MEMORANDUM

To: Eric Hooker, Matthew Cho, and Kristina Hatlelid, CPSC

From: Samantha Snow, Joei Robertson, and Elizabeth Martin, ICF
Mark Bradley and Lynne Haber, University of Cincinnati

Date: June 5, 2024

Re: Task 2A: Identify approaches to perform class-based dose-response analyses.
(CPSC Contract Number 61320622A0005, CPSC Order Number 61320623F2030)

---

1. Introduction

Under CO-9, subtask 2A seeks to identify approaches to perform class-based dose-response analyses. To locate approaches, a targeted literature search was conducted followed by title/abstract and full text screening to identify references of relevancy. This memo describes the data sources considered and a summary of the information found.

2. Literature Search

ICF and UC conducted a search for relevant methods papers for this task based on three data streams:

1) Literature identified via searches of scientific literature databases (i.e., PubMed and Web of Science) with no date restrictions;
2) Literature identified via a search of the grey literature, or publications produced outside of traditional scientific publishing avenues, including reports from government agencies and authoritative bodies not generally available in databases of published literature; and
3) Publications and reports identified by subject matter experts.

Three separate literature searches were conducted under subtask 2A to identify methods for various subtasks under CO-9. While our primary objective was to identify approaches for conducting class-based dose-response (subtask 2A), we simultaneously conducted a literature search to identify approaches for conducting class-based risk assessment (subtask 3A) and approaches for deriving toxicity reference values (TRVs) using new approaches/alternative methods (NAMs) data (subtask 1C). Given the similarities across these topics, we decided to conduct the literature searches for these topics under subtask 2A, perform a deduplication across the individual searches, and screen the papers in one title/abstract step (Section 3). The search strings and strategies for the three topics of interest are presented in Appendix A.

2.1. Literature Search Strategies

The following sections describe literature search strategies used for databases and additional sources. The literature search strategies included searches within core literature databases (i.e., PubMed and Web of Science) as well as relevant domestic and international non-periodical “grey” literature, such as technical reports, monographs, and symposium proceedings prepared by select committees or bodies (e.g., National Academy of Sciences).

2.2. Seed Studies

Subject matter experts identified 33 references to use as positive seed studies to help verify the accuracy and efficiency of our literature search (Attachment 1; Figure 1). These relevant references were identified from reference compilations identified by the report authors as part of other projects or from general monitoring of the literature. The goal was to identify key references related to the use of NAMs in risk assessment, with particular focus on dose-response, read across, and addressing data-poor chemicals. These studies were used as seed studies for the ensemble supervised clustering described in Section 3.1. Of the 33 seed studies, 7 studies were captured in our database search (Section 2.3) and 1 study was captured in our grey literature search (Section 2.4). Although the low capture rate for the seed studies in the initial literature search may raise concerns about the adequacy of the initial literature searches, the initial searches were heavily supplemented by expert review of the grey literature and tree searching, as described below. In addition, missing individual studies is of less concern for this task than for an assessment, since the goal was to capture the general approaches that are in the literature.

2.3. Database Searches

The literature searches for peer-reviewed publications were conducted on December 4, 2023 in PubMed and Web of Science with no date limit. There were three searches conducted on topics of interest:

1) Approaches for class-based dose-response (subtask 2A)
2) Approaches for class-based risk assessment (subtask 3A)
3) Approaches for deriving a TRV using NAMs data (subtask 1C)

Removing duplicates across the three searches resulted in a total of 4,672 unique references (Attachment 1; Figure 1). These references were moved to a topic extraction step (Section 3.1), before going forward into title/abstract screening using Litstream®.

2.4. Identification of Grey Literature

There were 11 agency websites identified by ICF and UC, and confirmed by CPSC, for review of relevant grey literature (Appendix A). Grey literature was pulled from these relevant websites using an ICF-developed Google webscraping tool. This tool crawled the Google results page for a search that had been limited to a specific domain and pulled links and metadata associated with the search results into an Excel file. Each search was limited to pull in only the top 10 hits per source (for 11 sources) per search string (for 3 topic specific search strings), meaning a maximum of 330 grey literature search results would be pulled by the webscraper.

A total of 259 references/websites were returned from the Google webscraping (Attachment 1; Figure 1). These items underwent a first pass screening by a subject matter expert to assign a priority rating (i.e., not relevant, maybe relevant, include) and to identify the topic or key take-away of the references that were classified as maybe relevant or include. If the initial webscraping tool identified a general website, that website was reviewed to identify any useful/relevant references, and those references were flagged for downloading if they were not included in the list from the webscraping tool. A second review was conducted of the literature marked as maybe relevant or include in order to identify the most recent relevant references. Key topics were those that addressed methods for read across, class-based assessments, cumulative risk/combined exposure, and use of NAMs in risk assessment. Please note that not all literature was reviewed in detail.

2.5. Expert Identified Literature

Key agency guidance documents that the authors were aware of but had been missed by the webscraping tool were downloaded and categorized as expert identified literature. In addition, subject matter experts may have identified additional peer-reviewed or grey literature sources during their synthesis writing (e.g., as a reference cited in a paper that was captured in our literature search) that resulted in them pulling in expert identified literature. A total of 36 expert identified literature were included in this report (Attachment 1; Figure 1).

3. Prioritization and Literature Screening

3.1. Prioritization of Literature for Screening

The peer-reviewed literature identified via database searches (n = 4,672) underwent a prioritization step prior to title/abstract screening. To prioritize these 4,672 studies, three unique approaches were applied: (1) keyword analysis, (2) unsupervised clustering, and (3) supervised clustering.

Keyword analysis enabled us to identify specific keywords that were directly relevant to the three topics of interest (i.e., approaches for class-based dose-response, approaches for class-based risk assessment, and deriving TRVs from NAMs data) within the title or abstract of a paper (Appendix Table B-1). This list was smaller than the broader set of terms used in the literature search strings (Appendix A). Based on keyword analysis, 1,489 papers contained one or more of the keywords.

The unsupervised clustering method groups papers that have similar keywords in both the title and abstract (Appendix B). In contrast to keyword analysis, which only reports the incidence of pre-specified words, unsupervised clustering analysis relies on a computational algorithm to group papers without any *a priori* information (Manning et al., 2008). Based on the unsupervised clustering analysis, only one group of papers (Cluster ID Group 6, Appendix Table B-2) was easily eliminated from consideration. This group focused on chemical measurement terms (e.g., spectrometry, extraction, analytical, liquid chromatography), implying primary literature focused on measurement of chemicals and not methods for our three topics of interest. These papers (n = 326) were removed from consideration and not included for title/abstract screening.

Supervised clustering uses expert identified seeds and multiple clustering algorithms to categorize papers by giving them an ensemble score to indicate likelihood of relevancy (Varghese et al., 2018) (Appendix B). For supervised clustering, 797 papers (i.e., the papers that were clustered into Cluster ID 4, 5, and 6, (Cawley et al., 2020) were considered to be similar to those identified by an expert (Appendix Table B-3).

When looking to prioritize papers, unsupervised clustering was too lax in the elimination of papers, while supervised clustering appeared to be too stringent. Therefore, we used keyword analysis to optimize inclusion of likely relevant literature and proceeded with title/abstract screening for the 1,489 papers identified using keyword analysis.

3.2. Title/Abstract Screening

A total of 1,489 studies identified using keyword analysis were uploaded into ICF's Litstream® software. A screening form was developed using inclusion/exclusion criteria to guide decisions about the relevance of the literature. Screeners reviewed the title and abstract to determine whether the reference should be included, excluded, or marked unclear. If a reference was marked as unclear, it underwent a QC step by a senior screener to make the final decision between inclusion and exclusion. For a reference to be included, it needed to apply to at least one of three categories: (1) describe an approach to perform class-based dose-response analyses, (2) describe an approach to perform class-based risk assessment, or (3) describe an approach to derive TRVs using NAMs data. As noted above, these categories were selected because they are the objectives of subtask 2A, subtask 3A, and subtask 1C, respectively.

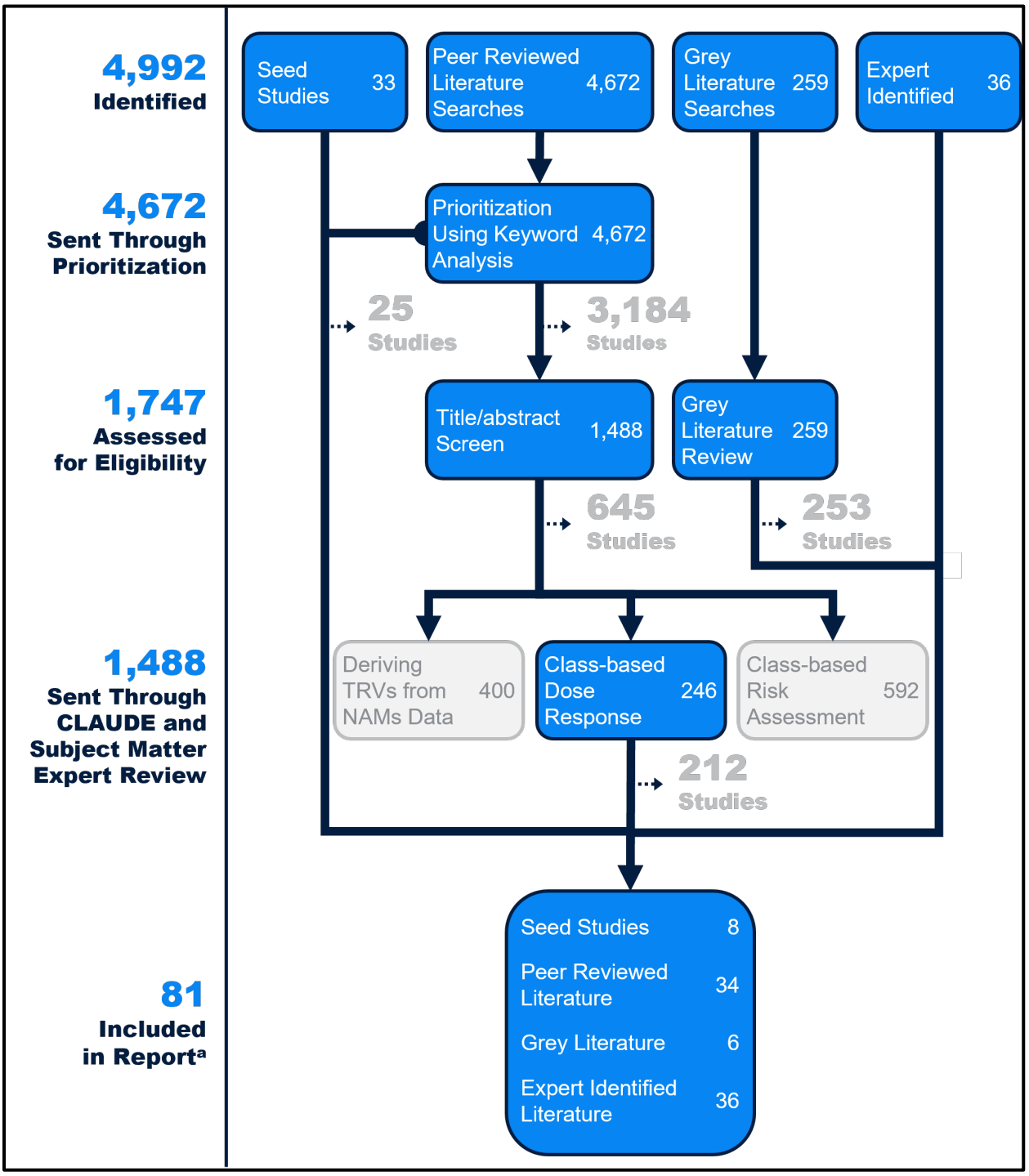
Each screener completed a pilot of 30 references, which was then QC'd. Feedback was provided to the screeners regarding any discrepancies between the primary screeners and senior screener. After successfully completing the pilot, screeners continued working until all 1,489 references were reviewed. Each reference was reviewed by only one screener following the pilot. At the completion of title/abstract screening, there were 246 references tagged to approaches for performing class-based dose-response (subtask 2A), 592 references tagged to approaches for performing class-based risk assessment (subtask 3A), and 400 tagged to approaches for deriving a TRV from NAMs data (subtask 1C). Because a reference could have been tagged to multiple subtasks, these numbers do not total the number of unique references. The total number of unique references that were included for at least one topic was 843 (Figure 1).

3.3. Full Text Screening

From title/abstract screening, 246 peer-reviewed studies were marked as 'include' for the topic of approaches to performing class-based dose-response (i.e., the topic of subtask 2A). Of those, 227 had PDFs that were retrievable. To prioritize full text PDFs, a generative artificial intelligence (GenAI) workflow utilizing the large language model Claude 3.0 was deployed. Specifically, PDFs were converted to text by Azure Document Intelligence as Claude cannot parse text directly from the PDF. Two PDFs were too large for conversion and were removed from subsequent analysis leaving 225 summarized studies. The two large PDFs were reviewed by subject matter experts to prevent loss of information. Next, a set of 15 prompt questions were tested against 10 expert curated PDFs (Appendix Table C-1). Parameters that were varied were preamble ("Acting as a toxicologist" vs "Acting as a risk assessor"), and Claude version (1.0, 2.0 and 3.0). Next, prompt questions were graded for accuracy (correct, mostly correct, incorrect, or hallucinating[1]). Prompt questions in which 90% of answers were not correct were removed from questioning. This left 10 prompt questions, which were revised as needed to help refine the AI response and were used to prioritize papers into bins based on report directives (Appendix Table C-2). Additionally, "Acting as a toxicologist" and Claude 3.0 were identified as the most accurate parameters and used for this prioritization (Appendix Table C-3).

[1] Hallucination = When a large language model makes up an answer that is completely unrelated to the source material.

Figure 1. Literature Search Results



[a] The total number of studies listed in the final blue box is higher than the total number of studies included in the report because some studies were identified from multiple sources. Specifically, one seed study was also captured in both our peer reviewed literature search and our grey literature search (Health Canada, 2021), and another seed study was captured in our peer reviewed literature search (Paul Friedman et al. (2020).

4. Approaches to Class-Based Dose-Response

Sections 4 and 5 of this report lay out the key steps in conducting a class-based dose-response, methods for conducting those steps, and issues to consider. These steps can be broken down as follows:

1. determination of class (chemical grouping),
2. identification of anchor or surrogate chemical(s),
3. identification of toxicity data for read across, including NAMs data in the case of OFR dose response,
4. conducting dose-response evaluation and calculation of relative potency,
5. identification of appropriate uncertainty factors, and
6. calculation of TRVs.

