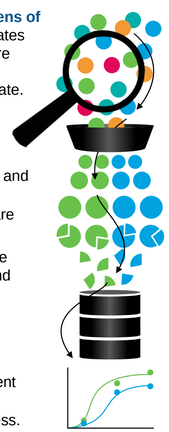
Visualizing the Evidence: Exploring and Explaining Your Data via Interactive Methods

Courtney Skuce, Alessandria Schumacher, George Agyeman-Badu, Pam Hartman, Kim Osborn | ICF

Data, Data, Everywhere…



Risk assessments begin with broad comprehensive literature searches which can produce **tens of thousands** of results that must be systematically screened and characterized. This also creates **large databases** of systematic review metadata (screening and tagging information, literature evaluation details) and relevant data extracted from the literature (exposure-response data, NOAELs and LOAELs, uncertainty ranges) that can be difficult and time-consuming to evaluate.

Data Visualization is ultimately one of the most useful methods risk assessors have in their toolbox to address these areas of need at every step of the systematic review process and in two contexts:

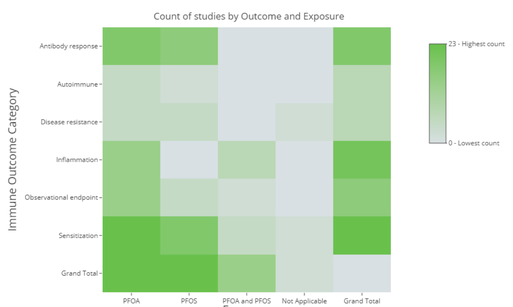
- **Exploring data** – Helping risk assessors explore complicated datasets to identify hazards and make decisions.
- **Explaining data** – Increasing the transparency and clarity with which data and analyses are presented to risk managers and the general public.

Evidence maps are one type of data visualization that can address these needs, and they are being used more often in public reports to visually represent systematic review databases and communicate characteristics of integrated evidence.

We present here:

1. How **adding an element of interactivity** to data visualizations can improve their ability to explore and explain your data;
2. **Examples of evidence maps** created using published assessment data[1] and three different tools and with varying degrees of interactivity, and
3. **Concepts to consider** while developing these visualizations to maximize their effectiveness.

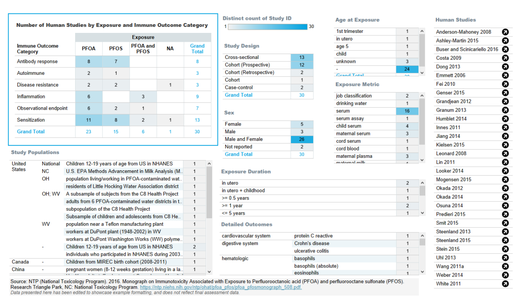
Example Evidence Maps



Created using R Shiny

Scan me!

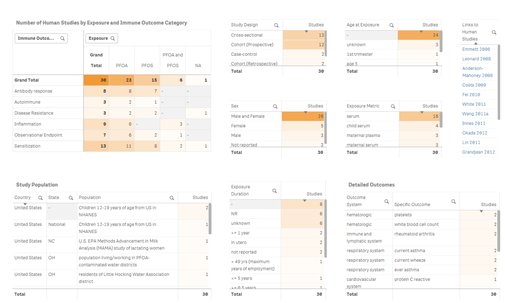
This is an interactive visualization. Use the sheets to see and interact with the 'raw data' as well as multiple visualizations and pivot tables.



Created using Tableau

Scan me!

This is an interactive visualization. Use the filters to 'drill down' into the details of the data, link to literature sources, and download the dataset.



Created using Qlik Sense[2]

Ask to see a demo of me!

Exploring and Explaining Your Evidence

All data visualizations ultimately serve two purposes: exploring, explaining, or both![3]

Exploratory Visualizations

You've collected so much data, now you need to know:

- What patterns are emerging?
- What story are the data telling?
- Do you have what you need to make actionable decisions?
- What conclusions can you make?

How are you currently exploring your evidence?

Adding **elements of interactivity** to your data visualizations can enhance your examination of your dataset by

- Amplifying your ability to 'drill down' into your data,
- Helping you easily view emerging patterns and compare groups, and
- Facilitating clear decision-making.

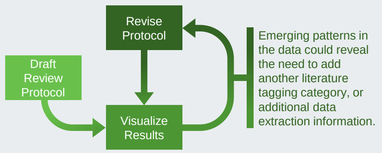
These evidence maps make it easy to discern:

Emerging patterns – Chemicals and/or outcomes with the most supporting studies

Data storytelling – Gaps in the literature

Possible conclusions – Number of studies examining exposures/outcomes under similar comparable conditions

Using visualizations like these throughout the literature characterization and data extraction processes can inform decisions to revise the review protocol:



Data visualization isn't often the first approach that comes to mind to explore risk assessment datasets, but it is a very effective tool for risk assessors to use to delve into their data and gain insights that allow them to better make decisions based on those data.

Explanatory Visualizations

You've characterized and integrated your evidence, now you need to support your findings:

- Will your audience understand your presentation of the data supporting your decisions?
- Does your audience have specific priorities to consider?
- Does your audience have meaningful access to your underlying data?

How are you currently explaining your evidence?

Leveraging **interactive visualizations** can help you reach your audience by

- Presenting complicated information in an intuitive and clear manner and
- Facilitating trust; the audience can interact with and discover the data for themselves to validate your conclusions.

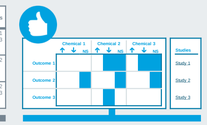
Considering your audience's needs and perspectives is crucial when crafting useful explanatory visualizations:

Expertise – Are you presenting data to risk managers? Or justifying decisions to the general public? How much does your audience know about your assessment? About exposure-response data?

Priorities – What's important to your audience? Are they interested in data supporting an exposure or outcome that could affect them? Will they have access to all the information important to them in your visualizations?

These evidence maps, compared to conventional summary tables, are an efficient means of presenting evidence because they provide audiences with meaningful access to the data used in the assessment, facilitating understanding and trust.

It is difficult to process data summarized in text and tables quickly. It's also more difficult to check the author's work.



Providing a visual makes the data easier to digest. Interactive features provide 'drill down' capabilities and immediate access to source studies.

Using interactive data visualizations tailored to your audience's priorities and expertise level is an excellent way to present findings and justify conclusions.

Contact: Courtney Skuce | courtney.skuce@icf.com

[1]NTP (National Toxicology Program). 2016. Monograph on Immunotoxicity Associated with Exposure to Perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS). Research Triangle Park, NC: National Toxicology Program. https://ntp.niehs.nih.gov/ntp/ohat/pfoa_pfos/pfoa_pfosmonograph_508.pdf. **Data presented here has been edited to showcase example formatting, and does not reflect final assessment data.**

[2]Funding for this work was partially provided by U.S. EPA under contract EP-C-14-001, work assignment 5-105. **The views in this poster are those of the authors and do not reflect the views or policies of the EPA.**

[3]Shander, B. (2014, June 5). Learning Data Visualization. [Video File]. Retrieved from Lynda.com.

SyRF: Systematic Review Facility

app.syrf.org.uk

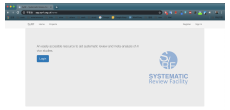
Jing Liao, Christopher Sena, Zsanett Bahor, Alexandra Bannach-Brown, Gillian Currie, Ezgi Tanriver Ayder, Qianying Wang, Emily Sena, Malcolm Macleod
CAMARADES, Clinical Center of Brain Science, University of Edinburgh

Systematic Review Completed in One Platform

SyRF web application (app.syrf.org.uk) facilities systematic review from search stage to meta analysis and graphing:

Systematic Review Customized for Your Project

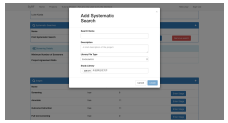
SyRF provides high flexibility and high elasticity in project design:



Free to use web application

Create your project private or public

Multiple users work on the same project at the same time



Directly retrieve data with search string or upload your own search

Automatically calculate agreement ratio : no need to reconcile separately

Design your annotation question with different types of answers

Group annotation information together for easy experiment design study

Tree Structure of annotation question design provides the maximum freedom and minimize the time answering unnecessary questions.

Systematic Review Powered by Text-Mining and Machine Learning

- References Classifiers built for your project
  - APIs built by NaCTeM and EPPI center are connected to SyRF
  - Sensitivity: 95%, specificity: 85%
  - Active learning with performance assessment in development

Automation of citation screening in pre-clinical systematic reviews, J. Liao, S. Ananiadou, L. G. Currie, B. E. Howard, A. Rice, S. E. Sena, J. Thomas, A.Varghese, M.R. Macleod bioRxiv 280131; doi: https://doi.org/10.1101/280131

- Risk of bias assessment for pre-clinical study
  - Blinding: accuracy 91% ~ 94%
  - Randomization: accuracy 67% ~ 86%
  - Sample Size Calculation: accuracy 96%~100%

Bahor Z, Liao J, Macleod MR, et al. Risk of bias reporting in the recent animal focal cerebral ischaemia literature. Clin Sci (Lond). 2017;131(20):2525–2532. Published 2017 Oct 13. doi:10.1042/CS20160722

- Drug and disease model extraction
  - Disease Model
  - Drug/Intervention extraction

Systematic Review Powered by Crowd Sourcing

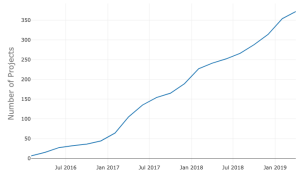
SyRF also has a suite enabling crowd sourcing systematic review:

Education website *syrf.org.uk* + Training web application *learn.syrf.org.uk* + System review web application *app.syrf.org.uk*

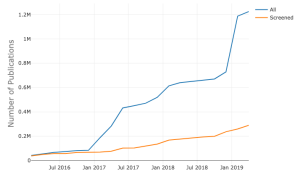
Systematic Review with Expert Assistance Available

SyRF is more than just an web application. SyRF is created by experts of systematic review and meta analysis CAMARADES in University of Edinburgh. Help is available through the format of web form, email, slack, weekly skype meeting, privately skype meeting by appointment.

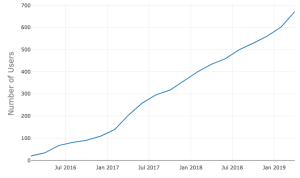
SyRF Usage Statistics



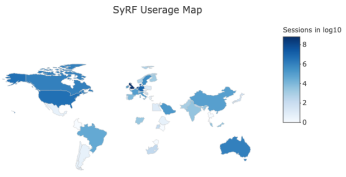
Over 350 Projects



Over 1.2 Million Publications



Over 650 registered users



Users from all over the world

Acknowledgement

SyRF was funded by NC3Rs. Thanks all members of CAMARADES for supporting the development of SyRF.

NC3Rs National Centre for the Replacement Refinement & Reduction of Animals in Research

Collaborative Approach to Meta Analysis and Review of Animal Data from Experimental Studies

Evaluating the consistency of heterogeneous results: important determinants of inconsistency

Barbara S. Glenn, Ph.D.; Elizabeth Radke, Ph.D.; and Andrew Kraft, Ph.D.

National Center for Environmental Assessment, U.S. Environmental Protection Agency, Washington, D.C., USA

Barbara Glenn | glenn.barbara@epa.gov | 202-564-5483

Background and Methods

Background/Aim

The analysis of study results across a set of studies is a powerful tool that can help with decisions about whether a potential bias is an important concern for an individual study, and to illuminate a pattern within apparently inconsistent effect estimates. The heterogeneity may stem from differing study designs examining varying outcome and exposure definitions and be influenced to varying degrees by sources of bias and other factors that affect the magnitude, direction, and precision of effect estimates. Influential aspects include potential bias (e.g., selection, information, confounding) and other quality aspects (e.g., sensitivity, precision). This type of analysis also can include factors, such as exposure levels, that are important for the interpretation of results. Studies of the association between indoor formaldehyde exposure and current asthma and pulmonary function were used as a case study to illustrate the impact of bias and other study attributes on the analysis of consistency across studies.

Methods:

Analyses of current asthma and pulmonary function endpoints were performed as part of a systematic evaluation of the literature database on studies examining the potential for respiratory and immune-mediated conditions in relation to formaldehyde exposure that was conducted through October 2016.

Criteria to evaluate risk of bias and sensitivity for the selected endpoints were developed using expert consultation or methodological reviews by professional organizations. The IRIS study evaluation tool included domains for participant selection, exposure, outcome, confounding, analysis, and sensitivity.

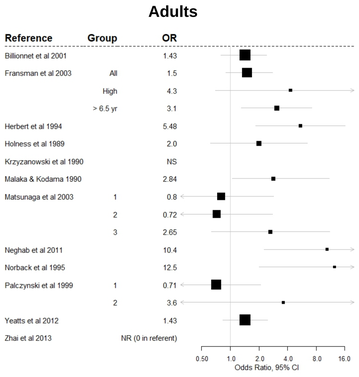
The consistency of results for current asthma was examined via forest plots presenting effect estimates (e.g., risk ratios, odds ratios) stratified by exposure levels (low vs high) and overall study confidence, and an analysis of potential confounding looking across study results was conducted for current asthma and pulmonary function endpoints.

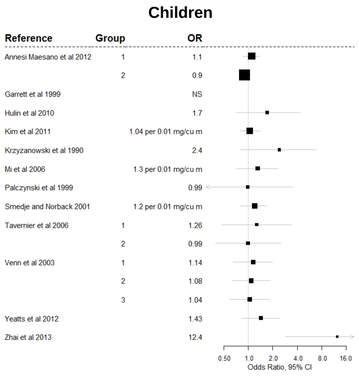
Current Asthma in Children and Adults

Current asthma is defined as a report of asthma symptoms during the last 12 months. Studies limited to "ever had asthma" were not included because the formaldehyde measures available did not reflect cumulative exposures that could be related to cumulative risk. The population relevant to the PECO for this analysis included children and adults in 22 studies of residential or school exposures and 5 occupational studies involving manufacture of pressed wood products, chemical production and embalming. These studies analyzed the variation in risk of prevalence of current asthma, incidence of asthma or asthma control or severity in relation to variation in formaldehyde at exposures above 0.010 mg/m^3 across a range spanning at least 0.01 mg/m^3. Of the 22 studies of residential or school settings, 4 were considered "not informative" for current asthma because the target population was under 5 years of age, an age range when asthma diagnoses are not specific. Three studies reported comparisons of mean formaldehyde concentrations in cases and controls and could not be included in the forest plots.

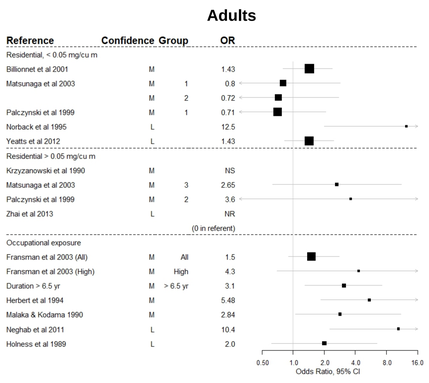
Sorting by Exposure Setting, Level, and Confidence for Studies of Current Asthma

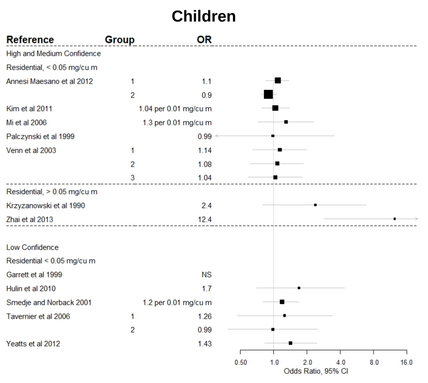
Studies ordered alphabetically





Studies ordered by exposure setting, level, and confidence





Conclusion

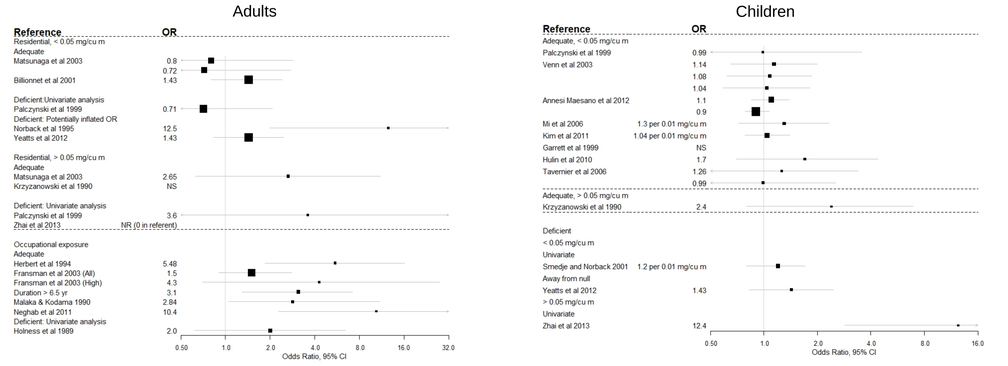
When studies are ordered alphabetically, results appear heterogenous, but when exposure levels and study confidence are considered, a pattern of increasing risk with increasing exposure levels is apparent among the high and medium confidence studies. No single domain limitation was a primary reason for the low confidence determinations, but collectively results of these studies are more variable.

Limitations of Low Confidence Studies

Adults	
Norback et al, 1995	Information bias: Most values < LOQ for formaldehyde Confounding: Unable to distinguish RR for VOCs with formaldehyde
Yeatts et al, 2012	Information bias: Analyses combined children and adults; mothers responded for children Confounding: Unable to distinguish RR for SO_2 with formaldehyde
Zhai et al, 2013	Selection bias: Participation rates not reported, but selection criteria were reported Information bias: Sampling period and protocol details not reported Confounding: Univariate, but magnitude of OR not likely explained by confounding Analysis: Small number of cases for analysis
Neghab et al, 2011	Selection bias: Lead time bias, Left truncation Information bias: Short formaldehyde sampling period; Asthma definition imprecise Confounding: Possible residual confounding for smoking
Holness et al, 1989	Selection bias: Lead time bias, Left truncation Information bias: Asthma definition imprecise Confounding: Univariate analysis
Children	
Smedje and Norback, 2001	Information bias: Exposure, uncertain concentration distribution, high proportion < LOD Confounding: No adjustment for coexposures, but results varied among exposures
Garrett et al, 1999	Selection bias: Potential household correlation of cases and controls Information bias: Asthma definition imprecise Analysis: Adjusted results reported as not signif
Hulin et al, 2010	Analysis: Small sample size; uncertain interpretation of urban/ rural stratified analyses
Tavernier et al, 2006	Selection bias: Missing data for 50% cases; not reported for controls Exposure: Distribution not reported Information bias: Asthma definition included questions not specific to asthma Analysis: Exposure levels by tertile not reported
Yeatts et al, 2012	Information bias: Analyses combined children and adults Confounding: Unable to distinguish RR for SO_2 with formaldehyde
Hsu et al, 2012	Not plotted. Selection bias: Low, differential participation rate Information bias: Short formaldehyde sampling period and protocol not reported Confounding: Univariate analysis Analysis: Limited
Hwang et al, 2011	Not plotted. Selection bias: High prevalence family history asthma in both groups Information bias: Asthma definition imprecise Analysis: Questions about analysis and distribution

Analysis of Confounding

Current Asthma: Sorting by Rating for Confounding



Conclusion

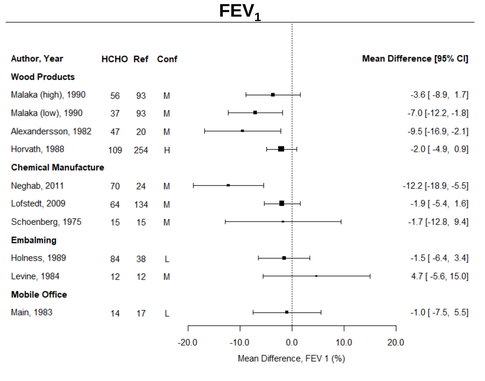
For studies in residential or school settings with lower exposure levels, a deficient rating for confounding with a predicted direction away from the null provides a potential explanation for some of the heterogeneity in odds ratios.

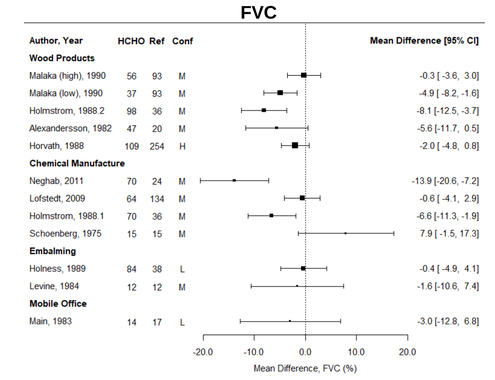
Occupational Studies of Pulmonary Function: Confounding

Forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC) were the most common measures analyzed by the studies of formaldehyde exposure. The population relevant to the PECO for this analysis included workers with occupational exposure to formaldehyde in studies where exposure was confirmed by air measurements, or involved professional categories of embalmers and anatomists/pathologists. A total of 21 studies involving manufacture of wood products, chemical production, embalming or offices in mobile trailers were identified, which analyzed variation in pulmonary function values in relation to variation in formaldehyde at exposures above 0.010 mg/m^3 across a range spanning at least 0.01 mg/m^3. Of the 21 studies, 5 were determined "not informative" because of one or more critical deficiencies; one reported additional analyses in the same cohort, and three presented longitudinal analyses or cross-shift changes, which are not shown in this example.

The occupational studies were limited by low sensitivity due to healthy worker survivor bias resulting from the cross-sectional analyses and loss of sensitive individuals before the studies began. Different analytic approaches were used making it challenging to examine results across studies in a graph using a single metric. However, most of the studies provided a mean value for exposed and referent groups as a percent of predicted adjusting for age, sex and height, and some expression of error. For these studies, forest plots were constructed using the mean difference of percent of predicted FEV_1 or FVC and confidence intervals. The mean difference is adjusted for smoking in these graphs only for three of the studies (Malaka et al., 1999; Holmstrom et al., 1988; Levine et al., 1984). However, most of the studies addressed smoking either in their designs or analyses, and the plotted results are in the same direction as the reported study results.

Cross-sectional analyses of pulmonary function endpoints in relation to occupational exposure





Notes on forest plots

- Difference in means of spirometry measurements taken before shift comparing exposed to unexposed. Means adjusted for age, height and sex.
- Study regression analyses also adjusted for smoking, and coefficients were in the same direction as those in forest plot.
- Two studies were not plotted because the studies reported only means of the unadjusted absolute values. Herbert et al. (1984) found a statistically significant decrease in FEV_1/FVC in analyses adjusted for age, height, sex and smoking, and Khamgaonkar et al. (1991) found a statistically significant decrease in FEV_1 and FVC in analyses adjusted for age, height, weight and sex.

Conclusion

Overall, mean values of pulmonary function among exposed workers were lower than those of unexposed comparison groups. The difference is not large, but is consistently observed in most of the studies, which were limited by a healthy worker survival bias, which may have attenuated the size of the observed difference. Concern for residual confounding is lessened because findings were consistent between the wood products and chemical manufacturing industries, which involve different coexposures.

U.S. Environmental Protection Agency
Office of Research and Development

#4 - Semi-Automated Data Extraction Workbench for Environmental Health

B. Howard, A. Maharana, A. Tandon, and *Ruchir Shah*
Sciome, LLC

Systematic review, already a cornerstone of evidence-based medicine, has recently gained significant popularity in several other disciplines including environmental health and evidence-based toxicology. One critical and time-consuming process that must occur during systematic review is the extraction of relevant qualitative and quantitative raw data from the free text of scientific documents. The specific data types extracted differ among disciplines, but within a given scientific domain, certain data points are extracted repeatedly for each review that is conducted. To that end, Sciome has begun research and development of a semi-automated data extraction workbench for use in this context. We are focusing our research on three specific goals. First, we are using deep learning to build novel data extraction models to extract data elements of interest to those conducting systematic reviews in the domain of environmental health. Second, we are building a web-based data extraction software platform specifically designed for usage in the domain of systematic review. And finally, we plan to introduce new protocols to standardize the inputs and outputs for data extraction software components. Here we report our results so far, including the performance of more than 20 novel data extraction components of relevance to environmental toxicology, created and tested on an annotated dataset from NTP. Performance varied widely among data types with some tasks inherently more difficult than others. For certain simple data types, like sex of the experimental animal, we achieved F-scores in excess of 95%; for more difficult entities, we were still often able to achieve an F-score of 65% or more, given sufficient training data. Because accurate data extraction can be a challenging problem, and given that current methods rarely achieve 100% accuracy, we are integrating our methods into a "human-in-the-loop" system that combines machine and human intelligence in a manner that is superior to using either in isolation. The system will: highlight extracted terms in a pdf; automatically populate extraction forms with extracted data; allow humans to intervene and correct the results; and learn from the corrections to continually update the model. The resulting system will make systematic reviews both more efficient to produce and less expensive to maintain, greatly accelerating the process by which scientific consensus is obtained in a variety of health-related disciplines having great public significance.

Evidence Synthesis and Integration in the IRIS Program

Xabier Arzuaga (based on Handbook materials developed by the IRIS Systematic Review Workgroup, particularly Barbara Glenn and Andrew Kraft)
National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency
Xabier Arzuaga | Arzuaga.Xabier@epa.gov

Introduction

Systematic reviews conducted as part of developing IRIS assessments (Figure 1) consist of structured processes for identifying the relevant evidence, evaluating individual studies, summarizing the relevant evidence (i.e., evidence synthesis), and arriving at summary conclusions regarding the overall body of evidence (i.e., evidence integration). These approaches were developed through discussions within EPA, and were informed by multiple reviews by the National Research Council (2011; 2014; 2018). In addition, IRIS assessments include quantitative toxicity values based on the evidence identified as most informative during the systematic reviews. The standard operating procedures, including frameworks and considerations for developing the different parts of the systematic reviews, are outlined in an internal document (IRIS Handbook; Figure 2).

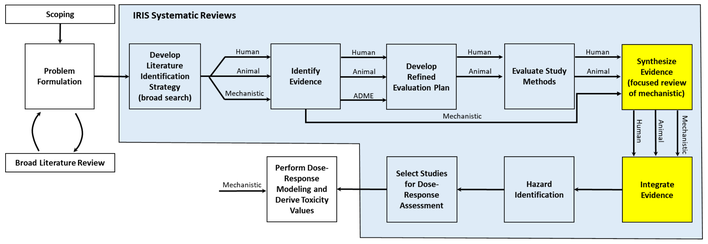


Figure 1. Systematic reviews in the IRIS Program: Figure adapted from the 2014 National Research Council review of the IRIS Program (adapted to show current workflows). Evidence synthesis and integration steps are highlighted.

Figure 2. IRIS Handbook: SOPs on approaches and considerations for applying principles of systematic review to IRIS assessments, including general frameworks, and examples. Evidence synthesis and integration steps are highlighted.

Overview of the Process

For each potential human health hazard, the evidence synthesis builds from the outcome-specific evaluations of individual studies, and discusses additional considerations across the sets of pertinent studies to summarize the available evidence in a manner that informs an evaluation of the body of evidence during evidence integration. Evidence integration is a two-step process based on structured, example-based frameworks for applying an adapted set of considerations described by Sir Bradford Hill (1965), first to each line of evidence, and then across all evidence. The general process is outlined in Figure 3.

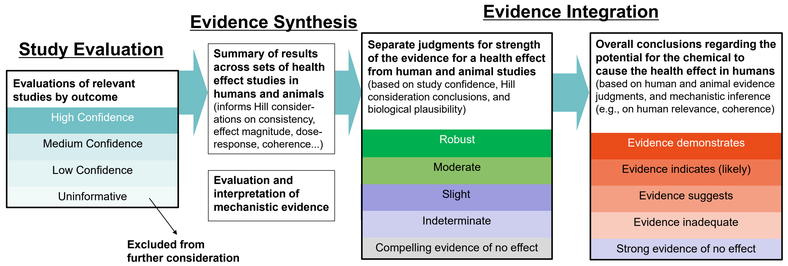


Figure 3. Outline of IRIS Evidence Synthesis and Integration. Human and animal evidence syntheses build from individual study evaluations and directly inform evidence integration across all lines of evidence.

Disclaimer: The views expressed are those of the authors and do not represent the views or policies of the U.S. EPA.

U.S. Environmental Protection Agency
Office of Research and Development

Evidence Synthesis

Summarize the information within each line of evidence (human, animal mechanistic), and analyze and present study results relevant to a given health effect to facilitate integration judgments.

- Narratives, not study summaries, focused on analyses that directly inform Hill considerations
- Human and animal health effect evidence is analyzed and synthesized separately. Mechanistic evidence is synthesized to inform the human and animal evidence conclusions (not shown).
- A primary goal of the evidence synthesis is to evaluate potential sources of heterogeneity across the study results (Figure 4), which informs evaluations of each Hill criterion.

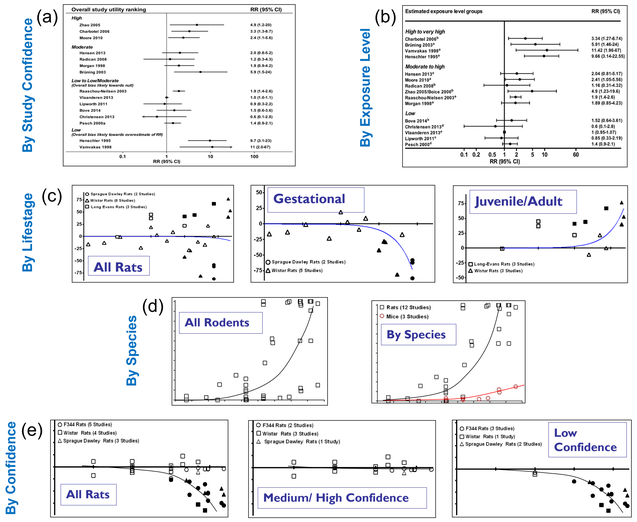


Figure 4. Evaluating Study Heterogeneity During Evidence Synthesis: (a) RoC Monograph on Trichloroethylene (2015); (b) EPA Toxicological Review of Trichloroethylene (2011); (c-e) "Edited" data from examples in draft IRIS assessments on hormones (c), pathology (d), and behavior (e).

Transitioning from Synthesis to Integration

The results of the analyses conducted during evidence synthesis inform an evaluation of each Hill consideration (Table 1) for the human and animal evidence relevant to a given health effect.

	Human Evidence Stream	Animal Evidence Stream
Individual Studies	• *High* or *medium* confidence studies provide stronger evidence within evaluations of each Hill consideration • Interpreting results considers biological as well as statistical significance, and findings across studies	
Consistency	• Different studies or populations increase strength	• Different studies, species, or labs increase strength
	• Analyze across study confidence, sensitivity, exposure levels/duration, lifestage, species or other factors • Unexplained inconsistency decreases evidence strength	
Dose-response	• Simple or complex (nonlinear) relationships within or across studies provide stronger evidence • Dose-dependence that is expected, but missing, can weaken evidence (after considering the findings in the context of other available studies and biological understanding)	
Magnitude, Precision	• Large or severe effects can increase strength; further consider imprecise findings (e.g., across studies) • Small changes don't necessarily reduce evidence strength (consider variability, historical data, and bias)	
Coherence	• Biologically related findings within an organ system, within or across studies, or across populations (e.g., sex) increases evidence strength (considering the temporal- and dose-dependence of the relationship) • An observed lack of expected changes (e.g., based on biological linkage) reduces evidence strength	
Mechanistic Evidence on Biological Plausibility	• Informed by mechanistic evidence on the biological development of the health effect or toxicokinetic/ dynamic knowledge of the chemical or related chemicals • Mechanistic evidence in humans or animals of precursors or biomarkers of health effects, or of changes in established biological pathways or a theoretical mode-of-action, can strengthen evidence • Lack of mechanistic understanding does not weaken evidence outright, but it can if well-conducted experiments exist and demonstrate that effects are unlikely	

Table 1. Factors that increase or decrease the strength of the human and animal evidence for a health effect. Expert judgments are organized using adapted Hill considerations (not shown are temporality- addressed during epidemiology study evaluation, and natural experiments- very rare that is important to highlight).

Evidence Integration

Develop summary judgments of the evidence relevant to a human health effect within the evidence integration narrative

- A two-step process (Figure 5) involving transparent and structured approaches for drawing summary conclusions (examples in Figure 6) across all lines of evidence.
- Evidence profile tables (Figure 7) document the primary decisions and rationales.

