study type.
4.3 Delivered doses/concentrations are reported (or data enabling calculation are available).
4.4 Dose/concentration levels, number of dose groups, and spacing between groups is appropriate to allow for a clear dose response (i.e., at least 3 dose levels/exposure concentrations).

5. Performance: Are there Differences in Experimental Conditions between Groups
5.1 Husbandry conditions are appropriate for the species and consistent across groups.
5.2 Exposure methods are consistent across treatment groups and throughout study such that the only difference was exposure.
5.3 Negative controls and any required vehicle and positive controls have produced expected and consistent results.
6. Assessment Methods: Confidence in Measurement of Response/Endpoint
6.1 The endpoint was assessed using valid, reliable, and sensitive methods.
6.2 The endpoint was measured consistently across treatment groups.
6.3 Assessors were blinded to treatment group assignments.
7. Data Analysis: Were Data Analyzed Appropriately and Data Removal/Animal Loss Addressed to Account for Potential Differences between Groups
7.1 Statistical methods were clearly reported and appropriate.
7.2 No removal from study or loss of animals OR removal or loss of animals adequately explained, and loss is small and consistent across groups. Missing data or loss will have no impact on study results.
8. Reporting and Bias
8.1 The nature and frequency of toxic responses noted in relation to dose/concentration and sex. Time of death was noted if it occurred, and necropsy was performed on all animals. Body weight and organ weights were reported for critical endpoints, where appropriate.
8.2 No evidence of data omission to bias study conclusions.
8.3 Funding sources were reported, and authors reported no conflict of interest.

Table C-2. Relevance of Response

1. Statistical Significance of Response
1. Endpoint demonstrates statistical significance (at an alpha value of at least p <0.05).
2. Effects are dose-dependent (or occur only at high-dose); repeat observations are internally consistent.
3. Effect in measured endpoint is large (i.e., effect size ≥0.8) or moderate (i.e., effect size ≥0.5 but <0.8).
2. Biological Significance: Confidence that Observed Effect is Important for an Adverse Health Outcome in Class or Species of Interest
2.1 Statistically significant response in a measurement that has direct biological relevance to disease (i.e., adverse physiologic response).
2.2 Statistically significant response in a measurement that has indirect biological relevance to disease (i.e., abnormal histopathology or biomarker with established predictability of disease).
3. Coherence: Confidence that Key Endpoint is Consistent with other Endpoints in the Study
3.1 Key endpoint is supported by complimentary assays within the study.
4. Exposure Relevance: Confidence in Accuracy of Internal Dose at Tissue/Organ of Interest Relative to Expected Exposure Conditions
4.1 Compound tested/vehicle used is similar to expected actual exposure conditions (i.e., relevant route, dosing regimen, compound/oxidation state, and no stabilizing solvents).
4.2 Preference given to studies conducted by oral, dermal, and inhalation routes of exposure.
4.3 Application of professional judgement to determine relevance of exposure protocol to TRV derivation; however, chronic and subchronic studies are generally preferred.

Appendix D

An Exemplar of Documenting the Literature Search Strategy

A very broad search on 3 June 2018, using DTIC's MultiSearch function used the single search term, dimethyl phthalate. This search identified 851 specific documents.

D-1. GENERAL APPROACH

Relevant biomedical, toxicological, and ecological databases (e.g., BIOSIS, Defense Technical Information Center's (DTIC) On-Line Multisearch, and TOXNET) were electronically searched on 4-6 January 2018 to identify primary peer-reviewed reports of studies and reviews on the toxicology of DEP. Separate searches were conducted for general toxicology and specific searches for birds, reptiles, amphibians, and wildlife. Each database was searched using keywords that included diethyl phthalate or its CAS No. 84-66-2 and terms that included toxicity, ecotoxicology, wild, wildlife, avian, bird, frog, amphibian, or reptile. Details of the literature search strategy will be documented under Appendix B in the WTA document.

The titles of articles identified in each search were reviewed for specific relevance. Potentially relevant articles focused on the toxicological effects of DEP on terrestrial vertebrates or its environmental fate. All potentially relevant articles were acquired as electronic files or by visiting the University of California, Davis, and the Johns Hopkins University School of Medicine libraries. Review articles provided additional articles that were not identified during searches of the initial databases.