Below we note key considerations for each of these topics, methods that are currently in use and/or under development, and potential for application to the OFRs.

4.1. Determination of Class for Read Across

4.1.1. Guidance for Grouping Chemicals

The goal of read across is to make conclusions about chemicals with little or no data by extrapolating from other chemicals that have substantial amounts of data. To do this, multiple chemicals are linked together into a group. Early grouping methods focused on physicochemical similarity, but considerations have been broadened to include a variety of biological areas of similarity, as discussed in the following paragraphs. There are two read across approaches: analog and category. In an analog approach, there is a target substance with unknown toxicity, and at least one data-rich structural analog is identified to extrapolate or estimate toxicity for the target substance. In contrast, the category approach groups multiple chemicals together by one or more shared characteristics, or by a known pattern of properties within the category (e.g., a physicochemical property trend). Class-based read across can use one of these approaches or a hybrid of both approaches to assess a class of chemicals.

Defining a chemical class has been the subject of various different approaches, all of which use the same principles for grouping chemicals for a read across category approach. Organizations such as the Organisation for Economic Co-operation and Development (OECD) (OECD, 2017), European Chemicals Agency (ECHA) (ECHA, 2017), and European Food Safety Authority (EFSA) (More et al., 2021) have published guides on how to effectively conduct read across using the two approaches above, which include guidance on how to group chemicals for a category approach. The guides share many similarities across organizations, but the differences in the purposes of the respective organizations are reflected in their guides. The Research Institute for Fragrance Materials (RIFM) frequently uses read across to assess the safety profiles of fragrance ingredients. Unlike OECD, ECHA, and EFSA, RIFM does not have a separate guidance document; instead, their methods are described on their website.

From these different organizations, we can see that there are unique considerations for dose-response assessment. OECD provides details on grouping chemicals together for a category read across approach in their Guidance on Grouping of Chemicals (OECD, 2017) by giving rationales for how to define a category, determine membership, and provides considerations for addressing structurally similar chemicals exhibiting different toxicities. ECHA's read across assessment framework (RAAF) is structured for the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) guidelines and ECHA's assessment requirements; the guidance on grouping of substances for categories can be used in the context of determining a class (ECHA, 2017). EFSA's publicly available guide for risk assessment of exposure to multiple chemicals provides guidance on chemical grouping (More et al., 2021). Unlike the other organizations, RIFM conducts read across and primarily publishes single chemical assessments using the analog method. However, their process for assessing similarity of chemicals (e.g., analogs and targets) can be applied to chemical grouping. Table 1 summarizes how each of these organizations' publications describe chemical grouping.

Table 1. Summary of Guidance Documents for Grouping Chemicals for Hazard Assessment

Organization	OECD	ECHA	EFSA	RIFM
Publication	Guidance on Grouping of Chemicals, Second Edition (OECD, 2017)	Read-Across Assessment Framework (RAAF) (ECHA, 2017)	Guidance Document on Scientific criteria for Grouping Chemicals into Assessment Groups for Human Risk Assessment of Combined Exposure to Multiple Chemicals (More et al., 2021)	RIFM.org[a]
Date	2017	2017	2021	Unknown[a]
Purpose	Provide a systematic approach to hazard assessment of groups of chemicals (vs. single chemicals) in effort to reduce animal testing.	Provide users methods for meeting ECHA's assessment requirements for REACH.	Provide harmonized methodologies for grouping chemicals for the purpose of combined exposure human risk assessment.	Description of how RIFM assesses the safety profiles of fragrance ingredients.
Summary of Chemical Grouping Guidance	Suggests a combination of structural similarities (e.g., common functional groups), similar toxicokinetics, shared MOA or other mechanistic similarities, shared AOP, and/or patterns in physicochemical properties as a function of structure.	Requires structural similarity, which can be common functional group(s), common precursors or products (i.e., shared metabolites), or a pattern in known biological or physicochemical properties within the group of chemicals.	Prioritizes similar MOA, common adverse effect endpoints, and similar exposure routes. Additional information (e.g., toxicokinetics, metabolism) recommended for refining the grouping.	Assessments are primarily for single chemicals using the analog method. Assigns chemical signatures (i.e., structural activity groups) to all chemicals, then clusters chemicals by signatures to determine the best target-analog matches using a decision tree approach.

[a] RIFM describes their method on their website blog.

4.1.2. Examples of Grouping Chemicals for Assessment

There are several examples of quantifying dose-response of groups of similar chemicals that could be considered class-based. These examples frequently rely on shared mode of action (MOA). Less often, the chemical grouping relies on a common health outcome target or category (e.g., liver effects), or assumes similarity based on heavily used structural grouping in publications (e.g., polycyclic aromatic hydrocarbons [PAHs]). However, the examples found in this literature review rarely took a quantifiable grouping approach or used multiple characteristics to evaluate the group as a class; instead they relied on a single common characteristic among them, with a few exceptions.

In our literature search, we found several examples of organophosphate pesticides (OPs) subject to class-based methods for dose-response analysis. OPs are good candidates because they are well studied, and many are empirically linked by their structure, mechanism (acetylcholinesterase inhibition), metabolism, and hazard (Brock et al., 2003; Chen et al., 2003; Li, 2018). Dioxins (Ring et al., 2023; Van den Berg et al., 2006), ethylene glycol alkyl ethers (Yamada et al., 2012), and PAHs (Li, 2021) have also been grouped together by a shared MOA. Dose-response for these groups is discussed in Sections 4.4.3 and 4.4.4.

In the absence of MOA data, early assessments supplemented structural approaches with other information on other endpoints, such as health effects or physicochemical properties. We found that prior research has used

chemical structure related to toxicological outcomes (Koleva et al., 2008; Yamada et al., 2012) as well as environmental transport and physicochemical properties (Rice et al., 2008) as an approach for grouping chemicals. Having a shared health effect outcome and/or metabolic mechanisms has been used for determining relative potency factors of polychlorinated biphenyls (PCBs) (Kodavanti, 2006), assessing three phosphate ester flame retardants to determine the total allowable concentration in drinking water (Berdasco & McCready, 2011), and linking α,β-unsaturated carbonyl compound structures to mutagenicity and acute aquatic toxicity (Koleva et al., 2008). Additionally, a strictly structural chemistry approach has been used to group chemicals based on quantitative structural similarity and/or physicochemical properties (Rice et al., 2008).

Based on our literature search, grouping chemicals by shared MOA for dose-response is highly desired. However, it is likely that the MOA of many/most of the OFR chemicals requiring assessment will not be known. Although varying degrees of MOA information may be available for data-rich members of a class, uncertainties exist regarding the degree to which such MOA information can be applied to data-poor target chemicals in a read across exercise. Given this challenge, other chemical characteristics need to be combined to group chemicals. The OECD, ECHA, EFSA, and RIFM read across guides emphasize that having multiple shared characteristics increases the grouping confidence, albeit with different priorities based on each institution's goals and priorities.

In order to create a multi-characteristic approach to grouping chemicals, ICF developed a process guide developed in CO-1, Call Order 61320622F2011, "Guidance Document for Conducting Class-based Hazard Assessment of Organohalogen Flame Retardants" under CPSC Blank Purchase Agreement (BPA 61320622A0005). This process involved evaluating structural fingerprints and then building five profiles for the chemicals: metabolite, physiologically based toxicokinetic (PBTK), physicochemical properties, mechanistic effects, and adverse effects. Not only is MOA considered within these profiles, but the profiles synergistically build confidence in chemical grouping and appropriately reduce confidence in outlier chemicals in the group. We implemented the guide in a case study on polyhalogenated organophosphate flame retardants (PHOPs). The list of PHOPs was based on the original list of National Academies of Sciences, Engineering, and Medicine (NASEM), as modified and enhanced by Bevington et al. (2022). Based on an analysis using the five profiles, five chemicals classified as PHOPs by Bevington et al. (2022) were determined to be outside the PHOP subclass based on their profiles. The remaining chemicals were considered within the same group for hazard assessment, with confidence varying from medium to very high despite the overall lack of data for many of the chemicals in the subclass. This approach can be considered in addition to the guides provided by OECD and other organizations described above when considering how to build confidence in a chemical group for class-based read across.

4.2. Identification of Surrogate/Anchor Chemicals

During class-based assessment, data-rich chemicals are used to extrapolate toxicity information for data-poor chemicals within that same chemical group. Data-rich chemicals are often termed "anchor chemicals," "index chemicals," or "surrogate chemicals" within this framework.

Selection of the anchor chemical depends on multiple factors such as:

1. availability of high-quality data for the relevant endpoint,
2. structural similarity between the anchor chemical and the rest of the chemicals,
3. similarity in physicochemical properties and metabolic pathways, and
4. similarity in toxicological effects between the chemicals within a group.

Ideally, a class-based dose-response analysis should be based on data from studies conducted using an exposure length and route of exposure consistent with the problem formulation, and from a common species and target organ for a group of chemicals in that class. Development of a class-based dose-response analysis for one route and study duration may lead to similar analyses for other, less data-rich combinations of route and study duration depending on the available data. It is worth nothing that risk assessment of chemicals as a group usually adds complexity to the evaluation methods and makes interpretation of the results difficult for individual compounds (Savitz & Hattersley, 2023). (Savitz & Hattersley, 2023)

4.3. Approaches Informed by Cumulative Risk Methods

As illustrated in Section 4.4, a common application of class-based approaches has been in the evaluation of the health impact of the combined exposure to multiple chemicals, also known as cumulative risk. The term cumulative risk can include risk from non-chemical stressors, but the focus in this document is on chemical stressors. ATSDR (2018) provides an extensive review of the approaches used by various agencies and organizations for assessing multiple chemicals, with most approaches falling into three main categories. The first involves evaluating the mixture as a mixture (e.g., gasoline). The other two approaches are component-based and are divided into assessments of toxicologically similar chemicals and toxicologically independent chemicals. If the components act independently and by different MOAs, risk is calculated by response addition. Toxicologically independent chemicals are not grouped together in a class, and so this approach will not be addressed further. A number of agency documents provide similar guidance regarding combined exposures (e.g., OECD (2018); RIVM (2015); U.S. Environmental Protection Agency (EPA) (2000, 2003)).

Dose addition is commonly used to assess chemicals that are toxicologically similar, and thus can be considered a type of class-based assessment. Dose addition assumes that "components of a mixture behave as concentrations or dilutions of one another, differing only in potency" (ATSDR, 2018). More pragmatically, dose addition is applied when the individual chemical components affect the same endpoint. A relative potency factor (RPF) expresses the potency of members of a related group of chemicals (e.g., a chemical class) relative to that of a data-rich index chemical. The index chemical is often the most toxic chemical in the class, but this is not a requirement.

In order to develop and apply RPFs, data from a single assay are needed for all of the chemicals of interest. That is, the chemicals need to be tested via the same exposure route and for the same duration, and the same endpoint is evaluated. Ideally, the same response level is used (e.g., ED10, LC50), although some earlier assessments compared no-observed-adverse-effect levels (NOAELs) as a less precise approach. For example, RPFs for organophosphate pesticides were based on a 10% reduction in brain acetylcholinesterase activity in rats in a 21-day oral study (U.S. EPA, 2006). US EPA has used the RPF approach for cumulative risk assessments for several other groups of mechanistically-related pesticides, including N-methyl carbamates, chloroacetanilides, and triazines.[2]

Toxicity equivalence factors (TEFs) are similar to RPFs, but with stricter similarity requirements. RPFs may be specific to an organ, route, or effect, while TEFs require similarity across all effects, organs, and routes. Few chemical groups meet the criteria for TEFs, but TEFs have been derived for dioxins and furans (Section 4.4.1) (Van den Berg et al., 2006).

In the context of a class-based assessment, the criteria for grouping chemicals as a class are comparable to those for grouping chemicals as toxicologically similar. The key difference is that a class-based assessment is likely to have less *in vivo* toxicity and MOA data, and so relies more on additional considerations for grouping (e.g., more systematic structural comparison, consideration of metabolites, and mechanistic data). In the case of the class-based assessment, the RPF approach is used as a quantitative read across approach, which may or may not be applied as part of a cumulative risk assessment. As for classical RPFs, the index chemical (or, in read-across parlance, the anchor chemical) is relatively data-rich. However, because there may not be a single assay covering all the data-poor chemicals in the class, it is theoretically possible that multiple anchor chemicals may be used in a class-based assessment. One chemical could be the anchor chemical for assay A, while a different chemical could be the anchor chemical for a different group of chemicals that were tested in related assay B but not assay A. In addition, most RPFs have been based on *in vivo* data (Section 4.4), but the class-based read across for the OFRs is likely to be based primarily on *in vitro* data, due to the minimal availability of *in vivo* data for these chemicals.

There are several assumptions related to using RPFs. A key one in the cumulative risk context is that additivity is generally assumed, in the absence of data on interactions. This assumption may or may not be relevant to class-based assessments of OFRs. Assuming additivity is a reasonable risk assessment approach for combined exposures

[2] US EPA cumulative assessment of risk from pesticides are located on this website: https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/cumulative-assessment-risk-pesticides.

in the range of or below the threshold for toxicity of individual components (ATSDR, 2018; U.S. EPA, 2000). A second key assumption is that the dose-response curves for the index (anchor) chemical are the same as those for similar chemicals. This assumption can be tested by comparing the dose-response curves for the endpoint of interest. Recognizing the uncertainties associated with these assumptions, the RPF approach is likely to be a key tool for conducting read across in a class-based analysis.

4.4. Examples of Class-Based Analyses

4.4.1. Phthalates

Several existing cumulative risk assessments on phthalates (CHAP, 2014; EFSA, 2019) and a proposed assessment approach (U.S. EPA, 2023) have assumed dose addition and used hazard indices with RPFs. For each, the endpoints of focus were the developing male reproductive system. This is not necessarily the most sensitive endpoint for each individual phthalate. For example, EFSA (2019) noted that diisononyl phthalate (DINP) affects the liver at lower doses than it affects fetal testosterone levels, but the shared effects on fetal testosterone warranted its inclusion. Similarly, the CHAP (2014) noted that DINP is a less potent anti-androgen than other phthalates but can still contribute to cumulative risk. Both organizations used *in vivo* data to calculate relative potency.