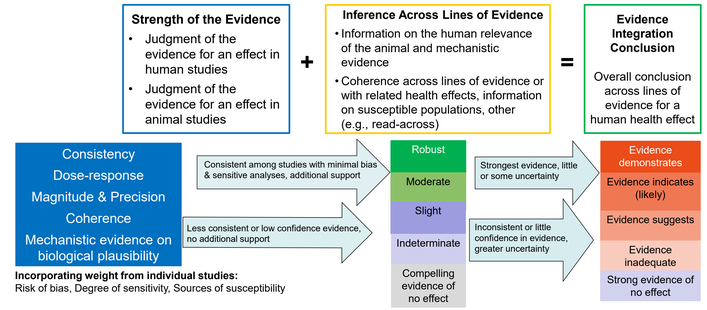


Figure 5. Evidence Integration Decision Process and Explanations

	Strength of the Evidence Judgments, Made Separately for Human and Animal Evidence
Robust	A set of consistent *high* or *medium* confidence, independent experiments reasonably ruling out alternative explanations; any conflicting set of studies is weaker. Additional criteria must also be met: **Human evidence**: Observed across populations, with clear dose-response evidence **Animal evidence**: Observed across labs or species, with multiple lines of additional support (e.g., pronounced severity or frequency; clear dose-response; coherence; a well-supported MOA).
	Overall Evidence Integration
Evidence Demonstrates	A very high level of certainty that exposure causes the health effect in humans: • The strongest evidence judgment (***robust***) for the human evidence stream • A moderately strong human evidence judgment (***moderate***) and the strongest animal evidence judgment (*robust*) alongside strong mechanistic evidence that MOAs and key precursors in animals are anticipated to occur in humans

Figure 6. Examples of Criteria for Evidence Integration Judgments (i.e., strongest judgments)

Step 1 – Evidence Integration of Human or Animal Evidence | **Step 2 – Overall Integration**

Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings	Within stream evidence judgments	Inference across evidence streams	Overall conclusion
[Health Effect or Outcome Grouping]						
Evidence from Human Studies (Route) • References • Study confidence • Study design description	• Consistency • Dose-response gradient • Coherence of observed effects • Effect size • Mechanistic evidence providing plausibility • Medium or high confidence studies	• Unexplained inconsistency • Imprecision • Low confidence studies • Evidence demonstrating implausibility	• Results across studies • Human mechanistic evidence informing biological plausibility	Describe strength of the evidence from human studies, and primary basis: ⊕⊕⊕ Robust ⊕⊕◯ Moderate ⊕◯◯ Slight ◯◯◯ Indeterminate – – – Compelling evidence of no effect	• Human relevance of findings in animals • Cross-stream coherence • Other inferences: ○ Information on susceptibility ○ MOA analysis inferences ○ Relevant information from other sources (e.g., read across)	Describe conclusion(s) for the integration of all available evidence: ⊕⊕⊕ Evidence demonstrates ⊕⊕◯ Evidence indicates ⊕◯◯ Evidence suggests ◯◯◯ Evidence inadequate – – – Strong evidence supports no effect Summarize the models and range of dose levels upon which the conclusions were primarily reliant
Evidence for an Effect in Animals (Route) • References • Study confidence • Study design description	• Consistency and/or Replication • Dose-response gradient • Coherence of observed effects • Effect size Mechanistic evidence providing plausibility • Medium or high confidence studies	• Unexplained inconsistency • Imprecision • Low confidence studies • Evidence demonstrating implausibility	• Results across studies • Animal mechanistic evidence informing biological plausibility	Describe strength of the evidence for an effect in animals, and primary basis: ⊕⊕⊕ Robust ⊕⊕◯ Moderate ⊕◯◯ Slight ◯◯◯ Indeterminate – – – Compelling evidence of no effect		

Figure 7. Evidence Profile Table (Template): Documents the story of the evidence and supporting rationale for evidence integration decisions (note: may be subdivided, e.g., by study design)

Transitioning from Integration to Dose-Response

Evidence integration directly informs study selection and toxicity value derivation (Figure 8).

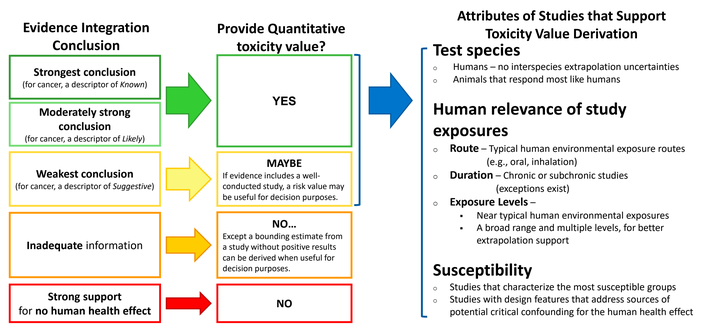


Figure 8. Considerations for Dose-Response: Note: study confidence informs study selection (not shown).

#6 - Research Update: Using SWIFT-Active Screener to Reduce the Expense of Evidence Based Toxicology

B. Howard, A. Tandon, J. Phillips, A. Maharana, and *Ruchir Shah*
Sciome LLC

Systematic review is a formal process used widely in evidence-based toxicology and environmental health research to identify, assess, and integrate the primary scientific literature with the goal of answering a specific, targeted question in pursuit of the current scientific consensus. We recently received Phase I SBIR funding to conduct research and development to enhance our web-based, collaborative systematic review software application, SWIFT-Active Screener. By employing a machine learning methodology called "Active Learning", and through a novel statistical method that can accurately estimate the percentage of relevant studies screened, Active Screener can significantly reduce the overall screening burden compared to traditional approaches. We first investigated several improvements to our statistical algorithms used for article prioritization and recall estimation (Aim 1 – Improved Statistical Models). The resulting refinements further improve the performance of our algorithms and address critical technical issues that previously limited the applicability of our methods. Secondly, we explored ways in which our models and methods can be improved to handle the scenario in which an existing systematic review is updated with new data several years after its initial publication (Aim 2 – New Methods for Systematic Review Updates). Finally, in order to ensure that our software is capable of supporting the full demand from our many users, we have reengineered the system to support hundreds to thousands of simultaneous screeners (Aim 3 - Software Engineering for Scalability, Usability). During this research, our methods and software have been rigorously tested on 26 different systematic review datasets, demonstrating robust performance of Active Screener's prioritization and recall estimation methods in a variety of real-world scenarios. For reviews with 5,000 or more documents, we report an average reduction in screening burden of 61% (to obtain 95% recall). Active Screener has been used successfully to reduce the effort required to screen articles for systematic reviews conducted at a variety of organizations, including NIEHS, EPA, USDA, TEDX, and EBTC. These early adopters have provided us with an abundance of useful data and user feedback, and we have identified several areas where we can continue to improve our methods and software. Several new features have been planned for the software, and it will be developed, improved and maintained for the foreseeable future.

Role of Semantics, Ontologies, and Adverse Outcome Pathways as a Point of Integration in Chemical Assessments

Michelle Angrish[1], George Woodall[1], Sean Watford[2,3,4,5], and Paul Whaley[6]
[1]NCEA, US EPA, RTP, Durham, NC, 27711, [2]NCCT, US EPA, RTP, Durham, NC, 27711, [3]ORAU, Oak Ridge, TN, 37830 , [4]Environmental Science and Engineering, GSGPH UNC-Chapel Hill, Chapel Hill, NC, 27599, [5]Currently Booz Allen Hamilton, Washington D.C., [6]Evidence Based Toxicology Collaboration, JHUBSPH, MD and Lancaster Environment Centre, Lancaster, UK.

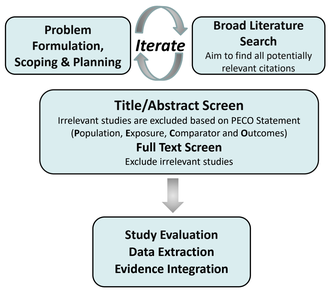
Michelle Angrish | angrish.michelle@epa.gov | 919-541-5381

Abstract

The quality and utility of literature based chemical assessments has been improved by leveraging the power of systematic review (SR) and systematic mapping (SM, also referred to as evidence mapping) approaches to aggregating and evaluating evidence of health risks posed by exposure to environmental chemicals. Taking maximal advantage of SRs and SMs is currently impeded by linguistic inconsistencies resulting from different communities using different vocabularies to describe common study characteristics, requiring the systematic reviewer to anticipate all the concepts, relationships, and words related to a science question when developing a search string sensitive enough to locate all potentially relevant studies. The state-of-the-art approach, to use dictionaries and thesauruses are useful for ensuring all semantically related terms are included in a search, but they do not offer the context necessary to capture relationships between concepts, e.g. according to biological organization such as gene expression. We are therefore exploring the use of ontologies and semantic mapping as a part of evidence integration in literature based chemical assessments. An ontology is a controlled vocabulary of precisely-defined terms and the specified relationships between them, interpretable by both humans and machines. Here we give an example of how literature prioritized for thyroid and neurological health outcome data extracted from human and animal literature studies can be matched to ontology concepts that serve as a point of integration in a semantic framework bounded by a structured Adverse Outcome Pathway (AOP) framework. When implemented, this ontological approach may solve the problem of a researcher needing perfect *a priori* knowledge of all relevant terms and relationships in order to query a database for comprehensive information about mechanisms of thyroid toxicity: this information is already provided in the database ontology.

Background

Systematic Review (SR, Figure 1 ▶) is a formal method used in literature based assessments meant to insure rigor and transparency.

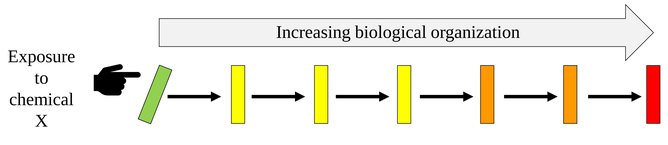


Systematic Mapping (SM, also referred to as evidence mapping) is a technique borrowing from SR principles and can be applied during the scoping, planning and problem formulation phase of a chemical assessment to summarize the characteristics of the evidence base. In chemical assessments, these characteristics are usually broad data categories such as evidence type, chemical, type of animal model or human population, outcome, etc. which are important to know before making decisions on how to approach the assessment, staffing needs, and to identify key data gaps.

Semantics is the study of linguistics, their meaning, relationship, and structure.

Ontologies are "A kind of controlled vocabulary of well-defined terms with specified relationships between those terms, capable of interpretation by both humans and computers" (whetzel et. Al, 2011).

Adverse Outcome Pathways (AOPs, Figure 2 below) are meant to describe how perturbation of a biological system leads to a particular adverse health outcome using components called molecular initiating events (MIEs, green bar below), Key Events (KEs, yellow and orange bars below), Key Event Relationships (KERs, arrows below), and Adverse Outcomes (AOs, red bar below) that are supported by scientific information.



Approach: Evidence Mapping and literature-based information integration into an Adverse Outcome Pathway

Information Identification Categorization and Labeling

1. Bound the literature search by well characterized molecular targets in the thyroid pathway linked to adverse outcomes.
2. Title and Abstract screen results for potentially relevant studies and tag for evidence stream and health outcome.
3. Subset studies tagged for Neurological health outcomes, screen for relevant studies, and tag for.
4. Create an evidence inventory from well characterized studies that includes information on well characterized thyroid pathway components.

Problem Formulation, Scoping & Planning — *Iterate* — **Broad Literature Search**
Aim to find all potentially relevant citations

Bound literature search by:
- NA^+/I^- (Sodium/Iodide) Symporter (NIS)
- Neurological Health Outcomes

Title/Abstract Screen
Irrelevant studies are excluded based on PECO Statement (**P**opulation, **E**xposure, **C**omparator and **O**utcomes)

Tag for evidence stream:
- Animal
- Human
- *In vitro*

Tag for health outcome
- Hepatic
- Cardiovascular
- **Neurological**
- Etc.

Full Text Screen
Exclude irrelevant studies
(**P**opulation, **E**xposure, **C**omparator and **O**utcomes)

Subset studies tagged for **Neurological** health outcomes, full text screen for relevance and tag for **NA^+/I^- (Sodium/Iodide) Symporter (NIS)**.

Evidence Inventory

Inventory reported evidence

Evidence Inventory

Annotate evidence inventory with ontologies amenable to AOP framework

Ontologies, Semantics and AOPs for Organizing Information

Ontology semantic mappings

National Cancer Institute (NCI): thyroxine (T4)	Protein Ontology (PO): Sodium Iodide Symporter (NIS)
+drug, food, chemical or biomedical material	+molecular entity
++physiology/regulatory factor	+++organic amino entity
+++hormone	++++amino acid chain
++++thyroid hormone	+++++protein
+++++thyroxine	++++++sodium/iodide cotransporter
synonym: thyroxin	synonym: NIS

Level of biological organization
Molecular
Cellular
Tissue
Organ
Individual
Population

Key Event
Level of biological organization
Component(s)

Key Event
NCI
Decreased thyroxine (T4)

Semantically index information

Process
PO
e.g. Sodium Iodide Symporter (NIS) mRNA

Component(s)
Process
Object
Context

Object
PO
e.g. Sodium Iodide Symporter (NIS)

Context
PO
Molecular
(e.g.) NIS protein levels

Feed annotated information into AOP

AOP

U.S. Environmental Protection Agency
Office of Research and Development

NASEM 2019

Strengthening the Evaluation of Mechanistic Evidence Categorized by the IARC 10 Key Characteristics of Carcinogens

Kirsten (Ke) Zu, Ph.D., Sc.D., M.P.H.[1]
Julie E. Goodman, Ph.D., DABT, FACE, ATS[1]
Robyn L. Prueitt, Ph.D., DABT[2]

[1]Gradient, Cambridge, MA; [2]Gradient, Seattle, WA

OBJECTIVE

Build off of the International Agency for Research on Cancer (IARC) 10 Key Characteristics of Carcinogens Framework to strengthen the evaluation of mechanistic evidence in assessments of potential carcinogens.

BACKGROUND

- IARC recently developed a framework for identifying and organizing mechanistic data around 10 key characteristics of carcinogens:
  1. **Electrophilic or metabolically activated**
  2. **Genotoxic**
  3. **Alters DNA repair or genomic stability**
  4. **Induces epigenetic alterations**
  5. **Induces oxidative stress**
  6. **Induces chronic inflammation**
  7. **Immunosuppressive**
  8. **Modulates receptor-mediated effects**
  9. **Causes immortalization**
  10. **Alters cell proliferation, death, or nutrient supply**
- This framework is useful for initial categorization of mechanistic evidence, but lacks certain aspects necessary for evaluating and integrating mechanistic evidence for human carcinogen hazard classifications.
  - The framework fails to provide explicit guidance for its implementation, describe how study quality and the human relevance of evidence should be evaluated and considered, indicate how the biological significance of mechanistic endpoints should be considered, address how mechanistic evidence should be integrated with other realms of evidence, and consider that some of the key characteristics of carcinogens are shared by non-carcinogenic substances.

METHODS/RESULTS

- Reviewed best practices for evaluating study quality and for evidence integration.
  - Surveyed over 50 weight-of-evidence frameworks, including several with guidance for evaluating study quality (*e.g.*, the National Toxicology Program Office of Health Assessment and Translation [NTP OHAT], Animal Research: Reporting of *In Vivo* Experiments [ARRIVE], United States Environmental Protection Agency [US EPA] Toxic Substances Control Act [TSCA], and Organisation for Economic Co-operation and Development [OECD] frameworks).
- Developed study quality criteria for *in vivo* and *in vitro* mechanistic studies **(Tables 1 and 2, respectively)**.
- Proposed criteria to determine the human relevance of mechanistic studies, based on the World Health Organization (WHO) International Programme on Chemical Safety (IPCS) cancer mode-of-action (MoA)/human relevance framework, and the biological significance of mechanistic results **(Table 3)**.
- Proposed a conceptual framework for evaluating and integrating mechanistic evidence in systematic reviews of potential carcinogens **(Figure 1)**.

Table 1 Quality Criteria for *In Vivo* Studies

Category	Aspect	Criteria for Higher Quality
Study Design	Study Objectives	Objectives/hypotheses are clearly described
	Negative Controls	Unexposed/vehicle control group was included
	Positive Controls	Positive control group was included or justification provided if a positive control was not required
	Randomization	Randomized assignment of animals to exposure and control groups
	Study Size	At least 2 exposed groups included to allow for the evaluation of dose-response relationships
	Size of Treatment Groups	At least 5 animals of one sex per study group per time point or power calculation reported to justify a smaller sample size
	Study Guidelines	Protocol followed OECD or similar study guidelines
	GLP Conditions	Compliant with GLP guidelines
Test Model	Animal Characteristics	Characteristics (*e.g.*, source, species, strain, age, stage, sex, weight, acclimation period) are reported and similar across study groups
	Animal Husbandry	Animal husbandry (*e.g.*, breeding program, food and water, light and dark cycle, cleaning cycle, environmental conditions) details are reported and consistent across study groups
	Housing Conditions	Housing conditions (*e.g.*, caging, bedding, number of animals per cage, environmental enrichment) are reported and consistent across study groups
	Attrition	All animals are accounted for and any exclusion of animals was adequately addressed and reasons documented
	Monitoring and Handling	Differences in monitoring or handling of animals were minimized across study groups
Test Substance	Substance Description	Test substance is clearly described (*e.g.*, identity, source, purity, stability, storage, vehicle)
Exposure	Exposure Route	Exposure route is clearly described and administration was consistent across study groups
	Exposure Monitoring	Exposure concentration was well characterized and monitored throughout the exposure period
Outcome	Blinding	Outcome assessors were blinded to the exposure status of each group
	Measurement Methods	Outcome was consistently assessed across study groups using well-established, sensitive methods: direct measurements of outcome or validated indirect measurements
	Assay Reproducibility	Assays are specified with enough detail to enable reproducibility of results
		Reported quantitative measures of reproducibility
	QA/QC	Implemented and reported appropriate QA/QC protocols for the collection, handling, and storage of biological specimens (if applicable)
	Time Points	Time points of outcome evaluations were consistent with study objectives
		Results reported for all time points
	Confounding and Variable Control	Identified and accounted for confounding factors and factors unrelated to exposure that may affect outcome
Statistical Analysis	Statistical Methods	Employed appropriate statistical approaches for study objectives
		Test assumptions were satisfied (*e.g.*, data distribution requirement, sparse data)
		Corrected for multiple comparisons (if applicable)

GLP = Good Laboratory Practice; OECD = Organisation for Economic Co-operation and Development; QA/QC = Quality Assurance/Quality Control.

Table 2 Quality Criteria for *In Vitro* Studies

Category	Aspect	Criteria for Higher Quality
Study Design	Study Objectives	Objectives/hypotheses are clearly described
	Negative Controls	Unexposed/vehicle control group was included
	Positive Controls	Positive control group was included or justification is provided if a positive control was not required
	Study Size	At least 2 exposed groups included to allow for the evaluation of dose-response relationships
	Number of Replicates	Number of replicates reported; at least triplicate measurements utilized to address variability
	Study Guidelines	Protocol followed OECD or similar study guidelines
	GLP Conditions	Compliant with GLP guidelines
Test Model	Test System Characteristics	Test system characteristics (*e.g.*, source, type, properties, number of cells) are reported and similar across study groups
	Cultivation and Maintenance	Media characteristics and conditions of cultivation and maintenance are reported and consistent across study groups
Test Substance	Substance Description	Test substance is clearly described (*e.g.*, identity, source, purity, stability, storage, vehicle)
Exposure	Exposure Administration	Route of administration is described and consistent across study groups
	Exposure Monitoring	Exposure concentration was well characterized and monitored throughout exposure period
Outcome	Blinding	Outcome assessors were blinded to the exposure status of each group
	Measurement Methods	Outcome was consistently assessed across study groups using well-established, sensitive methods: direct measurements of outcome or validated indirect measurements
	Assay Reproducibility	Assays are specified with enough detail to enable reproducibility of results
		Reported quantitative measures of reproducibility
	Time Points	Time points of outcome evaluations were consistent with study objectives
		Results reported for all time points
	Confounding and Variable Control	Identified and accounted for confounding factors and factors unrelated to exposure that may affect outcome
Statistical Analysis	Statistical Methods	Employed appropriate statistical approaches
		Test assumptions were satisfied (*e.g.*, data distribution requirement, sparse data)
		Corrected for multiple comparisons (if applicable)

Table 3 Criteria for Evaluating the Human Relevance[a] and Biological Significance of Mechanistic Studies

Criteria	Considerations
Is the weight of evidence sufficient to establish an MoA in animals?	**Postulated MoA:** List sequence of events on the path to cancer
	Key Events: List measurable events critical to cancer induction
	Dose-Response: Correlate the dose dependency of increases in a key event to increases in the severity of other key events and ultimate tumor incidence
	Temporality: Evaluate whether key events are apparent before tumor appearance and are temporally consistent with each other
	Strength, Consistency, and Specificity: Assess the repeatability of key events in different studies; whether there is a reduction of subsequent events or tumors when a key event is blocked; and whether key events and tumor response occur in same cell type
	Biological Plausibility and Coherence: Consider whether MoA is consistent with what is known about carcinogenesis in general and for the substance specifically
	Possible Alternative MoAs: Consider other plausible MoAs
	Uncertainties, Inconsistencies, and Data Gaps: Identify biological and database uncertainties and whether identified inconsistencies and data gaps are crucial as support for the MoA
	Conclusion About the MoA: Indicate level of confidence in the postulated MoA
Can the human relevance of the MoA be reasonably excluded on the basis of fundamental, qualitative differences in key events between experimental animals and humans?	Evaluate plausibility that key events in the animal MoA operate in humans
	Consider comparative developmental processes and their relative timing
	Assess factors that can modulate key events between species
Can the human relevance of the MoA be reasonably excluded on the basis of quantitative differences in either kinetic or dynamic factors between experimental animals and humans?	Evaluate differences in nature and time course of chemical uptake, distribution, metabolism, and excretion
	Consider the consequences of interaction of chemical with cells, tissues, and organs
	Assess the magnitude of exposure differences for observation of key events and tumors
What is the biological significance of specific endpoints or assays?	Assess predictive ability for cancer risk
	Consider the reversibility/reparability of effects
	Evaluate the magnitude of effect above natural variation and homeostasis

Notes: (a) Adapted from the WHO IPCS cancer MoA/human relevance framework (Boobis et al., 2006).

CONCLUSIONS

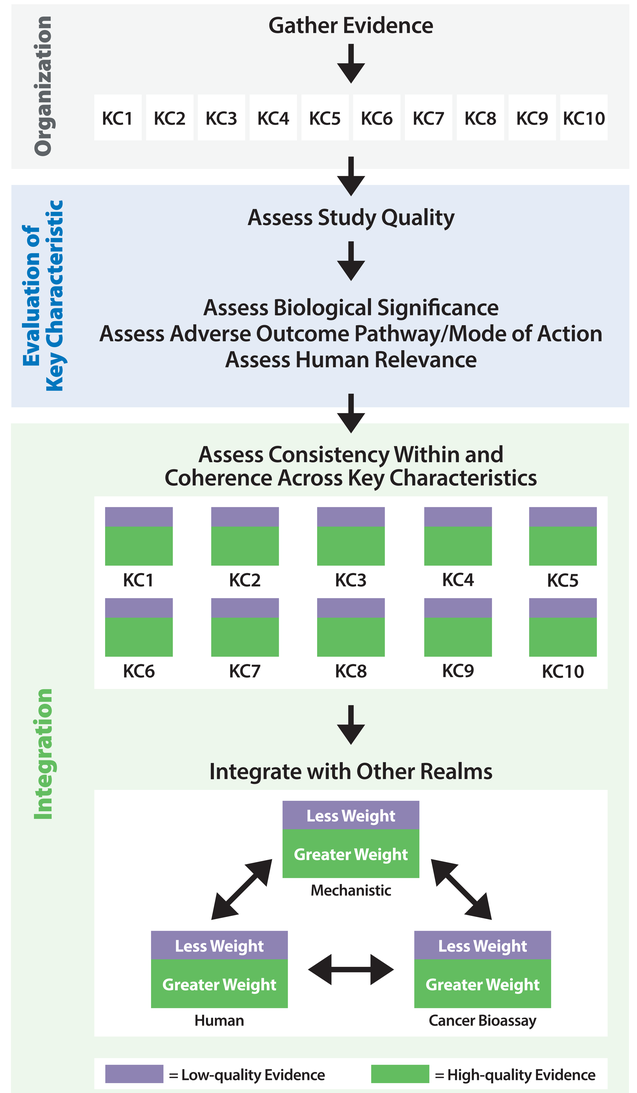
- The 10 Key Characteristics of Carcinogens Framework does not currently provide guidance for a standardized approach to evaluating mechanistic evidence.
- Consideration of study quality and relevance, and the biological significance of mechanistic evidence, in a systematic manner is critical for a robust assessment of a potential carcinogen.
- Mechanistic evidence must be evaluated concurrently with other realms of evidence to allow for scientifically defensible cancer hazard classifications.

ACKNOWLEDGMENTS

The work presented in this poster was completed with financial support provided to Gradient by the American Chemistry Council (ACC).

Figure 1 Incorporation of the IARC 10 Key Characteristics of Carcinogens Framework into Evaluations of *In Vivo* and *In Vitro* Mechanistic Evidence in Systematic Reviews of Potential Carcinogens

The 10 Key Characteristics (KC) of Carcinogens Framework can be used to organize mechanistic evidence. Once mechanistic evidence is categorized, the quality of mechanistic studies, biological significance of study results, pathway from upstream key events to tumor development (such as through development of an adverse outcome pathway and evaluation of MoA), and human relevance of the mechanistic evidence should be evaluated. The consistency across studies within each key characteristic and the coherence across the key characteristics should also be assessed. Mechanistic evidence must be integrated with other realms of evidence (human, cancer bioassay), with evidence from high-quality studies weighted more than evidence from low-quality studies.



Evidence Integration in Integrated Science Assessments (ISAs): A Case Study from the Draft Particulate Matter ISA

Michael J. Stewart, Ellen Kirrane, Thomas J. Luben, Jason Sacks, Barbara Buckley, Jennifer Nichols

National Center for Environmental Assessment | Environmental Media Assessment Group

Michael J. Stewart | stewart.michael@epa.gov | 919-541-7524

Background:

The National Center for Environmental Assessment (NCEA) develops Integrated Science Assessments (ISAs) as a key part of the Clean Air Act mandated reviews of the National Ambient Air Quality Standards (NAAQS), which are set for six criteria pollutants: particulate matter (PM), ozone, oxides of nitrogen, sulfur oxides, lead, and carbon monoxide. EPA establishes primary NAAQS to protect public health, including sensitive lifestages or populations, such as children or people with pre-existing disease. Secondary standards are established to protect against adverse ecological and other welfare effects. The ISAs identify, evaluate, integrate, and synthesize the comprehensive body of scientific evidence. This generally includes hundreds to thousands of studies spanning epidemiology, controlled human exposure, animal toxicology, dosimetry, exposure science, atmospheric science, welfare effects, and ecology. NCEA employs a weight of evidence framework in developing ISAs, integrating findings from the various lines of evidence and drawing conclusions on causality. More specifically, ISAs use a five-level hierarchical causal framework, incorporating aspects of the Hill criteria to assess causality (e.g., consistency, coherence, biological plausibility, temporality, etc.) and classify whether evidence is sufficient to conclude a "causal relationship", "likely to be a causal relationship", "suggestive of, but not sufficient to infer, a causal relationship", "inadequate to infer a causal relationship", or "not likely to be a causal relationship." Each level of the hierarchy represents the extent to which we can rule out chance, confounding or other biases. In ISAs, these causality determinations are presented both in a narrative form and in summary tables delineating the rationales and key evidence supporting the conclusion, reflecting the application of the framework and characterization of the evidence. In this case poster, an example from the draft PM ISA is presented, demonstrating the evaluation and integration of multiple lines of evidence underlying the conclusion that there is a "causal relationship" between short-term $PM_{2.5}$ exposure and cardiovascular effects.

ISA Development [1]

Literature Search and Study Selection
(See Figure III)

Evaluation of Individual Study Quality
After study selection, the quality of individual studies is evaluated by U.S. EPA or outside experts in the fields of atmospheric science, exposure assessment, dosimetry, animal toxicology, controlled human exposure, epidemiology, biogeochemistry, terrestrial and aquatic ecology, and other welfare effects, considering the design, methods, conduct, and documentation of each study. Strengths and limitations of individual studies that may affect the interpretation of the study are considered.

Develop Initial Sections
Review and summarize conclusions from previous assessments and new study results and findings by discipline and category of outcome/effect (e.g., toxicological studies of lung function or biogeochemical studies of forests)

Peer Input Consultation
Review of initial draft materials by scientists from both outside and within the U.S. EPA in public meeting or public teleconference.

Evaluation, Synthesis, and Integration of Evidence
Integrate evidence from scientific disciplines. Evaluate evidence for related groups of endpoints or outcomes to draw conclusions for specific health or welfare effect categories, integrating health or welfare effects evidence with information on mode of action and exposure assessment.

Development of Scientific Conclusions and Causal Determinations
Characterize weight of evidence and develop judgments regarding causality for health or welfare effect categories. Develop conclusions regarding concentration- or dose-response relationships, potentially at-risk populations, lifestages, or ecosystems.

Draft Integrated Science Assessment
Evaluation and integration of newly published studies

Clean Air Scientific Advisory Committee
Independent review of draft documents for scientific quality and sound implementation of causal framework during public meetings.

Public Comments
Comments on draft ISA solicited by the U.S. EPA

Final Integrated Science Assessment

U.S. Environmental Protection Agency
Office of Research and Development

This poster contains results from the external review Draft PM ISA. This information has been distributed solely for the purpose of predissemination peer review under applicable information quality guidelines. It has not been formally disseminated by EPA. It does not represent and should not be construed to represent any Agency determination or policy.

Aspects of Causality[1]

Aspect	Description
Consistency	An inference of causality is strengthened when a pattern of elevated risks is observed across several independent studies. The reproducibility of findings constitutes one of the strongest arguments for causality. Statistical significance is not the sole criterion by which the presence or absence of an effect is determined. If there are discordant results among investigations, possible reasons such as differences in exposure, confounding factors, and the power of the study are considered.
Coherence	An inference of causality from one line of evidence (e.g., epidemiologic, controlled human exposure, animal, or ecological studies) may be strengthened by other lines of evidence that support a cause-and-effect interpretation of the association. There may be coherence in demonstrating effects from evidence across various fields and/or across multiple study designs or related health endpoints within one scientific line of evidence. For example, evidence on welfare effects may be drawn from a variety of experimental approaches (e.g., greenhouse, laboratory, and field) and subdisciplines of ecology (e.g., community ecology, biogeochemistry, and paleontological/historical reconstructions).
Biological plausibility	An inference of causality is strengthened by results from experimental studies or other sources demonstrating biologically plausible mechanisms. A proposed mechanism, which is based on experimental evidence and which links exposure to an agent to a given effect, is an important source of support for causality.
Biological gradient (exposure-response relationship)	A well-characterized exposure-response relationship (e.g., increasing effects associated with greater exposure) strongly suggests cause and effect, especially when such relationships are also observed for duration of exposure (e.g., increasing effects observed following longer exposure times).
Strength of the observed association	The finding of large, precise risks increases confidence that the association is not likely due to chance, bias, or other factors. However, it is noted that a small magnitude in an effect estimate may or may not represent a substantial effect in a population.
Experimental evidence	Strong evidence for causality can be provided through "natural experiments" when a change in exposure is found to result in a change in occurrence or frequency of health or welfare effects.
Temporality of the observed association	Evidence of a temporal sequence between the introduction of an agent and appearance of the effect constitutes another argument in favor of causality.
Specificity of the observed association	Evidence linking a specific outcome to an exposure can provide a strong argument for causation. However, it must be recognized that rarely, if ever, does exposure to a pollutant invariably predict the occurrence of an outcome, and that a given outcome may have multiple causes.
Analogy	Structure activity relationships and information on the agent's structural analogs can provide insight into whether an association is causal. Similarly, information on mode of action for a chemical, as one of many structural analogs, can inform decisions regarding likely causality.