Additional focused searches on 3 June 2018, using the DTIC's MultiSearch function, which also searches records archived and found in PubMed (National Library of Medicine, National Institutes of Health) used the following terms including use of the wild-card (*) search operator:

diethyl phthalate + quail*. This search identified 4 documents.
diethyl phthalate + mallard*. This search identified 6 documents.
diethyl phthalate + bird*. This search identified 43 documents.
diethyl phthalate + avian. This search identified 17 documents.
diethyl phthalate + mouse. This search identified 22 documents.
diethyl phthalate + mice. This search identified 32 documents.
diethyl phthalate + rat. This search identified 33 documents.
diethyl phthalate + rats. This search identified 33 documents.
diethyl phthalate + mammal*. This search identified 38 documents.
diethyl phthalate + ecotox*. This search identified 33 documents.
diethyl phthalate + toxic*. This search identified 313 documents.
diethyl phthalate + amphib*. This search identified 57 documents.
diethyl phthalate + frog. This search identified 30 documents.
diethyl phthalate + reptil*. This search identified 37 documents.
diethyl phthalate + wildlife. This search identified 144 documents.

On 3 June 2018, a search of the EPA's online Ecotox database used the CAS No. 84-66-2. During this search, there were 40 hits. No references for amphibians, reptiles, or birds were identified. Twenty-seven mammalian references were found—predominantly mouse and rat as the standard test species in listed studies.

A search of the TOXLINE database [a sub-database of the National Library of Medicine's TOXNET Toxicology Data Network (https://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?TOXLINE)] on 3 June 2018, used the CAS No. 84-66-2 as the search term. A total of 1206 articles were identified. This search was refined with:

84-66-2 AND ecotox* resulted in 43 hits
84-66-2 AND reptil* resulted in no hits
84-66-2 AND amphib* resulted in 1 hit
84-66-2 AND frog resulted in 1 hit
84-66-2 AND avian resulted in 0 hits
84-66-2 AND mallard resulted in 0 hits
84-66-2 AND quail resulted in 0 hits
84-66-2 AND bird* resulted in 3 hits
84-66-2 AND wildlife resulted in 2 hits
84-66-2 AND mammal* resulted in 56 hits
84-66-2 AND toxicity resulted in 238 hits

Independent searches of the BIOSIS Citation Index (also known as Web of Science), on 3 June 2018, used many keyword combinations to capture articles that might have been missed in the broader searches. These combinations were:

diethyl phthalate AND ecotox* resulted in 12 hits
diethyl phthalate AND reptil* resulted in 0 hits
diethyl phthalate AND amphib* resulted in 2 hits
diethyl phthalate AND frog resulted in 2 hits
diethyl phthalate AND avian resulted in 0 hits
diethyl phthalate AND mallard resulted in 0 hits
diethyl phthalate AND quail resulted in 0 hits
diethyl phthalate AND bird* resulted in 0 hits
diethyl phthalate AND wildlife resulted in 5 hits
diethyl phthalate AND wild* resulted in 9 hits
diethyl phthalate AND toxic* resulted in 281 hits

The different searches defined above identified similar articles. Additional references were identified during the review of individual articles. A total of 108 articles were reviewed.

Updated searches were performed on 28 June 2019 for additional articles published between 2018 and 2019, using the following online references and search strings:

- **DTIC**
  84-66-2; Resulted in 3 hits. None were relevant to this WTA.
  Diethyl phthalate; Resulted in 10 hits. None were relevant to this WTA.

- **ECOTOX**
  84-66-2; Resulted in 0 hits.
  Diethyl phthalate; Resulted in 0 hits.

- **TOXLINE**
  84-66-2; Resulted in 86 hits. 4 were relevant to this WTA.
  Diethyl phthalate; Resulted in 86 hits. 4 were relevant to this WTA.

Appendix E

Example of Detail Required for a Data Extraction Spreadsheet of a Chemical of Toxicological Interest Derived from the Literature Search Strategy (as described in Appendix D)

Study Name	Route	Species	Exposure Duration	Domain	Critical Effect	NOAEL	LOAEL	Other Doses	(ppm) BMD	BMDL	POD	40h/w POD	Notes	Data used for BMD (N, Mean, SD)	BMD Model
Adams 1951	Inhalation	Rat, Guinea pig, Rabbit, Monkey	7h/d, ~5d/wk, 6- 8mo	Liver	Relative liver weight for rat and guinea pig	200	400	100	ND	ND	200	175	BMD not run b/c no SD reported		
Adams 1951	Inhalation	Rat, Guinea pig, Rabbit, Monkey	7h/d, ~5d/wk, 6- 8mo	Kidney	Relative kidney weight for rat	200	400	100	ND	ND	200	175	BMD not run b/c no SD reported		
Albee 2006	Inhalation	Rat	6h/d, 5d/w, 13w	Neuro	Hearing deficits/loss of cochlear hair cells	800	2500	250	ND	ND	800	600	BMD not run b/c no SD reported		
Arito 1994	Inhalation	Rat	8h/d, 5d/w, 2,4,6w	Neuro	Decreased wakefulness		50	100, 300	42	18.4	18.4	18.4	BMD did not adequately model the variance, but the modeled variance was more conservative than the actual variance. Thus, BMD was preferred instead of LOAEL.	0 (5, 103.798, 24.012), 50 (5, 90.526, 8.6846), 100 (5, 77.019, 5.597), 300 (5, 70.007, 15.543)	Modeled Variance, BMR1SD, Restricted, Exp4