4.4.2. Per- and Polyfluoroalkyl Substances (PFAS)

EFSA (2020) established a group tolerable weekly intake (group-TWI) for four PFAS based on immune effects. In the absence of appropriate data to derive RPFs, all four were assumed to be of equal potency. RIVM (2018) has derived RPFs for PFAS based on comparing the benchmark dose[3] (BMD) levels for liver hypertrophy. Similar to the case for phthalates, liver hypertrophy was the critical effect for most chemicals and was shared with several more for which it was not the critical effect. Several studies from the same group (Reardon et al., 2021; Rowan-Carroll et al., 2021) used 3D liver spheroids to compare PFAS potency and observed that PFAS with longer chains tended to be more potent. However, to our knowledge, relative potency calculations based on *in vitro* data have not yet been used in a regulatory context for PFAS.

4.4.3. Dioxins

Efforts to establish TEFs for dioxins and dioxin-like compounds used various increments, including the 2005 WHO TEFs rounded to half order of magnitude increments on a log scale (Van den Berg et al., 2006). This paper expressed a preference for calculating relative potency based on *in vivo* data, using the ED_{50} if dose-response curves were parallel. However, it noted that *in vitro* studies could be used in light of the importance of the aryl hydrocarbon (Ah) receptor-mediated MOA, if the concentration-response curves for 2,3,7,8-TCDD and the congener of interest were parallel. The relative potency values were then converted into the incremental TEFs. A 2022 re-evaluation of the 2005 WHO TEFs with an updated database derived "best estimate TEFs" (BE-TEFs) using consensus-based weighting and Bayesian dose-response modeling (DeVito et al., 2024; Ring et al., 2023; Wikoff et al., 2023). This effort did not round TEFs to half-log (with exceptions for mono-ortho PCBs) and allowed for estimates of uncertainty. Compared to the 2005 TEFs, the 2022 TEFs are expected to lead to lower total toxic equivalents (DeVito et al., 2024).

4.4.4. Polycyclic Aromatic Hydrocarbons (PAHs)

Early RPFs for PAHs were established semi-quantitatively, on an order of magnitude basis (U.S. EPA, Economy Division, 1998; 1993). A more recent attempt (U.S. EPA, 2010) calculated average RPFs based on available data from carcinogenicity or cancer-related studies and assigned a relative confidence to each average RPF. An attempt to build on the EPA's (2010) efforts with an updated literature review identified three additional publications covering two PAHs with data suitable for dose-response modeling and calculating RPFs (Haber et al., 2022). In the absence of oral data meeting study criteria, these estimates were based on BMD modeling of cancer data from

[3] A benchmark dose is the dose corresponding to a defined change in an endpoint and is determined by fitting flexible dose-response models to the empirical data.

dermal exposure, even though the route of interest is likely oral. Despite these and other uncertainties, these estimates were the best available given available data.

4.4.5. Chemicals Acting on the Thyroid

EFSA (2020) has conducted a cumulative risk assessment using dose addition for pesticides with chronic effects on the thyroid, focused on two endpoints: hypothyroidism, and parafollicular cell (C-cell) hypertrophy/hyperplasia/neoplasia. This effort derived RPFs using NOAELs for 128 pesticides (hypothyroidism) and 17 pesticides (C-cell effects). EFSA (2020) recommended refining the assessment using BMD modeling instead of NOAELs.

4.4.6. Organohalogenated Flame Retardants (OFRs)

NASEM (2019) determined that OFRs should not be treated as a single class, based on analogs identified using information on function, structure, and predicted bioactivity. They instead suggested 14 subclasses, outlined a scoping plan, and illustrated some initial aspects of this plan for two of the suggested subclasses (one identified as relatively data-rich, and another as relatively data-poor). NASEM also identified four likely scenarios into which subclasses may fall: having many data-rich members, having no relevant data on any subclass member, having sufficient data on one or two chemicals but few/no data on all others, and mixed data availability but heterogenous and/or inconsistent data. The purpose of the NASEM report was to provide guidance on how to conduct a class-based assessment, rather than to complete an assessment for any subclass.

CPSC has since refined these 14 subclasses (Bevington et al., 2022), and ongoing work has further investigated data availability (conducted under the CPSC co-sponsored scoping review project with the National Institute of Environmental Health Sciences (NIEHS) Division of Translational Toxicology (DTT)) and continued evaluation of the subclasses (conducted under CO-1; Call Order 61320622F2011). Results of the subclass evaluation for one subclass (PHOPs) are discussed in Section 4.5.1 of this report.

4.5. Consideration of Data for OFR Subclasses

4.5.1. Polyhalogenated Organophosphates (PHOPs)

Under a previous call order (CO-1; Call Order 61320622F2011), data on 14 OFR classes were analyzed based on chemical structure. Further, PHOPs were evaluated under five basic data profiles: metabolites, physiologically based toxicokinetics, physicochemical properties, mechanistic effects, and adverse effects. These profiles were used to assess the class as a whole, to identify chemicals that likely do not belong in the class, and to rate each chemical on confidence for its inclusion in the class. Broadly, hepatic, urinary (renal), endocrine and related (potentially thyroid in particular), nervous system, carcinogenicity, and acute toxicity were most commonly investigated. While many of the data evaluated were indicative of data availability (i.e., a study addressing this endpoint or target exists) rather than identified effects, these targets are generally in line with the initial scoping report for PHOPs (which did consider identified effects). Such targets may be useful to narrow the focus of dose-response analyses of PHOPs. Based on these data, the PHOPs subclass was also refined to exclude several chemicals: four chemicals collectively referred to as the bromo-phenyl phosphates, and a triazine-containing chemical. Finally, four anchor chemicals for the PHOP class were identified: tris(1,3-dichloro-2-propyl) phosphate (TDCPP), tris(2-chloroisopropyl) phosphate (TCPP), Tris(2-chloroethyl) phosphate (TCEP), and Tris(2,3-dibromopropyl) phosphate (TDBPP).

CO-1 also identified metabolites of PHOPs. Metabolites that are particularly data-rich and/or shared among multiple PHOPs may also be fruitful sources of toxicity data. The availability of such toxicity data has been identified from database sources, but specific outcomes have not yet been identified. While analog chemicals were also investigated under CO-1, none had sufficient toxicity data to be useful.

A key starting point for data for NAMs-based dose-response analyses of PHOPs is the curated high-throughput screening (cHTS) data available from Integrated Chemical Environment (ICE). Other databases may also be useful, but ICE stands out because it is curated and includes dose-response information.

Any references identified during prioritization and screening as being relevant to flame retardants will be reviewed to verify their relevance to PHOPs and the presence of NAMs-based dose-response data. A preliminary examination of references prioritized at the full text level by GenAI (Section 3.3) suggests relatively few will have data on PHOPs at all (regardless of usefulness of the data).

As part of subtask 1B of the present call order, CPSC staff provided data (availability) extracted as part of the DTT systematic literature review. This included both traditional animal data (subtask 1B), but also data potentially relevant to subtask 1C (deriving TRVs from NAMs data): from *in vitro* experiments, non-traditional animals, and traditional animals exposed via routes not typically used for risk assessment (e.g., injection). Since this information is based on data availability, it will need to be further screened for usefulness to the present task. However, it has been identified and separated from the animal data. Any papers containing dose-response data on animals exposed by non-typical routes were flagged as part of subtask 1B.

For integrating traditional (subtask 1A and subtask 1B) and NAMs-based (subtask 1C) TRVs on a class basis (subtask 2B), it may be worth checking the ToxVal database (available via the EPA CompTox Dashboard) for additional traditional TRVs that were not identified as part of subtask 1A.

4.5.2. Polyhalogenated Bisphenol Aliphatics and Functionalized (PHBAFs)

Aside from some very broad analyses done for all OFRs based on structure, the PHBAFs have not yet been analyzed at the greater level of detail used for the PHOPs – at the five different types of data considered in CO-1.

Many of the same kinds of information that are useful for the PHOPs (Section 4.5.1) will also be useful for PHBAFs. This information will likely cover fewer PHBAF chemicals than PHOP chemicals and will likely provide extensive data on only a single PHBAF chemical (tetrabromobisphenol A, TBBPA). Thus, a class-based analysis for PHBAFs is likely to be more difficult (or at least substantially different) than for PHOPs.

Any references identified during prioritization and screening as being relevant to flame retardants will be reviewed to verify their relevance to PHBAFs and the presence of NAMs-based dose-response data. A preliminary examination of references prioritized at the full text level by GenAI (Section 3.3) suggests relatively few will have data on PHBAFs (regardless of usefulness of the data).

5. General Approach for Calculating TRVs from NAMs Data

5.1. Types of NAMs Data

NAMs refer to the methods for evaluating the health effects of chemicals that refine or replace the use of mammalian animal testing. In this report we will specifically focus on methods that focus on replacement of *in vivo* mammalian testing. A variety of different types of NAMs have been developed, including alternative animal models (e.g., zebrafish, *daphnia*), *in vitro* methods (e.g., cell lines, cell reporter assay systems), *in chemico* methods (e.g., cell-free systems that utilize biochemical reactions), and *in silico* methods (e.g., computational methods that model biological events).

Additionally, use of molecular assays such as Next Generation RNA sequencing, reporter gene assays, or measurement of signaling molecules means that the outputs of these systems are generally molecular in nature rather than apical endpoints. As such, these methods commonly rely on points of departure (PODs) defined in terms of changes at the molecular level. Specifically, they seek to utilize molecular events, such as receptor binding, increases in gene expression, or measurements of reactive oxygen species, rather than whole organ toxicity or markers thereof, as was traditionally utilized in *in vivo* systems. This has resulted in the development of a number of new strategies for derivation of benchmark responses (BMRs) around these new data types.

5.1.1. High-Throughput Data

The data produced from many of these types of assays are high-throughput data, meaning that the data sets are complex in nature. For example, this may be transcriptomic data on a handful of chemicals, or a reporter assay in

an i*n vitro* system for thousands of chemicals. Other types of NAMs include computational models to predict chemical effects or assess the similarity of a group of chemicals, or assays investigating specific cellular events related to a MOA, such as receptor binding or inhibition of hormone production (Natsch et al., 2013). Broadly, the types of data generated from all these approaches are molecular in nature (i.e., reflecting the way in which a chemical interacts with a specific cellular substructure) and often high dimensional (i.e., the number of features is larger than the number of samples in the experiment).

To process these types of data, a large degree of computational power is required, as is subject matter expertise. Advances in this area include the development of computational dashboards and publicly accessible data sources that provide greater access and easier analytical pipelines for investigators with less background. There are some simpler assays of interest including reporter gene assays (e.g., yeast reporter assay for xenoestrogens (Stork et al., 2008); comet assays and hormesis (Calabrese & Selby, 2023); yeast estrogen screens (YES) (Ramirez et al., 2014), and behavioral assays in non-mammalian species (Barreto et al., 2023; Staal et al., 2018; Teuschler et al., 2005). However, these still require computational power as they measure a large number of chemicals or collect data from a vast number of outputs. Additionally, these assays suffer from a lack of comparability between labs as they are dependent upon relative measures (Tangri et al., 2013). However, some types of high dimensional data, specifically data generated by Next Generation RNA sequencing, are more comparable as they utilize read counts, which are a more absolute measure and have highly developed computational methods that address comparability concerns (Liao et al., 2013).

5.1.2. High Dimensional Data

A subset of high-throughput NAMs data are high dimensional data. Here, we are using the term “high dimensional data” to refer to “-omics” measurements (e.g., genomics, proteomics, transcriptomics). This data stream is useful as it can provide a large amount of data from a limited number of experiments. Additionally, it detects key molecular endpoints that can indicate toxicity on a short time scale relative to apical endpoints. Furthermore, -omics data can be particularly useful for class-based assessments, since it is possible to test a group of related chemicals and evaluate a wide range of endpoints in a short period of time. However, these techniques require computationally intensive methods of analysis and require expert curation efforts to understand the quality of data produced. Examples of efforts that have successfully curated data are ToxCast, ICE, Gene Expression Omnibus (GEO), and the Comparative Toxicogenomic Database (CTD). These curation efforts can be leveraged for class-determination, as demonstrated in CO-1, as well as potentially being leveraged for dose-response modeling.

5.2. Case Studies of High-Throughput Data Use

Examples of the use of high-throughput data in peer reviewed literature include mechanistic mapping of chemicals (De Abrew et al., 2019) and integration of pathway-based transcriptomic data for quantitative risk assessment (Thomas et al., 2013). Importantly, this data has also been used to group chemicals by MOA using techniques such as connectivity mapping (De Abrew et al., 2016) which is of direct relevance to the concept of class-based modeling. In a related concept, transcriptomic data has been used in unique ways to help tier chemical screening based on likelihood of toxicity (Aggarwal et al., 2022). Additionally, Pitzer et al. (2023) developed a map of a network of Adverse Outcome Pathways (AOPs) related to developmental neurotoxicity, with identified molecular initiating events (MIEs), key events (KEs), and assays related to the KEs. This is of interest as a key step in using NAMs data is connecting specific assays to either the broad MOA, or to KEs in AOPs. Such data can be generated relatively easily from a variety of *in vitro* systems (Rowan-Carroll et al., 2021). Newer advances for these systems include 3D cell cultures to map the impact of contaminants on different cell types (Chang et al., 2021). In general, these efforts show that MOA and POD calculations are possible with these data types. Further, the data can be used to determine BMDs for various cellular endpoints and determine PODs (Rowan-Carroll et al., 2021).

In addition to POD derivation, several groups have used NAMs data together with *in vitro* to *in vivo* extrapolation (IVIVE) and high-throughput exposure estimation to calculate margins of exposure (MOE) on a class basis, although they did not explicitly apply an uncertainty factor (UF)-type approach for determining the minimum MOE. As described in greater detail below (Section 5.3.3), Clewell et al. (2020) used data on inhibition of testosterone production in a rat Leydig tumor line to conduct a class-based MOE evaluation for phthalates. Similarly, Leonard et

al. (2016) estimated the MOE for thyroid peroxidase inhibitors using high-throughput *in vitro* data, and Bury et al. (2021) used MOA data to identify analogs and relevant NAMs assays to conduct a MOE evaluation of caffeine. Together, the studies show the importance of biological consideration in data analysis. At the broader scale, these methods for connecting assays to MOA and AOPs are informing efforts towards validating NAMs for regulatory purposes (ICCVAM, 2024).