ISAs Causality Framework[1]

WEIGHT OF EVIDENCE FOR CAUSAL DETERMINATION

	Health Effects	Ecological and Other Welfare Effects
Causal relationship	Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures (e.g., doses or exposures generally within one to two orders of magnitude of recent concentrations). That is, the pollutant has been shown to result in health effects in studies in which chance, confounding, and other biases could be ruled out with reasonable confidence. For example: (1) controlled human exposure studies that demonstrate consistent effects, or (2) observational studies that cannot be explained by plausible alternatives or that are supported by other lines of evidence (e.g., animal studies or mode of action information). Generally, the determination is based on multiple high-quality studies conducted by multiple research groups.	Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures. That is, the pollutant has been shown to result in effects in studies in which chance, confounding, and other biases could be ruled out with reasonable confidence. Controlled exposure studies (laboratory or small- to medium-scale field studies) provide the strongest evidence for causality, but the scope of inference may be limited. Generally, the determination is based on multiple studies conducted by multiple research groups, and evidence that is considered sufficient to infer a causal relationship is usually obtained from the joint consideration of many lines of evidence that reinforce each other.
Likely to be a causal relationship	Evidence is sufficient to conclude that a causal relationship is likely to exist with relevant pollutant exposures. That is, the pollutant has been shown to result in health effects in studies where results are not explained by chance, confounding, and other biases, but uncertainties remain in the evidence overall. For example: (1) observational studies show an association, but copollutant exposures are difficult to address and/or other lines of evidence (controlled human exposure, animal, or mode of action information) are limited or inconsistent, or (2) animal toxicological evidence from multiple studies from different laboratories demonstrate effects, but limited or no human data are available. Generally, the determination is based on multiple high-quality studies.	Evidence is sufficient to conclude that there is a likely causal association with relevant pollutant exposures. That is, an association has been observed between the pollutant and the outcome in studies in which chance, confounding, and other biases are minimized but uncertainties remain. For example, field studies show a relationship, but suspected interacting factors cannot be controlled, and other lines of evidence are limited or inconsistent. Generally, the determination is based on multiple studies by multiple research groups.
Suggestive of, but not sufficient to infer, a causal relationship	Evidence is suggestive of a causal relationship with relevant pollutant exposures but is limited, and chance, confounding, and other biases cannot be ruled out. For example: (1) when the body of evidence is relatively small, at least one high-quality epidemiologic study shows an association with a given health outcome and/or at least one high-quality toxicological study shows effects relevant to humans in animal species, or (2) when the body of evidence is relatively large, evidence from studies of varying quality is generally supportive but not entirely consistent, and there may be coherence across lines of evidence (e.g., animal studies or mode of action information) to support the determination.	Evidence is suggestive of a causal relationship with relevant pollutant exposures, but chance, confounding, and other biases cannot be ruled out. For example, at least one high-quality study shows an effect, but the results of other studies are inconsistent.
Inadequate to infer a causal relationship	Evidence is inadequate to determine that a causal relationship exists with relevant pollutant exposures. The available studies are of insufficient quantity, quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of an effect.	Evidence is inadequate to determine that a causal relationship exists with relevant pollutant exposures. The available studies are of insufficient quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of an effect.
Not likely to be a causal relationship	Evidence indicates there is no causal relationship with relevant pollutant exposures. Several adequate studies, covering the full range of levels of exposure that human beings are known to encounter and considering at-risk populations and lifestages, are mutually consistent in not showing an effect at any level of exposure.	Evidence indicates there is no causal relationship with relevant pollutant exposures. Several adequate studies examining relationships with relevant exposures are consistent in failing to show an effect at any level of exposure.

Sample Causality Text: Short-term Exposure to $PM_{2.5}$ and Cardiovascular Effects[2]

A large body of recent evidence confirms and extends the evidence from the previous ISA indicating that there is a "causal relationship" between short term $PM_{2.5}$ exposure and cardiovascular effects. In the current review, evidence supporting the causality determination includes generally positive associations reported from epidemiologic studies of hospital admissions and emergency department (ED) visits for cardiovascular related effects, and in particular, for ischemic heart disease and heart failure. Results from these observational studies are in agreement with experimental evidence from controlled human exposure and animal toxicological studies of endothelial dysfunction, as well as with endpoints indicating impaired cardiac function, increased risk of arrhythmia, changes in heart rate variability (HRV), increases in blood pressure (BP), and increases in indicators of systemic inflammation, oxidative stress, and coagulation. Results from observational panel studies, though not entirely consistent, also provide some evidence of increased risk of arrhythmia, decreases in HRV, increases in BP, and changes in cardiac electrophysiology. Thus, the combination of evidence from experimental and epidemiologic panel studies provides coherence and biological plausibility for the results from observational epidemiologic studies. Finally, epidemiologic studies of cardiovascular-related mortality provide additional evidence and contributes to the continuum of effects from biomarkers of inflammation and coagulation, subclinical endpoints (HRV, BP, endothelial dysfunction), ED visits and hospital admissions for outcomes such as ischemic heart disease (IHD) and congestive heart failure (CHF), and eventually death. The current body of evidence also reduces uncertainties from the previous review related to the potential for copollutant confounding and biological plausibility for cardiovascular effects following short term $PM_{2.5}$ exposure.

Sample Causality Table: Short-term Exposure to $PM_{2.5}$ and Cardiovascular Effects[2]

Rationale for Causal Determination	Key Evidence
Consistent epidemiologic evidence from multiple, high quality studies at relevant $PM_{2.5}$ concentrations	Increases in ED visits and hospital admissions for IHD and CHF in multicity studies conducted in the U.S., Canada, Europe, and Asia Increases in cardiovascular mortality in multicity studies conducted in the U.S., Canada, Europe, and Asia.
Consistent evidence from controlled human exposure studies at relevant $PM_{2.5}$ concentrations	Consistent changes in measures of endothelial dysfunction Generally consistent evidence for small increases in measures of blood pressure following CAPs exposure Additional evidence of conduction abnormalities, heart rate variability, impaired heart function, systemic inflammation/oxidative stress
Consistent evidence from animal toxicological studies at relevant $PM_{2.5}$ concentrations	Consistent changes in indicators of endothelial dysfunction. Additional evidence of changes in impaired heart function, conduction abnormalities/arrhythmia, heart rate variability, blood pressure, systemic inflammation/oxidative stress
Epidemiologic evidence from copollutant models provides some support for an independent $PM_{2.5}$ association	The magnitude of $PM_{2.5}$ associations remain positive, but in some cases are reduced with larger confidence intervals in copollutant models with gaseous pollutants. Further support from copollutant analyses indicating positive associations for cardiovascular mortality. Recent studies that examined potential copollutant confounding are limited to studies conducted in Europe and Asia. When reported, correlations with gaseous copollutants were primarily in the low to moderate range ($r < 0.7$).
Consistent positive epidemiologic evidence for associations between $PM_{2.5}$ exposure and CVD ED visits and hospital admissions across exposure measurement metrics	Positive associations consistently observed across studies that used ground-based (i.e., monitors), model (e.g., CMAQ, dispersion models) and remote sensing (e.g., AOD measurements from satellites) methods, including hybrid methods that combine two or more of these methods.
Epidemiologic evidence supports a log-linear, no-threshold concentration-response (*C-R*) relationship	
Generally consistent evidence for biological plausibility of cardiovascular effects	Strong evidence for coherence of effects across scientific disciplines and biological plausibility for a range of cardiovascular effects in response to short-term $PM_{2.5}$ exposure. Includes evidence for reduced myocardial blood flow, altered vascular reactivity, and ST segment depression.
Uncertainty regarding geographic heterogeneity in $PM_{2.5}$ associations	Multicity U.S. studies demonstrate city-to-city and regional heterogeneity in $PM_{2.5}$-CVD ED visit and hospital admission associations. Evidence supports that a combination of factors including composition and exposure factors may contribute to the observed heterogeneity.

* CMAQ= Community Multiscale Air Quality Modeling System; AOD= Aerosol Optical Depth; CAPs = Concentrated Ambient Particles

References:

1. Preamble to the ISA: https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=310244
2. ISA for PM (External Review Draft): http://cfint.rtpnc.epa.gov/ncea/prod/recordisplay.cfm?deid=341593

Systematically Evaluating and Integrating Evidence in National Ambient Air Quality Standards (NAAQS) Reviews

Julie E. Goodman, Ph.D., DABT, FACE, ATS[1]
Giffe Johnson, Ph.D.[2]
Robyn L. Prueitt, Ph.D., DABT[3]
Kirsten (Ke) Zu, Ph.D., Sc.D., M.P.H.[1]

[1]Gradient, Cambridge, MA; [2]The National Council for Air and Stream Improvement, Inc., Newberry, FL; [3]Gradient, Seattle, WA

NASEM 2019

OBJECTIVE

Provide recommendations for systematically evaluating and integrating evidence on cancer in National Ambient Air Quality Standards (NAAQS) reviews.

BACKGROUND

- The Clean Air Act mandates that the United States Environmental Protection Agency (US EPA) set primary and secondary NAAQS for six criteria air pollutants (*i.e.*, particulate matter [PM], ground-level ozone, nitrogen dioxide, sulfur dioxide, carbon monoxide, and lead).
- Every 5 years, US EPA is required to review the available science and data to determine whether it should maintain or revise the NAAQS for each of these pollutants.
- As part of this process, US EPA generates Integrated Science Assessments (ISAs), in which causal relationships between criteria pollutants and health effects are assessed using a framework specifically developed for NAAQS reviews.
- The NAAQS causal framework includes a list of quality aspects for evaluating individual studies of health effects.
  - However, it is not a complete list, there are no quality criteria for *in vitro* studies, there are no criteria for evaluating study relevance, and the quality evaluation does not inform evidence integration.

Table 1 Quality Criteria for Epidemiology Studies of Air Pollutants and Health Effects

Category	Aspect	Criteria for Higher Quality
Study Design	Study Objectives	Objectives/hypotheses are clearly described
	Study Design	Panel, case-crossover, cohort, or nested case-control study
	Study Location	Multiple cities
	Sample Size	Power calculation is presented to indicate sufficient sample size[a]
	Study Duration	Conducted over multiple years
Study Population	Participant Characteristics	**Characteristics (*e.g.*, age, race, sex, eligibility criteria) are reported**
	Inclusion/Exclusion Criteria	Clearly reported and consistent with study objectives
	Recruitment/Participation	Representative of the target population
		High participation rate, not dependent on exposure or outcome
	Underlying Health Conditions	Ascertained by independent clinical assessment or self-report of physician's diagnosis
	Follow-up	Minimal or non-differential loss to follow-up
Pollutant	Pollutant Description	Clearly described (*e.g.*, size of PM fraction)
	Pollutant Source	Source-related indicators were evaluated
Exposure Assessment	Measurement Methods	Utilized and compared with more than one exposure assessment method
		Used well-established, sensitive methods: direct measurements of exposure or indirect measurements that have been validated
	Exposure Window	Assessed relevant exposure windows; multiple exposure windows were investigated if the relevant exposure windows were unknown
	Spatial Variability	Sufficiently captured the spatial variability of the exposure
		Exposure was estimated from the closest central site monitor (limit of distance dependent on pollutant) or from averaging concentrations from multiple monitors, when only using monitoring data
		Exposure was estimated from modeling data with sufficient spatial resolution
	Temporal Variability	Used time-varying or multiple lags of exposure estimates
	Temporality	**Exposure occurred BEFORE the outcome**
Outcome Assessment	Blinding	Assessors of outcome were blinded to exposure levels
	Measurement Methods	Used well-established, sensitive methods: direct measurements of outcome or indirect measurements that have been validated
		Clinical measurements: administered or overseen by medical professionals
		Bioassay measurements: provided sufficient details to allow for reducing the assay and quantitative measures of reproducibility
	QA/QC	**Implemented and reported appropriate QA/QC protocols for the collection, handling, and storage of biological specimens, if applicable**
	Ascertainment	Relied on medical records and insurance claims with ICD codes, cancer registries, or self-reports of physician's diagnoses in validated questionnaires
	Adjustment of Acute Effects	Adjusted for short-term exposure if the health effects of long-term exposure were assessed by acute events
	Time Points	Time points of outcome evaluation are consistent with study objectives
		Reported effect estimates for all *a priori* lag times in studies of short-term exposure
Confounding	Confounders	Identified and adjusted for potential confounders and primary covariates (*e.g.*, temporal trends, meteorology, socioeconomic status, age, race, sex, medication use, smoking status, and other chemical exposures)
		Used valid and reliable measurements of these factors
	Copollutant	Copollutant (two-pollutant) modeling was conducted
		Correlations were observed between the pollutant of interest and copollutants were considered
Statistical Analysis	Descriptive Statistics	**Summary statistics for the study population presented**
	Univariate Analyses	**Univariate analyses with pollutant of interest, covariates, and copollutants were conducted and results are presented**
	Multivariate Analyses	Employed appropriate statistical models for multivariate analyses
		Model assumptions were tested and satisfied
		Multiple-comparison-corrected, if applicable
	Sensitivity Analyses	Sensitivity analyses were conducted and results were stable

Notes: ICD = International Classification of Diseases; QA/QC = Quality Assurance/Quality Control. ***Bolded text*** *indicates criteria that are not from the NAAQS framework. (a) The PM ISA indicates that studies with large sample sizes are more reliable, but does not specify what is considered a large sample size.*

METHODS/RESULTS

- Reviewed best practices for evaluating study quality and for evidence integration.
  - Surveyed over 50 weight-of-evidence frameworks, including several that provided guidance for evaluating study quality (*e.g.*, the National Toxicology Program Office of Health Assessment and Translation [NTP OHAT], Strengthening the Reporting of Observational Studies in Epidemiology [STROBE], Organisation for Economic Co-operation and Development [OECD] frameworks).
- Modified the NAAQS framework for evaluating the health effects of air pollutants using these best practices.
  - Developed general quality criteria for epidemiology studies **(Table 1)** and demonstrated how to adapt these to address a specific topic, using a case study of PM exposure and lung cancer **(Table 2)**. Also, developed general quality criteria for experimental studies (*i.e.*, controlled human exposure and animal toxicology studies) and *in vitro* studies (available upon request).
  - Proposed criteria to determine the human relevance of mechanistic studies, based on the World Health Organization (WHO) International Programme on Chemical Safety (IPCS) human relevance framework (available upon request).
- Proposed a conceptual framework for evidence integration:
  - Bodies of evidence are organized by an outcome and receive an initial confidence rating based on study design **(Table 3)**. Each body of evidence is evaluated for additional domains that increase or decrease confidence **(Figure 1)**. Causal conclusions are based on the levels of confidence in human, animal, and mechanistic data **(Table 4)**.

Table 2 Quality Criteria for Epidemiology Studies of PM and Lung Cancer

Category	Aspect	Criteria for Higher Quality
Study Design	Study Objectives	Objectives/hypotheses are clearly described
	Study Design	Cohort or nested case-control study
	Study Location	Multiple cities
	Sample Size	Power calculation is presented to indicate sufficient sample size[a]
	Study Duration	Sufficient follow-up time (at least 10-20 years)
Study Population	Participant Characteristics	**Characteristics (*e.g.*, age, race, sex, eligibility criteria) are reported**
	Inclusion/Exclusion Criteria	Clearly reported and consistent with study objectives
	Recruitment/Participation	Representative of the target population
		High participation rate, not dependent on exposure or outcome
	Underlying Health Conditions	Ascertained by independent clinical assessment or self-report of physician's diagnosis
	Follow-up	Minimal or non-differential loss to follow-up
Pollutant	Fraction Size	Clearly described (*e.g.*, size of PM fraction)
	Particulate Source	Source-related indicators were evaluated
Exposure Assessment	Measurement Methods	Utilized and compared more than one measurement method
		Indirect measurements of exposure (*e.g.*, modeling) are validated
	Exposure Window	**Latency of lung cancer (*i.e.*, at least 10-20 years prior to diagnosis) is accounted for**
	Spatial Variability	$PM_{2.5}$ – Community-level modeled concentrations, measurements from multiple central site monitors[b]
		$PM_{10/2.5}$ or UFP – Finer spatial resolution for modeled concentrations or validated central site measurements
	Temporal Variability	**Used time-varying exposure estimates**
	Temporality	**Exposure occurred BEFORE the outcome**
Outcome Assessment	Blinding	Assessors of outcome were blinded to exposure levels
	Ascertainment	Relied on medical records and insurance claims with ICD codes, cancer registries, or self-reports of physician's diagnoses in validated questionnaires
	Incidence	**Cancer incidence was assessed**
	Clinical Information	**Clinical information on lung cancer (*e.g.*, histological type, stage) was collected**
Confounding	Smoking	**Multiple metrics for smoking behaviors and history were adjusted for (*e.g.*, intensity, duration, years since quitting)**
		Exposure to secondhand smoking was adjusted for
	Other Covariates	Key potential confounders were adjusted for (*e.g.*, age, socioeconomic status, race, family history, occupational exposures)
	Copollutant	Correlations were observed between the pollutant of interest and copollutants were considered
Statistical Analysis	Descriptive Statistics	**Summary statistics for the study population (*e.g.*, number of participants, age, race, socioeconomic status, smoking, number of lung cancer cases, person-time) are presented by exposure categories**
	Univariate Analyses	**Univariate analyses with PM, covariates, and copollutants were conducted and results are presented**
	Multivariate Analyses	Employed appropriate statistical models (*e.g.*, logistical regression, Cox proportional hazard)
		Model assumptions were tested and satisfied
		Multiple-comparison-corrected, if applicable
	Sensitivity Analyses	Sensitivity analyses were conducted and results were stable

Notes: $PM_{10/2.5}$ = Particulate Matter with a Diameter Less than or Equal to 10/2.5 µm; UFP = Ultrafine Particles. ***Bolded text*** *indicates criteria that are not from the NAAQS framework. (a) The PM ISA indicates that studies with large sample sizes are more reliable, but does not specify what is considered a large sample size. (b) The PM ISA indicates that measurements from a single central site monitor are sufficient to capture the spatial variability of within-community $PM_{2.5}$ concentrations.*

ACKNOWLEDGMENTS

The work presented in this poster was completed with financial support provided to Gradient by the National Council for Air and Stream Improvement, Inc. (NCASI) and the American Petroleum Institute (API).

CONCLUSIONS

Evaluating the quality of individual studies is an integral part of assessing causal relationships between air pollutants and health effects, including cancer. The relevance of mechanistic evidence to humans must be determined. The results of the quality evaluation must inform evidence integration.

Table 3 Initial Confidence Rating for Study Designs

Study Design	Controlled Exposure	Exposure Prior to Outcome	Individual Outcome Data	Comparison Group Used	Initial Confidence Rating
Human Controlled Exposure	Likely	Likely	Likely	Likely	High
Experimental Animal	Likely	Likely	Likely	Likely	High
Cohort	Unlikely	May or May Not	Likely	Likely	Low to Moderate
Case-control	Unlikely	May or May Not	Likely	Likely	Low to Moderate
Cross-sectional	Unlikely	Unlikely	Likely	Likely	Low
Ecologic	Unlikely	May or May Not	Unlikely	Likely	Very Low to Low
Case Series/Report	Unlikely	May or May Not	Likely	Unlikely	Very Low to Low

Figure 1 Confidence Assessment in Each Body of Evidence

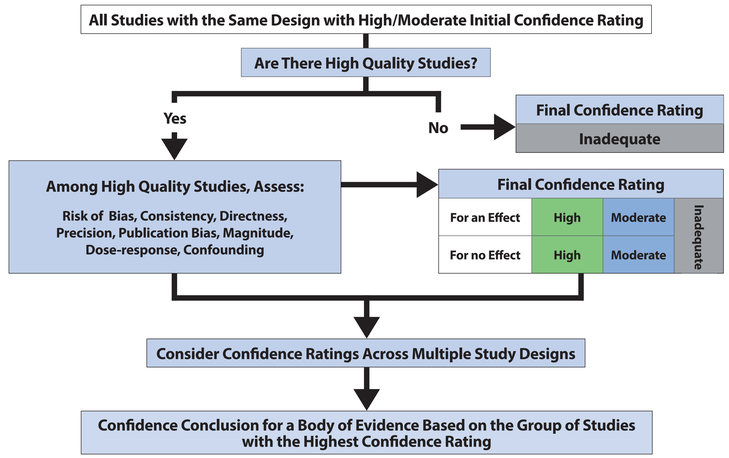


Table 4 Causal Classification

	Human Evidence	Animal Evidence	Causal Conclusion for An Effect		
			No/Inadequate Mechanistic Data	Strong Mechanistic Data Supporting An Effect	Strong Mechanistic Data Supporting No Effect
Confidence in Evidence for An Effect	High	High	Causal		Suggestive
		Moderate			
		Inadequate			
	Moderate	High	Suggestive		Inadequate
		Moderate			
		Inadequate			
	Inadequate	High	Suggestive		
		Moderate	Inadequate	Suggestive	
		Inadequate	Inadequate		
Confidence in Evidence for No Effect	High	High	Unlikely Causal		Unlikely Causal
		Moderate			
		Inadequate	Unlikely Causal	Inadequate	
	Moderate	High	Inadequate		Unlikely Causal
		Moderate			Inadequate
		Inadequate			
	Inadequate	High	Inadequate		Unlikely Causal
		Moderate			Inadequate
		Inadequate			

Note: This does not address discordant human and animal evidence. If this occurs, further evaluation will be needed to assess causation.

Modeling mechanistic processes from source to outcome to support evidence integration and inform risk assessment

David E. Hines[1], Rory B. Conolly[1], Annie M. Jarabek[2]

U.S. Environmental Protection Agency, Office of Research and Development; Research Triangle Park, NC;[1] National Health and Environmental Effects Research Laboratory; [2] National Center for Environmental Assessment

David E. Hines | hines.david@epa.gov | 919-541-1469

Introduction

- Evidence integration in current IRIS assessments considers the contributions of human health, animal, and mechanistic data streams according to PECO criteria in a hierarchical and parallel approach. (*Fig. 1*)

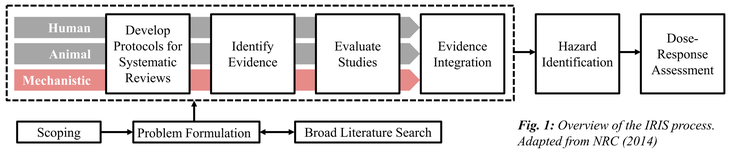


Fig. 1: *Overview of the IRIS process. Adapted from NRC (2014)*

- The NAS has emphasized the use of mechanistic process models of pathogenesis to evaluate relationships among biomarkers (exposure/effect/susceptibility) as well as modernizing risk predictions using exposure science and computational models.
- We propose mechanistic data should serve as a scaffold for the use of process models when integrating evidence across human health and ecological endpoints. *(Fig. 2)*

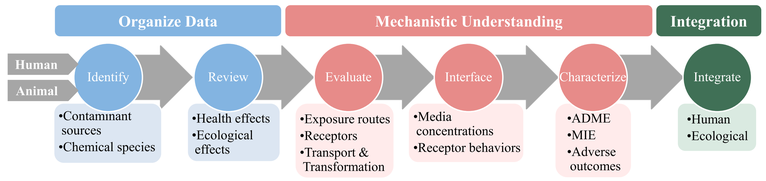


Fig. 2: *Mechanistic workflow for evidence integration*

Case Study Example

- We demonstrate how the Aggregate Exposure Pathway (AEP) and Adverse Outcome Pathway (AOP) frameworks create a source to outcome continuum using a case study of the perchlorate anion (ClO_4^-).

Teeguarden et al. (2016), Ankley et al. (2010)

Quantitative AEP (A)

- Constructed a six compartment fate-and-transport network for the hypothetical site. *(Fig. 3)*
- Considered inputs to site from surface water, atmospheric deposition, and groundwater contamination.
- Literature values were used to restrict parameter ranges.
- A Monte Carlo approach (n=10,000) was used to estimate variability in the exposure network based on variability in literature data.

Exposure Scenarios (B)

- Behavioral assumptions:
  - Groundwater from well
  - Media (Surface water)
  - Grass (95%) Surface water (5%)
- Contamination input scenarios:
  - **Mild**: Inputs from literature, similar to published concentrations
  - **Moderate**: 10x Mild scenario inputs
  - **High**: 100x Mild scenario groundwater inputs

Analyses (C D E)

- Estimated external exposure and source apportionment using Network Environ Analysis

Fath and Patten (1999)

- Linked **AEP network** to multispecies **AOP network** using previously published PBPK models
- Estimate hazard index (HI) using EQ. 1

$$HI = \sum_{i=1}^{n} \frac{E_i}{AL_i}$$

EQ 1: *i is each exposure source, E is the exposure level, and AL is the acceptable limit of exposure. AL was the lowest reported LOAEL for each species*

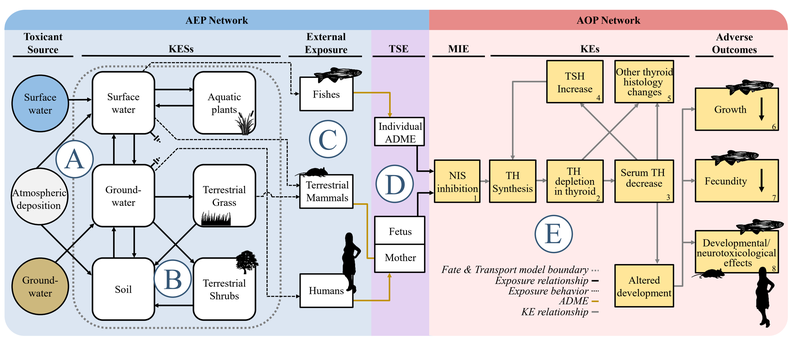


Fig. 3*: Joint AEP-AOP construct for the ClO_4^- case study. Detailed description of AOP network in Hines et al. (2018).*

U.S. Environmental Protection Agency
Office of Research and Development

Quantitative Case Study

B Exposure Estimation

	Mild			Moderate			High		
	1%	Med	99%	1%	Med	99%	1%	Med	99%
	<0.1	0.1	0.3	0.4	1.0	2.7	4.0	10.3	25.4
	0.1	0.5	1.4	1.2	5.4	14.1	2.1	9.8	28.6
	3.1	15.8	34.2	89.5	159.4	395.7	369.5	2990.1	8072.4

Tab. 1*: Lower confidence interval (1%), median (Med), and upper confidence interval (99%) external toxicant doses predicted for human, fish, and small mammals in each scenario.*

Table units are µg/kg/d.

C Source Apportionment

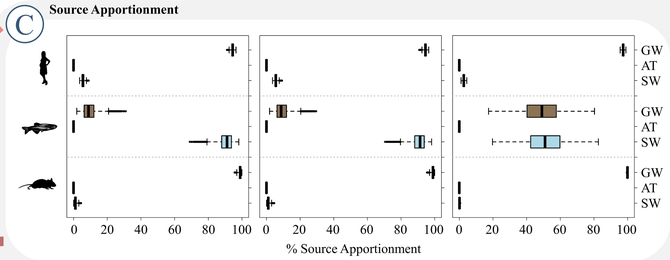


Fig. 4: *Species-specific source apportionment*

Groundwater (GW)
Atmosphere (AT)
Surface water (SW)

D MIE Activation

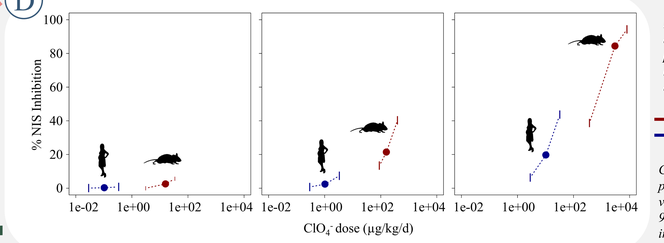


Fig. 5: *MIE activation. No published PBPK model was available for fish.*

Merrill et al. (2003)
Lumen et al. (2013)

Circles show median prediction and vertical points show 99% confidence interval.

E Evaluation of Risk

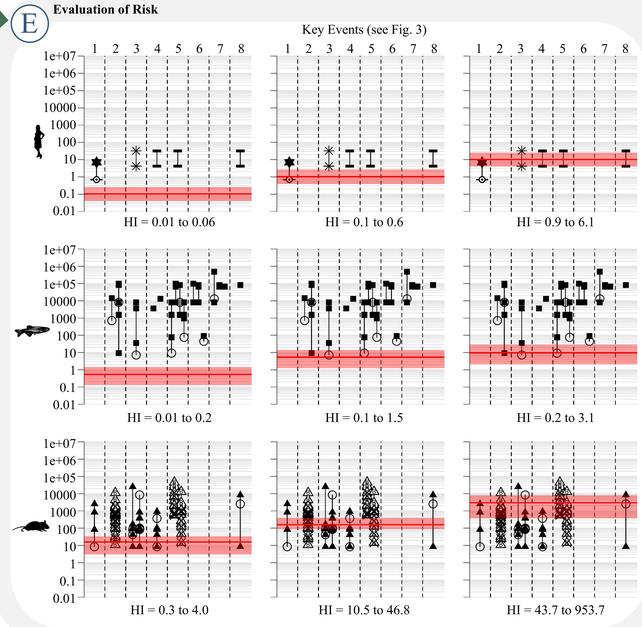


Fig. 6: *Predicted internal doses at the hypothetical site under different contamination scenarios compared with dose-response toxicity data. See KE Hines et al. (2018) for KE details.*

Y-axis shows toxicant dose in µg/kg/d on a log scale

- LOEL/LOAEL
- Point of departure
- Reference dose
- Model prediction
- Model extrapolation
- NOEL/NOAEL
- BMD
- BMDL
- Median internal dose
- Internal dose range prediction

Units: µg/kg/d

Disclaimer: The views expressed in this poster are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Discussion

- The source to outcome case study demonstrates how a workflow for using a mechanistic scaffold can facilitate evidence integration. (Fig. 7)

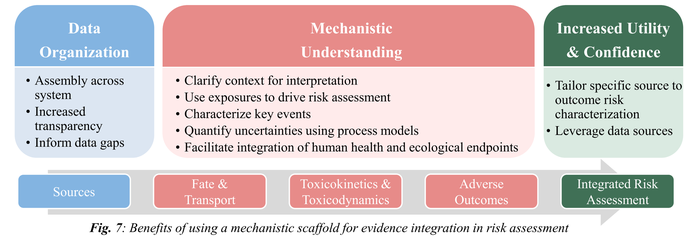


Fig. 7*: Benefits of using a mechanistic scaffold for evidence integration in risk assessment*

- The AEP and AOP frameworks facilitate exposure driven risk assessments in support of assessments required by the new TSCA
- Mechanistic approaches to data integration can act as an organizing framework to inform ontologies or evidence maps, leverage data sources, and facilitate quantitative characterization of key events in pathogenesis.
- Explicit elucidation of key events and parameters supports transparency in risk assessments.
- Risk assessments based on exposure use cases and toxicity pathways involved in pathogenesis allow for more targeted assessment and increased confidence.

Conclusions

A mechanistic scaffold informs problem formulation, aids evaluation of study quality criteria, and facilitates evidence integration to support source-to-outcome risk assessments that are:

1) Exposure driven to target specific use-cases
2) Quantitative for key events in relevant AOPs
3) Capable of characterizing human health and ecological endpoints

Literature Cited & Abbreviations

Ankley, G. T.; Bennett, R. S.; Erickson, R. J.; Hoff, D. J.; Hornung, M. W.; Johnson, R. D.; Mount, D. R.; Nichols, J. W.; Russom, C. L.; Schmieder, P. K.; Serrrano, J. A., **2010**. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem. 29 (3), 730-741

Fath, B.D. and Patten, B.C., **1999**. Review of the foundations of network environ analysis. Ecosystems, 2(2), pp.167-179.

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 endpoints for Cumulative Risk Assessment (CRA). Environ. Sci. Technol. 52, 839-849.

Lumen, A., Mattie, D.R., Fisher, J.W., **2013**. Evaluation of perturbations in serum thyroid hormones during human pregnancy due to dietary iodide and perchlorate exposure using a biologically based dose-response model. Toxicological Sciences 133(2), 320-341.

Merrill, E.A., Clewell, R.A., Gearhart, J.M., Robinson, P.J., Sterner, T.R., Yu, K.O., Mattie, D.R. and Fisher, J.W., **2003**. PBPK predictions of perchlorate distribution and its effect on thyroid uptake of radioiodide in the male rat. Toxicological Sciences, 73(2), pp.256-269.

NRC (National Research Council), **2014.** *Review of EPA's integrated risk information system (IRIS) process*. National Academies Press.

Teeguarden, J.G., Tan, Y., Edwards, S.W, Leonard, J.A., Anderson, K.A., Corley, R.A., Kile, M.L, Simonich, S.M., Stone, D., Tanquay, R.L., Waters, K.M., Harper, S.L., Williams, D.E., **2016**. Completing the link between exposure science and toxicology for improved environmental health decision making: The aggregate exposure pathway framework. Environmental Science & Technology 50, 4579-4586.