Study Name	Route	Species	Exposure Duration	Domain	Critical Effect	NOAEL	LOAEL	Other Doses	(ppm) BMD	BMDL	POD	40h/w POD	Notes	Data used for BMD (N, Mean, SD)	BMD Model
Bushnell 1997	Inhalation	Rat	1h/d, 3d/w, 2w	Neuro	Sensitivity to light stimulus (visual learning)	800	1200	400, 1600, 2000, 2400	1681	1366	1366		This exposure requires large extrapolation from 3h/wk to 40h/wk. This exposure may not qualify for an occupational setting. Did not adjust POD because effects showed greater reliance on exposure concentration than duration.	0 (11 0.582, 0.081), 400 (11 0.549, 0.081), 800 (11, 0.558, 0.113), 1200 (11, 0.494, 0.096), 1600 (11, 0.496, 0.087)	Constant Variance, BMR1SD Up, Restricted, average of Exp3, Ply3, and Pow
Carney 2006	Inhalation	Rat	6h/d, 7d/w from GD6-20	Develop	Cardiac Defects	600		50, 150	ND	ND	600	630	No LOAEL determined		
Crofton 1997	Inhalation	Rat	6h/d, 5d/w, 1d, 1w, 4w, 13w	Neuro	Auditory threshold at 13wks	1600	2400	800, 3200	1695	1367	1367	1025	Considered using BMR of 15dB (as recommended by publication), 10dB (as performed by EPA), and 1SD (which is our default). BMD1SD gave the most conservative BMDL.	0 (10, 7.749, 5.588), 800 (10, 7.965, 4.330), 1600 (10, 12.468, 7.264), 2400 (10, 28.701, 7.473), 3200 (10, 41.212, 6.635)	Constant Variance, BMR1SD, Restricted, average of Hill and Exp5
Healy 1982	Inhalation	Rat	4h/d, GD8-21	Develop	Resorptions, fetal weight		100		ND	ND	100	70	One dose		

Study Name	Route	Species	Exposure Duration	Domain	Critical Effect	NOAEL	LOAEL	Other Doses	(ppm) BMD	BMDL	POD	40h/w POD	Notes	Data used for BMD (N, Mean, SD)	BMD Model
Jaspers 1993	Inhalation	Rat	18h/d, 5d/w, 3w	Neuro	Auditory threshold	1500	3000		ND	ND	1500	3375	Cannot run BMD- Low dose gives response below control		
Kaneko 2000	Inhalation	MRL Mice	4h/d, 6d/w, 8w	Immune	B cell function was used b/c the histology was not quantified, and the T cell staining lacks clear relevance and shows an effect at only one dose.		500	1000, 2000	ND	ND	500	300	No NOAEL determined. Considered BMD, but none of the curves adequately fit the data ($p<0.1$)		
Kjellstrand 1983b	Inhalation	Mouse	Cont (30, 120d). Int (1-16h/d, 7d/w, 30, 120d)	Liver	Relative liver weight (continuous exposure, 30 d)		37	75, 150, 300	ND	ND	37	155	No NOAEL determined. The multiple control groups make BMD analysis difficult to interpret. It's also unclear whether error is CI or SEM.		
Kumar 2000a	Inhalation	Rat	4h/d, 5d/w, 12 or 24 wks	Repro	Sperm count and mobility		376		ND	ND	376	188	One dose		

Study Name	Route	Species	Exposure Duration	Domain	Critical Effect	NOAEL	LOAEL	Other Doses	(ppm) BMD	BMDL	POD	40h/w POD	Notes	Data used for BMD (N, Mean, SD)	BMD Model
Kumar 2001a	Inhalation	Rat	4h/d, 5d/w, 8,12,24 wks	Liver	Liver weight (sig inc but data not shown), enlarged fatty hepatocytes (via histology but no stats)		376		ND	ND	376	188	One dose		
Land 1981	Inhalation	Mouse	4h/d, 5d	Repro	Abnormal epidiymal sperm	200	2000		804	575	575	287		0 (15, 1.42, 0.310), 200 (10, 1.68, 0.538), 2000 (5, 2.43, 0.335)	Constant Variance, BMR1SD, Restricted BMDL identical for Lin Ply2, Ply3, and Pow

Note:
Caution is recommended when using assumptions required in the determining daily oral dose from feeding and drinking water studies as these assumptions often implicitly contain significant variation. The above Table is an exemplar and was derived from a toxicological report on TCE.