5.3. Derivation of BMRs and PODs with NAMs data

As the case studies above demonstrate, there is general agreement that connecting the data output to biological relevance is of key importance. However, the large amount of data resulting from high-throughput NAMs assays, such as transcriptomics assays, means that data processing to group or identify relevant changes is an important early step. Additionally, another key decision is to determine how to define the benchmark response (BMR), that is, the degree of change that determines the POD. The general consensus for apical endpoints is that the ideal approach for continuous endpoints is to determine the BMR based on the degree of change that is adverse (IPCS, Haber et al., 2018; 2020). At this time, this determination of adversity for NAMs data has not been made. Current proposed approaches to defining the BMR for apical endpoints in the absence of knowledge about the adverse degree of change include a change in the mean by one control standard deviation (SD), or a percentage change in the response (e.g., 10% change in gene expression). The appropriate approach to the determination of the BMR for NAMs has been the subject of substantial research, as described below and reviewed by Farmahin et al. (2017).

For NAMs that look at a single outcome (e.g., a reporter gene assay, receptor-binding assay), the NTP (2018) review on genomic dose-response modeling proposed the use of EPA's default for continuous endpoints – a change in the mean by one control SD. Unfortunately, for many measures, the response data are highly variable and, as such, have limited comparability between different assay runs. This means that defining the BMR solely in terms of the control SD would lead to markedly different responses at the BMR for different runs of the same assay. Normalization by controls leading to a percentage change has been used as a way to address this issue (Mpindi et al., 2015). However, this approach is also limited, because there is still a large amount of variability between the same assay conducted across labs.

Because sequencing data has well developed computational methods and can be verified by orthologous methods (i.e., performing qPCR to verify changes observed in sequencing data), variability is less problematic for analyzing transcriptomics data. As such we will focus on describing how the BMR has been defined for transcriptomic data. Rowan-Carroll et al. (2021) defined transcriptomics BMRs as a one control SD change in gene expression for differentially expressed genes while Yang et al. (2007) based their BMR calculation on the 95% lower confident limit on the dose causing a change in the mean of 1.349 control SD. Another approach was to define the BMR as a 10% change in gene expression (Chen et al., 2022). Reardon et al. (2023) (described in depth in Section 5.3.1) recommended the use of the one control SD change in their comprehensive approach for deriving PODs from NAMs data. This approach of basing a BMR for a continuous endpoint on the degree of change of the mean is similar to EFSA's approach for apical endpoints (EFSA, 2022), but EFSA uses a 5% change in the mean.

5.3.1. Approaches to Identification of the POD for NAMs Data

The majority of work on identifying PODs for NAMs data has been on transcriptomic data. As such, this section describes different approaches for calculating PODs for transcriptomic data. There are two major approaches for conducting dose-response analyses of transcriptomic data. The first is gene-set based, in which investigators filter for genes that are dose-responsive by a trend test at the gene or pathway level (NTP, 2018) or absolute change in gene expression (Rowan-Carroll et al., 2021). The second is pathway-based, in which genes are grouped by biological pathways and then the BMD is defined relative to the pathway. Farmahin et al. (2017) describes 11 different approaches for gene-set/pathway-based PODs. To do this, they reviewed published microarray data from studies on six chemicals in rats exposed orally for 5, 14, 28, and 90 days. Four of the studies evaluated the liver, while one study each evaluated the thyroid and bladder (one chemical per study). Generally, the authors found good agreement between apical endpoint BMDs and molecular BMDs. Additionally, they showed that approaches that utilize the top 20 genes (whether calculated by relative fold change, most enriched pathways, or greatest number of enriched pathways) had the greatest agreement with the apical response calculation. Additional work

has shown that when comparing *in vivo* exposures, transcriptomic BMDs (tBMDs) are concordant with BMDs for apical endpoints (LaRocca et al., 2020). With regards to time, Farmahin et al. (2017) found the greatest agreement between the calculated molecular PODs and the corresponding apical PODs at 28 days, followed by 5, 14, and finally 90 days.

In recent work addressing this issue, Reardon et al. (2023) released a best practices approach for calculating transcriptomic points of departure (tPODs) from *in vitro* studies. They used a data set built from 117 unique chemicals, three cell types, and durations ranging from 12 hours to 14 days. They evaluated seven approaches for calculating tPODs, all using a BMR of one control SD, and compared the results with PODs derived using traditional methods. They found that use of the most sensitive pathway-level gene sets (e.g., using KEGG, GO, or REACTOME grouping) resulted in the most sensitive calculation of the tPOD (Reardon et al., 2023). Their final recommendation is the use of the lowest consistent response dose (LCRD) method and one control SD of change for the BMR for tPOD derivation coupled with high-throughput toxicokinetics (httk) to perform IVIVE and derive a human equivalent dose. Transcriptional BMDs derived from *in vitro* data tended to be lower than the lowest apical BMD, but by less than a factor of 10 for most combinations of chemical, analysis method, and time points; all of the ratios greater than 10 were for one chemical.

5.3.2. In Vitro to In Vivo Extrapolation (IVIVE)

The goal of a pharmacokinetic model is to predict the concentration of a chemical within the body over time from a given external exposure. This is important in the determination of health effects because tissue deposition and metabolism are key drivers in likelihood of presentation of a health effect, and the actual dose-response is determined by the dose to the tissue. As such, absorption, distribution, metabolism, and excretion should be taken into account during POD determination. Specifically, extrapolation between internal doses and external doses is critical for IVIVE during high-throughput toxicity testing. Additionally, IVIVE is of importance when considering whether, for example, an oral administered dose is reflective of the tissue dose specific to observed health effects. More recent work has integrated differential gene expression analysis with traditional physiologically-based pharmacokinetic (PBPK) models to convert *in vitro* concentrations to human equivalent doses (HED) during IVIVE (Chen et al., 2022).

The specific parameter of interest when considering the application of pharmacokinetic modeling during IVIVE is concentration at steady state (C_{ss}) following exposure. At least one study that has included liver C_{ss} has shown that this is a key parameter for converting toxicogenomic data into a POD for use in risk assessment (Chen et al., 2022). Furthermore, C_{ss} in blood or liver can be calculated during IVIVE to determine oral equivalent doses for setting TRVs (Abdo et al., 2015a). Often this is calculable in ICE utilizing the package httk. The calculations in httk use two chemical-specific parameters: intrinsic clearance rate in liver cell culture (C_{int}) and unbound fraction of parent compound in the blood (fu_b). These parameters can be measured or estimated *in silico.* A number of standard physiological parameters are also part of the calculation (e.g., body weight, cardiac output [$Q_{cardiac}$], blood flow through the heart and lungs, glomerular filtration rate [GFR]). The output of httk is a C_{ss}, which can then be used within the same package and computational environment to perform IVIVE to move from an *in vitro* dose to a proposed *in vivo* administered dose.

5.3.3. Considerations for Class-Based POD Derivation

There has been limited use of NAMs in class-based dose-response assessments to date. However, some studies provide a helpful conceptual framework for how these data might be integrated. Hines et al. (2018) created a framework for integrating non-traditional animal model data (e.g., zebrafish data) with traditional rodent data. Such a framework is useful as it identifies PODs across a range of species and identifies species agnostic MIEs, KEs, and AOPs. Such ideas could be broadly applied to a class of chemicals (as opposed to individual chemicals) and should be considered when evaluating chemicals with similar MOAs. In addition, Clewell et al. (2020) applied the Hines et al. (2018) framework to inform selection of relevant *in vitro* assays for a cumulative risk assessment of two phthalates. An aggregate exposure assessment was conducted and a PBPK model was used to predict internal concentrations of metabolites. Based on the AOP, RPFs were calculated for the metabolites based on inhibition of testosterone production in a rat Leydig tumor line. This paper is noteworthy as one of the few identified that calculated RPFs based on *in vitro* data, and because the NAM was related to biological activity, rather than

transcriptional changes. More recently, when comparing NAMs-based PODs to traditional PODs across 448 studies, Paul Friedman et al. (2020) showed that chemotypes link chemicals for which NAMs-based PODs are greater than apical PODs. Specifically, the authors demonstrated that of the PODs derived from NAMs at the 95th percentile credible interval, 89% had a POD that was less than that defined by the apical endpoint. The remaining 11% of chemicals with NAM PODs that were greater than traditional PODs were enriched for chemical structural features associated with organophosphate and carbamate insecticides.

Similarly, when considering applications of PBPK approaches to class-based assessment, there are some additional considerations that should be taken. First, chemical characteristics play a key role in determining absorption and distribution around specific tissue compartments. The lipophilicity as well as structure of a compound can help to predict absorption, deposition, and potential metabolic enzymes that might be induced. These can be used to choose appropriate values for each of the required coefficients during modeling with httk. Second, chemicals that share a structure (i.e., chemotype) are likely to act by the same or similar MOA. For example, Paul Friedman et al. (2020) showed in a chemotype analysis that gene-expression signatures often share commonalities within a single chemical class (i.e., organophosphates). These findings suggest that chemical characteristics and chemical structure can help with extrapolation from data-rich to data-poor chemicals during IVIVE.

Taken together, these studies help lay the groundwork for how class-based dose-response assessment might look when using NAMs data: focusing on molecular PODs derived primarily from *in vitro* data and compared using a RPF approach.

5.4. Calculation of a Reference Dose

Once a human equivalent concentration is derived by IVIVE from the POD, a toxicity reference value (TRV) must be established next. To do this, uncertainty factors (UFs) are applied. Below we discuss UFs for NAMs data as well as data-derived approaches for addressing variability and uncertainty.

5.4.1. Use of Uncertainty Factors for NAMs Data

In contrast to the extensive literature on choice of BMR that was discussed in Section 5.3, the literature on choice of UFs for NAMs data is relatively sparse. Most of the studies discussed in Section 5.3 evaluated the choice of BMR solely based on the impact on the POD. Comparisons or benchmarking, when done, were generally relative to an estimated exposure. That is, authors typically evaluated the MOE between the POD and the estimated exposure (after appropriate IVIVE), as part of a general screening/prioritization exercise.

Publications that did consider the use of UFs in calculating reference doses (RfDs) from *in vitro* data have typically used the classical UF structure, modified for use with *in vitro* data (Dourson et al., 2021; Health Canada, 2021). Specifically, Health Canada (2021) considered three areas of uncertainty – human variability, deriving the bioactivity-based POD, and the use of cell-based assays. Dourson et al. (2021) used the five areas of uncertainty applied by the U.S. EPA (i.e., interspecies, intraspecies, LOAEL to NOAEL, subchronic to chronic, and database), but focused on the human variability UF.

As an alternative to developing a NAMs-based RfD, the bioactivity exposure ratio (BER) can be calculated based on the ratio of an administered equivalent dose (AED, calculated using IVIVE from *in vitro* data) to estimated human exposure, analogously to the calculation of the MOE. Health Canada (2021) used the classical UF framework (as applied by Health Canada) to consider in some depth what BER is sufficient. The analysis examined a group of 46 chemicals that had TRVs derived by traditional methods to compare the POD based on NAMs data with the POD based on the traditional approach. Endpoints evaluated included results for repeat dose, developmental, and reproductive toxicity studies. The PODs were calculated using ToxCast data, based on the 5th percentile from the distribution of AC_{50} values from active assay endpoints. The authors noted that their comparison was between data from human cells and *in vivo* rodent models, and so differences in the PODs could have a variety of sources.

Interspecies UF - Much of the *in vitro* data comes from human cells or cell lines, and so this UF would not be needed. No publications were found in the literature surveyed regarding how to consider this area of uncertainty with data from rodent cells containing key human genes, or with NAMs other than human cells (e.g., zebrafish). It has also been suggested that this UF may be replaced by a UF addressing the use of *in vitro* data (see below).

Intraspecies UF – This UF addresses human variability in kinetics and dynamics (response). As noted in Section 5.4.2, this variability can be characterized by large banks of human cells from a variety of donors, but in doing so it is important to ensure that the key determinants of variability have been identified and captured with the experimental setup. In addition, there is the potential for *in vivo* toxicokinetic variability beyond that captured in the *in vitro* system and IVIVE calculation.

LOAEL to NOAEL UF – Because the POD is based on BMD modeling, rather than a NOAEL/LOAEL approach, this UF would not be needed.

Bioactivity UF – Health Canada (2021) noted that using the ToxCast test battery results in uncertainty regarding whether all possible effects have been tested (i.e., incomplete biological space), and so this UF overlaps with EPA's database UF. Uncertainties in IVIVE were also grouped with this UF.

Subchronic to Chronic UF – *In vitro* assays generally involve acute exposures, often on the order of days, but in general no longer than 28 days, and so there is uncertainty regarding how well such assays model the impact of chronic exposure. However, the early changes measured in transcriptomic assays may manifest later as chronic toxicity. In addition, as noted in Section 5.3.1, Farmahin et al. (2017) found that transcriptional data from 28-day oral studies did a better job of predicting apical PODs than did comparable data from 90-day studies. This area of uncertainty was noted by Dourson et al. (2021) as an area for future research. Health Canada (2021) did not explicitly address this area of uncertainty but appears to address the need for this UF in its comparison of traditional and bioactivity-based PODs.

Cell-Based Limitations UF – Health Canada (2021) included an extensive discussion of these uncertainties, including lack of full metabolic competence, the use of cultured cell lines with altered morphology, growth rates and responses, and the use of monocultures.

***In Vitro* to *In Vivo* UF –** As part of a workshop on the use of NAMs to evaluate respiratory irritation (Haber et al., 2024), some participants suggested that this UF could replace the interspecies UF. An advantage of this approach, compared with the Health Canada (2021) approach, is the potential to break this area of uncertainty into toxicokinetic and toxicodynamic components, each of which can be analyzed and quantified. It was noted that the issue of incomplete biological space might be part of this UF or part of a database UF.

Health Canada (2021) applied their UF approach in determining the sufficiency of a calculated BER in a risk-based screening tool. Based on a comparison of traditional and bioactivity-based PODs, they proposed an UF of 3 for the bioactivity UF. A factor of 3 was proposed for the cell-based UF. This UF took into consideration the redundancies in endpoint evaluation in ToxCast and the existence of other assays in addition to cell-based assays in ToxCast, as well as the overlap in uncertainties between this UF and the other two UFs used in their analysis. A factor of 10 was used for inter-individual human variability, recognizing that some of this variability is accounted for in the httk model used for IVIVE. A total factor of 100 was thus used to determine the adequacy of the BER.