Abbreviations: ADME, Absorption, Distribution, Metabolism and Elimination; AEP, Aggregate Exposure Pathway; AOP, Adverse Outcome Pathway; BMD, Benchmark Dose; BMDL, Benchmark Dose confidence interval; HI, Hazard Index; IRIS, Integrated Risk Information System; KE, Key Event; KES, Key Exposure State; LO[A]EL, Lowest Observed [Adverse] Effect Level; NAS; National Academy of Sciences; NIS, Sodium Iodide Symporter; NO[A]EL, No Observed [Adverse] Effect Level; PBPK, Physiologically Based Pharmacokinetic; PECO, Population, Exposure, Comparators, Outcomes; TH, Thyroid Hormone; TSE Target Site Exposure; TSCA, Toxic Substances Control Act

Application of the EFSA/ECHA Endocrine Disruption Guidance as a Framework for Evidence Integration in a Weight-of-Evidence (WoE) Analysis for Oxybenzone (BP-3)

Susan Borghoff, Seneca Fitch, Janice Britt, Kara Franke, Daniele Wikoff

Introduction

- Oxybenzone, also known as benzophenone-3 (BP-3), is a common chemical ultraviolet (UV) light filter used in sunscreens and other personal care products.
- Selected studies and reports have suggested that BP-3 may exhibit endocrine disruptor properties; however, a systematic review within a structured assessment framework according to specific guidelines for identifying potential endocrine disruptors across the estrogen, androgen, thyroid, and steroidogenesis (EATS) pathways has not been conducted.
- ECHA/EFSA (2018) "Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009" provides an evidence-to-decision framework for determining if a substance is considered to have endocrine disruptor (ED) properties. The approach involves integration of lines-of-evidence (LoE) that characterize modulation of endocrine pathways relative to adverse effects.

Objectives

Implement systematic review methodology to evaluate if oxybenzone has endocrine disrupting properties in humans under the conditions of use as a UV filter.

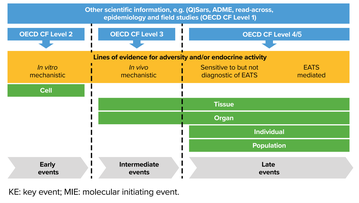
Methods

- A systematic review is in the process of being implemented according to that described in the protocol, "A Hypothesis-Driven Weight-of-Evidence Analysis to Evaluate Potential Endocrine Activity of Oxybenzone Following Exposure from UV Filters in Humans." Available: https://zenodo.org/record/2636528#.XOWm9FNKhTY
- PECO (Population, Exposure, Comparator, Outcome)
  - **Primary question:** Does oxybenzone have endocrine disrupting properties in humans under the conditions of use as a UV filter?
  - **Sub-questions**, consistent with ECHA/EFSA (2018) guidance:
    - Does oxybenzone modulate estrogen, androgen, thyroid, or steroidogenesis pathways?
    - Are any observed EATS modalities associated with adverse effect(s) (i.e., are reproductive, developmental, or carcinogenic effects a consequence of an endocrine mode of action)?
- A comprehensive literature search was conducted to identify relevant literature based on inclusion/exclusion criteria described in the protocol. Multiple databases were utilized, including PubMed, Embase, NTP CEBs, CompTox Dashboard; database-specific syntax was developed for each as warranted.
  - ToxCast/Tox21 (v3) data were downloaded and activity determined based on criteria previously defined by the U.S. EPA (Judson et al., 2016). The results of an assay are only considered to be active when the hit call is 1 and when the Z-score is ≥ 3, which signifies that assay activity was well below the cytotoxic concentration range.
  - Only completed datasets (defined as having a peer-review publication, authoritative report, or public presentation in a scientific forum) were included for this analysis.
- Title and abstract screening, as well as full text screening, were conducted in DistillerSR.
- Extraction was conducted by endpoint and EATS pathway
  - Templates were developed as part of the ECHA/EFSA guidance; however, these draft templates were refined and/or new templates developed to include high-throughput (HTS) data such as ToxCast/Tox21as well as epidemiological studies
  - Each assay/endpoint was assigned a study level (1-5), representing an increasing order of biological complexity used to evaluate strength of evidence for ED (OECD 150 GD).
    - **Level 1:** Existing data, new non-test information, e.g., PC properties, read across, QSAR, ADME, non-standardized tests, epidemiological data, etc
    - **Level 2:** *In vitro* endocrine mechanistic pathways
    - **Level 3:** *In vivo* endocrine (i.e. EATS specific) mechanistic
    - **Level 4:** *in vivo* adverse effects on endocrine-relevant endpoints
    - **Level 5:** *in vivo* comprehensive data on adverse effects on endocrine-relevant endpoints over life cycle of organism

Increasing Confidence

  - Individual assays were critically appraised and/or are in the process of being appraised for validity using a variety of approaches.
    - Though systematic review is recommended as the method for facilitating assessments, the ECHA/EFSA Guidance does not provide granular processes for critical appraisal of individual studies via risk of bias; rather, emphasis (confidence) is placed on guideline-based assays diagnostic of EATS.
    - The internal validity of epidemiological studies was evaluated using the National Toxicology Program's Office of Health Assessment and Translation's Risk of Bias Tool.
    - The validity of *in vivo* and *in vitro* studies is currently being evaluated using ToxRTool and SciRAP. For the purposes of this assessment, reliability of *in vitro* was based on the inclusion of a positive control as well as assay-account for cytotoxicity (those not meeting these criteria were not carried forward to integration herein).
- Structured evidence integration was conducted using ECHA/EFSA (2018), with considerations of processes used by the USEPA Endocrine Disruptor Screening Program as well as concepts from OECD Guideline 150. This involved using an adverse outcome pathway approach using the structured placement of studies by level as described by EFSA/ECFA (**Figure 1**).
  - Endpoints (individual assays) were structured in the AOP based on line of evidence (LoE) (i.e., multiple endpoints/assays may relate to a single LoE, such as receptor binding). *Note only the estrogen LoE has been completed herein.*
- Following completion of LoE for all modalities, a final WoE assessment will be conducted by integrating the LoEs for each pathway with considerations for assay level, reliability, consistency, magnitude, dose-response, and dose relevance to determine the plausibility of an endocrine mode of action (MoA).

Figure 1. General organizational approach for integrating evidence in an adverse outcome by level of assay in lines of evidence analyses (adapted from EFSA/ECHA 2018).



KE: key event; MIE: molecular initiating event.

Results

Evidence Identification and Extraction

- More than 60 studies/sources were identified in the peer review literature and authoritative databases (**Figure 2**).
  - Included epidemiological, experimental animal, and *in vitro* studies.
  - Each study contained 1 to >20 relevant endpoints (relating to assays), often spanning across at least 2 pathways.
- Information was extracted by endpoint and by EATS pathway per the guidance and assigned a study level (1-5), representing an increasing order of biological complexity used to evaluate strength of evidence for ED.
- Data was extracted for >200 endpoints across pathways; the majority of the evidence base consisted of level 2 (e.g., receptor binding and transactivation assays) through level 4 assays (e.g., repeated dose toxicity studies and non-guideline reproductive and developmental studies with focus on critical windows of exposure).
  - **For the estrogen pathway, data included:**
    - **Level 1:** QSAR analysis, epidemiological data
    - **Level 2:** Grouped ToxCast/Tox21 (26 ER assays, 1 aromatase assay, 10 steroidogenesis assay), ER competitive binding (2 assays), ER agonist transactivation (2 assays), ER agonist cell proliferation (1 assay), ER antagonist transactivation (1 assay); in total, 7 LoE evaluated from 43 assays reported in 5 studies/sources
    - **Level 3:** 1 LoE evaluated from 5 uterotrophic assays reported in 4 studies/sources
    - **Level 4:**
      - 20 LoE evaluated for EATS-mediated parameters from 122 measurements from 19 different studies reported in 6 sources
      - 18 LoE evaluated for sensitive to EATS parameters from 47 measurements from 17 different studies reported in 4 sources
    - **Level 5:** No qualifying studies

Figure 2. Flow diagram for implementation of search through integration.

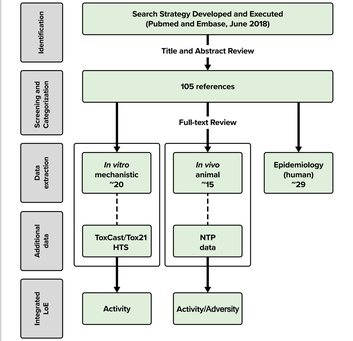
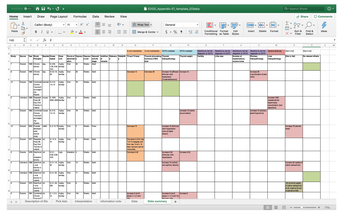


Figure 3. Example of extraction template recommended by EFSA/ECHA (figure adapted from ECHA/EFSA 2018).

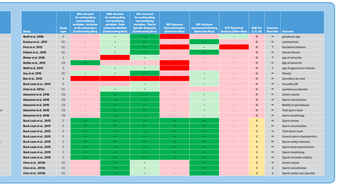


Appraisal, Synthesis, and Integration: Lines of Evidence Analysis for Estrogen Pathway

Level 1 Data:

- The majority of epidemiological studies report a lack of activity or adversity, as indicated by a lack of change in direction of the outcomes reported across the evidence base (**Table 1**).
  - A lack of associations in estradiol level changes was consistently reported in three studies; no other hormone changes were consistently reported across studies (though separate studies reported changes in some of the hormones measured).
  - A lack of changes was consistently reported across apical outcomes; for those outcomes with reported changes, none were repeated in multiple studies.
  - Formal critical appraisal of epidemiological studies demonstrated high risk of bias in key domains—primarily exposure and confounding—across most of the evidence base, resulting in Tier III categorizations for most of the epidemiological evidence base (**Table 1**).
- Weak activity predicted in QSAR ER model (weak as described by model output)

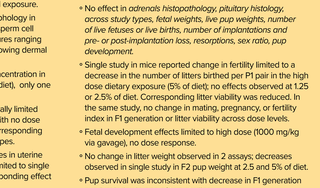
Table 1.
Summary findings and critical appraisal summary for key domains by endpoint in epidemiological studies. Endpoint direction based on author report. Study types: CT = controlled trial, C = cohort, CS = cross-sectional, CC = case-control, NCC = nested case control. Key risk of bias domains shown in the table: "–/red shading" = definitely high risk of bias, "-/pink shading" = probably high risk of bias, "+/light green shading" = probably low risk of bias, "++/green shading" = definitely low risk of bias per OHAT (2015). Overall tiers based on key domains; Tier I being most reliable based on validity and Tier III the least.



Level 2 and 3 Data (Table 2)

- Several *in vitro* studies from the peer-reviewed literature did not account for cytotoxicity and thus were not carried forward to the integration. *Formal critical appraisal of study validity beyond the cytotoxicity criteria is under way.*
- No activity in HTS assays across ER binding, agonist, and antagonist assays, along with lack of activity to inhibit aromatase or disrupt steroidogenesis.
- *In vitro* studies from NTP and studies in the literature reported that BP-3 does not compete with estradiol for binding to ER, ER agonist activity is negative to weak, and ER antagonism is inactive when accounting for cytotoxicity.
- No activity was observed in uterotrophic assays to assess estrogen agonist or antagonist activity when administered either orally or SC to rats or mice. An increase in the wet uterus weight at a dose level above the limit dose (rendering the high dose difficult to interpret) was noted in one nonguideline study conducted in an immature rat model.

Table 2. Assembly and integration of LoE for estrogen activity.



Integrated Activity Conclusion: ER agonist assays were mostly inactive. Some assays inconsistently reported weak activity. ER antagonism was inactive. All endpoints in uterotrophic assays were negative/inactive up to the limit dose of 1000 mg/kg-d in the diet, which confirms a lack of activity across the estrogen pathway.

Level 4 and 5 Data

- For estrogen, all data were Level 4 and contained endpoints that were both EATS mediated (**Table 3**) as well as sensitive to EATS.
- EATS mediated (**Table 3**)
  - Multiple studies/assays consistent demonstrated a lack of effect on: *epididymis histopathology, sperm motility, testis histopathology, uterus histopathology, prostate histopathology, mammary gland histopathology, and thyroid histopathology.* Assay conditions involved dietary exposures up to 50000 ppm for 13 weeks in rats and mice.
  - *Epididymis weight*—Several studies reported lack of effect; studies reporting effects were inconsistent between / species and doses. No corresponding effect on epididymis histopathology
  - *Estrous cycle*—Several studies show lack of effect to estrous cycle length; effects limited to high dose group following dietary exposure. No effects following dermal exposure.
  - *Sperm morphology*—No effect on sperm morphology in 5/6 studies; single study reporting change in sperm cell abnormalities in mice following dietary exposures ranging from 3125—50000 ppm in diet. No effect following dermal exposure.
  - *Sperm concentration*—Decrease in sperm concentration in 4/6 studies, mainly high dose (50000 ppm in diet), only one study observed a dose response.
  - *Testis weight*—Changes in testis weight generally limited to highest dose in dietary exposure studies, with no dose response. No effects in dermal studies. No corresponding effect on testis histopathology across study types.
  - *Uterus weight*—Inconsistent findings in changes in uterine weight; changes (associated with exposure) limited to single study and high dose (50,000 ppm). No corresponding effect on uterine histopathology across study types.
  - *Ovarian weight*—Changes in ovarian weights are not consistent, not attributed to chemical treatment (e.g, low dose only), and are associated with high dose (e.g., 50000 ppm in diet). No corresponding effect on ovarian histopathology across study types.
  - *Prostate weight*—No effect following dermal exposure. Dietary exposure study reported change following high dose exposure (single dose study). No corresponding change in prostate histopathology across study types.
  - *Seminal vesicles weight*—No effect following dermal exposure. Dietary exposure study reported change following high dose exposure (single dose study). No corresponding change in seminal vesicles histopathology across study types.
  - Database was considered limited for anogenital distance and nipple retention.
- Sensitive to EATS
  - Most LoE limited to a single study.
  - No effect in *adrenals histopathology, pituitary histology, across study types, fetal weights, live pup weights, number of live fetuses or live births, number of implantations and pre- or post-implantation loss, resorptions, sex ratio, pup development.*
  - Single study in mice reported change in fertility limited to a decrease in the number of litters birthed per P1 pair in the high dose dietary exposure (5% of diet); no effects observed at 1.25 or 2.5% of diet. Corresponding litter viability was reduced. In the same study, no change in mating, pregnancy, or fertility index in F1 generation or litter viability across dose levels.
  - Fetal development effects limited to high dose (1000 mg/kg via gavage), no dose response.
  - No change in litter weight observed in 2 assays; decreases observed in single study in F2 pup weight at 2.5 and 5% of diet.
  - Pup survival was inconsistent with decrease in F1 generation survival index in one of two studies.

Adversity: In rodent models exposed to oxybenzone levels in the diet < 10,000 ppm, there were no consistent adverse effects across study types in endpoints mediated through a disruption in the estrogen pathway.

Table 3. Assembly and integration of LoE for estrogen adversity. LoE based on EFSA/ECHA guidance. Some LoE contain multiple studies/assays represented by a single role; for brevity, multiple studies/assays are combined when the findings were all similar (no effect). For each LoE, the effect dose (lowest observed effect level) or no effect level (Italics) is provided. LoE with insufficient data are not shown (anogenital distance, nipple retention)

Overall LoE Conclusion for Estrogen:

Lack of activity in uterotrophic assay demonstrates that weak agonist activity observed *in vitro* does not occur *in vivo*; this is supported by inactivity across most *in vitro* assays and epidemiological data. In rodent models exposed to oxybenzone levels in the diet < ~10,000 ppm, there were no consistent adverse effects across study types in endpoints mediated through a disruption in the estrogen pathway. Based on experimental evidence evaluated herein, with context provided from epidemiological studies, oxybenzone would not be considered an ED as it relates to the estrogen pathway at current exposure levels as a UV filter.

Conclusions

Methodology

- Use of the ECHA/EFSA guidance in WoE analyses, along with templates modified to incorporate HTS and epidemiological data, allow for systematic and transparent organization and integration of heterogenous data in characterizing ED potential.
  - Utilization of an AOP-approach allows for some aspects of construct and external validity to be operationalized during evidence integration.
  - For BP-3, no AO was identified (associated with the estrogen LoE); thus the AOP approach allowed for organization of evidence in a manner that provided important context for interpretation of *in vitro* assays in particular.
- Evaluation of ED properties cannot be characterized alone by single *in vitro* assays, but instead requires evaluation by LoE across endocrine pathways to evaluate individual hypotheses. Each LoE itself is based on weight of evidence, providing transparency in evaluating the totality of data.
- The templates from EFSA/ECHA are useful, though refinements are needed to provide sufficient information (e.g., dose, assay conditions, etc.) to appraise studies and provide sufficient information to integrate findings.

Oxybenzone

- The evidence base for oxybenzone is relatively large and quite heterogenous, containing a diverse array of assays that contribute to understanding the potential for ED activity and adversity.
  - Although it does not include any Level 5 assays; i.e., updated guideline 2-generation DART or the EOGRD assay, there are a number of validated, guideline-based studies available to inform these lines of evidence and conduct an analyses.
- Confidence in guideline type studies is typically greater than other evidence from the peer-reviewed literature in which aspects of study validity (e.g., *in vitro* assays which did not account for cytotoxicity, epidemiological studies in which there is a high risk of bias for exposure and confounding domains) limit the reliability.
  - The majority of guideline studies are based on dietary exposures to very high doses of BP-3 (e.g., 50000 ppm or ~18,500 mg/kg-day in mice), thus limiting the generalizability (i.e., external validity) in much of the evidence base. Additional work is ongoing to characterize the construct and dose relevance of these assays relative to dermal exposure to BP-3 in humans.
- Collectively, the receptor binding assays for BP-3 suggest weak - if any - activity as a ER agonist or AR antagonist; though a lack of activity was observed in Uterotrophic and Hershberger assays, strengthening the findings that BP-3 does not exhibit estrogen agonist and androgen antagonist activity *in vivo*. Across the evidence base, changes in apical outcomes associated with estrogen agonism and androgen antagonism were limited to high dose exposure and were not consistently reflected across study types. Taken together, evidence support that dietary exposure to <~10,000 ppm BP-3 are not associated with adverse effects on the reproductive system in rats. And subsequently, BP-3 is unlikely to have ED properties related to estrogen and androgen pathways at current exposure levels as a UV filter.
  - These interim findings will be expanded to include other EATS modalities, as well as include formal critical appraisal of *in vivo* and *in vitro* data, and inclusion of evidence via an updated literature search as part of completing the systematic review.

Acknowledgements and Disclosures

The authors gratefully acknowledge the technical support of Grace Chappell for her assistance with the HTS data and Mina Suh for her assistance with extraction and appraisal of epidemiological data.

This work is funded by Johnson & Johnson. Consistent with guidance provided by the IOM (2011), the sponsors were given the opportunity to review and provide input on the protocol and subsequently the synthesis and integration of material presented in this poster with the purpose of allowing input on the clarity of reporting.

References available upon request

Using In vitro ToxCast Assays to Evaluate Mechanistic Plausibility and Build Confidence in the Selection of Analogues for Quantitative Read-Across: A Case Study on *p,p'*-Dichlorodiphenyldichloroethane

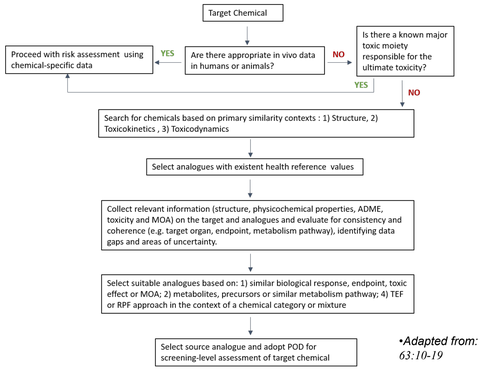
Lucina E. Lizarraga, Jeffry L. Dean, J. Phillip Kaiser, Scott C. Wesselkamper, Jason C. Lambert, Elizabeth O. Owens, Belinda Hawkins, Q. Jay Zhao.
National Center for Environmental Assessment (NCEA), U.S. Environmental Protection Agency, Cincinnati, Ohio 45268, USA

Lucina E. Lizarraga | Lizarraga.Lucina@epa.gov | 513-487 2648

Overview

- Deriving human health reference values for environmental chemicals has traditionally relied on toxicity data from humans and/or experimental animals
- In the absence of in vivo toxicity data, new approach methodologies such as read-across can be used to fill data gaps for a target chemical using known information from a source analogue
- A read-across approach illustrated below (Figure 1) was applied to assist in screening-level assessment of noncancer oral toxicity for the target, *p,p'*-DDD, a data-poor chemical known to occur at contaminated sites in the U.S.

Figure 1: Read-across Approach



- Analogues were identified and evaluated for similarities in structure and physicochemical properties, toxicokinetics, and toxicodynamics (toxicity and in vitro bioactivity) with respect to the target chemical
- The primary focus of this investigation was to evaluate the integration of mechanistic evidence from in vitro high-throughput screening (HTS) assays from ToxCast in support of the similarity justification for the selection of analogues for quantitative read-across

•Adapted from: Wang et al., 2012, Regul Toxicol Pharmacol 63:10-19

Structural and Toxicity Similarity Comparisons

Identification of Structural Analogues of *p,p'*-DDD

Table 1. Structural Analogues of p,p'-DDD

	Target Chemical	Analogues[a]		
Name	p,p'-Dichlorodiphenyl dichloroethane (p,p'-DDD)	p,p'-Dichlorodiphenyl trichloroethane (p,p'-DDT)	p,p'-Dichlorodiphenyl dichloroethylene (p,p'-DDE)	p,p'-Dimethoxydiphenyl trichloroethane (Methoxychlor)
CASRN	72-54-8	50-29-3	72-55-9	72-43-5
Structure				
ChemIDplus similarity score (%)	100	77	67	65
DSSTox similarity score (%)	100	96	61	52

[a]Analogues represent a set of structurally similar chemicals identified using two publicly available similarity databases (ChemIDplus and DSSTOX) prefiltered on the basis of availability of health reference values for non-cancer oral toxicity from regulatory agencies, including ATSDR (2002a, b) and U.S. EPA (2017 b, c).

Putative Toxicity Targets for *p,p'*-DDD and Analogues Include the Liver and Reproductive System in Animals

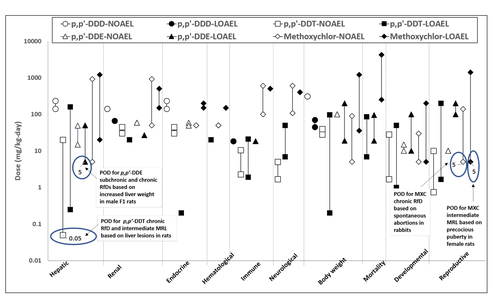


Figure 1. Comparison of Health Effects and Associated Effect Levels for Non-Cancer Oral Toxicity. Range of effect levels (no-observed-adverse-effect levels [NOAEL] and lowest-observed-adverse-effect levels [LOAEL]) for noncancer endpoints for the target and analogues from repeated-dose animal toxicity studies via oral administration reported by ATSDR (2002a, b) and U.S. EPA (2017 b, c). Circles note points-of-departure (PODs) used in the derivation of oral reference doses (RfDs) and minimal risk levels (MRLs) for these chemicals (ATSDR, 2002a,b; U.S. EPA 1987c, 1999, 2017a).

Bioactivity Similarity Comparisons Evaluating Mechanistic Plausibility for Liver and Reproductive Toxicity

p,p'-DDD and Analogues Exhibit Similarities in Cell-specific Responses and Target Gene Pathways in *In Vitro* ToxCast Assays Conducted in Human Liver Cells

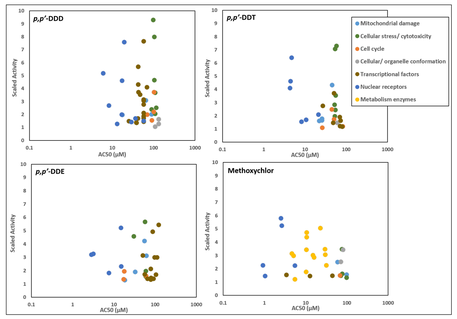


Figure 2. Bioactivity data for *p,p'*-DDD and Analogues in ToxCast Assays Conducted in Human Hepatoma HepG2 Cells and Primary Human Hepatocytes. Scatterplots show AC50 and scaled activity values for *p,p'*-DDD, *p,p'*-DDT, *p,p'*-DDE and methoxychlor from *in vitro* assays visualized according to the type of biological response or biological target. AC50 values refer to the concentration that elicits half maximal response and the scaled activity refers to the response value divided by the activity cutoff. Metabolism enzyme-related assays were conducted in human primary hepatocytes and all other *in vitro* assays were measured in HepG2 cells. Assays for which chemicals were inactive are not displayed. Data were sourced from the EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) (U.S. EPA, 2017a).

(Lizarraga et al., 2019, *Regul Toxicol Pharmacol* 103:301-313)

p,p'-DDD and Analogues Exhibit Similar Estrogenic and Anti-Androgenic Activities in *In Vitro* ToxCast Assays and Model Predictions for the ER and AR Across Multiple Tissues and Cell lines

Table 2. ToxCast Bioactivity Summary and Model Prediction Scores (AUC values) for ER and AR activities[a]

	p,p'-DDD	*p,p'*-DDT	*p,p'*-DDE	Methoxychlor
ER assays				
Active/Total Assays (%)	7/18 (39)	11/18 (61)	8/18 (44)	14/18 (78)
AC50 values (μM)	Range = 14.0 - 32.4 Median = 18.7	Range = 3.3 - 59.8 Median = 6.1	Range = 3.5 - 46.2 Median = 16.5	Range = 0.9 - 44.2 Median = 4.6
Agonist activity AUC value (95% CI)[b]	0.0715 (0.0342-0.0738)	0.190 (0.181-0.231)	0.0679 (0.0614-0.0963)	0.254 (0.247-0.260)
Antagonist activity AUC value (95% CI)	0	0	0	0
AR assays				
Active/Total Assays (%)	4/11 (36)	3/11 (27)	4/11 (36)	3/11 (27)
AC50 values (μM)	Range = 31.0 - 62.8 Median = 44.8	Range = 17.8 - 72.0 Median = 47.0	Range = 7.0 - 58.7 Median = 29.6	Range = 29.3 - 40.8 Median = 34.2
Agonist activity AUC value (95% CI)	0	0	0	0
Antagonist activity AUC value (95% CI)	0.0973 (0.0649-0.124)	0.0642 (0.0318-0.108)	0.251 (0.234-0.291)	0.0429 (0.0364-0.0465)

[a]Data were sourced from Judson et al. (2015) and Kleinstreuer et al. (2016). [b] 95% CI for the ER activity model were sourced from a subsequent publication to the Judson et al., (2015) study (Watt and Judson, 2018).
Abbreviations: AUC = area under the curve score ranging from 0–1. An AUC value of 0 indicates that the chemical is inactive; CI = confidence interval.

(Lizarraga et al., 2019, *Regul Toxicol Pharmacol* 103:301-313)

p,p'-DDD and Analogues Exhibit Similar Upregulation of Steroid/Xenobiotic-sensing Nuclear Receptors in *In Vitro* ToxCast Assays Conducted in Hepatoma HepG2 Cells

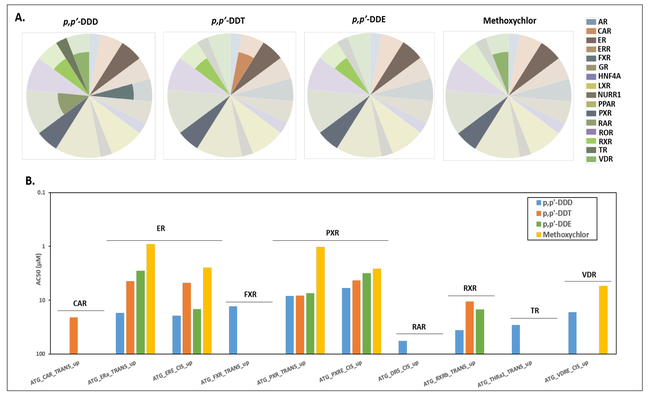


Figure 3. ToxCast Assays Evaluating Regulation of Nuclear Receptor Activity for *p,p'*-DDD and Analogues in Human Hepatoma HepG2 Cells. Panel A shows radar plots for *p,p'*-DDD, *p,p'*-DDT, *p,p'*-DDE and methoxychlor, summarizing active calls from nuclear receptor assays conducted in HepG2 cells and mapped to specific target genes. The shaded area of the pie slice represents the number of active assays as a proportion of total assays. The width of the slice refers to the proportion of assays within a given target gene. Bar graphs compare AC50 values (concentration at half maximal response) for active assays (panel B). The scale for the AC50 values is shown in reverse order to visualize the most sensitive nuclear receptor activities (the higher bar indicates a lower AC50 value). Data were sourced from the EPA's CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) (U.S. EPA, 2017a).
Abbreviations: AR, androgen receptor [■]; CAR, constitutive androgen receptor [■]; ER, estrogen receptor [■]; ERR, estrogen-related receptor [■]; FXR, farnesoid X receptor [■]; GR, glucocorticoid receptor [■]; HNF4A, hepatocyte nuclear factors 4 alpha [■]; LXR, liver X receptor [■]; NURR1, nuclear receptor related-1 protein [■]; PPAR, peroxisome proliferator-activated receptor [■]; PXR, pregnane X receptor [■]; RAR, retinoid acid receptor [■]; ROR, RAR-related orphan receptor [■]; RXR, retinoid X receptor [■]; TR, thyroid hormone receptor [■]; VDR, vitamin D receptor [■].

(Lizarraga et al., 2019, *Regul Toxicol Pharmacol* 103:301-313)

Summary and Conclusion

- The current read-across approach relies on the evaluation and integration of evidence across three primary similarity contexts (structure, toxicokinetics and toxicodynamics) for the selection of a suitable source analogue for screening-level quantitative assessment of the target, *p,p'*-DDD (Table 3)
- Analysis of ToxCast assays reveal similarities between *p,p'*-DDD and analogues in *in vitro* responses related to mitochondrial damage, cellular stress/cytotoxicity and the upregulation of specific steroid/xenobiotic-sensing nuclear receptors (Figures 2 and 3) that are relevant to their mechanism of hepatotoxicity
- ToxCast assays and model predictions suggest that *p,p'*-DDD and analogues may act as ER agonists and AR antagonists (Table 2), coinciding with the estrogenic and anti-androgenic reproductive effects observed *in vivo*
- Coherence across *in vivo* toxicity and *in vitro* bioactivity similarity comparisons help reduce uncertainties associated with toxicity data gaps for the target
- These findings demonstrate the utility of integrating evidence from HTS data platforms to support mechanistic conclusions and increase confidence in the application of read-across in quantitative risk assessment

Evidence Integration

Table 3. Using Evidence Integration to Identify Suitable Source Analogues for Read-across

Similarity Context	Summary of Findings	Evidence Integration Conclusions
Structure and physicochemical properties	• *p,p'*-DDD and identified analogues (*p,p'*-DDT and *p,p'*-DDE and methoxychlor) demonstrate similarities in basic structural features (chlorinated diphenylalkane structure) • *p,p'*-DDT and *p,p'*-DDE also share key functional groups (*p,p'*-chlorine substituents) and physicochemical properties important for bioavailability (lipophilicity and low BCF values) with *p,p'*-DDD	• *p,p'*-DDT is selected as a suitable source analogue for the assessment of non-cancer oral toxicity of *p,p'*-DDD based largely on toxicokinetic similarities, with supportive information from *in vivo* toxicity testing, structural similarity evaluations and *in vitro* bioactivity from HTS assays
Toxicokinetics	• *p,p'*-DDT is a metabolic precursor of *p,p'*-DDD and both chemicals show similarities in toxicokinetics (Absorption, Distribution and Metabolism [ADME]) in humans and experimental animal models (preferential partitioning into fat, similar metabolism and excretion pathways and prolonged elimination rates) • Other analogues demonstrate differences in ADME in comparison to the target. *p,p'*-DDE is less metabolically active; methoxychlor is metabolized differently and appears to be less bioaccumulative	
Toxicodynamics	• Consistency and coherence across health effects in experimental animals for non-cancer oral toxicity among the analogues point to putative toxicity targets for *p,p'*-DDD (primarily liver and reproductive toxicity) • Similarities in *in vitro* bioactivity profiles from ToxCast assays between the target and analogues with respect to cell-specific responses and target gene pathways provide mechanistic plausibility for the liver and reproductive effects associated with this group of chemicals	

U.S. Environmental Protection Agency
Office of Research and Development

Targeted Mechanistic Evidence Synthesis to Inform Evidence Integration Decisions on the Potential Human Carcinogenicity of Naphthalene Exposure

Ingrid L. Druwe[1], Janice S. Lee[1], Kristina Thayer[1], John Bucher[2], Erin E. Yost[1]

1. US Environmental Protection Agency, National Center for Environmental Assessment, Research Triangle Park, NC

2. National Toxicology Project, National Institutes of Environmental Health Sciences, Research Triangle Park, NC

www.epa.gov

Ingrid L. Druwe | druwe.ingrid@epa.gov | 919-541-2452

Background

Naphthalene has been demonstrated to cause respiratory tumors in rats and mice, but the few available epidemiologic studies are inadequate to evaluate the potential for naphthalene to cause cancer in humans. In lieu of human studies, mechanistic information may be used to inform the potential carcinogenicity of naphthalene for human health risk assessment.