5.4.2. Data-Derived Approach

Rusyn and Chiu (2022) have noted the potential to characterize human variability using cells from large numbers of donors, often combined with Bayesian modeling. In order to use this approach, Rusyn and Chiu (2022) noted the importance of distinguishing individual dose-response (where the magnitude or severity of response increases with dose) from the population dose-response (in which the fraction of the population that responds increases with dose). For example, Abdo et al. (2015b) evaluated the cytotoxic response to 179 chemicals in more than 1,000 lymphoblastoid cell lines from the 1000 genomes project, with the goal of characterizing population variability. The authors found that the concentration causing a cytotoxic response in the most sensitive 1% was within about a factor of three of the cytotoxic concentration in the median individual for about half of the tested chemicals. For most chemicals, but not all, this ratio was within a factor of 10. Overall, they found that the *in vitro* toxicodynamic variability distribution was very similar to the *in vivo* distribution found by IPCS (2014) based on 34 chemicals.

Several issues need to be considered in applying the sort of analysis conducted by Abdo et al. (2015b) in a risk assessment context. Ensuring that the donor population sampled adequately captures the key determinants of toxicity is an important aspect of this approach. The 1000 genomes project is representative of nine geographically (international) and ancestry-diverse populations. The large numbers and diversity of origin make it more likely that

key aspects of diversity are captured, even in the absence of mechanistic data. The authors appropriately interpreted the variability (after adjusting for technical variability) as reflecting toxicodynamic variability. Consistent with the chemical-specific adjustment factor (CSAF) paradigm (IPCS, 2005), Abdo et al. (2015b) characterized toxicodynamic uncertainty based on the ratio between the concentration causing the median response and that for a low percentage at the sensitive end of the spectrum. Although the most sensitive 5% is often used, IPCS suggests that alternative percentiles such as 1% could be used. Use of the 5th percentile would have meant that a factor of three would cover a smaller percentage of the population. An advantage of the approach used by Abdo et al. (2015b) is that it allows for the identification of genetic variations associated with greater variability, and for screening chemicals that may have greater than default variability in population response. Finally, Abdo et al. (2015b) was a proof-of-concept analysis; application of this approach to characterize human variability using *in vitro* data would require analyses with cells from the target organ. To that end, Rusyn and colleagues have conducted numerous studies with induced pluripotent stem cell (iPSC)-derived cells. For example, Blanchette et al. (2019) used iPSC-derived myocytes to evaluate one aspect of cardiotoxicity and found that the results with a sample size of 27 was consistent with clinical data. These latter results suggest that meaningful variability data can be obtained without needing to use the full power of the 1000 genomes study, although it is not clear if the relevant situations have been defined.

Overall, these data-derived approaches show promise for characterizing human variability and are being used in screening assessments. However, such data are unlikely to be available for the OFRs or for subclasses of OFRs, in light of the absence of *in vitro* data for many of these chemicals. Nonetheless, characterizing human variability in response to an anchor chemical in a class could provide useful information for the anchor chemical(s) in a subclass.

6. Risk Assessment for Non-Threshold Endpoints

Dose-response for carcinogens is typically based on the mode(s) of carcinogenic action, which may include mutagenicity, mitogenesis, inhibition of cell death, cytotoxicity with reparative cell proliferation, immune suppression, or other MOAs (Davidsen et al., 2023; U.S. Environmental Protection Agency (EPA), 2005). Some studies have also grouped chemicals based on key characteristics that they exhibit, including genotoxicity; alteration of cell proliferation, cell death, or nutrient support; modulating receptor-mediated effects; inducing oxidative stress; inducing epigenetic alterations; inducing chronic inflammation; being electrophilic or metabolically activated; altering DNA repair or causing genomic instability; causing immortalization; or suppressing the immune system (Atwood et al., 2019).

RPFs and TEFs, described more fully in Section 4.3, may be utilized for dose-response assessment for well-defined classes of carcinogens that share a common MOA for the same endpoint, as EPA did for dioxin-like compounds and some carcinogenic PAHs. Characteristics that may inform the relative potency of the anchor chemical to other chemicals in the class include relative toxicological outcomes; relative metabolic rates; relative absorption rates; quantitative structure activity relationships (QSARs), particularly for mutagenic, electrophilic, or DNA-reactive chemicals; or receptor binding characteristics (U.S. EPA, 2005).

While a linear approach is used to model dose-response for mutagenic MOAs and as a general default for other MOAs, nonlinear approaches can be used when supported by sufficient evidence (U.S. EPA, 2005). Linear approaches may overestimate risk, and specific models, such as an equivalency iterative algorithm for lognormal functions, have been proposed to better estimate risk from carcinogen mixtures with additive effects and the same MOA (Li, 2021).

In a cumulative risk context, cancer risk is often calculated assuming independent action, that is, applying response addition. However, in a class-based context, relative potency can be used to relate cancer potency of data poor chemicals to that of an index chemical, as described in Section 4.4.3.

7. Discussion

In this report we describe methods that are integral to class-based assessments, how these methods are being applied to NAMs data, and potential ways in which these methods may be applied to OFR chemicals in the PHOP and PHBAF chemical subclasses. Overall, we found that a variety of organizations apply the same general principles to chemical grouping, although they differ in some of the specifics. Category approaches to read across are most similar to class-based dose-response and are based on similarities among multiple chemicals. Once classes are established, methods informed by cumulative risk assessment, such as RPFs, can be applied to make inferences between data-rich and data-poor chemicals. One or more data-rich chemical(s) in a class may be the anchor chemical or index chemical to which other chemicals are compared. While most of the literature on RPFs is based on *in vivo* data, these methods have also been applied in limited cases to NAMs with a focus on derivation of PODs from transcriptomic data.

The first major takeaway of this report is that class-based assessments are of greater value when chemical classification is based on both biological and chemical properties. This is illustrated by the results of CO-1 Call Order 61320622F2011, "Guidance Document for Conducting Class-based Hazard Assessment of Organohalogen Flame Retardants" under CPSC Blank Purchase Agreement (BPA 61320622A0005) in which ICF demonstrated that five PHOPs from the NASEM list were outside the initial PHOP subclass. Interestingly, four of these PHOPs were bromophenyl phosphates, and one was a triazine phosphate. It is likely that these structural differences from the other (chlorinated) PHOPs contributed to the observed differences from other class members. From these findings regarding class-based assessments and linking chemicals by class, we will move forward with the reclassified list of PHOPs and will consider MOAs when performing dose-response assessment.

Shared biology at the level of health outcomes, MOA, or, ideally, AOPs, are important in transitioning from qualitative groupings into classes to a class-based dose-response assessment. Specifically, consideration of shared biology was important for the case studies discussed in Sections 4.4 and 5 that illustrate the development of RPFs or TEFs for chemical groups or classes. Additionally, in Section 5.2 we show that the literature demonstrates that NAM-based PODs are generally consistent with traditional PODs based on apical endpoints. In addition to how chemicals are classified, this literature screen led to information about various new experimental approaches using the NAMs paradigm. This paradigm allows for the production of more chemical data across a shorter time period and more cheaply than conducting studies of apical endpoints *in vivo*. One major downside of these models is that some assays lack comparability between labs, chemicals, and cell lines. For example, assay comparability is limited for fluorescent and reporter gene assays as they are dependent upon relative rather than absolute measures. More work is needed in order to be able to compare the results of reporter gene assays, yeast2hybrid assays, and luciferase assays conducted in different laboratories. In contrast to these challenges, transcriptomic data provides a very robust data set. Not only is this one of the most abundant data streams currently available, it also benefits from a high level of comparability because it is dependent on read count as an absolute unit. While there are experimental conditions that can influence this metric, post-processing and statistical tools have been well developed to address these issues. As such, we propose moving forward with class-based dose-response for the OFR subclasses based upon transcriptomic data.

To this end, we present a number of methods by which transcriptomic data can be parsed during dose-response modeling. Overall, we propose utilizing the framework developed by Reardon et al. (2023). This approach utilizes pathway level gene sets to identify transcriptomic PODs based on the LCRD and a one control SD change for the BMR. As LCRD is based upon a consistent response in non-outlier biological features, it addresses a key need for biological plausibility. This basis in biological mechanism will be critical for comparing across a chemical class and will allow for the calculation of RPF values from the transcriptomic data and for extrapolating from data–rich class members to data-poor ones. It is possible that different anchor chemicals will be used for different data-poor chemicals, depending on which chemicals have transcriptional data for the same *in vitro* assays. It is important to note that there has been no consensus reached at this time in the literature as experts continue to debate the merits of these approaches for conducting the modeling and determining the BMR (Gant et al., 2023). Therefore, prior to making regulatory decisions, consultation with experts should be undertaken.

Unlike the approach for calculating transcriptomic-based PODs, there are relatively few publications on the application of UFs to calculate TRVs. Instead, most publications use a MOE type approach to prioritize chemicals or a screening approach. No consensus has been reached to date on best practices for determination of UFs from *in vitro* or other NAMS endpoints, although there is more clarity related to some of the UFs. Interspecies uncertainty is not an issue when using NAMs from human cell lines, and the use of a BMD removes the need for extrapolating from a LOAEL. Human variability can be addressed with a standard UF for the OFRs, although a data derived approach can be developed based on iPSCs. There is less clarity on how to address extrapolation to chronic duration, and whether a UF is needed to account for IVIVE and/or the gaps in typical *in vitro t*est batteries. The element of subtask 1C in which NAMs-based TRVs are compared with existing apical-derived TRVs identified in subtask 1A and derived in subtask 1B can help to supplement the Health Canada (2021) analysis in addressing these issues. An alternative approach would be to use NAMs data to calculate relative potency, but then apply the RPF to TRVs developed based on classical apical endpoints. This approach would take advantage of the NAMs data to address data gaps, without the added uncertainties of developing TRVs directly from NAMs data.

In summary, our overall proposal is to (1) use the re-aligned PHOPs class members to (2) identify transcriptomic data sets across various data sources to (3) perform an RPF dose-response analysis in support of CPSC's goal of class-based dose-response assessment.

Appendix A. Literature Search Strings and Strategy

A.1. Peer Database Searches

All peer database searches were executed on 12/04/2023.

A.1.1. Approaches for Class-Based Dose-Response (Subtask 2A)

A.1.1.1. PubMed

Search String:

("Dose-Response Relationship, Drug"[Mesh] OR Dose-Response*[tiab] OR Dose-group*[tiab] OR Dose-Range*[tiab])

AND

("chemical class" OR “Chemical classes” OR "Chemical Grouping" OR “Chemical Groupings” OR "Chemical Group" OR “Chemical Groups” OR “Surrogate chemical” OR “Surrogate Chemicals” OR “Read Across” OR “Chemical Mixture” OR “Chemical Mixtures” OR “Combined exposure” OR “combined exposures” OR “aggregate exposure” OR “Aggregate exposures” OR “Cumulative Exposure” OR “Cumulative Exposures” OR ("Class-based"[tiab] AND Chemical*))

Limits: None

A.1.1.2. Web of Science

Search String:

TS=("Dose-Response Relationship” OR Dose-Response* OR Dose-group* OR Dose-Range*)

AND

TS=("chemical class" OR “Chemical classes” OR "Chemical Grouping" OR “Chemical Groupings” OR "Chemical Group" OR “Chemical Groups” OR “Surrogate chemical” OR “Surrogate Chemicals” OR “Read Across” OR “Chemical Mixture” OR “Chemical Mixtures” OR “Combined exposure” OR “combined exposures” OR “aggregate exposure” OR “Aggregate exposures” OR “Cumulative Exposure” OR “Cumulative Exposures” OR ("Class-based" AND Chemical))

Limits: None

A.1.2. Approaches for Class-Based Risk Assessment (Subtask 3A)

A.1.2.1. PubMed

Search String:

("Risk Assessment/methods"[Mesh] OR "Risk Assessment/standards"[Mesh] OR Method*[ti] OR Approach*[ti])

AND

("chemical class" OR “Chemical classes” OR "Chemical Grouping" OR “Chemical Groupings” OR "Chemical Group" OR “Chemical Groups” OR “Surrogate chemical” OR “Surrogate Chemicals” OR “Read Across” OR “Chemical Mixture” OR “Chemical Mixtures” OR “Combined exposure” OR “combined exposures” OR “aggregate exposure” OR “Aggregate exposures” OR “Cumulative Exposure” OR “Cumulative Exposures” OR ("Class-based"[tiab] AND Chemical*))

Limits: None

A.1.2.2. Web of Science

Search String:

TS=("Risk Assessment" AND (Standard* OR Method*))

AND

TS=("chemical class" OR “Chemical classes” OR "Chemical Grouping" OR “Chemical Groupings” OR "Chemical Group" OR “Chemical Groups” OR “Surrogate chemical” OR “Surrogate Chemicals” OR “Read Across” OR “Chemical Mixture” OR “Chemical Mixtures” OR “Combined exposure” OR “combined exposures” OR “aggregate exposure” OR “Aggregate exposures” OR “Cumulative Exposure” OR “Cumulative Exposures” OR ("Class-based" AND Chemical*))

Limits: None

A.1.3. Approaches for Deriving a Toxicological Reference Value (TRV) using New Alternative Methods (NAMs) Data (Subtask 1C)

A.1.3.1. PubMed

Search String:

(“Toxicological Reference Value” OR "Toxicological Reference Values" OR "Toxicity Reference Value" OR “Toxicity Reference Values” OR “Reference Dose” OR “Reference Doses” OR “Reference Dosage” OR “Reference Concentration” OR “Reference Concentrations” OR “AccepTable Daily Concentration” OR “AccepTable Daily Concentrations” OR “AccepTable Daily Intake” OR “AccepTable Daily Intakes” OR “Minimal risk level” OR “Minimal risk levels” OR “Cancer potency estimate” OR “Cancer potency estimates” OR “Cancer Potency Estimation” OR “Cancer Potency Estimations” OR “slope factor” OR “slope factors” OR “relative hazard potency” OR “relative hazard potential” OR “margin of exposure” OR “margins of exposures” OR “threshold of toxicological concern” OR “Tolerable Daily Intake” OR “Tolerable Daily Intakes” OR “Inhalation Unit Risk” OR (TRVs AND Toxicol*[tiab]))