Multiple modes of action (MOAs) for naphthalene-induced carcinogenesis have been proposed based on animal and in vitro studies, including genotoxicity, cytotoxicity, and sustained regenerative cell proliferation. While these proposed MOAs may differ in specific key events, the formation of toxic naphthalene metabolites and the biological relevance of these toxic metabolites to humans has emerged as a key component in answering the question of applicability of carcinogenic risk to humans. There is a great deal of similarity between the rodent and human naphthalene metabolic pathways; however, the activity of the enzymes involved in naphthalene metabolism and therefore the number of metabolites and stereoisomers of the produced metabolites may differ between rodents to humans.

Here, concurrent with a broad systematic review of health effects related to naphthalene exposure, animal and in vitro studies of the available mechanistic evidence was analyzed to (1) integrate the available evidence in vitro models on the formation and toxicity of each of the key toxic metabolites of naphthalene and (2) determine the biological plausibility that each of these key metabolites could be generated in human tissue and increase human oncogenic risk.

Methods

Literature Search and Tagging: Mechanistic studies were identified by tagging studies during screening of the broad literature search focused on the potential human health impacts associated with napthalene exposure.

Study evaluation: Studies tagged as mechanistic were evaluated using the SciRAP web tool (www.scirap.org) for either in vivo or in vitro study evaluation for factors rated to reporting quality, methodological quality, and relevance. SciRAP was selected for this evaluation because it has both in vivo and in vitro study evaluation tools available.

Evidence synthesis: For the specific question of metabolic relevance, we used the metabolic pathway for napthalene (developed from rodent models) as a scaffold and then evaluated studies that addressed the applicability of this metabolic pathway to humans, focusing on three key napthalene metabolites (Figure 1): 1S,2R-naphthalene oxide, 1,2-naphthoquinone, and 1,4-naphthoquinone. Studies that had deficiencies in reporting critically important study details (e.g., missing experimental exposure details) were excluded.

The evidence regarding the formation, toxicity, and human relevance of these three key naphthalene metabolites was integrated in a tabular format describing the formation and toxicity of each metabolite, factors that increase strength of evidence, and factors that decrease strength of evidence (Table 1).

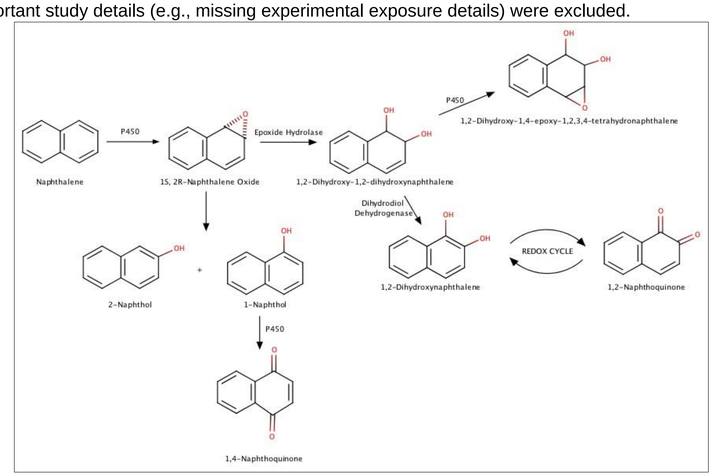


Figure 1. Naphthalene Metabolic Pathway

SciRAP Study Evaluation Results

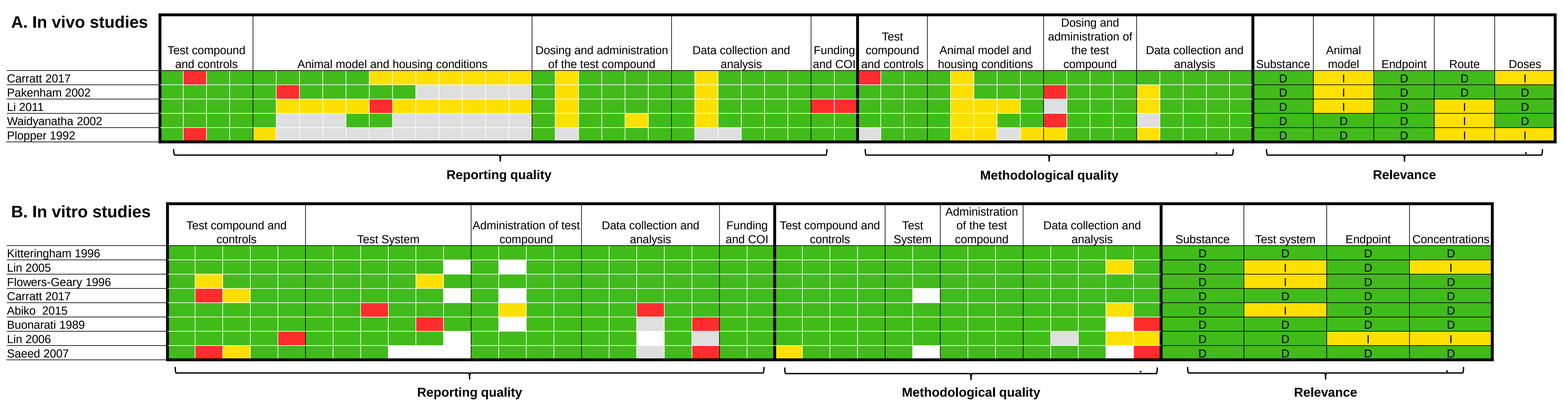


Figure 2. Representative study evaluation results. Representative studies examining three key naphthalene metabolites of interest (see Figure 1) were evaluated using SciRAP tool (n= 5 in vivo studies, n= 8 in vitro studies). For reporting and methodological quality criteria, green = fulfilled, yellow = partially fulfilled, red = not fulfilled, gray = not determined, and white = not applicable. For relevance categories, green (D) indicates that the study design was directly relevant to human health, and yellow indicates that the study design was indirectly relevant to human health.

Evidence Synthesis

*Identification number in EPA's Health & Environmental Research Online (HERO) database

	References [HERO ID*]	Factors that increase strength	Factors that decrease strength	Summary of evidence
1S,2R-Naphthalene oxide	**In vivo** • Plopper, 1992 [1469611] • Waidyanatha, 2002 [1469054] • Li, 2011 [1005231] **In vitro** • Buonarati,1989 [94674] • Buckpitt, 1992 [067441] • Lanza, 1999 [1489430] • Wilson, 1995	• No serious reporting or methodological quality limitations • Metabolite formation and cytotoxicity observed in models with greater directness (nonhuman primates and humanized mice) [Buckpitt, 1992; Li, 2011]	• Indirectness in some studies (studies in isolated rodent primary hepatocytes; route of in vivo exposure i.p. [Plopper, 1992] • Inconsistency (potential lack of metabolite formation and cytotoxicity in vitro) [Lanza, 1999; Wilson, 1995]	• CYP450 activity varies across species and determines severity of cytotoxicity produced by 1,2-naphthalene oxide [Buonarati, 1989; Plopper, 1992] • 1,2-naphthalene oxide is produced as two isomers: 1S,2R- (predominant human form) and 1R,2S. Animal studies suggest the 1S,2R isomer's cytotoxicity is > the 1R, 2S isomer [Buckpitt, 1992]. Conversely, in vitro assays in lymphoblastoid cells showed that napthalene oxide was not genotoxic in a sister chromatid exchange (SCE) assay [Wilson, 1995]. • Human CYP2A13 and 2F1, which catalyze the formation of 1,2-naphthalene oxide, were demonstrated to bioactivate naphthalene and induce toxicity in humanized transgenic mice [Li 2011] at occupationally relevant exposure levels. Conversely, microsomal assays found that recombinant human CYP2F1 had <0.1% the rats of metabolism observed with the mouse orthologue [Lanza, 1999].
1,2-Naphthoquinone	**In vivo** • Waidyanatha, 2002 [1469054] • Carratt, 2017 [345264] **In vitro** • Abiko, 2015 [4331236] • Carratt, 2017 [345264] • Flowers-Geary, 1996 [1012266] • Kitteringham, 1996 [1469475] • Saeed, 2007 [517040] • Wilson, 1996 [081049]	• No serious reporting or methodological quality limitations • Multiple positive mutagenicity assays including salmonella and SCE assays [Flowers-Geary, 1996]. • Cytotoxicity observed [Carrat, 2017; Kitterhangham 1996;]	• Indirectness in in vitro studies that observed effects (direct incubation with DNA and/or in vitro studies; mutagenicity assays were all tested in conditions that did not have an exogenous metabolic system) [Wilson, 1996; Saeed, 2007] • Mutagenesis assay information all came from a single source [Flowers-Geary 1996]	• 1,2-naphthoquinone produces cytotoxicity and increased formation of reactive oxygen species [Carratt, 2017; Kitteringham, 1996]. • 1,2-naphthoquinone forms adducts with proteins and DNA adducts that are linked to mutagenicity, chromosome aberrations, tumor promotion, and cancer [Abiko, 2015, Waidyanatha, 2002; Saeed, 2007; Flowers-Geary, 1996]. • In addition, 1,4-Naphthoquinone produced a dose dependent increase in SCE in vitro [Wilson, 1996]
1,4-Naphthoquinone	**In vivo** • Waidyanatha, 2002 [1469054] **In vitro** • Abiko, 2015 [4331236] • Lin, 2005 [148718] • Lin, 2006 [1468615] • Destephano-Shields, 2010 [1467694] • Wilson, 1996 [081049]	• No serious reporting or methodological quality limitations • Directness in the study by DeStephano-Shields, 2010 adducts formed in non-human primates after in situ exposure	• Indirectness in some studies that observed effects (direct incubation with DNA in vitro; proteomics study; route of exposure in vivo) [Lin, 2005; Lin, 2006].	• 1,4-naphthoquinone leads to protein and DNA adduct formation that are linked to chromosome aberrations, tumor promotion, and cancer [Abiko 2015; Lin 2005, Lin 2006, Waidyanatha, 2002] • In addition, 1,4-Naphthoquinone produced a dose dependent increase in SCE in vitro [Wilson, 1996]

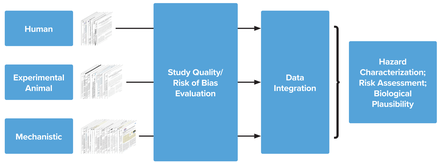
Table 1 Evidence profile table describing a summary of the toxicological evidence for each of the known naphthalene metabolites

- The available evidence showed that 1S,2R-naphthalene oxide (the prevalent naphthalene metabolite in humans) is a highly reactive metabolite that is more toxic and metabolized more slowly than the 1R,2S enantiomer more commonly observed in mice, which may allow it more time to produce cytotoxicity.
- 1S,2R-naphthalene oxide can be metabolized to 1,2-naphthoquinone or 1,4-naphthoquinone (Figure 1), which have been shown to elicit cytotoxicity. These quinone metabolites can bind to proteins and have been demonstrated in situ and across species (including non-human primate tissue) to form protein adducts. In addition, these quinones may also undergo protein adduction and disrupt normal cellular function by binding to CYP450 enzymes and to proteins involved in cell signaling and transduction.
- The electrophilic nature of 1,2- and 1,4-napthoquinone cause these metabolites to undergo 1,4-Michael addition and covalently bind to DNA, forming depurinating N3Ade and N7Gua adducts as well as stable adducts. Therefore, it is biologically plausible for the reactive naphthalene metabolites 1,2- and 1,4-naphthoquinone to form depurinating and stable DNA adducts.

Tiered approach to integrating evidence streams assessing *in utero* trichloroethylene exposure and development of congenital heart defects (TCE-CHD)

Urban J[1]; Wikoff, D[2]; Haws, L[1]

[1]ToxStrategies, Austin, TX 78759. [2]ToxStrategies, Asheville, NC 28804.



Introduction

- Mechanistic data can be useful for interpreting findings in human and animal studies, helping elucidate biological pathways that may help determine biological plausibility of an adverse effect and potentially inform human relevance. Compared to epidemiology and animal toxicology evidence, however, the challenge of objectively evaluating study quality and applicability is considerably more complex for mechanistic datasets due to the considerably more heterogeneous and often non-standard, alternative experimental approaches.
- Efforts are underway to develop a systematic approach for the evaluation and integration of mechanistic data with human observational and animal experimental study data to facilitate use in the risk assessment process, particularly in cases where the human and animal literature may be limited, weak, and/or conflicting.
- The body of evidence relevant to the TCE-CHD hypothesis — i.e., *in utero* exposures to trichloroethylene (TCE) increase risk of congenital heart defects (CHDs) — presents a relevant dataset for developing a data integration approach for mechanistic data. In this case, prior systematic evaluation demonstrated that the epidemiological and animal evidence is not suitable for development of noncancer toxicity values based on CHDs as a result of a variety of shortcomings and limitations in the underlying evidence base (Wikoff et al., 2018). However, mechanistic studies were not assessed in this prior systematic evaluation due to limitations in the available tools for doing such at the time.

Objective

To identify, appraise, and integrate the mechanistic evidence stream to assess the larger TCE-CHD database.

Table 1. Study Quality Evaluation for *In vitro* Experiments—Key Metrics for TCE-CHD Database

TSCA Metric No.	Metric Title	Metric Description
8	**Preparation and Storage of Test Substance**	Did the study characterize preparation of the test substance and storage conditions? Were the frequency of preparation and/or storage conditions appropriate to the test substance stability and solubility (if applicable)?
11	**Exposure Duration**	Was the exposure duration (e.g., minutes, hours, days) reported and appropriate for this study type and/or outcome(s) of interest?
16	**Outcome Assessment Methodology**	Did the outcome assessment methodology address or report the intended outcome(s) of interest? Was the outcome assessment methodology (including endpoints and timing of assessment) sensitive for the outcome(s) of interest (e.g., measured endpoints that are able to detect a true effect)?
22	**Data Analysis**	Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)?
24	**Cytotoxicity**	Were cytotoxicity endpoints defined, if necessitated by study type, and were methods for measuring cytotoxicity described and commonly used for assessments?

Methods

- **Development of TCE-CHD Evidence Base (Literature Search):**
  - Using handsearching and reference chasing, mechanistic studies were identified from recent comprehensive reviews conducted systematically (Makris et al., 2016; Wikoff et al., 2018). Additional PubMed and Embase searches were also conducted using the same search syntax was conducted to capture relevant studies published since Wikoff et al. (2018). Searches were executed October 30, 2018.
  - Mechanistic studies were categorized based on the assay type(s) to accommodate the TSCA study quality tool: *in vivo* (animals exposed), *in vitro* (cell culture, *in ovo*, *ex ovo*, *ex vivo*).
- **Critical Appraisal Tool**
  - **TSCA Study Quality Evaluation:** Primary tool applied to TCE-CHD assays. Mechanistic datasets were scored based on two study categories (*in vivo* and *in vitro*) with specific evaluation and scoring metrics (n=24 and 25, respectively), with each metric scored on 1 of 4 criteria; overall study quality is determined by weighted scoring calculations and categorizations. A pilot study of 4 studies identified five key metrics applied for screening purposes (**Table 1**).
- **Study Quality Assessment Procedure**
  - Quality assessments were conducted by two PhD scientists with experience reviewing mechanistic (GC, JU) studies. In cases of conflict, a third scientist (DW) was consulted to facilitate a consensus solution.
- **Mechanistic Data Integration and Body of Evidence Assessment**
  1. Datasets were evaluated using a tiered approach (**Figure 1**).
     - **Hazard:** Adverse endpoints related to heart development reported among the various mechanistic models would provide initial support for TCE as a potential CHD hazard. Studies meeting all TSCA quality standards provide stronger support than those determined to be unreliable due to poor quality. Studies that contradict adverse endpoint datasets weaken hazard potential.
     - **Quantitative Risk:** Only studies that meet all TSCA quality metrics qualify for consideration. Elements relevant to dose-response consideration included: if effect was adverse or not; if endpoint could be phenotypically anchored to CHD (e.g., is gene(s) under study known to be linked to heart defects?); if model is amenable to extrapolating dose-effect level to human maternal exposure levels (i.e., are PBPK models available for non-mammalian models?).
     - **AOP:** The putative adverse outcome pathway (AOP) for CHDs posited by Makris et al. (2016) (**Figure 2**) was used to structure the AOP-based assessment (despite lack of adverse outcomes observed in experimental animal studies).
  2. The confidence-rating factors for mechanistic datasets proposed by OHAT (2015) were also adopted to assess the body of evidence

Figure 1. Three-tiered integration approach for mechanistic evidence base

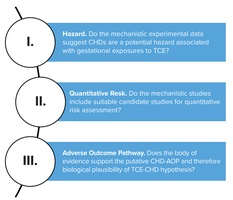
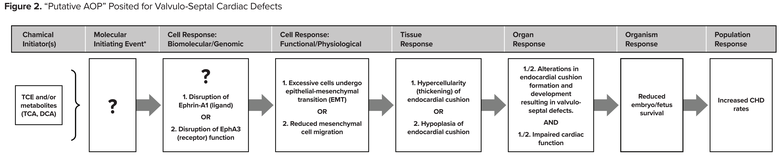


Figure 2. "Putative AOP" Posited for Valvulo-Septal Cardiac Defects



*Makris et al. (2016) note there is no MEI in this AOP, and speculate on what the subsequent KE might be, suggesting that the ephrin-EPH system could "be of high relevance". There are no TCE/TCE metabolite data that indicate a potential MEI or subsequent KE for this theoretical pathway.

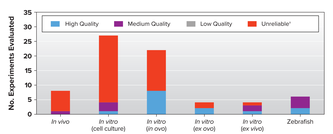
Results

- **TCE-CHD Mechanistic Evidence Base (Table 2)**
  - Consists of 22 studies reporting on 71 relevant experiments.
  - Majority of the experiments were comprised of traditional cell or tissue culture models (27 assays, mostly originating from rat and mouse tissues, with some human) and *in ovo* models (22 assays). The remaining models included *in vivo* (8 assays from rat and mouse), ex *ovo* & *ex vivo* (4 whole embryo assays each for chicken and rat), and recombinant zebrafish (6 assays) experiments.
  - The endpoints were fairly heterogeneous as well (presented by biological organization):
    - **Cell biomolecular effects** (models representing all species in TCE-CHD database): About ½ of all assays (35/71) examined effects on gene expression, epigenetics, protein expression/interaction/activity.
    - **Cell physiology effects** (human, bovine, rat, chicken cells only): A small fraction of all assays (6/71) examined physiological effects at the cellular level (morphology, function, proliferation).
    - **Tissue level effects** (chicken only): A small fraction of all assays (8/71) examined the morphology/cell makeup at the heart tissue level.
    - **Organ effects** (rat, chicken, zebrafish): Several assays (11/71) examined effects on the heart.
    - **Organism effects** (rat, mouse, chicken, zebrafish): Several assays (11/71) examined effects on the survival/viability of the developing fetus, embryo or larva.
- **Critical Appraisal of Mechanistic Datasets (note critical appraisal of studies from other evidence streams [i.e., human and *in vivo* animal] previously conducted by Wikoff et al., 2018):**
  - Quality rankings based on the TSCA tool varied by study model (**Figure 3**).
  - Aspects that commonly differentiated studies within the tool included reporting on the preparation and storage of the test substance (Metric 8), some element of data analysis (Metrics 22 and/or 23), and reporting on cytotoxicity (Metric 24, only relevant to cell culture experiments) (**Figure 4**).
  - Study quality categorizations were overall similar for the subset of experiments assessed using SciRAP and ToxRTool (*data previously reported*).
  - Seven studies of heterogeneous model/design reported mechanistic datasets that met EPA TSCA study quality standards, with highly variable dose responses reported (**Table 3**).

Table 2. TCE-CHD Mechanistic Evidence Base

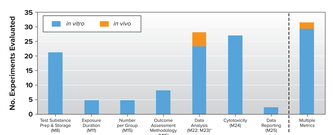
Reference	Number of assays relevant to TCE-CHD	≥1 Assay(s) meeting all study quality metrics?	% TCE-CHD assays meeting all study quality metrics
Wirbisky et al. (2016)	6 (zebrafish)	Yes (6 of 6)	100%
Drake et al. (2006a)	4 (*in ovo*)	Yes (4 of 4)	100%
Drake et al. (2006b)	4 (*in ovo*)	Yes (4 of 4)	100%
Jiang et al. (2016)	4 (*in vitro* – human)	Yes (4 of 4)	100%
Saillenfait et al. (1995)	3 (ex vivo – rat)	Yes (3 of 3)	100%
Mishima et al. (2006)	2 (*ex ovo*)	Yes (2 of 2)	100%
Caldwell et al. (2010)	3 (*in vivo* - mouse)	Yes (1 of 3)	33%
Boyer et al. (2000)	4 (*in vitro*)	No	0%
Bross et al. (1983)	2 (*in ovo*)	No	0%
Caldwell et al. (2008)	2 (*in vitro* - rat)	No	0%
Collier et al. (2003)	2 (*in vivo* - rat)	No	0%
Elovaara et al. (1979)	2 (*in ovo*)	No	0%
Harris et al. (2018)	1 (*in vitro* – human) 1 (*in ovo*) 1 (*ex ovo*)	No	0%
Loeber et al. (1988)	1 (*in ovo*)	No	0%
Makwana et al. (2010)	3 (*in ovo*)	No	0%
Makwana et al. (2013)	2	No	0%
Ou et al. (2003)	4 (*in vitro* – bovine)	No	0%
Palbykin et al. (2011)	3 (*in vivo* – rat) 4 (*in vitro* – rat)	No	0%
Rufer et al. (2010)	3 (*in ovo*)	No	0%
Selmin et al. (2005)	1 (ex vivo – rat) 4 (*in vitro* – rat)	No	0%
Selmin et al. (2008)	2 (*in vitro* – mouse)	No	0%
Selmin et al. (2014)	1 (*ex ovo*) 2 (*in vitro* – rat)	No	0%
Total (22 References)	71 Assays	7 References, 24 assays	34%

Figure 3.
TCE-CHD Mechanistic Studies by Model Type* and Study Quality Category.



* The TSCA *in vitro* study quality metrics were used to evaluate experiments designed around cell culture (*in vitro*"), chicken egg ("*in ovo*"), whole embryo culture (chicken "*ex vivo*" and rodent "ex vivo") and zebrafish models. ^ Did not meet all quality metrics.

Figure 4.
The most common mechanistic assay deficiencies.



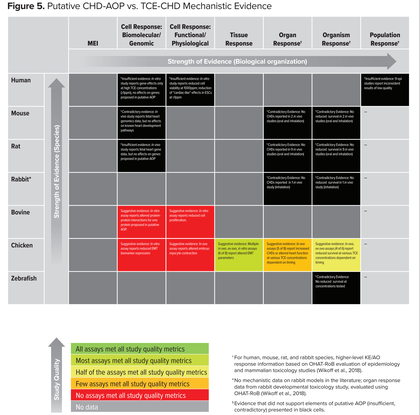
* Data Analysis is covered in TSCA Metric 22 for *in vitro* study quality metrics and TSCA Metric 23 for *in vivo* study quality metrics.

Table 3. Summary of TCE-CHD assays that met study quality metrics

Reference	Type	Assay Model	Exposure	No. Endpoints - description	Reported Outcome
Caldwell et al. (2010)	*in vivo*	Mouse – Fetuses (GD10) from dams	Drinking water: 0, 10ppb TCE from GD1-10.	1 - Microarray gene expression	LOEL: 10ppb TCE. Altered gene expression in GD10 fetal hearts; no genes directly relevant to CHD.
Drake et al. (2006a)	*in ovo*	Chicken - fertilized egg	4x yolk injections (via shell hole) at HH stages 3+, 6, 13 and 17. "cumulative" doses of 0, 0.4, 4, 8, 40, or 400 ppb TCE.	4 - Cell apoptosis and proliferation in OFT & AV cushions (HH18), embryo survival (HH18 & HH24) and heart function (HH18, HH21, HH23)	NOAEL: 0.4ppb TCE; LOAEL (p<0.05): 8ppb TCE increased cell proliferation. No effects on apoptosis, embryo survival or heart function.
Drake et al. (2006b)	*in ovo*	Chicken - fertilized egg	4x yolk injections (via shell hole) at HH stages 13, 15, 17, & 20 : "cumulative" doses of 0, 0.4, 8, or 400 ppb TCE.	4 - Cell apoptosis and proliferation in OFT & AV cushions (HH24), embryo survival (HH24 & HH30) and heart function (HH24)	NOAEL: 0.4ppb TCE; LOAEL (p<0.05): 8ppb TCE increased cell proliferation, decreased embryo survival and reduced heart blood flow.
Jiang et al. (2016)	*in vitro*	Human - H9 human embryonic stem cells	Cultured in medium containing 0, 100, 1000, or 10,000ppb TCE for 21 days.	4 - Cytostructure, % beating cells, beat rates, PCR gene expression (related to cardiocellular development)	NOAEL: 100ppb TCE; LOAEL (p<0.05): 1000ppb TCE increased cell area, reduced % beating cells and beat rate; altered gene expression
Mishima et al. (2006)	*ex ovo*	Chicken - Whole embryo culture (HH14)	Cultured in medium containing 0, 10,000, 80,000 ppb TCE at HH14	2 – Cell population in AV cushion, embryo viability (HH17)	NOAEL: 10,000ppb TCE; LOAEL (p<0.05): 80,000 ppb TCE cell population reduction in AV canal cushion. Embryo survival not affected.
Saillenfait et al. (1995)	ex vivo	Rat - Whole embryo culture (GD10)	Cultured in medium containing 0, 2,500, 5,000, 10,000, 15,000 or 30,000 µM TCE for 46 hrs.	3 – Cardioteratogenicity, embryo growth, embryo viability	NOAEL (p<0.05): 2,500 µM TCE; LOAEL (p<0.05): 5,000 µM TCE reduction in embryo growth, increase in # of malformations. No heart defects reported; survival LOAEL: 30,000 µM based on lack of heart beat.
Wirbisky et al. (2016)	zebrafish	Zebrafish (recombinant) - Tg(fli1:EGFPy)1 embryos, (1-5 hpf)	Cultured in medium containing 0, 10, 100 or 500ppb TCE until 72 or 96 hpf.	6 - Larvae survival, vasculature, heart actin, heart mitochondria, microarray gene expression, PCR gene expression	LOAEL (p<0.05): 100ppb TCE for all endpoints except survival (NOAEL: 500ppb TCE). Expression altered for 70 genes, but none directly associated with CHD.

- **Mechanistic Evidence Integration:**
  1. A limited number of TCE-CHD mechanistic datasets (mostly *in ovo*, *ex ovo*) suggest CHDs as a potential hazard for consideration
     - Most of these mechanistic datasets (2/3rds) were of unreliable study quality, but even accounting for these to address hazard (i.e., complete evidence base) studies reported conflicting data and/or mechanistic data that did not indicate a direct adverse effect in the model (e.g., gene expression, protein expression/interactions).
     - None of the three *in vivo* (mouse, rat) mechanistic studies reported adverse effects related to CHD at GD10 in fetuses/dams (i.e., early heart tube defects; secondary effects potentially related to CHDs such as altered development, reduced viability, reduced litter sizes) exposed to TCE via drinking water.
     - Chicken embryo only *in vitro* animal model where TCE-CHD observed; appears to be uniquely sensitive to CHDs
     - Observations potentially relevant to CHD hazard were primarily observed in chick embryos (*in ovo*, *ex ovo*), with less direct, but still suggestive evidence also reported in a single human ESCs and a single zebrafish embryo study (**Table 3**).
  2. None of the mechanistic experiments would be suitable as candidate studies for quantitative risk assessment.
     - The 7 mechanistic studies with datasets that met all study quality metrics included one human model (embryonic stem cells), two rat models (*in vivo*, *ex vivo*), three chicken models (*in ovo*, *ex ovo*), and a recombinant zebrafish model.
     - None of the molecular (i.e., gene expression changes) or cellular responses reported among the reliable studies have not been phenotypically anchored to a CHD-relevant adverse events.
     - Given the lack of CHD response consistently observed in mammalian models, it would not be reasonable to extrapolate the quantitative dose-response CHD data reported in chicken models as they involved exposure routes (e.g., yolk injections) that are not relevant as candidate studies in developing toxicity values.
  3. Overall, TCE-CHD mechanistic database only provided limited support for the putative CHD-AOP proposed by Makris et al. (2016), which is contradicted by "high confidence" mammalian evidence (Wikoff et al., 2018) (**Figure 5**)
     - None of the TCE-CHD mechanistic datasets identified a molecular initiating event (MIE), a necessary element for AOP validation.
     - The *in vivo* mammalian TCE-CHD database do not support any of the key events (KEs) or adverse outcome (AO) critical to the putative AOP.
     - The key event (KE) and adverse outcome (AO) data relevant to the putative CHD-AOP are limited to non-mammalian models (chickens, zebrafish) which are of questionable relevance for assessing development of cardiac malformations in humans.
     - The quality of much of the TCE mechanistic database illustrates a largely unreliable evidence stream in the context of human health risk assessment.

Figure 5. Putative CHD-AOP vs. TCE-CHD Mechanistic Evidence



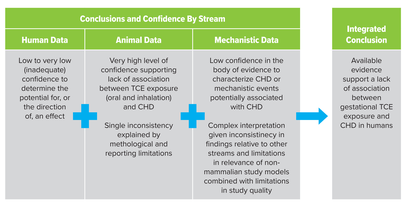
*For human, mouse, rat, and rabbit species, higher-level KE/AO response information based on OHAT-RoB evaluation of epidemiology and mammalian toxicology studies (Wikoff et al., 2018).

^No mechanistic data on rabbit models in the literature; organ response data from rabbit developmental toxicology study, evaluated using OHAT-RoB (Wikoff et al., 2018).

‡Evidence that did not support elements of putative AOP (insufficient, contradictory) presented in black cells.

- **Mechanistic Body of Evidence Assessment (OHAT Framework):**
  - Heterogeneity of mechanistic studies and results nullified rating factors used to consider increased evidence base confidence: effects were of variable magnitudes when reported; dose-response relationships were either not observed, or inconsistent in nature; effects inconsistent between models).
  - Validity (largely unreliable study quality) and indirectness (effects reported in models of little relevance for human risk assessment) reduced overall confidence in the TCE-CHD mechanistic body of evidence.
  - Overall, OHAT framework indicates a low confidence level in the TCE-CHD mechanistic body of evidence.
- Integration into larger TCE-CHD Body of Evidence (**Figure 6**): Considered together, the available human, animal, and mechanistic study data support a lack of association between *in utero* TCE exposure and CHDs.
  - Human studies ➔ Low confidence in evidence stream associating *in utero* TCE exposure with increased risk of CHDs (Wikoff et al., 2018).
  - Animal studies ➔ High confidence in evidence stream for TCE-CHD null hypothesis, i.e., no association of gestational TCE exposure and increased CHD risk (Wikoff et al., 2018). Only study to show dose response effect was determined to be unreliable (Dawson/Johnson study).
  - Mechanistic studies ➔ Low confidence in evidence stream: inconsistency and relevance of outcomes and non-mammalian models are difficult to interpret given the lack of effect in experimental animal models (mammalian).

Figure 6. TCE-CHD Evidence Stream Summaries and Integrated Conclusion



Conclusions

- Consideration of the type of outcome assessed (e.g., gene expression, *in ovo* development), the study model (e.g., chicken eggs, rat whole culture embryos, zebrafish larvae, human embryonic stem cells), as well as the plausibility of findings in a biological construct (e.g., adverse outcome pathway type of construct) were critical to integrating the evidence. The few mechanistic studies that were of sufficient quality were limited in their applicability due to heterogeneous models of questionable relevance to human physiology and exposure timing/dosing.
- Regardless of the integration approach, the same result is achieved for the TCE-CHD mechanistic database: the available evidence does not support the association between *in utero* TCE exposure and increased risk of CHDs. Mechanistic data would have been more informative had more relevant experiments (e.g., human cell types, physiologically relevant doses, phenotypical anchoring of gene expression data) been utilized. In the larger picture, however, the weak epidemiology database and lack of CHD evidence in mammalian studies indicate that even a more complete mechanistic TCE-CHD picture would be insufficient to elevate CHDs to an endpoint of relevance in TCE risk assessment.
- The use of multiple tools and approaches for evaluating the quality of study data and multiple approaches for integrating mechanistic data into the larger body of evidence, can increase confidence in systematic review findings and provide an understanding of the practical application of available approaches.

References Available upon request.