AND

("New Approach" OR “New Approaches” OR "Alternative method" OR "Alternative methods" OR "alternative methodology" OR "Novel Method" OR "novel methods" OR "novel methodology" OR "New method" OR "new methods" OR "new methodology" OR NAMs[ti] OR "Alternative data" OR "Novel data" OR "unique data" OR "alternate data" OR "in vitro data" OR "mechanistic"[tiab] OR "computational method" OR "computational methods" OR "computational methodology" OR “Read Across” OR transcriptomic* OR QIVIVE OR “Quantitative in vitro to in vivo extrapolation” OR IVIVE OR “in vitro to in vivo extrapolation”)

Limits: None

A.1.3.2. Web of Science

Search String:

TS=(“Toxicological Reference Value” OR "Toxicological Reference Values" OR "Toxicity Reference Value" OR “Toxicity Reference Values” OR “Reference Dose” OR “Reference Doses” OR “Reference Dosage” OR “Reference Concentration” OR “Reference Concentrations” OR “AccepTable Daily Concentration” OR “AccepTable Daily Concentrations” OR “AccepTable Daily Intake” OR “AccepTable Daily Intakes” OR “Minimal risk level” OR “Minimal risk levels” OR “Cancer potency estimate” OR “Cancer potency estimates” OR “Cancer Potency Estimation” OR “Cancer Potency Estimations” OR “slope factor” OR “slope factors” OR “relative hazard potency” OR “relative hazard potential” OR “margin of exposure” OR “margins of exposures” OR “threshold of toxicological concern” OR “Tolerable Daily Intake” OR “Tolerable Daily Intakes” OR “Inhalation Unit Risk” OR (TRVs AND Toxicol*))

AND

TS=("New Approach" OR “New Approaches” OR "Alternative method" OR "Alternative methods" OR "alternative methodology" OR "Novel Method" OR "novel methods" OR "novel methodology" OR "New method" OR "new methods" OR "new methodology" OR "Alternative data" OR "Novel data" OR "unique data" OR "alternate data" OR "in vitro data" OR "mechanistic" OR "computational method" OR "computational methods" OR "computational methodology" OR “Read Across” OR transcriptomic* OR QIVIVE OR “Quantitative in vitro to in vivo extrapolation” OR IVIVE OR “in vitro to in vivo extrapolation”)

Limits: None

A.2. Grey Literature Google Webscraping

All grey literature searches were executed on 12/04/2023. Table A-1 lists the grey literature sources that were searched using the Google webscraping tool.

Table A-1. Grey Literature Sources

Source	Type	URL Domain
US Environmental Protection Agency (EPA)	Federal, Regulatory	epa.gov
National Institute of Environmental Health Sciences (NIEHS)	Federal, Research	niehs.nih.gov
Centers for Disease Control and Prevention (CDC)	Federal, Regulatory	cdc.gov
National Technical Reports Library (NTRL)	Federal, Repository	ntrl.ntis.gov
National Academies of Sciences, Engineering, and Medicine (NASEM)	NGO, Research	nationalacademies.org
National Institute for Public Health and the Environment (RIVM)	International, Regulatory	rivm.nl
European Chemicals Agency (ECHA)	International, Regulatory	echa.europa.eu
European Food Safety Authority (EFSA)	International, Regulatory	efsa.europa.eu
Organisation for Economic Co-operation and Development (OECD)	International, Research	oecd.org
Health Canada (HC)	International, Regulatory	canada.ca
Japanese Ministry of Health	International, Regulatory	mhlw.go.jp

The below search strings were used. Due to strict character limits in Google searching, strings were edited from the peer database strings.

A.2.1.1. Approaches for Class-Based Dose-Response (Subtask 2A)

(Dose Response* OR Dose group* OR Dose Range*) AND ("chemical class" OR “Chemical classes” OR "Chemical Grouping" OR "Chemical Group" OR “Surrogate chemical” OR “Read Across” OR “Chemical Mixture” OR “Combined exposure” OR “combined exposures” OR “aggregate exposure” OR “Cumulative Exposure” OR ("Class-based" AND Chemical))

A.2.1.2. Approaches for Class-Based Risk Assessment (Subtask 3A)

("Risk Assessment" AND (Standard* OR Method*)) AND ("chemical class" OR “Chemical classes” OR "Chemical Grouping" OR "Chemical Group" OR “Surrogate chemical” OR “Read Across” OR “Chemical Mixture” OR “Combined exposure” OR “combined exposures” OR “aggregate exposure” OR “Cumulative Exposure” OR ("Class-based" AND Chemical))

A.2.1.3. Approaches for Deriving a Toxicological Reference Value (TRV) using New Alternative Methods (NAMs) Data (Subtask 1C)

("Toxicological Reference Value" OR "Reference Dose" OR "Daily Concentration" OR "Minimal Risk Level" OR "Cancer Potency" OR "Slope factor") AND ("New Approach" OR "Alternative Methods" OR "Alternative Data" OR "In vitro data" OR "mechanistic" OR "computational methods")

Appendix B. Prioritization of Literature

B.1. Prioritization Methods

B.1.1. Keyword Analysis Tool

ICF's Keyword Analysis Tool (KAT) searches for the occurrence of one or more keywords within either a title or abstract of a bibliographic reference. The tool generates an RIS or CSV file with a frequency array by keyword for each reference.

The KAT module searches RIS files for the occurrence of one or more keywords within either the title or abstract and provides summary statistics (e.g., keyword percent occurrence, number of documents with keyword, number of unique documents with keyword). These statistics can be used to iteratively refine literature searches, such as identifying terms that contribute to off-topic hits, or to help prioritize screening, such as eliminating or including studies with a given keyword.

Table B-1 provides the terms used to perform Keyword Analysis on the literature corpus.

Table B-1. Keyword Analysis Terms

Keyword Analysis Terms
risk assessment
read across
risk assessments
hazard assessment
nams
risk assessment methods
dose response model
relative potency factors
risk assessment approach
risk assessment approaches
relative potency factor
dose response models
class based
dose response modeling
ivive
risk assessment method
risk assessment framework
hazard models
in vitro in vivo extrapolation
surrogate chemical
exposure response model
toxicity equivalence factors
surrogate chemicals
exposure response models
risk method
dose response approaches
hazard modeling
anchor chemical
new approaches methodologies
risk assessment standard
risk methods

Keyword Analysis Terms
anchor chemicals
toxicity equivalence factor
dose response approach
exposure response approach
risk assessment standards
exposure response approaches
new approaches methodology

B.1.2. Unsupervised Clustering

Topic Extraction is an automated technique to cluster references based on keywords they have in common within their titles and abstracts. Topic Extraction requires no a priori knowledge and commonality among keywords is determined by the literature; keywords are not pre-populated. References are clustered into groups based on similarities of keywords, each reference is assigned to a single cluster and the shared keywords for each cluster are provided in a table. Subject matter experts review keywords and associated references and may assign priority levels to each cluster or look further into clusters of interest.

Table B-2 provides the clusters identified by Topic Extraction and their associated keywords.

Table B-2. Unsupervised Clustering: Topic Extraction

Cluster ID	Number of References	Keywords
1	133	['cas', 'sensitization', 'local respiratory', 'photoallergenicity', 'endpoints evaluated', 'local respiratory toxicity', 'respiratory toxicity', 'phototoxicity', 'repeated dose', 'phototoxicity photoallergenicity', 'skin sensitization', 'material', 'fragrance', 'endpoints', 'repeated', 'safety assessment', 'safety', 'local', 'respiratory', 'evaluated']
2	425	['cancer', 'cumulative', 'lung', 'cumulative exposure', 'lung cancer', 'risk', 'workers', '95', 'years', 'mortality', 'ci', 'response', 'dose', 'cohort', 'dose response', 'smoking', '95 ci', 'study', 'occupational', 'age']
3	725	['mixtures', 'mixture', 'risk', 'chemicals', 'assessment', 'chemical', 'risk assessment', 'effects', 'toxicity', 'chemical mixtures', 'health', 'data', 'approach', 'based', 'dose', 'environmental', 'human', 'action', 'individual', 'addition']
4	556	['assessment', 'data', 'risk', 'risk assessment', 'based', 'approach', 'models', 'chemicals', 'approaches', 'model', 'human', 'read', 'vitro', 'chemical', 'toxicity', 'used', 'using', 'health', 'use', 'vivo']
5	411	['activity', 'inhibitors', 'compounds', 'chemical', 'inhibition', 'acid', 'cells', 'dependent', 'cell', 'dose', 'receptor', 'classes', 'different', 'potent', 'induced', 'microm', '10', 'chemical classes', '50', 'binding']
6	326	['ms', 'method', 'mass', 'samples', 'spectrometry', 'chromatography', 'mass spectrometry', 'extraction', 'liquid', 'compounds', 'determination', 'analytical', 'analysis', 'liquid chromatography', 'detection', 'gc', 'sample', 'lc', 'ms ms', 'quantification']
7	608	['cells', 'cell', 'effects', 'rats', 'induced', 'combined', 'treatment', 'dna', 'dose', 'increased', 'expression', 'mg', 'kg', 'activity', 'combined exposure', 'study', 'mice', 'doses', 'combination', 'exposed']

B.1.3. Supervised Clustering

Supervised clustering requires a small set of known relevant studies (or "seed studies") to add to results when clustering. Clustering with seeds takes the guesswork out of prioritizing clusters and generates unbiased forecasts of retrieval accuracy if seeds are randomly selected.

Ensemble supervised clustering works by using six clustering algorithms. A document is tagged as relevant if any one of the six models finds it to be relevant. An ensemble score is generated by adding the result of the six models. A document with an ensemble score of six indicates the document was found to be relevant by all six models.

Table B-3 provides the results of supervised clustering.

Table B-3. Supervised Clustering

Cluster ID	Number of seeds in cluster	Unique references in cluster
0	1	2,404
1	1	793
2	1	458
3	1	219
4	2	189
5	1	208
6	27	400
Total	34	4,671

Appendix C. Generative Artificial Intelligence (GenAI) for Full Text Screening

C.1. Prompts Used for Full Text Screening

A set of 15 prompts were initially developed to categorize 10 seed studies according to topic. Subject matter experts graded these prompts for accuracy against the seed studies and removed 5 of the prompts for providing inaccurate responses. This left 10 prompts that were used to review the 225 peer reviewed papers identified as containing potential methods on class-based dose-response.

Table C-1 provides an overview of the initial 15 prompts that were used for testing. Table C-2 provides a list of the final 10 prompts that were used to screen the identified references. Table C-3 lists the various parameters used for the GenAI.

Table C-1. Draft Prompt Questions

Topic	Prompt Question
Framework	Is this a paper presenting a framework or method of dose-response for chemicals in a class-based manner (e.g., using common/related health effects)?
Applicable to OFR Chemicals	Does this paper present a method of dose-response directly applied to PHOPs or PHBAFs chemicals as part of a class?
Case Study	Is this a paper presenting a case study of class-based dose-response (e.g., using common/related health effects)?
Target for Toxic Effects	What target tissue or system does this paper assess? (if any)
Surrogate	Does this paper discuss methods or criteria for choosing a surrogate chemical for a class-based assessment?
BMR	Does this paper discuss methods or issues related to the choice of benchmark response?
AOP	Does this paper discuss methods or issues related to creating or using quantitative AOPs to inform for dose-response assessment?
MOE and BER	Does this paper discuss methods or issues related to class-based margins of exposure (including Bioactivity - Exposure Ratio - BER)?
PBPK TKTD Uncertainty	Does this paper discuss using a PBPK (TKTD) model to address class-based toxicokinetics or toxicodynamics?
IVIVE	Does this paper discuss methods or issues related to in vitro to in vivo extrapolation?
NAMs Uncertainty	Does this paper discuss methods or issues related to applying uncertainty factors based on the use of NAMs data in dose-response analyses?
Class-Based Uncertainty	Does this paper discuss methods or issues related to applying uncertainty factors based on the conduct of a class-based assessment (e.g., read across)?
Mixtures	Does this paper consider any of the topics (AOP, MOE, Uncertainty, framework, etc) listed above but simply for mixtures rather than (structural) chemical classes?
RPF TEF	Does the paper contain anything novel about the application of relative potency factors (RPFs) or toxicity equivalent factors (TEFs)?

Topic	Prompt Question
Other	Does this paper discuss methods for conducting class-based dose-response for another topic not listed above?

Table C-2. Final Prompt Questions

Topic	Prompt Question
Framework	Does this paper present a framework, method, decision tree, toolbox, or workflow for deriving toxicity reference values (TRV) for chemicals using in vitro data, New Approaches Methods (NAMs), toxicogenomics, early mechanistic endpoints, or nonmammalian data?
Applicable to OFR Chemicals	Does this paper present data on flame retardants?
Surrogate	Does this paper discuss methods or criteria for choosing a surrogate or anchor chemical for a class-based assessment or grouping or read-across?
BMR	Does this paper discuss methods or issues related to the establishment of a point of departure (POD), benchmark response (BMR), AC50 (50% activity concentration, or concentration at which 50% of maximum activity is observed), ACC (activity concentration at cutoff), or administered equivalent dose (AED)?
AOP	Does this paper discuss methods or issues related to filling in quantitative Adverse Outcome Pathways (AOPs) with New Approach Methods (NAMs) data (or otherwise using AOPs and NAMs data to inform deriving a toxicity reference value (TRV))?
IVIVE	Does this paper discuss methods or issues related to in vitro to in vivo extrapolation (IVIVE) or administered equivalent dose (AED)?
NAMs UF	Does this paper discuss methods or issues related to deriving uncertainty factors for New Approach Methods (NAMs) data (duration, use of NAMs data, etc.)?
NAMs RA	Does this paper discuss methods or issues related to applying New Approach Methods (NAMs) in risk assessment?
NAM Type	What types of New Approach Methods (NAMs) are used in this paper?
RPF TEF	Does the paper contain anything about the application of relative potency factors (RPFs) or toxicity equivalent factors (TEFs), or any other similar method of comparing potency of related chemicals?

Table C-3. Final Parameters Used for GenAI Full Text Screening

Parameter	Final Details
GenAI	Claude version 3.0
Preamble	Please act as a toxicologist. Wrap each response in an XML tag based on the prefix 'response' concatenated with the question number.
Post-amble	If yes, justify your response with quotes. If you don't know, say "I don't know".

Appendix D. References

Abdo, N., Wetmore, B. A., Chappell, G. A., Shea, D., Wright, F. A., & Rusyn, I. (2015a). In vitro screening for population variability in toxicity of pesticide-containing mixtures. *Environment international*, *85*, 147-155. https://doi.org/10.1016/j.envint.2015.09.012

Abdo, N., Xia, M., Brown, C. C., Kosyk, O., Huang, R., Sakamuru, S., Zhou, Y. H., Jack, J. R., Gallins, P., Xia, K., Li, Y., Chiu, W. A., Motsinger-Reif, A. A., Austin, C. P., Tice, R. R., Rusyn, I., & Wright, F. A. (2015b). Population-based in vitro hazard and concentration-response assessment of chemicals: The 1000 genomes high-throughput screening study. *Environmental health perspectives*, *123*(5), 458-466. https://doi.org/10.1289/ehp.1408775

Agency for Toxic Substances and Disease Registry (ATSDR). (2018). *Framework for assessing health impacts of multiple chemicals and other stressors*. https://www.atsdr.cdc.gov/interactionprofiles/ipga.html

Aggarwal, M., Chikwana, E., & Corvaro, M. (2022). A proposed NAM-based tiered phototoxicity testing and human risk assessment framework for agrochemicals. *Regulatory toxicology and pharmacology : RTP*, *135*, 105250. https://doi.org/10.1016/j.yrtph.2022.105250

Atwood, S. T., Lunn, R. M., Garner, S. C., & Jahnke, G. D. (2019). New perspectives for cancer hazard evaluation by the report on carcinogens: A case study using read-across methods in the evaluation of haloacetic acids found as water disinfection by-products. *Environmental health perspectives*, *127*(12), 125003. https://doi.org/10.1289/ehp5672

Barreto, A., Santos, J., Almeida, L., Tavares, V., Pinto, E., Celeiro, M., Garcia-Jares, C., & Maria, V. L. (2023). First approach to assess the effects of nanoplastics on the soil species Folsomia candida: A mixture design with bisphenol A and diphenhydramine. *NanoImpact*, *29*, 100450. https://doi.org/10.1016/j.impact.2023.100450

Berdasco, N. A. M., & McCready, D. (2011). Risk assessment and class-based evaluation of three phosphate esters. *Human and Ecological Risk Assessment*, *17*(2), 367-380, Article Pii 936270088. https://doi.org/10.1080/10807039.2011.552394

Bevington, C., Williams, A. J., Guider, C., Baker, N. C., Meyer, B., Babich, M. A., Robinson, S., Jones, A., & Phillips, K. A. (2022). Development of a flame retardant and an organohalogen flame retardant chemical inventory. *Scientific Data*, *9*(1), 295. https://doi.org/10.1038/s41597-022-01351-0

Blanchette, A. D., Grimm, F. A., Dalaijamts, C., Hsieh, N. H., Ferguson, K., Luo, Y. S., Rusyn, I., & Chiu, W. A. (2019). Thorough QT/QTc in a dish: An in vitro human model that accurately predicts clinical concentration-QTc relationships. *Clinical pharmacology and therapeutics*, *105*(5), 1175-1186. https://doi.org/10.1002/cpt.1259

Brock, W. J., Rodricks, J. V., Rulis, A., Dellarco, V. L., Gray, G. M., & Lane, R. W. (2003). Food safety: Risk assessment methodology and decision-making criteria. *International journal of toxicology*, *22*(6), 435-451. https://doi.org/10.1177/109158180302200605

Bury, D., Alexander-White, C., Clewell, H. J., 3rd, Cronin, M., Desprez, B., Detroyer, A., Efremenko, A., Firman, J., Hack, E., Hewitt, N. J., Kenna, G., Klaric, M., Lester, C., Mahony, C., Ouedraogo, G., Paini, A., & Schepky, A. (2021). New framework for a non-animal approach adequately assures the safety of cosmetic ingredients - A case study on caffeine. *Regulatory toxicology and pharmacology : RTP*, *123*, 104931. https://doi.org/10.1016/j.yrtph.2021.104931

Calabrese, E. J., & Selby, P. B. (2023). Comet assay and hormesis. *Environmental pollution (Barking, Essex : 1987)*, *341*, 122929. https://doi.org/10.1016/j.envpol.2023.122929

Cawley, M., Beardslee, R., Beverly, B., Hotchkiss, A., Kirrane, E., Sams, R., 2nd, Varghese, A., Wignall, J., & Cowden, J. (2020). Novel text analytics approach to identify relevant literature for human health risk assessments: A pilot study with health effects of in utero exposures. *Environment international*, *134*, 105228. https://doi.org/10.1016/j.envint.2019.105228

Chang, Y., Rager, J. E., & Tilton, S. C. (2021). Linking coregulated gene modules with polycyclic aromatic hydrocarbon-related cancer risk in the 3D human bronchial epithelium. *Chemical research in toxicology*, *34*(6), 1445-1455. https://doi.org/10.1021/acs.chemrestox.0c00333

Chen, J. J., Chen, Y. J., Teuschler, L. K., Rice, G., Hamernik, K., Protzel, A., & Kodell, R. L. (2003). Cumulative risk assessment for quantitative response data. *Environmetrics*, *14*(4), 339-353. https://doi.org/10.1002/env.587

Chen, Q., Chou, W. C., & Lin, Z. (2022). Integration of toxicogenomics and physiologically based pharmacokinetic modeling in human health risk assessment of perfluorooctane sulfonate. *Environmental science & technology*, *56*(6), 3623-3633. https://doi.org/10.1021/acs.est.1c06479

Chronic Hazard Advisory Panel on Phthalates and Phthalate Alternatives. (2014). *Report to the U.S. Consumer Product Safety Commission by the Chronic Hazard Advisory Panel on Phthalates and phthalate Alternatives*. https://www.cpsc.gov/s3fs-public/CHAP-REPORT-With-Appendices.pdf

Clewell, R. A., Leonard, J. A., Nicolas, C. I., Campbell, J. L., Yoon, M., Efremenko, A. Y., McMullen, P. D., Andersen, M. E., Clewell, H. J., 3rd, Phillips, K. A., & Tan, Y. M. (2020). Application of a combined aggregate exposure pathway and adverse outcome pathway (AEP-AOP) approach to inform a cumulative risk assessment: A case study with phthalates. *Toxicology in vitro : an international journal published in association with BIBRA*, *66*, 104855. https://doi.org/10.1016/j.tiv.2020.104855

Davidsen, J. M., Cohen, S. M., Eisenbrand, G., Fukushima, S., Gooderham, N. J., Guengerich, F. P., Hecht, S. S., Rietjens, I., Rosol, T. J., Harman, C. L., & Taylor, S. V. (2023). FEMA GRAS assessment of derivatives of basil, nutmeg, parsley, tarragon and related allylalkoxybenzene-containing natural flavor complexes. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*, *175*, 113646. https://doi.org/10.1016/j.fct.2023.113646

De Abrew, K. N., Kainkaryam, R. M., Shan, Y. K., Overmann, G. J., Settivari, R. S., Wang, X., Xu, J., Adams, R. L., Tiesman, J. P., Carney, E. W., Naciff, J. M., & Daston, G. P. (2016). Grouping 34 Chemicals Based on Mode of Action Using Connectivity Mapping. *Toxicological sciences : an official journal of the Society of Toxicology*, *151*(2), 447-461. https://doi.org/10.1093/toxsci/kfw058

De Abrew, K. N., Shan, Y. K., Wang, X., Krailler, J. M., Kainkaryam, R. M., Lester, C. C., Settivari, R. S., LeBaron, M. J., Naciff, J. M., & Daston, G. P. (2019). Use of connectivity mapping to support read across: A deeper dive using data from 186 chemicals, 19 cell lines and 2 case studies. *Toxicology*, *423*, 84-94. https://doi.org/10.1016/j.tox.2019.05.008

DeVito, M., Bokkers, B., van Duursen, M. B. M., van Ede, K., Feeley, M., Antunes Fernandes Gáspár, E., Haws, L., Kennedy, S., Peterson, R. E., Hoogenboom, R., Nohara, K., Petersen, K., Rider, C., Rose, M., Safe, S., Schrenk, D., Wheeler, M. W., Wikoff, D. S., Zhao, B., & van den Berg, M. (2024). The 2022 World Health Organization reevaluation of human and mammalian toxic equivalency factors for polychlorinated dioxins, dibenzofurans and biphenyls. *Regulatory Toxicology and Pharmacology*, *146*, 105525. https://doi.org/10.1016/j.yrtph.2023.105525

Dourson, M., Ewart, L., Fitzpatrick, S. C., Barros, S. B. M., Mahadevan, B., & Hayes, A. W. (2021). The future of uncertainty factors with In vitro studies using human cells. *Toxicological Sciences*, *186*(1), 12-17. https://doi.org/10.1093/toxsci/kfab134

Economy Division. (1998). *Selected non-heterocyclic polycyclic aromatic hydrocarbons - Environmental Health Criteria 202*. https://wedocs.unep.org/20.500.11822/29533

EFSA Panel on Contaminants in the Food Chain, Schrenk, D., Bignami, M., Bodin, L., Chipman, J. K., del Mazo, J., Grasl-Kraupp, B., Hogstrand, C., Hoogenboom, L., Leblanc, J.-C., Nebbia, C. S., Nielsen, E., Ntzani, E., Petersen, A., Sand, S., Vleminckx, C., Wallace, H., Barregård, L., Ceccatelli, S., . . . Schwerdtle, T. (2020). Risk to human health related to the presence of perfluoroalkyl substances in food. *Efsa Journal*, *18*(9), e06223. https://doi.org/10.2903/j.efsa.2020.6223

EFSA Panel on Food Contact Materials, E., Aids, P., Silano, V., Barat Baviera, J. M., Bolognesi, C., Chesson, A., Cocconcelli, P. S., Crebelli, R., Gott, D. M., Grob, K., Lampi, E., Mortensen, A., Rivière, G., Steffensen, I.-L., Tlustos, C., Van Loveren, H., Vernis, L., Zorn, H., Cravedi, J.-P., . . . Castle, L. (2019). Update of the risk assessment of di-butylphthalate (DBP), butyl-benzyl-phthalate (BBP), bis(2-ethylhexyl)phthalate (DEHP), di-isononylphthalate (DINP) and di-isodecylphthalate (DIDP) for use in food contact materials. *Efsa Journal*, *17*(12), e05838. https://doi.org/10.2903/j.efsa.2019.5838

European Chemicals Agency (ECHA). (2017). *Read-Across Assessment Framework (RAAF)*. European Chemicals Agency. https://doi.org/doi/10.2823/619212

European Food Safety Authority (ESFA). (2022). Guidance on the use of the benchmark dose approach in risk assessment. *Efsa Journal*. https://doi.org/10.2903/j.efsa.2022.7584

European Food Safety Authority (ESFA), Craig, P. S., Dujardin, B., Hart, A., Hernandez-Jerez, A. F., Hougaard Bennekou, S., Kneuer, C., Ossendorp, B., Pedersen, R., Wolterink, G., & Mohimont, L. (2020). Cumulative dietary risk characterisation of pesticides that have chronic effects on the thyroid. *Efsa Journal*, *18*(4), e06088. https://doi.org/10.2903/j.efsa.2020.6088

Farmahin, R., Williams, A., Kuo, B., Chepelev, N. L., Thomas, R. S., Barton-Maclaren, T. S., Curran, I. H., Nong, A., Wade, M. G., & Yauk, C. L. (2017). Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. *Archives of Toxicology*, *91*(5), 2045-2065. https://doi.org/10.1007/s00204-016-1886-5

Gant, T. W., Auerbach, S. S., Von Bergen, M., Bouhifd, M., Botham, P. A., Caiment, F., Currie, R. A., Harrill, J., Johnson, K., Li, D., Rouquie, D., van Ravenzwaay, B., Sistare, F., Tralau, T., Viant, M. R., van de Laan, J. W., & Yauk, C. (2023). Applying genomics in regulatory toxicology: a report of the ECETOC workshop on omics threshold on non-adversity. *Archives of Toxicology*, *97*(8), 2291-2302. https://doi.org/10.1007/s00204-023-03522-3

Haber, L., Bradley, M., Buerger, A., Behrsing, H., Burla, S., Clapp, P., Dotson, S., Fisher, C., Genco, K., Kruszewski, F., McCullough, S., Page, K., Patel, V., Pechacek, N., Roper, C., Sharma, M., & Jarabek, A. (2024). New Approach Methodologies (NAMs) for the in vitro assessment of cleaning products for respiratory irritation: Workshop report. Submitted.

Haber, L. T., Dourson, M. L., Allen, B. C., Hertzberg, R. C., Parker, A., Vincent, M. J., Maier, A., & Boobis, A. R. (2018). Benchmark dose (BMD) modeling: Current practice, issues, and challenges. *Critical reviews in toxicology*, *48*(5), 387-415. https://doi.org/10.1080/10408444.2018.1430121

Haber, L. T., Pecquet, A. M., Vincent, M. J., & White, L. M. (2022). The long goodbye: Finally moving on from the relative potency approach to a mixtures approach for polycyclic aromatic Hhdrocarbons (PAHs). *International Journal of Environmental Research and Public Health*, *19*(15), 9490. https://www.mdpi.com/1660-4601/19/15/9490

Health Canada. (2021). *Science approach document - Bioactivity exposure ratio: Application in priority setting and risk assessment*. https://www.canada.ca/en/environment-climate-change/services/evaluating-existing-substances/science-approach-document-bioactivity-exposure-ratio-application-priority-setting-risk-assessment.html

Hines, D. E., Edwards, S. W., Conolly, R. B., & Jarabek, A. M. (2018). A case study application of the aggregate exposure pathway (AEP) and adverse outcome pathway (AOP) frameworks to facilitate the integration of human health and ecological end points for cumulative risk assessment (CRA). *Environmental Science & Technology*, *52*(2), 839-849. https://doi.org/10.1021/acs.est.7b04940

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). (2024). *Validation, qualification, and regulatory acceptance of new approach methodologies*.