Evidence Integration in Deriving Toxicity-Based Benchmarks for Trichloroethylene

Thomas E. Sussan, Mark S. Johnson, Glenn J. Leach

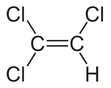
Toxicology Directorate, US Army Public Health Center, APG, MD 21010

Abstract

An important public health function within the Army is balancing the critical mission of national defense with the risks associated with exposure to various substances by Soldiers, workers, and their families. Developing toxicity-based benchmarks for risk assessment requires the integration of evidence from human, laboratory animal, and mechanistic studies, each with varying study designs that are collected independently and often by disparate means. Here, we use the development of an occupational exposure level (OEL) for trichloroethylene as an example of a process for assessing the weight of evidence of various toxicity endpoints. Following collection of relevant studies via a systematic literature search, we developed a quantitative process for evaluating the controlled animal data with respect to study quality, strength of effects, relevance, data consistency, and risk of bias. Studies were then graphically compared within each non-cancer health effect domain (neurological, kidney, liver, immunological, reproductive, and developmental) to establish points of departure (PODs) for each class of health effects based on data that is robust and relevant. An iterative process was then used to incorporate human health data and mechanistic data that considered mode of action, plausibility, and human relevance of these PODs. Physiologically-based pharmacokinetic (PBPK) modeling was then used to estimate human equivalent concentrations (HECs), and uncertainty factors were employed using a Bayesian approach to establish an OEL for each health effect domain. Potential cancer risks were also evaluated and this dose-response was estimated at various risk levels for the purpose of comparison to the non-cancer OEL.

Introduction

American National Standards Institute



- TCE is a volatile industrial and commercial solvent that has been used widely throughout the Department of Defense (DOD), for example, as a metal degreaser or cleaner for the maintenance of engines and weapon systems.
- Workers may be exposed to TCE via either direct contact during work tasks or via indirect exposure due to vapor intrusion. The primary route of exposure for workers is inhalation.
- Thousands of *in vivo*, *in vitro*, and epidemiological studies demonstrate widespread effects due to TCE exposure.
- Exposure standards endorsed by various governmental regulatory agencies span a 250,000-fold range:

  OSHA Permissible Exposure Limit: 100 ppm
  NIOSH Recommended Exposure Limit: 25 ppm
  ACGIH Threshold Limit Value: 10 ppm
  EPA Regional Screening Level- Composite Worker Air (non-cancer: 0.0016 ppm; $1x10^{-6}$ cancer risk: 0.0006 ppm)
  EPA Reference Concentration: 0.0004 ppm

- EPA has occasionally directed evacuation of DoD workplaces based on environmental standards, leading to confusion among commanders. A single occupational exposure level (OEL) is needed that applies to all workers, including those impacted by vapor intrusion.

Objective

To develop an occupational exposure level (OEL) for TCE inhalation based on an assessment of all of the current toxicological evidence, utilizing methods that maximize scientific robustness and transparency, while balancing timeliness (**Figure 1**).

Methods

- Initial steps included problem formulation, development of study inclusion/exclusion criteria, literature search/screen (**Figure 2**), and hazard characterization. Hazard characterization included human observational studies (occupational and residential) and controlled experimental studies in mammalian models exposed to TCE via all physiologically relevant routes of exposure. Additional lines of evidence, such as mechanistic data and experimental studies using non-physiologically relevant exposures, were identified via subsequent targeted literature searches, and may have been considered as supporting evidence to inform coherence, relevance, and plausibility (**Table 1**).
- A quantitative study evaluation tool was developed to assess the quality, relevance, strength of results, and consistency of controlled laboratory studies (**Figure 3**). Repeat-dose non-cancer studies were evaluated using this tool, with greatest emphasis placed on inhalation studies. Oral studies were evaluated only if the inhalation database for each toxicity category was determined to be insufficient.
- Critical effects from each study were stratified into non-cancer toxicity categories (neurological, kidney, liver, immunological, reproductive, and developmental). Points of departure (PODs) were derived for each critical effect (**Figure 4**).
- PODs were compared within each toxicity category, and selection of an overall POD for each toxicity category was determined after consideration of critical effects, study evaluation score, qualitative study assessment, and supporting human and mechanistic evidence (**Figure 5 and Table 1**).
- Key studies were analyzed via physiologically-based pharmacokinetic (PBPK) modeling to derive human equivalent concentrations (HECs). Uncertainty factors were applied to HECs using a Bayesian approach to derive candidate occupational exposure levels (OELs) for each non-cancer toxicity category (**Table 2**).
- Cancer screening levels were derived for either kidney cancer alone or for the combined risk of kidney cancer plus non-Hodgkin lymphoma using a no-threshold linear dose response extrapolation (**Table 2**).

Results

Figure 1. Overview of the process for deriving an occupational exposure level for TCE, including integration of toxic endpoints across multiple study designs.

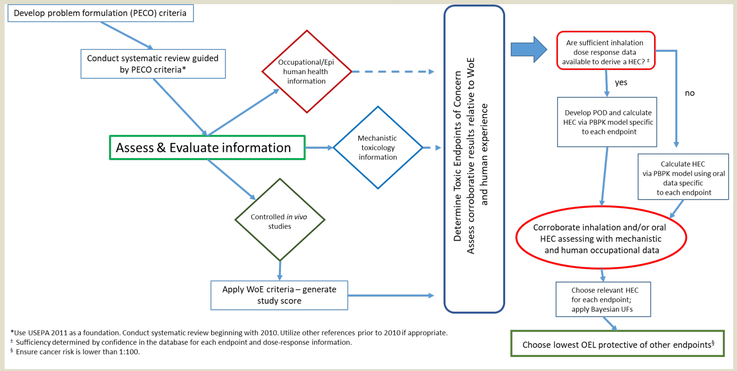
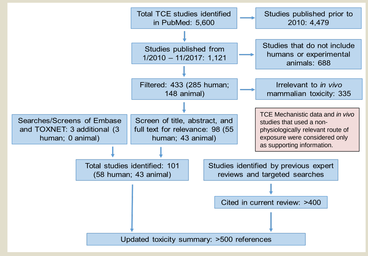


Figure 2. Flow chart of literature identified and included in the toxicity assessment, based on a systematic literature search and other targeted searches. Studies published prior to 2010 were obtained from expert reviews. Mechanistic data was included in targeted searches that informed coherence and plausibility.



Domain	Description	Points
	Study Quality (40 points)	
Purity	Assesses confidence that response is due to compound of interest rather than impurity/contaminant.	10
Dose Confirmation, Stability, and Homogeneity	Assesses verification of dosing accuracy and homogeneity and potential for issues of chemical instability.	10
Test Methodology	Assesses study design and risk of bias.	20
	Strength of Results (25 points)	
Dose Response	Assesses coherence of response across doses. Dose response incorporates aspects of study design (i.e., number/interval of doses tested) and robustness of the data.	15
Statistical Significance	Statistical confidence in outcome based on either p-value for alternative hypothesis or sample size for null effects. Statistical confidence incorporates magnitude/precision of response and sample size.	10
	Relevance/Applicability to Hypothesis (25 points)	
Exposure Characterization	Assesses relevance of exposure route, absorption rate, and chemical form toward informing the weight of evidence.	10
Biological Significance	Assesses importance of health outcome and directness of measured biological endpoint to predict the health outcome.	15
	Consistency of Outcome (10 points)	
Coherence	Assesses consistency of key endpoint with other endpoints within the study.	5
Reproducibility	Assesses consistency of endpoint with identical or complimentary endpoints in other studies.	5

Figure 3. Overview of quantitative assessment tool designed to evaluate controlled animal studies on a 100 point scale. Qualitative annotations and sub-scores derived from tool were also considered in weight of evidence assessment.

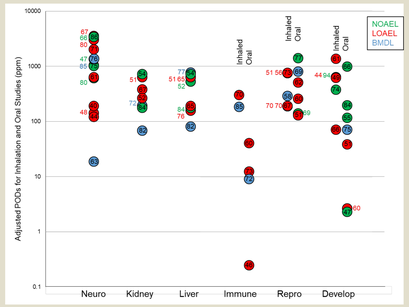


Figure 4. Top Panel- Points of departure (Y-axis) and total score (inside/adjacent to circle) for all 63 critical endpoints, stratified by health effect category. Bottom Panel- Summary of total scores.

Figure 5. Comparison of all experimental studies included in POD assessment. All exposure concentrations tested in each study (adjusted to 40 h/wk) are plotted, including those concentrations determined to be the NOAEL (green circles), LOAEL (red circles), and BMDL (if determined- blue lines). Black dots indicate other treatment exposures. Annotations within the graph include the critical effect and overall study applicability score. Bolded studies were determined to be those most informative for deriving the overall POD, and these studies were analyzed via PBPK modeling. Pink boxes identify supporting human occupational evidence. Boxes overlaying a graph indicate exposure ranges where human effects have been observed.

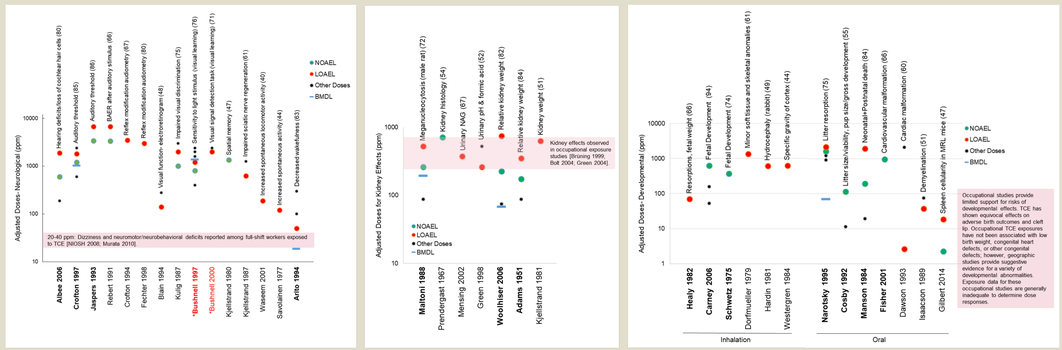


Table 1. Examples of incorporation of mechanistic evidence into the toxicological assessment.

Effect	In Vivo Observation	Mechanistic Insight	Implication
Kidney Toxicity	Kidney toxicity is more commonly reported (1) in rats than in mice and (2) in males than females.	β-lyase-dependent metabolism of DCVC is an important mediator of kidney toxicity; and β-lyase is localized exclusively to the proximal tubular epithelium of the kidney, which is the primary site of TCE-induced toxicity. Specific activity of β-lyase in humans is comparable to that of mice, which are both only about 10% of the activity observed in rat kidney. Additionally, β-lyase activity is higher in male rats compared to female rats or in mice of either sex. Glutathione conjugation is a relatively minor metabolic pathway in the kidney, and thus the onset of kidney damage is likely to only occur when the oxidative cytochrome P450-mediated pathway becomes saturated in the liver. Saturation of this pathway has been observed at high doses in rats but not in mice or humans.	These species/sex-specific differences may underlie the observations that male rats are most susceptible to TCE-induced kidney toxicity, and it suggests that rats may be more sensitive than humans. Saturation of oxidative metabolism in the liver has been observed at high doses in rats but not in mice or humans, which further suggests that rats are particularly susceptible to nephrotoxicity and only at high doses.
Liver Toxicity	Liver toxicity is more commonly reported (1) in mice than in rats and (2) in males than females.	Mice have a faster rate of oxidative TCE metabolism than rats, and rats also experience a dose-dependent saturation of oxidative TCE metabolism that was not observed in mice. Thus, at elevated exposures, mice generate significantly higher levels of oxidative TCE metabolites than rats. Even among different mouse strains, significant variability in the rate of oxidative TCE metabolism has been observed, suggesting that only specific strains of mice will display elevated susceptibility to liver toxicity. The metabolism of TCE is estimated to be approximately 20 times slower in humans than in rats and 60 times slower in humans compared to mice.	Humans are likely to show decreased susceptibility to liver toxicity compared to either mice or rats.
Kidney Cancer	The dose-response curve of kidney cancer is unclear.	Among individuals with exposure to high concentrations of TCE, increased somatic mutations have occasionally been reported in the tumor suppressor gene Von Hippel-Lindau (VHL), a key event for initiation/progression of renal cell carcinoma. Suggestive evidence for mutagenicity is based on multiple occupational and in vivo animal studies; although in vitro studies have shown that the mutagenic effects of TCE may be mediated by its metabolites. The overall body of evidence for mutagenicity is weak, and consensus for the mutagenicity of TCE in mammalian cells has not been established. A threshold response may be supported based on the assumption that the etiology of kidney cancer is mediated by secondary metabolites of the GSH conjugation pathway that are not generated at lower exposure levels. A threshold response is also supported by studies in rodents, which have only shown kidney tumors after exposure to relatively high levels of TCE, where it has also been noted that any incidence of kidney tumors is likely to be preceded by severe and chronic kidney injury that is likely to exhibit a threshold response.	Potential modes of action for TCE-mediated kidney cancer may include both non-threshold and threshold key events. Due to the uncertain mode of action for TCE-induced kidney cancer, a non-threshold linear regression response was used. A non-threshold dose-response is generally more conservative than a threshold dose-response, which is true in this instance as well. However, we recognize the uncertainty in this approach and acknowledge the possibility that this approach may over-estimate the risk.
Lung Cancer	Increased lung tumors have been observed in chronic inhalation studies of mice, but are not generally reported in humans.	The ability of the human lung to metabolize TCE is approximately 600-fold less than that in the mouse. Expression of CYP2E1 is markedly lower in rat lungs than in mouse lungs, and it is expressed only marginally in human lungs. Comparisons of species-specific metabolism of CYP2E1 and GST include lower level, or potentially no activity, in human lung compared to rodents.	Species-specific differences in metabolism of TCE in lung epithelial cells suggest that mice may be uniquely sensitive to this effect due to elevated generation of the TCE metabolite chloral in mouse lung epithelium. Thus, lung cancer is not relevant to humans.
Liver Cancer	Evidence for liver cancer is weak/inconclusive in humans but has been reported in chronic inhalation and oral studies of male mice.	Peroxisome proliferator-activated receptor alpha (PPARα) is a key mediator of TCE-induced liver toxicity due to its effects on cellular proliferation and apoptosis after activation by TCA. The mode of action of TCA as a rodent liver carcinogen is principally as a liver peroxisome proliferator in specific strains of mice. Peroxisome proliferation is not induced in human hepatocytes after treatment with TCA, and mice expressing human PPARα show diminished hepatotoxicity after exposure to other peroxisome proliferators. There is significant lack of concordance in the human and rodent hepatocytes to PPARs, with humans being much less sensitive.	Mice show substantially higher responsiveness to peroxisome proliferators than primates, and thus the human relevance of this mode of action is questionable. The weak carcinogenic activity in the liver of chloral hydrate in male B6C3F1 mice combined with lower rates of oxidation and higher rates of conjugation in humans compared with mice indicate that the mode of action for mice is not relevant to humans.

Table 2. Points of departure, human equivalent concentrations and occupational exposure levels for each health effect category.

Domain	POD (ppm)	HEC (ppm)	OEL (ppm)	OEL (mg/m^3)	Critical Effect
Neurological	18	18	3.8	20	Wakefulness
Kidney	68	74	16	86	Kidney weight
Liver	160	180	6.9	37	Liver weight
Immunological	12	23	0.9	4.8	Plaque forming assay
Reproductive	130	140	5.1	27	In vitro fertilization of oocytes
Developmental	70	78	2.9	16	Fetal resorptions, weight
Cancer (10^{-2} Risk) 30y			13	70	Kidney
Cancer (10^{-2} Risk) 45y			8.9	48	Kidney
Cancer (10^{-2} Risk) 30y			3.6	19	Kidney and NHL
Cancer (10^{-2} Risk) 45y			2.5	13	Kidney and NHL
Cancer (10^{-3} Risk) 30y			1.3	7.0	Kidney
Cancer (10^{-3} Risk) 45y			0.9	4.8	Kidney
Cancer (10^{-3} Risk) 30y			0.36	1.9	Kidney and NHL
Cancer (10^{-3} Risk) 45y			0.25	1.3	Kidney and NHL
Cancer (10^{-4} Risk) 30y			0.13	0.70	Kidney
Cancer (10^{-4} Risk) 45y			0.09	0.48	Kidney
Cancer (10^{-4} Risk) 30y			0.036	0.19	Kidney and NHL
Cancer (10^{-4} Risk) 45y			0.025	0.13	Kidney and NHL

Conclusions

- *In vivo*, *in vitro*, and epidemiological studies were integrated and used to support derivation of an occupational exposure limit.
- Dose-response data were strongest for *in vivo* controlled laboratory studies, and these studies served as the primary basis for determining causal relationships between TCE and toxicity (inhalation and oral).
- Derivation of non-cancer OELs were developed from inhaled and oral PODs converted to inhaled HECs via PBPK modeling; derivation of OELs incorporated application of uncertainty factors using a Bayesian approach. Human and mechanistic data provided supporting evidence that informed coherence, relevance, and plausibility of the *in vivo* laboratory findings.
- Cancer risks were derived from from human data supported by laboratory animal and mechanistic data.
- An OEL of 0.9 ppm was selected based on the most sensitive non-cancer effect (Immunosuppression). This is supported by a 10^{-3} cancer risk in the kidney assuming a 45 year occupational exposure to TCE.

Reference

Sussan, TE, Leach GJ, Covington, TR, Gearhart, JM, Johnson, MS. (2019). Trichloroethylene: Occupational Exposure Level for the Department of Defense. Aberdeen Proving Ground, MD.

ToxStrategies

A Fit-for-Purpose Framework for Use of Systematic Methods in Risk Assessment

Daniele Wikoff[1], R. Jeffrey Lewis[2], Neeraja Erraguntla[3], and Jennifer E. Foreman[2]

[1]ToxStrategies. [2]ExxonMobil Biomedical Sciences, Inc. [3]American Chemistry Council.

Introduction

- Systematic review is being adopted globally in the fields of toxicology, epidemiology, and risk assessment.
- Best practices are still under development; processes and tools established for evidence-based medicine (EBM) need to be refined and/or new processes and tools are needed for the practice of evidence-based toxicology (EBT).
- Practitioners recognize particular challenges related to:
  - Heterogenous datasets including *in vivo* (multiple species, including humans), *in vitro*, and *in silico* study designs.
  - Identification, evaluation, and integration of mechanistic data which are not direct measures of an outcome or population (but are critical to the practice of chemical health risk assessment).
  - Application of systematic review to risk assessment, which involves more than assessing a "yes/no" question regarding hazard.
  - The need to evaluate very broad evidence bases that include multiple outcomes assessed via diverse endpoints across study designs (which also relates to challenges in specifically characterizing a whole assessment *a priori*).
  - Time and financial resources needed to develop knowledge and tools, as well as to conduct the review.

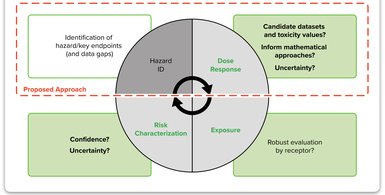
Objective

To address recognized challenges, we propose a framework for using evidence-based methods to facilitate the risk assessment process in a "fit for purpose" manner to determine hazard, develop toxicity values, and characterize uncertainty.

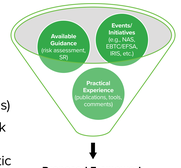
Methods

- Confirmed the need for a specific framework (**Figure 1**)
  - Surveyed available guidance for systematic review; focus of available frameworks is on hazard identification.
  - Available/draft guidance related to systematic review and risk assessment do not fully address elements beyond hazard identification (e.g., candidate study selection, quantitative uncertainty, relevance to humans).
  - Confirmed that a framework was needed to be "fit for risk assessment"

Figure 1. Representation of risk assessment components addressed by the proposed approach.



- Considered a diverse array of resources, with heavy reliance on the evolving landscape, and practical experience:
  - Events and Initiatives (e.g., EBTC/EFSA Colloquium, NAS meetings, journal standards)
  - Available Guidance for both risk assessment (USEPA, EFSA, WHO/IPCS, NAS) and systematic review (e.g., NTP-OHAT, Navigation Guide, EFSA, USEPA IRIS and TSCA)
  - Practical Experience (e.g., protocols, publications, public comments)

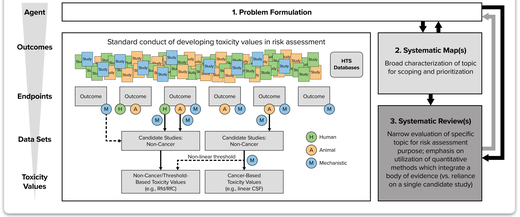


- Considered common challenges in practice to determine desired aspects of the framework
  - Reflect stepwise and adaptive nature of conducting risk assessment
  - Accommodate existing risk assessment methodologies and concepts (e.g, MoA, WoE, BMD, etc.)
  - Appraise all aspects of study validity (not just internal validity) for studies critical to decision-making
  - Better use of all available data (e.g., extrapolation methods, quantitative methods, uncertainty analyses)
  - Readily facilitate expert judgement using *a priori*, structured methods
  - Include emphasis on topic-specific refinements

Results

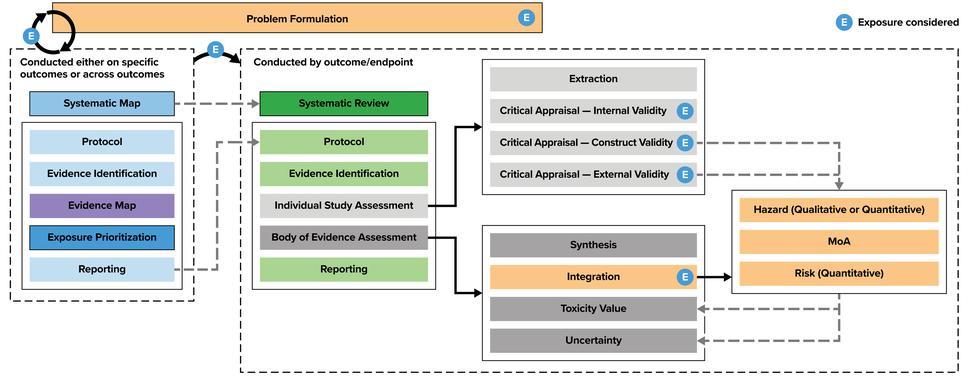
- The overall intent of the framework is to use multiple evidence-based methodologies to facilitate the risk assessment process. This is accomplished via three main components (problem formulation, systematic mapping, and systematic review(s)), thus allowing for an adaptive risk assessment process. **Figure 2**.
  - Problem formulation, a well-established component in the practice of evidence-based methods, is the step which defines the question(s) to be evaluated, the rationale for the question(s), as well as the basis for how the assessment(s) will be conducted. *This allows for both the utility and feasibility to be considered in determining specific objectives related to risk assessment.*
  - Systematic map(s) involve a broad characterization of a topic, allowing for well-informed scoping and prioritization of outcomes (and endpoints) to be considered for systematic review. *This allows for all data to be considered systematically, and allows for a transparent method to identify the subset of the data (likely specific outcomes or endpoints) to be carried forward to systematic review in support of developing toxicity values.*
  - Systematic review(s) involve a rigorous assessment of a narrow topic, including critical appraisal of study validity and structured methods for developing conclusions. *This allows for transparency in identification and selection of candidate studies with consideration for study quality and relevance, utilization of quantitative methods which rely on a body of evidence (vs. a candidate study approach; evaluation of MoA), and directly provides data for quantitative uncertainty assessment.*

Figure 2. Overall structure is designed to complement the standard conduct of developing toxicity values in risk assessment.



- The framework involves stepwise application of each of the three main components. Each component consists of multiple elements (e.g., protocol, evidence identification, critical appraisal, integration, etc.). **Figure 3**.
  - Mapping can be carried out on specific outcomes or across outcomes, whereas systematic reviews are conducted on an individual outcome(s) or endpoint(s) (and thus multiple systematic reviews may be conducted).
  - Some elements did not require significant modification from existing methodologies where as others were refined (or added) to fit the needs of chemical health risk assessment (relative to that established for clinical medicine):
    - Critical appraisal of individual studies includes consideration of internal, external, and construct validity (vs. only internal validity in existing EBM).
    - Body of evidence evaluation assessment provides a structured approach for hazard and mode of action (MoA) in quantitatively characterizing risk via dose-response, utilizing elements of the systematic mapping and review to facilitate development of the toxicity value, as well as quantitatively characterizing uncertainty.

Figure 3. Proposed framework for the use of evidence-based methods in support of risk assessment.



- *In the manuscript (currently under development) each element, or box on the figure, is accompanied by a series of steps to complete the element. In such, steps unique to making the systematic method "fit for purpose" to risk assessment are differentiated.*
- Each element involves evidence from human, animal, and mechanistic streams, though the identification and integration of each is dependent on assessment objectives.
- Unique to this framework is the consideration of exposure in each main component, providing important context to facilitate risk-based decisions (**Figure 3; Figure 4**).

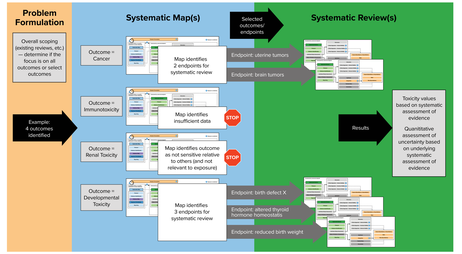
Figure 4. Exposure considerations related to the use of evidence-based methods in support of risk assessment. Figure demonstrates examples of how exposure aspects can be considered in guiding problem formulation, developing conclusions in systematic mapping, and refining appraisal tools and techniques to best inform risk-based conclusions.

Problem Formulation	Systematic Map(s)	Systematic Review(s)
Confirm the need for assessment; guide what is assessed • Is there exposure? • Receptors/routes/levels of interest or relevance?	Confirm the relevance/rank the importance of outcomes selected (provides rationale for sensitive endpoints) • Is there toxicity at relevant exposure levels (based on MOE/MOS)? • Are certain endpoints of greater interest based on receptor/routes?	Informs appraisal and decision criteria • What aspects are important to consider as part of critical appraisal (internal validity)? • What aspects are important to consider in determining model and study relevance (external validity)?

Example

- In practice, application of this framework is likely to involve an in-depth problem formulation phase followed by the conduct of a series of maps and review in order to accommodate the needs of risk assessment (**Figure 5**).
  - By using the step-wise approach, many different "exit points" are available, ensuring that time and resources are spent in a meaningful way.

Figure 5. Example of framework implementation. Pending the scope characterized in Problem Formulation, multiple systematic maps and/or multiple systematic reviews may be conducted.



Conclusions

The proposed framework allows for an assessment to be adaptive to the needs of a risk assessor while still adhering to the principles of evidence-based methodology. The framework also allows for integration of mechanistic data based on established risk assessment approaches and provides a practical approach for expediting evidence integration in context of developing toxicity values.

Acknowledgements and Disclosures

Financial support for D.W. was provided by the American Chemistry Council (travel support) and the American Chemistry Council's Center for Advancing Risk Assessment Science and Policy (ARASP) (conduct of the work). J.L. and J.F. are members of ARASP. N.E. is the ARASP manager.

1

TERA

TOXICOLOGY EXCELLENCE FOR RISK ASSESSMENT

Evidence-Based Dose Response Assessment for Thyroid Tumorigenesis: A tale of two MOAs... ...and what can be done about it

Michael Dourson[1], Rick Hertzberg[2], Bruce Allen[3], Lynne Haber[1,4], Ann Parker[1,4], Oliver Kroner[1,5], Andy Maier[1,6], and Melissa Kohrman[1,6]

[1]Toxicology Excellence for Risk Assessment, [2]Emory University, [3]Independent Consultant, [4]now at University of Cincinnati, [5]now at City of Cincinnati, [6]now at Cardno ChemRisk

2

Hazard Identification 1

- Acrylamide not shown to cause cancer in humans (Marsh et al., 2007; Swaen et al., 2007; Mucci & Adami, 2005).
- Six long-term experiments in rats show tumors (Johnson et al., 1986; Friedman et al.,1995; and NTP, 2012). All experiments used the same strain of rats.
- Two long term studies also available in mice (NTP, 2012).
- Four long term studies are also available in mice and rats for glycidamide, a known metabolite of acrylamide *in vivo*.

TERA

3

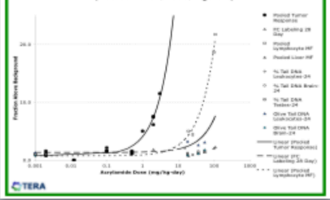
Mode of Action 1

- Acrylamide is genotoxic, but not directly mutagenic. Acrylamide also causes growth stimulation and oxidative stress, the latter of which can lead to mutations and other genotoxicity.
- Mutagenicity and genotoxicity from acrylamide exposure are not dose response concordant---only seen at doses higher than those that caused tumors (Figure 1).
- Moreover, *in vivo* mutations do not occur in the rat mammary gland or testes (Table 1 and Figure 2 of Mei et al., 2010).
- Accordingly, it is unlikely that the tumors evoked are solely caused by either mutagenicity or genotoxicity. However, unmeasured mutagenicity might be occurring at low doses and might be responsible for some of the low dose tumors (Figure 2).

TERA

4

Figure 2. Multiples above background for thyroid tumors, DNA labeling & several measures of in vivo genotoxicity (Dourson et al., 2008, Figure 5).



TERA

5

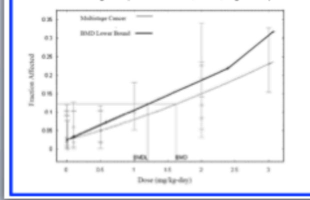
Dose Response Assessment 1

- EPA (2005) suggests "decoupling" data when several modes of action occur in different parts of the dose response curve. Thus,
  - a mutagenic, non-threshold, linear mode of action may be occurring at doses of less than 1 mg/kg-day, and
  - a growth stimulation, threshold, non-linear mode of action likely dominates at doses in excess of 1 mg/kg-day.
- EPA (2005) suggests that selection of a point of departure be close to the lower range of data of interest. Since controls had dose, a BMR lower than 5% is reasonable; a 2% extra risk is:
  - ~double the background rate of the pooled data, and
  - 1/3 of the highest low dose response rate of ~6%.
- Thus, a 2% BMR is comfortably within the interpolation range allowing for a stable estimate for the point of departure.

TERA

6

Figure 3a. Multistage model fitted to pooled-all thyroid tumor data, showing little change in slope between the low and high dose regions (Dourson et al., 2008, Figure 6a).



7

Abstract

- Acrylamide is commonly found in various foods.
- Cancer studies in rats report increases in many types of tumors. We find that thyroid tumors are most sensitive because they are repeatable, relevant and potent.
- Both mutagenicity & growth-stimulation occur; mutagenicity determines low-dose; growth stimulation at higher doses.
- We "decouple" these MOAs and determine that the probit model best reflects the underlying "decoupled" biology.
- We identify a cancer slope factor of 0.030 $(mg/kg\text{-}day)^{-1}$ for low dose, and a RfD of 0.05 to 0.02 mg/kg-day for high dose.

TERA

8

Hazard Identification 2

"Risk analysis shall be based on the most sensitive study deemed to be of sufficient quality" (California, 2011).

A risk assessment perspective of "most sensitive study" for determining the risk in humans includes studies that are:

- Repeatable. A sensitive study shows effects that are found in comparable studies of the same or other species/strains.
- Relevant. A sensitive study shows effects in one species that are expected to be found in humans.
- Potent. A sensitive study shows effects that occur at the lowest dose or concentration.

Most sensitive study ≠ only the most potent study

(See Appendix to this presentation)

TERA

9

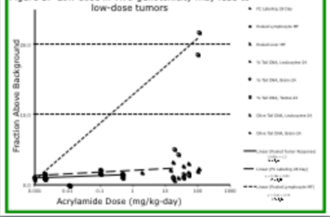
Mode of Action 2

- The weight of scientific evidence supports growth stimulation as contributing to thyroid tumors.
  - Khan et al. (1999) showed statistically significant morphological changes in the thyroid consistent with stimulation after acrylamide for 2 or 7 days.
  - Lafferty et al. (2004) observed three measures of growth stimulation in the thyroid after acrylamide for 7, 14 & 28 days; DNA labeling was statistically significant by pairwise comparison. Chico-Galdo et al. (2007) results suggest that this DNA labeling was growth stimulation.
  - Friedman et al. (1999) showed thyroid hormones to be statistically significantly decreased in males after 28 days of exposure (trend test & pairwise comparison). Females appeared to be affected, but less so.
  - Johnson et al. (1986) showed thyroid hyperplasia in both male and female rats after 2 years of exposure (statistically significantly-trend test).
- Comparison with Table 3 of EPA (1998) also supports this choice.