International Programme on Chemical Safety (IPCS). (2005). *Chemical-specific adjustment factors for interspecies differences and human variability: Guidance document for use of data in dose/concentration–response assessment*. http://apps.who.int/iris/bitstream/10665/43294/1/9

International Programme on Chemical Safety (IPCS). (2014). *Guidance document on evaluating and expressing uncertainty in hazard characterization* (Harmonization Project Document 11). http://www.who.int/ipcs/methods/harmonization/uncertainty_in_hazard_characterization.pdf?ua=1

International Programme on Chemical Safety (IPCS). (2020). Chapter 5: Dose–response assessment and derivation of health-based guidance values. In *EHC 240: Principles for Risk Assessment of Chemicals in Food* (Second ed.). https://cdn.who.int/media/docs/default-source/food-safety/publications/chapter5-dose-response.pdf?sf

Kodavanti, P. R. (2006). Neurotoxicity of persistent organic pollutants: Possible mode(s) of action and further considerations. *Dose-response : a publication of International Hormesis Society*, *3*(3), 273-305. https://doi.org/10.2203/dose-response.003.03.002

Koleva, Y. K., Madden, J. C., & Cronin, M. T. (2008). Formation of categories from structure-activity relationships to allow read-across for risk assessment: toxicity of alpha,beta-unsaturated carbonyl compounds. *Chemical research in toxicology*, *21*(12), 2300-2312. https://doi.org/10.1021/tx8002438

LaRocca, J., Costa, E., Sriram, S., Hannas, B. R., & Johnson, K. J. (2020). Short-term toxicogenomics as an alternative approach to chronic in vivo studies for derivation of points of departure: A case study in the rat with a

triazole fungicide. *Regulatory toxicology and pharmacology : RTP*, *113*, 104655. https://doi.org/10.1016/j.yrtph.2020.104655
Leonard, J. A., Tan, Y. M., Gilbert, M., Isaacs, K., & El-Masri, H. (2016). Estimating margin of exposure to thyroid peroxidase inhibitors using high-throughput in vitro data, high-throughput exposure modeling, and physiologically based pharmacokinetic/pharmacodynamic modeling. *Toxicological sciences : an official journal of the Society of Toxicology*, *151*(1), 57-70. https://doi.org/10.1093/toxsci/kfw022
Li, Z. (2018). Introducing relative potency quotient approach associated with probabilistic cumulative risk assessment to derive soil standards for pesticide mixtures. *Environmental pollution (Barking, Essex : 1987)*, *242*(Pt A), 198-208. https://doi.org/10.1016/j.envpol.2018.06.076
Li, Z. J. (2021). An equivalency iterative algorithm for cancer risk assessment of chemical mixtures with additive effects. *Chemosphere*, *263*, Article 128131. https://doi.org/10.1016/j.chemosphere.2020.128131
Liao, Y., Smyth, G. K., & Shi, W. (2013). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923-930. https://doi.org/10.1093/bioinformatics/btt656
Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071
More, S. J., Bampidis, V., Benford, D., Bragard, C., Hernandez-Jerez, A., Bennekou, S. H., Halldorsson, T. I., Koutsoumanis, K. P., Lambré, C., Machera, K., Naegeli, H., Nielsen, S. S., Schlatter, J. R., Schrenk, D., Silano, V., Turck, D., Younes, M., Benfenati, E., Crépet, A., . . . Comm, E. S. (2021). Guidance Document on Scientific criteria for grouping chemicals into assessment groups for human risk assessment of combined exposure to multiple chemicals. *Efsa Journal*, *19*(12), Article e07033. https://doi.org/10.2903/j.efsa.2021.7033
Mpindi, J. P., Swapnil, P., Dmitrii, B., Jani, S., Saeed, K., Wennerberg, K., Aittokallio, T., Östling, P., & Kallioniemi, O. (2015). Impact of normalization methods on high-throughput screening data with high hit rates and drug testing with dose-response data. *Bioinformatics (Oxford, England)*, *31*(23), 3815-3821. https://doi.org/10.1093/bioinformatics/btv455
National Academies of Sciences, E., & Medicine. (2019). *A Class Approach to Hazard Assessment of Organohalogen Flame Retardants*. The National Academies Press. https://doi.org/doi:10.17226/25412
National Toxicology Program (NTP). (2018). *NTP research report on National Toxicology Program approach to genomic dose-response modeling: Research Report 5*. National Toxicology Program. https://doi.org/10.22427/ntp-rr-5
Natsch, A., Ryan, C. A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F., & Kern, P. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *Journal of applied toxicology : JAT*, *33*(11), 1337-1352. https://doi.org/10.1002/jat.2868
OECD. (2017). *Guidance on grouping of chemicals, second edition*. https://doi.org/10.1787/9789264274679-en
OECD. (2018). *Considerations for assessing the risks of combined exposure to multiple chemicals, series on testing and assessment No. 296*. https://www.oecd.org/chemicalsafety/risk-assessment/considerations-for-assessing-the-risks-of-combined-exposure-to-multiple-chemicals.pdf
Paul Friedman, K., Gagne, M., Loo, L. H., Karamertzanis, P., Netzeva, T., Sobanski, T., Franzosa, J. A., Richard, A. M., Lougee, R. R., Gissi, A., Lee, J. J., Angrish, M., Dorne, J. L., Foster, S., Raffaele, K., Bahadori, T., Gwinn, M. R., Lambert, J., Whelan, M., . . . Thomas, R. S. (2020). Utility of in vitro bioactivity as a lower bound estimate of in vivo adverse effect levels and in risk-based prioritization. *Toxicological sciences : an official journal of the Society of Toxicology*, *173*(1), 202-225. https://doi.org/10.1093/toxsci/kfz201
Pitzer, E. M., Shafer, T. J., & Herr, D. W. (2023). Identification of neurotoxicology (NT)/developmental neurotoxicology (DNT) adverse outcome pathways and key event linkages with in vitro DNT screening assays. *Neurotoxicology*, *99*, 184-194. https://doi.org/10.1016/j.neuro.2023.10.007
Ramirez, T., Buechse, A., Dammann, M., Melching-Kollmuß, S., Woitkowiak, C., & van Ravenzwaay, B. (2014). Effect of estrogenic binary mixtures in the yeast estrogen screen (YES). *Regulatory toxicology and pharmacology : RTP*, *70*(1), 286-296. https://doi.org/10.1016/j.yrtph.2014.07.006
Reardon, A. J. F., Farmahin, R., Williams, A., Meier, M. J., Addicks, G. C., Yauk, C. L., Matteo, G., Atlas, E., Harrill, J., Everett, L. J., Shah, I., Judson, R., Ramaiahgari, S., Ferguson, S. S., & Barton-Maclaren, T. S. (2023). From vision toward best practices: Evaluating in vitro transcriptomic points of departure for application in risk

assessment using a uniform workflow [Original Research]. *Frontiers in toxicology*, *5*. https://doi.org/10.3389/ftox.2023.1194895

Reardon, A. J. F., Rowan-Carroll, A., Ferguson, S. S., Leingartner, K., Gagne, R., Kuo, B., Williams, A., Lorusso, L., Bourdon-Lacombe, J. A., Carrier, R., Moffat, I., Yauk, C. L., & Atlas, E. (2021). Potency ranking of per- and polyfluoroalkyl substances using high-throughput transcriptomic analysis of human liver spheroids. *Toxicological Sciences*, *184*(1), 154-169. https://doi.org/10.1093/toxsci/kfab102

Rice, G., MacDonell, M., Hertzberg, R. C., Teuschler, L., Picel, K., Butler, J., Chang, Y. S., & Hartmann, H. (2008). An approach for assessing human exposures to chemical mixtures in the environment. *Toxicology and applied pharmacology*, *233*(1), 126-136. https://doi.org/10.1016/j.taap.2008.05.004

Ring, C., Blanchette, A., Klaren, W. D., Fitch, S., Haws, L., Wheeler, M. W., DeVito, M., Walker, N., & Wikoff, D. (2023). A multi-tiered hierarchical Bayesian approach to derive toxic equivalency factors for dioxin-like compounds. *Regulatory Toxicology and Pharmacology*, *143*, 105464. https://doi.org/10.1016/j.yrtph.2023.105464

RIVM. (2015). *Risk assessment of substances in combined exposures (mixtures)* (RIVM Letter report 2015-0189). https://www.rivm.nl/bibliotheek/rapporten/2015-0189.pdf

RIVM. (2018). *Per- and polyfluoroalkyl substances (PFASs) in food contact material* (RIVM Letter report 2018-0181). https://www.rivm.nl/bibliotheek/rapporten/2018-0181.pdf

Rowan-Carroll, A., Reardon, A., Leingartner, K., Gagné, R., Williams, A., Meier, M. J., Kuo, B., Bourdon-Lacombe, J., Moffat, I., Carrier, R., Nong, A., Lorusso, L., Ferguson, S. S., Atlas, E., & Yauk, C. (2021). High-throughput transcriptomic analysis of human primary hepatocyte spheroids exposed to per- and polyfluoroalkyl substances as a platform for relative potency characterization. *Toxicological sciences : an official journal of the Society of Toxicology*, *181*(2), 199-214. https://doi.org/10.1093/toxsci/kfab039

Rusyn, I., & Chiu, W. A. (2022). Decision-making with new approach methodologies: Time to replace default uncertainty factors with data. *Toxicological sciences : an official journal of the Society of Toxicology*, *189*(1), 148-149. https://doi.org/10.1093/toxsci/kfac033

Savitz, D. A., & Hattersley, A. M. (2023). Evaluating chemical mixtures in epidemiological studies to inform regulatory decisions. *Environmental Health Perspectives*, *131*(4), 045001. https://doi.org/doi:10.1289/EHP11899

Staal, Y. C. M., Meijer, J., van der Kris, R. J. C., de Bruijn, A. C., Boersma, A. Y., Gremmer, E. R., Zwart, E. P., Beekhof, P. K., Slob, W., & van der Ven, L. T. M. (2018). Head skeleton malformations in zebrafish (Danio rerio) to assess adverse effects of mixtures of compounds. *Archives of toxicology*, *92*(12), 3549-3564. https://doi.org/10.1007/s00204-018-2320-y

Stork, L. G., Gennings, C., Carter, W. H., Teuschler, L. K., & Carney, E. W. (2008). Empirical evaluation of sufficient similarity in dose-response for environmental risk assessment of chemical mixtures. *Journal of Agricultural Biological and Environmental Statistics*, *13*(3), 313-333. https://doi.org/10.1198/108571108x336304

Tangri, S., Vall, H., Kaplan, D., Hoffman, B., Purvis, N., Porwit, A., Hunsberger, B., Shankey, T. V., & Group, o. b. o. I. I. W. (2013). Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS – part III – analytical issues. *Cytometry Part B: Clinical Cytometry*, *84*(5), 291-308. https://doi.org/10.1002/cyto.b.21106

Teuschler, L. K., Gennings, C., Hartley, W. R., Carter, H., Thiyagarajah, A., Schoeny, R., & Cubbison, C. (2005). The interaction effects of binary mixtures of benzene and toluene on the developing heart of medaka (Oryzias latipes). *Chemosphere*, *58*(9), 1283-1291. https://doi.org/10.1016/j.chemosphere.2004.09.075

Thomas, R. S., Himmelstein, M. W., Clewell, H. J., 3rd, Yang, Y., Healy, E., Black, M. B., & Andersen, M. E. (2013). Cross-species transcriptomic analysis of mouse and rat lung exposed to chloroprene. *Toxicological sciences : an official journal of the Society of Toxicology*, *131*(2), 629-640. https://doi.org/10.1093/toxsci/kfs314

U.S. Environmental Protection Agency (EPA). (2000). *Supplementary guidance for conducting health risk assessment of chemical mixtures* (EPA/630/R-00/002). https://cfpub.epa.gov/ncea/raf/chem_mix.htm

U.S. Environmental Protection Agency (EPA). (2003). *Framework for cumulative risk assessment* (EPA/630/P-02/001F).

U.S. Environmental Protection Agency (EPA). (2005). *Guidelines for carcinogen risk assessment* (EPA/630/P-03/001F). https://www.epa.gov/sites/default/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf

U.S. Environmental Protection Agency (EPA). (2006). *Organophosphorus cumulative risk assessment - 2006 update* (738R06041). https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P1010KM6.txt

U.S. Environmental Protection Agency (EPA). (2010). *Development of a relative potency factor (Rpf) approach for polycyclic aromatic hydrocarbon (PAH) mixtures (External Review Draft, Suspended)* (EPA/635/R-08/012A).

U.S. Environmental Protection Agency (EPA). (2023). *Draft proposed approach for cumulative risk assessment of high-priority phthalates and a manufacturer-requested phthalate under the Toxic Substances Control Act* (EPA-740-P-23-002). https://www.epa.gov/system/files/documents/2023-02/Draft%20Phthalate%20CRA%20Approach.pdf

U.S. Environmental Protection Agency (EPA), & Environmental Criteria Assessment Office. (1993). *Provisional guidance for quantitative risk assessment of polycyclic aromatic hydrocarbons*.

Van den Berg, M., Birnbaum, L. S., Denison, M., De Vito, M., Farland, W., Feeley, M., Fiedler, H., Hakansson, H., Hanberg, A., Haws, L., Rose, M., Safe, S., Schrenk, D., Tohyama, C., Tritscher, A., Tuomisto, J., Tysklind, M., Walker, N., & Peterson, R. E. (2006). The 2005 World Health Organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. *Toxicological Sciences*, *93*(2), 223-241. https://doi.org/10.1093/toxsci/kfl055

Varghese, A., Cawley, M., & Hong, T. (2018). Supervised clustering for automated document classification and prioritization: a case study using toxicological abstracts. *Environment Systems and Decisions*, *38*(3), 398-414. https://doi.org/10.1007/s10669-017-9670-5

Wikoff, D., Ring, C., DeVito, M., Walker, N., Birnbaum, L., & Haws, L. (2023). Development and application of a systematic and quantitative weighting framework to evaluate the quality and relevance of relative potency estimates for dioxin-like compounds (DLCs) for human health risk assessment. *Regulatory Toxicology and Pharmacology*, *145*, 105500. https://doi.org/10.1016/j.yrtph.2023.105500

Yamada, T., Tanaka, Y., Zhang, H., Hasegawa, R., Sakuratani, Y., Mekenyan, O., Yamazoe, Y., Yamada, J., & Hayashi, M. (2012). A category approach to predicting the hemolytic effects of ethylene glycol alkyl ethers in repeated-dose toxicity. *The Journal of toxicological sciences*, *37*(3), 503-515. https://doi.org/10.2131/jts.37.503

Yang, L., Allen, B. C., & Thomas, R. S. (2007). BMDExpress: A software tool for the benchmark dose analyses of genomic data. *BMC genomics*, *8*, 387. https://doi.org/10.1186/1471-2164-8-387