TERA

10

Figure 3. Low dose *in vivo* genotoxicity may lead to low-dose tumors



11

Dose Response Assessment 2

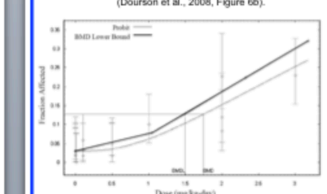
Using EPA (2005) guidelines, we compared different mathematical models to fit these "decoupled" data.

- Multistage model did not fit "decoupled" data as well (Figure 3a);
- Weibull model fit these "decoupled" data well with fixed power of 2, but EPA software did not allow a positive value for control doses with power unfixed; controls animals had a positive dose (Twaddle et al., 2004).
- Probit model fit these "decoupled" data well, showing a linear response for tumors in the low dose range and a curvilinear upward trend for tumors from growth stimulation in the high dose range (Table 4; Figure 3b);
- When low dose "decoupled" data only were considered, a weighted linear regression and a multistage model both confirmed the use of the probit model to generate a point of departure for the low dose extrapolation (Table 5; Figure 3c)

TERA

12

Figure 3b. Probit model fitted to pooled-all thyroid tumor data, showing differing slopes between the low and high dose regions (Dourson et al., 2008, Figure 6b).



13

Background

- Previous assessments were found to be generally conservative, "screening level" assessments.
- We undertook a more comprehensive risk assessment that benefited from the following:
  - new science on acrylamide's toxicity and its mode of action; and
  - new guidelines for the evaluation of cancer risk developed by the U.S. Environmental Protection Agency (EPA, 2005).
- Key determinants for evaluating tumor response are:
  - consistently observed tumors that can be statistically combined among studies using EPA research and guidelines, and
  - probable relevance to humans from experimental animals (note: when little is known such relevance is assumed).

TERA

14

Hazard Identification 3

- Thyroid tumors in rats exposed to acrylamide were repeatable and statistically significant in all six experiments (Table 1).
- Although, scientists have not identified any chemical that has caused thyroid tumors in humans and the rat thyroid is different from the human thyroid, we conclude that these rat thyroid tumors could not be discounted. Therefore, they are considered as relevant to humans.
- The type of thyroid tumors formed in rats is generally recognized as resulting from growth stimulation and/or mutation, and these modes of action also operate in humans.

TERA

15

Mode of Action 3

- Tumors evoked by acrylamide exposure were generally benign, occurred late in life, and were more often in hormonally-active organs, in all six rat experiments.
- Such tumor appearance is more consistent with manners of tumor formation that are not from direct mutation. These observations also mean it is unlikely that direct mutations are causing all of the tumors in these experiments.
- Thus, both a mutagenic and non-mutagenic manners of tumor formation are likely to contribute to thyroid tumors (Table 2).

TERA

16

Table 2. Key events, related evidence and plausibility of tumor modes of action.

Event	Summarized evidence in animals (see text for additional details)	Is this key event in the MOA plausible in humans?	Taking into account kinetic and dynamic factors, is this key event in the MOA plausible in humans
An heritable mutation in thyroid follicular cell DNA	• Glycidamide is directly mutagenic. • DNA-glycidamide adducts are found in multiple tissues, but overall tumor response does not match distribution. • Mutations lead tumors in timing of response, but not in dose response.	The MOA is plausible in humans, but not without subsequent growth stimulation.	This MOA is plausible and comparative kinetic information between experimental animals and humans allows specific choices of adjustment factors.
Follicular cell growth stimulation	• Increased DNA synthesis in thyroid at multiple time points; increased mitotic index; decreased apoptotic index. • Decreased thyroid colloid and increase follicular cell height after 2 or 7 days. • Thyroid hormone decrease & thyroid weight increase after 28 days. • Equivocal changes in liver weights occur after long-term exposure, suggesting increase catabolism of thyroid hormones; other possible sites of action are thyroid microfilament development & hypothalamus toxicity.	The MOA is known to occur in humans.	This MOA is plausible, but less likely to occur in humans given the same dose of acrylamide as in rats, based on well-recognized differences in thyroid hormone homeostasis. A conservative assumption is to use a factor of 1-fold when extrapolating rat data to humans.

TERA

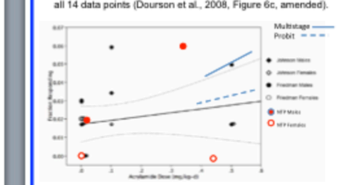
17

Dose Response Assessment 3

- The probit model Benchmark dose (BMD) of 2% extra risk is 0.81 mg/kg-day, associated with a slope of 0.025 $(mg/kg\text{-}day)^{-1}$.
- Slope value adjusted by 1.2 for known kinetic rat and human differences; not further adjusted for dynamic variability because:
  - Williams (1995) states that thyroid tumors in humans do not form in the presence of mutagens if TSH-stimulated growth is prevented;
  - EPA (1998) considers an adjustment factor of 1 for chemicals having a growth stimulation mode of action, unless specific data suggest otherwise;
  - Analysis of published studies comparing human & animal tumor slope factors show that the most likely value for an overall factor is roughly 1-fold;
  - Goodman and Wilson (1991) consider the best estimate of the interspecies factor to be log normally distributed around a value of 1.
- The adjusted slope factor is 0.030 $(mg/kg\text{-}day)^{-1}$.

18

Figure 3c. Weighted linear regression on low-dose, pooled data with 95% confidence curves for the model. Data jittered to show all 14 data points (Dourson et al., 2008, Figure 6c, amended).



TERA

19

TERA

TOXICOLOGY EXCELLENCE FOR RISK ASSESSMENT

Mission is to support the protection of public health by:

- Developing, reviewing and communicating risk assessment values and analyses;
- Improving risk methods through research; and
- Educating risk assessors, managers, and the public on risk assessment issues

- TERA is a 501c3 nonprofit organization

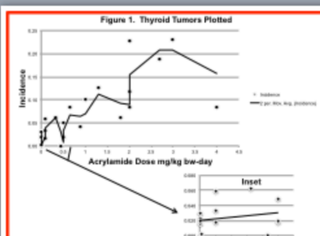
20

Table 1. Thyroid tumors in Johnson et al. (1986), Friedman et al. (1995), & NTP (2012) drinking water studies.

Acrylamide dose mg/kg-day	Johnson Males Incidence	Johnson Males Fraction	Johnson Females Incidence	Johnson Females Fraction	Friedman Males Incidence	Friedman Males Fraction	Friedman Females Incidence	Friedman Females Fraction	NTP Males Incidence	NTP Males Fraction	NTP Females Incidence	NTP Females Fraction
0.0013			Control doses in red						1:47	0.02		
0.0015											0:48	0.00
0.002	1:60	0.017	1:58	0.017	3:100	0.030	1:50	0.020				
0.002					3:102	0.029	1:50	0.020				
0.012	0:58	0.0	0:59	0.0								
0.10	2:59	0.034	1:59	0.017	12:203	0.059						
0.33									3:48	0.06		
0.44											0:48	0.00
0.50	1:59	0.017	1:58	0.017	5:101	0.050						
0.66									4:47	0.09		
0.88											2:48	0.04
1.0							18:100	0.18				
1.3									6:48	0.13		
1.8											3:48	0.06
2.0	7:59	0.12	5:60	0.083	17:75	0.23						
2.7									9:48	0.19		
3.0							23:100	0.23				
4.0											4:47	0.09

Shaded values indicate authors designation of statistical significance.

21



TERA

22

TERA 22

23

Table 4. Probit model estimates of slope factors (SF) for rat thyroid tumor data. BMD/L values in mg/kg-day; SF values in $(mg/kg\text{-}day)^{-1}$ (Dourson et al., 2008, Table 6).

Data set	BMD 0.2	SF at BMD 0.2	BMDL 0.2	SF at BMDL 0.2
Pooled female	0.82	0.024	0.67	0.030
Pooled male	0.97	0.021	0.58	0.034
Pooled all	0.81	0.025	0.69	0.029
	BMD 10	**SF at BMD 10**	**BMDL 10**	**SF at BMDL 10**
Pooled female	1.8	0.057	1.5	0.069
Pooled male	1.7	0.059	1.3	0.079
Pooled all	1.7	0.057	1.5	0.067

TERA

24

Table 5. Comparison of 2% BMD and BMDL from various models for acrylamide induced thyroid tumors in rats & projected slope factors.

Data Set	Model	BMD_{02}	SF BMD_{02}	$BMDL_{02}$	SF $BMDL_{02}$	BMD/BMDL
All pooled tumors	Multistage	0.39	0.052	0.23	0.087	1.7
Low-dose pooled tumors	Multistage	0.80	0.025	0.23	0.088	3.5
All pooled tumors	**Probit**	**0.81**	**0.025**	**0.69**	**0.029**	**1.2**
All pooled tumors	Weibull	0.82	0.024	0.72	0.028	1.1
Low-dose pooled tumors	Weighted linear regression	0.92	0.022	0.33	0.061	2.8

P-values for all regressions exceed 0.1 and thus are acceptable; AICs and residuals for the multistage and probit models with all pooled tumors were comparable, however, BMD/BMDL ratios (above) and visual fit differ (see Figure 2a,b). Bold printed row indicates model chosen for extrapolation to humans.

TERA

Evidence Integration Using AOP Networks: Assessing Human Health Risks Associated With Hydrogen Sulfide Exposure

Katy O. Goyak[1], R. Jeffrey Lewis[1]

[1]ExxonMobil Biomedical Sciences Inc., Annandale, NJ

Introduction

- Hydrogen sulfide is a naturally occurring constituent in crude oil and a byproduct of certain refinery processing steps.
- In acute, high exposure conditions, a series of hallmark biological effects are known to occur, starting with olfactory paralysis and progressing to failure of the respiratory drive, coma, and death.
- Deriving a safe dose to protect against low-level, long-term exposure to hydrogen sulfide is subject to uncertainty surrounding the relevance of the critical effect in animals (olfactory nasal lesions) and its protection of the early hallmark effect in humans (olfactory paralysis), given the significant physiological differences in breathing between rodents and humans.
- The current study integrates the animal and human data via an Adverse Outcome Pathway (AOP) network, developed from 32 studies and comprised of 5 AOPs sharing 2 molecular initiating events (MIEs) and 7 potential outcomes.

Methods

Biological effects induced by hydrogen sulfide were identified by a review of the primary literature cited in 4 regulatory and/or advisory agency reviews on hydrogen sulfide were used (1-4).

1. Effects were categorized initially by level of biological organization.
2. Effects were than organized into linear pathways leading to the hallmark effects defined in the reviews as relevant for regulatory risk assessment, consistent with the AOP Developer's Handbook (5).
3. Each AOP was assessed according to its biological plausibility and empirical evidence.

When the supporting evidence for a key event in relation to the entire AOP was weak based on the hydrogen sulfide data alone, supplemental information for the key event was identified using the peer-reviewed literature and the AOP-Wiki.

The AOPs were further assessed for temporal, dose-response, and incidence concordance data shown here are focused on AOP1 due to its strong supporting evidence and effects at the lowest doses (6).

AOP NETWORK DEVELOPMENT

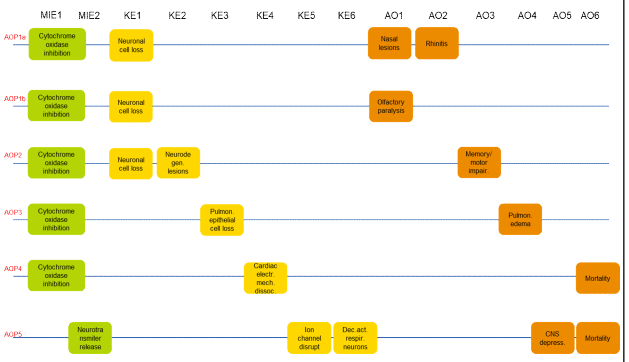


Figure 1: AOP network describing the mechanisms leading to five hallmark pathological effects of H2S exposure. Abbreviations: impair., impairment; pulmon., pulmonary; elect. mech. dissoc., electromechanical dissociation; dec. act. respir. neurons, decreased activation of respiratory neurons; CNS depress., central nervous system depression.

Objectives

- Integrate mechanistic evidence from multiple data streams to identify an appropriate point of departure for derivation of occupational exposure limits that is relevant to humans and protective of the early hallmark effects known to occur under acute conditions.

Results

MAPPING EMPIRICAL EVIDENCE TO THE NETWORK

Figure 2: Empirical evidence showing lowest effect level per reference for all key events in the AOP network. Primary references are cited in references 1-4 or are listed separately as references (7-9).

BIOLOGICAL PLAUSIBILITY OF KEY EVENT RELATIONSHIPS

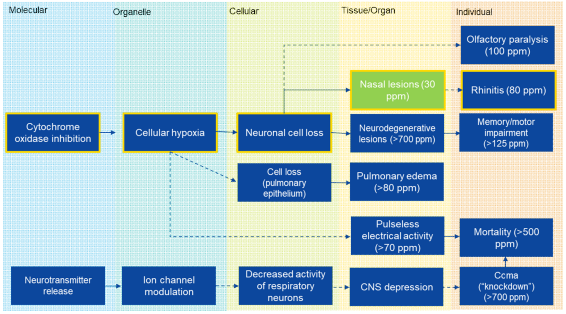


Figure 3: Comparative AOP assessment. Lowest effect levels are reported (if known) for key events or adverse outcomes. Dotted lines: weak or moderate biological plausibility; solid lines: strong biological plausibility; yellow outline: identified as AOP leading to the critical effect (defined by low dose and strong biological plausibility; highlighted in green).

CONCORDANCE OF KEY EVENT RELATIONSHIPS

Temporality

Dose (ppm)	KE1: Cytochrome oxidase inhibition, Acute (3-4 hour)	KE2: Neuronal cell loss (olfactory epithelium), Subchronic	KE3: Nasal lesion (cell less and basal cell hyperplasia), Subchronic	KE4: Rhinitis, Subchronic
0				
1	-			
10	+	-	-	-
30		++	++	-
50	+			
80		+++	++	+*
100	+			
200	++		++	
400	+++		+++	
500		+*	+*	
765	+++	+*		

Figure 4: Dose-response and temporal concordance analysis of AOP1a (cytochrome oxidase inhibition leading to rhinitis). +: 0-33% inhibition or severity; ++: 34-66% inhibition or severity; +++: 67-100% inhibition or severity; *observed and reported qualitatively, green box: no adverse effect level in animal studies commonly used as point of departure in regulatory assessments (1, 3, 4).

CONCLUSIONS

- A comparative dose assessment of subchronic animal data identified AOP1a (nasal lesions at 30 ppm) a critical pathway with strong biological plausibility and experimental evidence, with effects at the lowest dose relative to other outcomes (neurological effects, >125 ppm; pulmonary edema, >80 ppm; cardiovascular related mortality, >500 ppm; mortality via central nervous system depression, >500 ppm).
- Strong dose-temporal concordance occurs among the 4 proposed key events: cytochrome oxidase inhibition (>10 ppm), neuronal cell loss (>30 ppm), olfactory nasal lesions (defined as both neuronal cell loss and basal cell hyperplasia; >30 ppm), and rhinitis (80 ppm).
- We conclude that AOP1a leads to the critical effect in animals (olfactory nasal lesions) and likely manifests as olfactory paralysis in humans (AOP1b), suggesting that this point of departure is relevant for humans and protective against low-dose effects.

REFERENCES

1. Agency for Toxic Substances and Disease Registry (ATSDR). 2016. Toxicological profile for Hydrogen Sulfide / Carbonyl Sulfide. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service.
2. ACGIH. 2010. Hydrogen Sulphide. Documentations of the Threshold Limit Values and Biological Exposure Indices, Seventh Edition. Cincinnati, Ohio. American Conference of Governmental Industrial Hygienists (ACGIH)
3. SCOEL. 2007. Recommendation from the Scientific Committee on Occupational Exposure Limits (SCOEL) for Hydrogen Sulphide. European Commission SCOEL/SUM/124.
4. EPA US. 2003. Toxicological review of hydrogen sulphide (CAS No. 7783-06-4). US Environmental Protection Agency EPA/635/R-03/005.
5. OECD. 2018."Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways", OECD Series on Adverse Outcome Pathways, No. 1, OECD Publishing, Paris, https://doi.org/10.1787/5jlv1m9d1g32-en.
6. Meek, M.E., et al., 2014. Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. 34(6): p. 595-606.
7. Anantharam P, et al., 2017. Characterizing a mouse model for evaluation of countermeasures against hydrogen sulfide–induced neurotoxicity and neurological sequelae. Annals of the New York Academy of Sciences 1400:46-64.
8. Haouzi P, Sonobe T, Judenherc-Haouzi A., 2016. Developing effective countermeasures against acute hydrogen sulfide intoxication: challenges and limitations. Annals of the New York Academy of Sciences 1374:29-40.
9. Wu N, et al., 2011. Myocardial and lung injuries induced by hydrogen sulfide and the effectiveness of oxygen therapy in rats. Clinical Toxicology 49:161-166.

Complex Evidence Integration Using Evidence-Based Tables: A Case-Study Using the NTP Cancer Hazard Assessment of Night Shift Work and Light at Night Related to Circadian Disruption

Suril Mehta[1], Pamela Schwingl[2], Gloria Jahnke[1], Stanley Atwood[2], Sanford Garner[2], Ruth Lunn[1]

[1]Office of the Report on Carcinogens, DNTP, NIEHS, Research Triangle Park, NC, [2]Integrated Laboratory Systems, Morrisville, NC

Background

- Cancer hazard evaluations can be complex, requiring an accounting for a breadth of data, assessing multiple exposure-response relationships, and integrating evidence across multiple data streams (i.e., animal, human, and mechanistic studies).
- Evidence-based figures and tables can be used to visually collate and compare relevant evidence across studies to reach hazard conclusions.
- NTP's draft cancer hazard assessments of both night shift work (NSW) and light at night (LAN) highlight the use of evidence-based tables and figures to help reach an overall weight-of-evidence hazard conclusion.
- As a case study, we use NSW and breast cancer incidence for this method of evidence integration.

Establish framework ("PECO")

Evidence "Population"	Exposure (E) vs. Comparison (C)	E vs. C or Intermediate Outcome	Intermediate Outcome	Outcome
Human	NSW vs. DSW	CD		
Animals	Simulated NSW vs. SL	CD		
Human	NSW vs. DSW		KC	
Animals	Simulated NSW vs. SL		KC	
Human		CD		Cancer*
Animals				Cancer
Animals/ *In vitro*		CD/melatonin clock gene	KC/hallmarks of cancer	
Humans	NSW vs. DSW			Cancer
Animals	Simulated NSW vs. SL			Cancer

*Includes breast, prostate, colorectal, female hormonal, and lung cancers.

Note: Circadian disruption can be considered an intermediate in the pathway and is evaluated as an exposure and an outcome.

Abbreviations for poster

CD = circadian disruption, DSW = day shift work, KC = key characteristics of carcinogens, LAN = light at night; MOA= mode of action; NSW = night shift work, PECO = population, exposure, comparison/comparator, outcome; SL = standard lighting conditions.

NTP Draft Monograph

- To access NTP's Draft Monograph on Night Shift Work and Light at Night, please visit:

https://ntp.niehs.nih.gov/pubhealth/roc/listings/shiftwork

Evaluate by evidence stream

- We evaluated evidence from animal, human, and mechanistic studies for multiple exposure-intermediate-outcome relationships.
- Due to the complexity of the carcinogenicity pathway, multiple data types had to be integrated prior to consideration of toxicology and epidemiology data.

Example: Human epidemiology studies of NSW and breast cancer

Reference	Ever worked	Duration	Frequency, cumulative duration	Younger age	Receptor positive
Moderate to strong evidence of a positive association - informative studies					
Wegrzyn *et al.* 2017 (NHS2)					
Davis *et al.* 2001					
Grundy *et al.* 2013					
Hansen & Lassen 2012					
Hansen & Stevens 2012					
Lie *et al.* 2011, Lie *et al.* 2013					
Menegaux *et al.* 2013, Cordina-Duverger *et al.* 2016					
Some evidence for a positive association - informative studies					
Knutsson *et al.* 2013					
Fritschi *et al.* 2013, 2017					
Papantoniou *et al.* 2015					
Pesch *et al.* 2010, Rabstein *et al.* 2013					
Some evidence for a positive association - lower utility studies					
Åkerstedt *et al.* 2015					
UK EPIC Oxford, Travis *et al.* 2016					
Million Women, Travis *et al.* 2016					
Tynes *et al.* 1996					
Hansen 2001					
Wang *et al.* 2015					
No evidence of a positive association					
Li *et al.* 2015					
Vistisen *et al.* 2017					
Pronk *et al.* 2010					
O'Leary *et al.* 2006					

Blue: positive; Tan: null; Purple: negative

Assessing strengths, weaknesses and confidence: Series of evidence-based tables

Table 1: Detailed analysis of data for specific evidence stream

Exposure	Outcome	Type of studies	Strengths & Limitations	Assessment
Night shift work	Breast cancer	13 human case-control studies	Detailed exposure assessment	Consistency of evidence
Night shift work	Breast cancer	9 cohort studies	Left truncation, survivor bias	Supporting evidence
Night shift work	CD: Melatonin	Human cross-sectional	Large number of studies	Consistency of evidence

Table 2: Mechanistic-related data

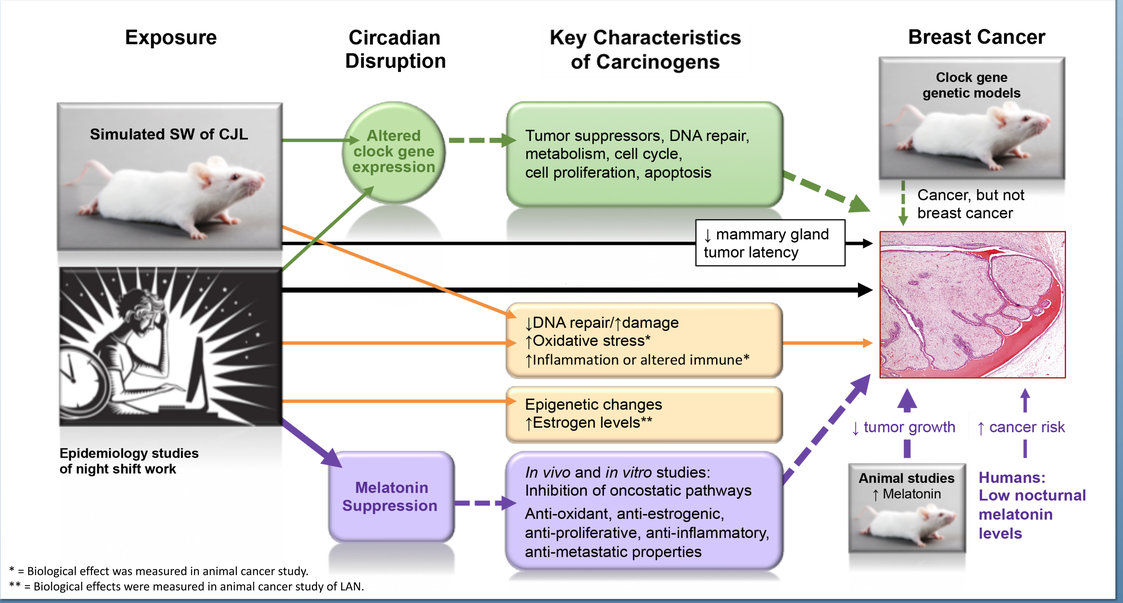
MoA	Out-come	Evidence stream	Confidence of the evidence	Assessment
Melatonin suppression	Cancer	▪ NSW: Melatonin human epidemiology studies ▪ LAN: Experimental studies in humans and animals ▪ Melatonin: Mechanistic and cancer studies	▪ Strong evidence: LAN & NSW are associated with ↓ or altered melatonin ▪ Strong evidence: melatonin ↓ tumor growth and is oncostatic	Strong evidence: melatonin is involved in carcinogenicity Indirect NSW; direct LAN

* Streamline examples: See monograph for complete entries.

Table 3: Integrate evidence across evidence streams

Exposure	Endpoint	Evidence stream or approach	Confidence in the evidence	Draft conclusions
Night shift work	Breast cancer	**Human epidemiology studies** ▪ 21 studies of independent populations ▪ Pooled analysis of 5 case-control studies **Experimental animal studies** ▪ Simulated shift work and jet lag in susceptible transgenic mice **Mechanistic and biomonitoring data** ▪ Melatonin suppression Circadian disruption Biological effects associated with cancer ▪ Other effects: sleep deprivation and vitamin D deficiency	▪ Strong evidence that persistent night shift work (frequent and long-term night shift work, especially among women who began night shift work at a younger age) is associated with an increased risk of breast cancer. ▪ Some evidence that simulated shift work or chronic jet lag decreased mammary-gland tumor latency or increased mammary-gland tumor multiplicity. ▪ Indirect evidence that melatonin suppression contributes to breast cancer development in night shift workers. ▪ Strong but indirect evidence that circadian disruption contributes to breast cancer development. ▪ Night shift work is associated with effects that are consistent with several of the key characteristics of carcinogens and also consistent with effects mediated by melatonin and altered clock-gene expression. Epigenetic effects may be a key early step responsible for altered gene expression. ▪ Role of vitamin D and sleep in night shift work is unclear.	***Human carcinogen*** *Persistent night shift work that causes circadian disruption* Strong evidence of carcinogenicity of persistent night shift work from studies in humans. Strong toxicological and mechanistic data providing evidence that circadian disruption plays a role in the cancer pathway in humans. Risk patterns in human cancer studies — younger age, hormone-receptor positive — supported by mechanistic data and biology of breast cancer development. Exposure to LAN may contribute to cancer risk, but data are inadequate to evaluate a direct association.

Visualize findings using conclusions from evidence-based tables



* = Biological effect was measured in animal cancer study.
** = Biological effects were measured in animal cancer study of LAN.

Conclusions

- Evidence-based tables and graphics provide transparency and a systematic structure for reaching cancer hazard conclusions.
- Tables provide information regarding the approaches used to evaluate the relationship, strengths and limitations of the studies, an assessment of confidence in the evidence, and integration of the evidence.
- Given multiple disciplines involved with this complex evaluation, our evidence-based tables and figures allowed for a collaborative approach to collate, compare and triangulate across evidence streams and the causal pathway.
- Evidence integration also facilitated the contextualization of the hazard exposure.

Mechanistic evidence integration case study: using ten key characteristics of carcinogens and a systematic review approach for antimony trioxide (Sb_2O_3) cancer hazard identification

Amy Wang[1], Joanne Trgovcich[2], Kristine L. Witt[1], Andrew Ewens[3], Jessica Geter[4], Sanford Garner[3], Gloria Jahnke[1], Stephanie L. Smith-Roe[1], Ruth Lunn[1]

[1]NTP, National Institute of Environmental Health Sciences (NIEHS), [2] ICF, [3] ILS, [4] Formerly ILS

Goal

Apply an approach that uses ten key characteristics of carcinogens (KCs) to evaluate mechanistic information of substances' carcinogenicity to identify human cancer hazard

Background

Evaluating a substance's broad and voluminous possible carcinogenic mechanism is challenging. To search mechanistic information unbiasedly (without perceived emphasis), we had developed search strings[1] for

PubMed Scopus Clarivate Analytics = Web of Sciences

on 10 key characteristics of carcinogens (KCs)[2]

	The ability of a substance to
KC1	act as an electrophile either directly or after metabolic activation
KC2	be genotoxic
KC3	alter DNA repair or cause genomic instability
KC4	induce epigenetic alterations
KC5	induce oxidative stress
KC6	induce chronic inflammation
KC7	be immunosuppressive
KC8	modulate receptor-mediated effects
KC9	cause immortalization
KC10	alter cell proliferation, cell death, or nutrient supply

Systematic review is an approach that aims to answer a specific question while minimizing bias. Advantages of systematic review (vs. descriptive literature review):

- useful for handling inconsistent results
- use a pre-determined protocol (e.g., search terms, inclusion and exclusion criteria)
- consistent evaluation of study quality (e.g., risk of bias, study utility)
- more comprehensive
- more transparent

Sb_2O_3 as a case study in NTP Report on Carcinogens (RoC)

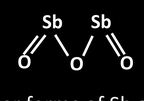
NTP Report on Carcinogens (RoC)[3] is mandated by Congress to identify cancer hazards for people living in the US

- Overall evaluation is based on human cancer studies, animal cancer studies, and mechanistic and other relevant information

Listing	Listing criteria
Known to be human carcinogen	*Sufficient* evidence of cancer in humans (including mechanistic information in humans)
Reasonably anticipated to be human carcinogen	*Limited* evidence of cancer in humans, OR
	Substance belong to a class whose member is listed in RoC, OR
	Convincing mechanistic information the substance would likely cause cancer in humans

- RoC is used by the public and various agencies for decision making

Antimony trioxide is used in making flame retardants, polyethylene terephthalate (PET) plastics, specialty glass, and paints



Occupational exposure is the main source of exposure

- Other forms of Sb at high temperature could turn into Sb_2O_3
- People can be exposed to Sb_2O_3 from breathing contaminated air (e.g., house dust from wear and tear of flame-retardant treated textiles, traffic pollution)

- Level of evidence of cancer in humans is *inadequate*
- Level of evidence of cancer in experimental animal is *sufficient*

Mice had increased incidences of

Lung tumors	
Benign	Alveolar/bronchiolar adenoma (F)[a]
Malignant	Alveolar/bronchiolar carcinoma (M & F)[a]
Combined	Alveolar/bronchiolar adenoma or carcinoma (F)[a]

Skin tumors	
Benign	Fibrous histiocytoma (M)[a]
Combined	Fibrous histiocytoma or fibrosarcoma (M)[a]

Lymphoma	
Malignant	Lymphoma (F)[a]

Rats had increased incidences of

Lung tumors	
Benign	Alveolar/bronchiolar adenoma (M* & F)[a]
Combined	Alveolar/bronchiolar adenoma or carcinoma (M*)[a]
Benign	Bronchiolar/alveolar adenoma (F)[b]
Malignant	Squamous-cell carcinoma (F)[b]
Malignant	Scirrhous carcinoma (F)[b]
Malignant	Scirrhous carcinoma (F)[c]

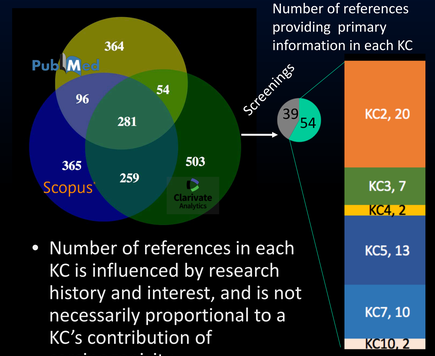
Adrenal gland tumors	
Benign	Pheochromocytoma (M & F)[a]
Combined	Pheochromocytoma (F)[a]

*M: carcinogenicity in male rats based on multiple factors

[a]NTP 2017 TR-590 [b]Groth et al. 1986 [c]Watt 1983 [d]Newton et al. 1994 Rat study reported no increase in tumors

Mechanistic information evaluation method

(1) Search and screen literature



- Number of references in each KC is influenced by research history and interest, and is not necessarily proportional to a KC's contribution of carcinogenicity

(2) Study evaluation

For genotoxicity studies, as an example, we considered

- Substance: identity, purity, solubility, etc.
- Study design and report: model system (e.g., humans, animals, in vivo, in vitro, biochemical, in silico), exposure route, directness (of measurement to interested events), sensitivity (e.g., detection method, group size), etc.
- Study utility to inform carcinogenicity

(3) Synthesis of mechanistic information by KCs

All relevant data, including literature, Tox21, and omic data, are considered.

Results

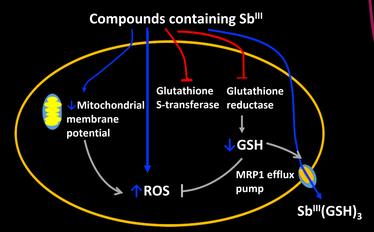
KC1 Sb_2O_3 interacts with

- Peptides (e.g., GSH)
- Proteins/ enzymes (including zinc finger)

Sb^{III} is highly reactive to sulfhydryl groups (thiols), especially vicinal thiol groups

KC5 Sb_2O_3 ↓ antioxidants (e.g., GSH)

- Sb^{III} compounds directly inhibit redox enzymes



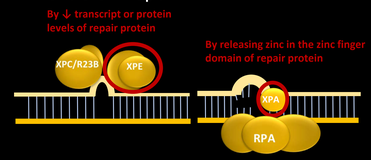
→ Compounds containing Sb^{III} leads to

⊣ Compounds containing Sb^{III} inhibits

→ Possible contribution to

KC8 **Antimony(III) potassium tartrate** prevents decrease in epidermal growth factor receptor (EGFR) → prevent cell differentiation → could preserve proliferation potential

KC3 Sb_2Cl_3 interferes with nucleotide excision repair



KC2 Sb_2O_3 is genotoxic

	In vitro	*In vivo*
Any DNA damage (prokaryotes)	+	+
Any DNA damage (eukaryotes)	+	+
Chromosomal aberrations	+	-[a]
Micronucleus induction	+[b]	+
Sister chromatid exchange	+	No data
Any mutation (prokaryotes)	-	No data
Any mutation (eukaryotes)	-	-*

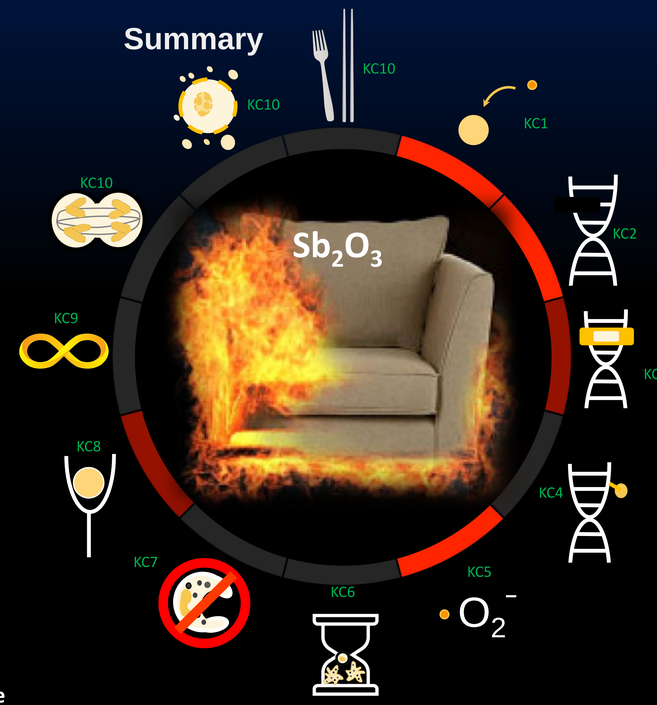
+ positive - negative

[a] Negative in rats; uncertain in mice due to severe study limitations.

[b] Correction from public comment version monograph

* mutations seen in Sb_2O_3-induced lung tumors

Summary



Direct evidence from Sb_2O_3

Direct evidence from inorganic compounds containing Sb^{III}

No direct evidence suggesting a major role of $Sb^{III}{}_2O_3$ carcinogenicity

- Based on *sufficient* animal evidence and supporting mechanistic information NTP recommends Sb_2O_3 be listed as *reasonably anticipated to be a human carcinogen* [4]
- KCs provides an unbiased approach for searching possible mechanisms

Next steps

Develop a more structured framework for mechanistic information evaluation. It may include

- guiding questions for study quality evaluation
- guidelines for mechanistic evidence synthesis
- descriptors and criteria to determine level of evidence for mechanistic information

References

1. ROC handbook, including search strings https://ntp.niehs.nih.gov/pubhealth/roc/handbook/index.html
2. Smith MT et al 2016. Environ. Health Perspect. 124:713-21
3. NTP Report on Carcinogens https://ntp.niehs.nih.gov/pubhealth/roc/index.html
4. A final decision for listing in the RoC has not been made.
5. Antimony Trioxide RoC Monograph https://ntp.niehs.nih.gov/pubhealth/roc/listings/antimonyt/index.html

Using study evaluation to inform evidence integration: Application in a systematic review of hexavalent chromium male reproductive outcomes

Erin E. Yost[1], Xabier Arzuaga[2], Alan Sasso[2], Catherine Gibbons[2]

1. US Environmental Protection Agency, National Center for Environmental Assessment, Research Triangle Park, NC
2. US Environmental Protection Agency, National Center for Environmental Assessment, Washington, DC

www.epa.gov

Erin E. Yost ❙ yost.erin@epa.gov ❙ 919-541-3906

Background

Study evaluation is used in systematic reviews to identify the strengths and weaknesses of the evidence base in a consistent and transparent manner. These evaluations can be used to inform evidence integration by identifying factors that may affect the reliability and interpretability of the results. Here, we describe how this principle was applied in a systematic review of the male reproductive effects of hexavalent chromium [Cr(VI)].

Methods

Literature search and screening: This evaluation of male reproductive effects was conducted as part of a systematic review of the health effects of Cr(VI) exposure. Studies were identified by searching three online databases (PubMed, Web of Science, Toxline) through May 2018. Title/abstract screening followed by full-text screening was used to identify animal studies meeting the following PECO (**P**opulation, **E**xposure, **C**omparators, **O**utcomes) criteria:

- **P**: Nonhuman mammalian animals (whole organism) of any life stage
- **E**: Any exposure to Cr(VI) by oral or inhalation routes
- **C**: Concurrent vehicle control or untreated control group
- **O**: All cancer outcomes; noncancer outcomes in relevant target systems

The literature search identified 23 animal toxicology studies that examined effects on the male reproductive system. Studies included evaluation of:

- Male fertility
- Sperm parameters
- Reproductive hormones
- Reproductive organ weights
- Anogenital distance (AGD)
- Sexual behavior

Study evaluation: Each of these studies was evaluated by at least two independent reviewers for reporting quality, risk of bias, and sensitivity using the domain-based approach outlined in Figure 1. Based on the results of the evaluation, each study was rated overall as *high* confidence, *medium* confidence, *low* confidence, or *uninformative*. Evaluations were performed on an outcome-specific basis, as the utility of a study may vary across outcomes.

Figure 1. Study evaluation process

Judgment	Interpretation
Good	Appropriate study conduct relating to the domain & minor deficiencies not expected to influence results.
Adequate	A study that may have some limitations relating to the domain, but they are not likely to be severe or to have a notable impact on results.
Deficient	Identified biases or deficiencies interpreted as likely to have had a notable impact on the results or prevent reliable interpretation of study findings.
Critically Deficient	A serious flaw identified that is interpreted to be the primary driver of any observed effect or makes the study uninterpretable. Study is not used without exceptional justification.

Animal Study Evaluation Domains

- Reporting Quality
- Selection or Performance Bias
  - Allocation
  - Observational Bias / Blinding
- Confounding/Variable Control
- Selective Reporting and Attrition
- Exposure Methods Sensitivity
  - Chemical Administration and Characterization
  - Exposure Timing, Frequency, and Duration
- Outcome Measures and Results Display
  - Endpoint Sensitivity and Specificity
  - Results Presentation

Overall study rating

Rating	Interpretation
High	No notable deficiencies or concerns identified; potential for bias unlikely or minimal; sensitive methodology.
Medium	Possible deficiencies or concerns noted, but resulting bias or lack of sensitivity would be unlikely to be of a notable degree.
Low	Deficiencies or concerns were noted, and the potential for substantive bias or inadequate sensitivity could have a significant impact on the study results or their interpretation.
Uninformative	Serious flaw(s) makes study results unusable for hazard identification

Evidence synthesis: Evidence was synthesized across studies, using the following considerations to articulate the strengths and weaknesses of the dataset: consistency, biological gradient (dose-response), strength (effect magnitude) and precision, biological plausibility, and coherence. Careful examination was given to the potential impacts of risk of bias and sensitivity on the conclusions. Relevant mechanistic data identified in the literature search was considered as part of the weight of evidence for biological plausibility. Based on this synthesis, the evidence was assigned a conclusion of *robust*, *moderate*, *slight*, *indeterminate*, or *compelling evidence of no effect*.

Results

Table 1. Study evaluation results. These results represent the composite ratings for male reproductive outcomes within each evaluation domain; there were some instances where outcomes within the same study were rated differently due to outcome-specific concerns, in which case an average rating (representative of most outcomes) is shown here. In addition to the 15 studies shown in this table, 8 studies were considered *uninformative* due to serious flaws in the study design (e.g., use of wild-caught animals) or reporting (e.g., data could not be interpreted) and were excluded from consideration.

Reference	Study description			Study evaluation										
	Species/ Strain	Exposure life stage and duration	Route of exposure	Reporting quality	Test animal allocation	Blinding of investigators	Confounding / variable control	Reporting or attrition bias	Characterization of exposure	Utility of exposure design	Sensitivity, specificity, and usability of results	Presentation of results	Overall confidence
NTP 1997	Mouse (BALBC)	Reproductive Assessment by Continuous Breeding	Diet	G	G	NR	G	G	G	G	G	G	High
NTP 1996a	Mouse (BALBC)	Adult males and females; 3, 6, or 9 weeks	Diet	G	G	G	G	A	G	G	G	G	High
NTP 1996b	Rat (Sprague-Dawley)	Adult males and females; 3, 6, or 9 weeks	Diet	G	G	G	G	G	G	G	G	A	High
NTP 2007	Rat (F344/N), Mouse (B6C3F1, BALB/c, C57BL-6)	Adult males and females; 3 months	Drinking water	G	G	A	G	A	G	G	G	G	High
Elbetieha et al. 1997	Mouse (Swiss)	Adult males and females; 12 weeks	Drinking water	G	NR	NR	D	A	D	G	G	A	Low
Bataineh et al. 1997	Rat (Sprague-Dawley)	Adult males; 12 weeks	Drinking water	D	A	NR	D	D	D	G	A	G	Low
Yousef et al. 2006	Rabbit (NZ white)	Adult males; 10 weeks	Oral gavage	G	A	NR	D	D	D	G	G	D	Low
Li et al. 2001	Rat (Wistar)	Adult males; 6 weeks	Oral feeding	D	NR	NR	D	A	D	A	A	A	Low
Rasool et al. 2014	Mouse (strain not reported)	Adult males; 30 or 60 days	Oral (unspecified)	D	A	NR	D	A	D	G	D	D	Low
Wang et al. 2015	Rat (Sprague-Dawley)	Adult males; 4 weeks	Drinking water	G	A	NR	G	G	A	G	D	A	Low
Kumar et al. 2017	Rat (Wistar)	F1 offspring; GD 9–14	Drinking water	A	NR	NR	G	D	D	A	A	D	Low
Al-Hamood et al. 1998	Mouse (BALBC)	F1 offspring; GD 12–PND 20	Drinking water	G	D	NR	D	D	D	G	G	D	Low
Glaser et al. 1986	Rat (Wistar)	Adult males; 18 months	Inhalation	A	A	NR	G	A	A	A	D	A	Low
Glaser et al. 1985	Rat (Wistar)	Adult males; 28 or 90 days	Inhalation	A	A	NR	G	A	A	G	D	A	Low
Kim et al. 2004	Rat (Sprague-Dawley)	Adult males; 90 days	Inhalation	G	A	NR	G	A	A	A	A	D	Low

Abbreviations: Gestation day (GD); Postnatal day (PND); Good (G); Adequate (A); Deficient (D); Not Reported (NR)

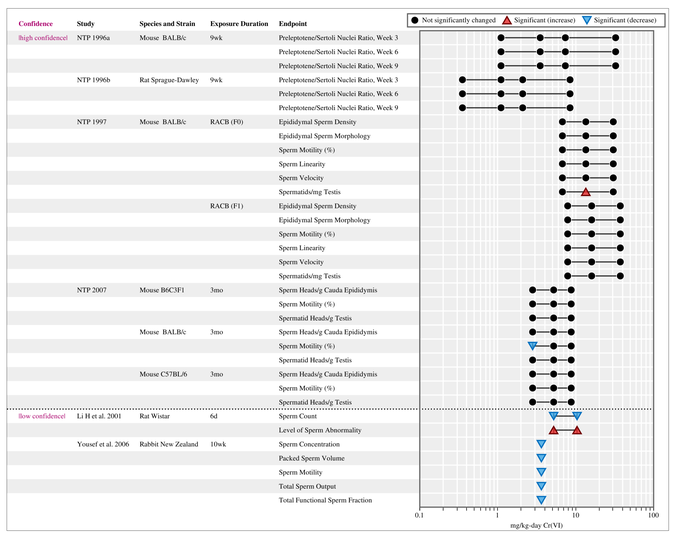


Figure 2. Summary of effects on sperm. Data is shown for all studies for which the ingested dose of Cr(VI) could be calculated. Decreased sperm count, mobility, and viability were also observed in the *low confidence* study by Kumar et al. 2017, but the ingested dose of Cr(VI) could not be calculated based on the reported information.

Summary of Effects:

- *High confidence* subchronic oral exposure studies in rats and mice (NTP 1996a, 1996b, 2007) and a continuous breeding study in mice (NTP 1997) generally indicated that the male reproductive system is not affected by Cr(VI) exposure.
- *Low confidence* oral exposure studies consistently observed effects on sperm quality and quantity, testicular histopathology, male reproductive organ weights, hormone levels, sexual behavior, and AGD.
- As an example, Figure 2 summarizes effects on sperm parameters across studies.
- Biological plausibility for male reproductive effects of Cr(VI) exposure was supported by mechanistic studies (in vivo and in vitro) demonstrating oxidative stress and apoptosis in male reproductive tissues, altered steroidogenic signaling, disruption of the blood-testis barrier, and alterations in meiosis.
- No effects were observed in three *low confidence* inhalation studies.

Summary of effects in high vs. low confidence studies

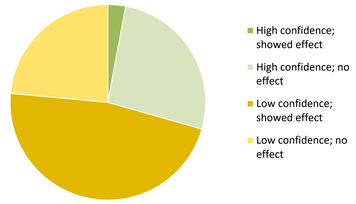


Figure 3. Incidence of outcomes indicative of male reproductive effects across *high* and *low confidence* Cr(VI) animal studies. One *high* confidence study observed decreased testis weight, but otherwise *high* confidence studies found no evidence of male reproductive effects. Comparative, male reproductive effects was frequently observed in *low* confidence studies.

Integration of evidence

It was concluded that animal toxicology studies along with supportive data from mechanistic studies provide *slight* evidence that Cr(VI) is a male reproductive toxicant. The rationale for this conclusion is documented in an evidence profile table (Table 2). Relatively severe male reproductive effects were observed across multiple *low* confidence studies and are supposed by mechanistic evidence. However, similar effects were not observed in *high* confidence studies, and concerns were raised about the potential impact of bias on the interpretation of the results in *low* confidence studies. Fertility (ability to produce offspring) was not affected in any studies but this did not affect overall conclusions, since rodents can remain fertile after large reductions in sperm count.

Table 2: Evidence profile table for Cr(VI) male reproductive effects

	Studies [confidence]	Factors that increase strength	Factors that decrease strength	Summary of findings
Fertility and Fecundity	NTP 1997 [high] Bataineh et al. 1997 [low] Elbetieha et al. 1997 [low] Al-Hamood et al. 1998 [low]		• Only study that observed an effect is considered *low* confidence • No effects observed in *high* confidence studies	No effects on male fertility (ability to produce offspring) were observed across studies in rats or mice, although one *low confidence* study in rats observed decreased fetal viability following paternal exposure (Elbetieha et al. 1997).
Sperm evaluation	NTP 1996a [high] NTP 1996b [high] NTP 1997 [high] NTP 2007 [high] Kumar et al. 2017 [low] Li et al. 2001 [low] Yousef et al. 2006 [low]	• Dose-response gradient • Biological plausibility (mechanistic evidence of oxidative stress, effects on blood-testis barrier, and altered meiosis)	• No effects observed in *high* confidence studies • Studies that observed effects were all considered *low* confidence	No effects on sperm parameters were observed in *high* confidence studies in rats or mice, whereas *low* confidence studies in rats and rabbits reported decreased sperm quality and quantity.
Histopathology	NTP 2007 [medium] Kumar et al. 2017 [low] Li et al. 2001 [low] Rasool et al. 2012 [low] Wang et al. 2015 [low] Glaser et al. 1985 [low] Kim et al. 2004 [low]	• Coherence with effects on sperm • Dose-response gradient • Biological plausibility (mechanistic evidence of oxidative stress and effects on blood-testis barrier)	• No effects observed in *high* confidence studies • Studies that observed effects were all considered *low* confidence	No histopathological effects were reported in the *high confidence* study in rats and a variety of mouse strains by NTP 2007, whereas three *low* confidence studies in rats and mice observed histopathological changes in the testis and seminiferous tubules.
Organ weight	NTP 1996a [high] NTP 1996b [high] NTP 1997 [high] NTP 2007 [high] Al-Hamood et al. 1998 [low] Bataineh et al. 1997 [low] Elbetieha et al. 1997[low] Kumar et al. 2017 [low] Yousef et al. 2006 [low] Wang et al. 2015 [low] Kim et al. 2004 [low] Glaser et al. 1986 [low]	• Coherence with decreased testosterone	• Unexplained inconsistency • Most studies that observed effects were considered *low* confidence	Decreased testis weight was observed in one out of three mouse strains in the *high confidence* study by NTP 2007, and decreased testis and accessory male reproductive organ weights were observed in four *low confidence* studies in rabbits (Yousef et al. 2006), rats (Bataineh et al. 1997, Kumar et al. 2017), and mice (Elbetieha et al. 1997). No effects were observed in the remaining 7 studies.
Hormones	Yousef et al. 2006 [low] Kumar et al. 2017 [low]	• Consistency • Biological plausibility (mechanistic evidence of decreased steroidogenesis)	• Few studies • Only *low* confidence studies available	Decreased testosterone was observed in rabbits exposed as adults, and decreased testosterone and gonadotropins were observed in F1 rats that had been exposed during gestation.
Sexual behavior	Bataineh et al. 1997 [low] Yousef et al. 2006 [low]	• Consistency	• Few studies • Only *low* confidence studies available	Decreased mounts, increased ejaculation latency and post-ejaculation interval, and decreased percentage of males ejaculating were observed in rats exposed as adults. Increased reaction time to mounting was observed in rabbits.
AGD	Kumar et al. 2017 [low]	• Coherence with decreased testosterone	• Single study • *Low* confidence	Decreased AGD was observed in F1 rats, which is consistent with the observation of decreased testosterone in these animals.

Disclaimer: The views expressed are those of the authors and do not necessarily represent the views/policies of the US EPA. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

U.S. Environmental Protection Agency
Office of Research and Development

#23 - Systematized Review Approaches in the Assessment of Tobacco Toxicants: Acrolein as a Case Study

***Mary Kushman*, R. Phillip Yeager, Susan Chemerynski, Roxana Weil, Xin Fu, and Hans Rosenfeldt**
ToxStrategies, Inc.

Research questions in tobacco science, such as understanding the biological effects of smoke toxicants like acrolein, benefit from rapid, transparent, and reproducible methods of evidence synthesis and integration. A systematized review for acrolein toxicity, conducted in approximately seven months, analyzed, synthesized, and integrated scientific evidence using a more-rapid approach over a traditional systematic review, while still addressing a specific research question. The key question of interest was whether acrolein is a major driver of tobacco smoke-related toxicity, especially that associated with noncancer respiratory disease resulting from exposure to cigarette smoke. A search strategy using publicly available databases was executed initially, employing pre-specified selection criteria to extract data from reviews and regulatory documents, from which a preliminary mode of action was built. Data from primary research articles were then evaluated to better inform the assessment of acrolein toxicity relevant to toxic endpoints, dose-response relationships, mechanisms or modes of action, with information from different streams of evidence integrated to support the role of acrolein in tobacco smoke-related respiratory diseases. Key cellular processes in the proposed MOA for acrolein toxicity, inflammation and necrosis, were presented as narrative summaries, with accompanying graphical and tabular representation. This serves as an example of how scientific evidence can be evaluated, integrated, and presented in a timely yet reproducible and rigorous, stepwise manner.

This information is not a formal dissemination of information by FDA and does not represent agency position or policy.

Accelerating Chemical Assessments : A Case Study in Automatic Evidence Extraction from Text

Catherine Blake[1] and Jodi Flaws[2]

[1] School of Information Sciences and Department of Computer Science, [2] Department of Comparative Biosciences, College of Veterinary Medicine, University of Illinois at Urbana-Champaign.

Motivation

The manual processes used to extract mechanistic evidence from studies is one of the most time consuming steps when conducting a chemical assessment. Our goal is to automate evidence extraction in order to reduce the time to conduct a review and/or increase the scope of a review.

Method

Explicit claims (Blake, 2010) were identified from abstracts (n=3078) collected in a previous study (Korhonen et all 2012). Prior mode of action annotations were used to identify keywords using Shannon's measure of entropy and expert review. The number of supporting claims, where the MOA has increased (e.g. improve, extend), where there was some effect but the claim is neutral (e.g. change, effect), and where the MOA has decreased (e.g. reduce, inhibited) are reported.

Abstracts from Medline were pre-processed to identify sections and sentences. A dependency parse was generated (Manning, 2014) for each sentence in order to attend to elliptical coordinated compound noun phrases using the method described in (Blake and Rindflesch, 2017).

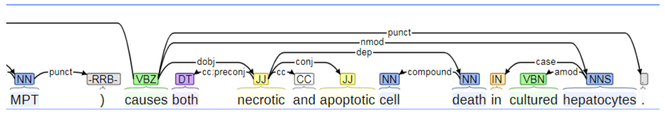


Figure 1 – Partial dependency parse for the sentence from PMID 12700412 "Onset of the mitochondrial permeability transition (MPT) causes both necrotic and apoptotic cell death in cultured hepatocytes.". The coordinated noun phrases would be resolved to produce "apoptotic cell death" and "necrotic cell death"

Explicit claims involve an agent, the nature of the change and an object (Blake, 2010). Explicit claims that include a keyword associated with cell proliferation and cell death were identified automatically. Claims were characterized as supporting if they increased the MOA, neutral if no directionality was provided and refuting if they decreased the MOA. Negation was detected.

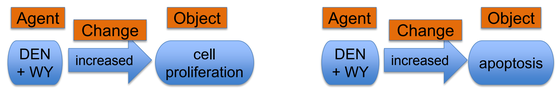


Figure 2 – Explicit claim with MOA from "DEN + WY increased both cell proliferation and apoptosis in both the wild-type and p50 +/- mice; DEN treatment alone has no effect.."

Abstracts were collected from an earlier study (Korhonen et al 2012) that characterized evidence and modes of action for 7 chemicals: 4-aminobiphenyl, asbestos, ethylene oxide, formaldehyde, genistein, methylene chloride, pyridine. This pilot study considers two of the non-genotoxic modes of action reported - cell proliferation and cell death.

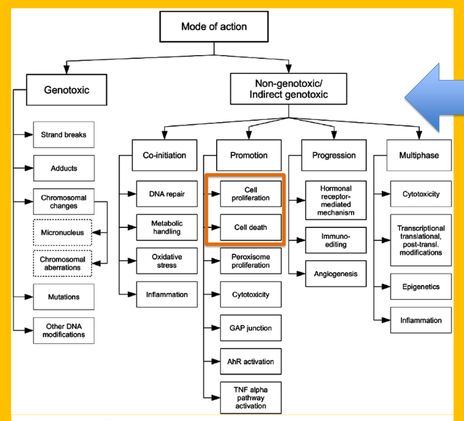
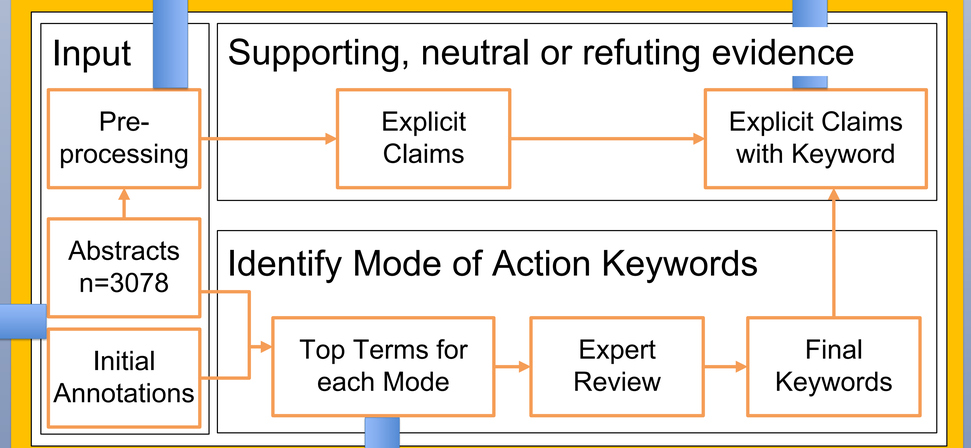


Figure 3 – The Modes of Action established in Korhonen et al (2012). Modes explored in this pilot study are shown in red.



Shannon entropy was used with the initial set of abstracts (e.g. cell proliferation versus not cell proliferation) to establish keywords for each mode of action, which were then reviewed by an expert. Noun phrases that include a keyword were identified in all abstracts from the collection.

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i).$$

Results

Refuting evidence
CaN inhibitor cyclosporine A (CsA) reduced cell growth in these cell lines.

Negated supporting evidence
We observed no treatment-related increases in cellular proliferation.

Cell Proliferation

Supporting		Neutral		Refuting	
Neg		Neg		Neg	
8	168	5	53	2	91

Neutral evidence
Binary mixtures of the compounds produced effects on cell proliferation and on each of the responsive protein ions that were fully consistent with concentration additivity.

Of the 327 cell proliferation claims, 99 (30.3%) directly contradicted the premise that cell proliferation increased because the evidence was either negated or reported a decrease.

Just over half (168, 51.4%) of the evidence supports the premise that cell proliferation actually increased.

Negated refuting evidence
Application of PCB 52 in calcium-free medium reduced the calcium accumulation, but did not reduce cell death.

Supporting evidence
Activation of protein kinase G is sufficient to induce apoptosis and inhibit cell migration in colon cancer cells.

Negated neutral evidence
The HDAC inhibitor-induced apoptosis appears to be p53 independent, because no change in apoptotic cell death was observed in H1299 cells that expressed exogenous wild-type p53 (H1299 cells express no endogenous p53 protein).

Cell Death

Supporting		Neutral		Refuting	
Neg		Neg		Neg	
5	416	8	74	3	153

Of the 659 cell death claims, 158 (24.0%) directly contradicted the premise that cell death had increased because the evidence was either negated or reported a decrease in cell death.

About two-thirds (416, 63.1%) of the evidence supports the premise that cell death increased.

Conclusions and Future Work

Results show that simply reporting a mode of action should not be interpreted as evidence that the MOA has increased. Explicit claims from the Claim Framework provide the granularity necessary to differentiate between supporting, neutral, and refuting claims for a given MOA. Further work is required to differentiate between claims made as background knowledge and the results from current experiments.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1535167.

Works Cited

Blake, C., Beyond genes, proteins, and abstracts: Identifying scientific claims from full-text biomedical articles. *Journal of Biomedical Informatics*, 2010. 43(2): p. 173-189.

Blake, C. and T. Rindflesch, Leveraging syntax to better capture the semantics of elliptical coordinated compound noun phrases. *Journal of Biomedical Informatics*, 2017. 72(120-131).

Manning, C.D., et al., The Stanford CoreNLP Natural Language Processing Toolkit in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2014, Association of Computational Linguistics: Baltimore, Maryland, USA. p. 55-60.

Korhonen, A., et al., Text mining for literature review and knowledge discovery in cancer risk assessment and research. *PLoS One*, 2012. 7(4): p. e33427.

ILLINOIS

EXPOSURE EVIDENCE INTEGRATION IN SYSTEMATIC REVIEW

Dream big.
Then call ICF.

Exposure Data Extraction in litstream™

Kevin Hobbie, Ashley Williams, Tom Feiler, Cara Henning and Heidi Hubbard | ICF

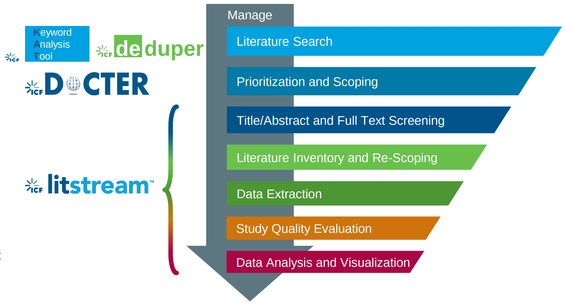
Contact
Kevin Hobbie
kevin.hobbie@icf.com

Systematic Reviews and the ICF Tools

Integrating Exposure Data Across Diverse Data Streams

Systematic review has grown out of the fields of healthcare and toxicology to become a standard practice in risk assessment. As such, best practices in systematic review and evidence integration are now required to be adapted to the field of exposure assessment.

Exposure studies and data are frequently more heterogeneous in nature than toxicology data and differ in structure and format from epidemiological studies. This requires new methodologies for systematically extracting, evaluating, and integrating exposure data.



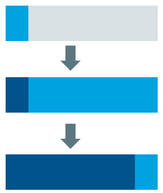
Flexible Extractions Forms for Exposure Data Curation

Flexible Forms Extraction Steps in litstream™

Due to the non-standardized format of most exposure data, the ability to create custom forms designed to meet the needs of individual assessments rather than relying on a set of pre-existing templates is frequently required.

Flexible extraction forms in ICF's litstream™ systematic review management application can be quickly and easily customized to accommodate the capture of differing streams of exposure data ranging from environmental monitoring data; biomonitoring data; data generated from laboratory experiments; and modeling data estimates for media concentrations, intake, and dose.

Blueprint Feature Workflow



Pilot step. Senior reviewer reviews a set of the studies, and the results are used to train other project members.

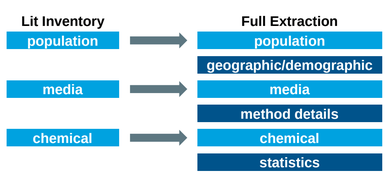
Main step. Pilot results are brought into the step and then the other studies can be assigned and completed.

QA step. A set of studies can be checked by a senior reviewer with other final results retained.

Features of Flexible Forms

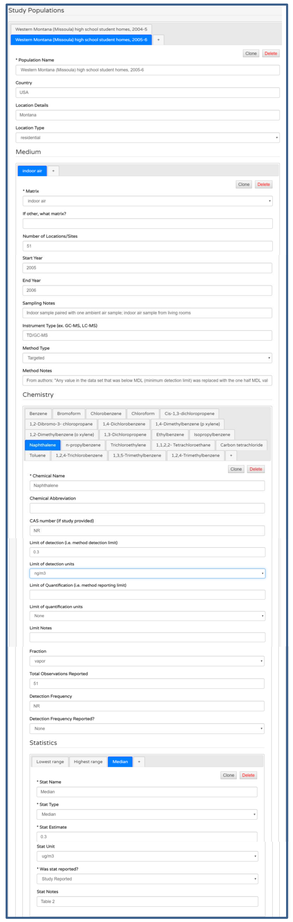
- **Assignment-based structure.** Extraction tasks are assigned to specific individuals, and studies receive a final result when the final task is finished.
- **Visual schema builder.** Assessors are able to design and test the layout and behavior of flexible extraction forms using a visual schema builder. Data containers are objects holding fields with text, numeric or boolean data. Lists can be employed to nest data where there is a many-to-one relationship.
- **Flexibility of ad-hoc edits.** Additional changes can be made by project admins outside of assignments when needed to overwrite final results.
- **Specific functionality for pilot and QA steps.** litstream™ integrates a blueprint feature used to document pilot steps, main extraction and downstream QC steps. This feature documents task assignment at each step while migrating data to downstream steps.
- **Extraction Lit Inventory.** Using the blueprint feature while conserving a nested data schema, users can catalog literature in inventory step then move targeted studies into downstream steps for full extraction where data from the inventory are prepopulated to save time.

Extraction Light Workflow



litstream™

Monitoring Data Curation



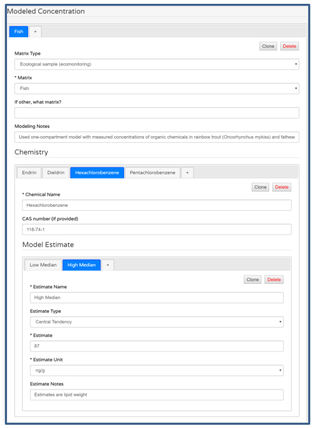
Data Schema

- Study
- Population
- Media
- Chemical
- Statistic

Considerations

- Flexibility to handle many media types encountered in:
  - Biomonitoring (ecological and human)
  - Environmental monitoring
- Capture enough metadata to explore data without creating an overburden
  - Geographic and temporal annotation
  - Population/receptor descriptions and classification
  - Detailed chemical identification
  - Sample collection and analysis details

Modeled Concentration Data Curation



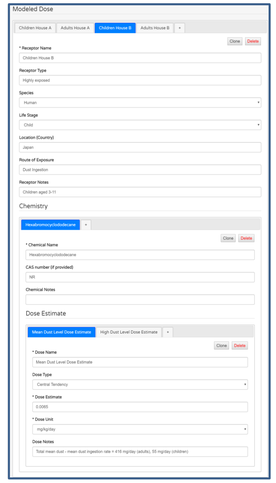
Data Schema

- Study
- Media
- Chemical
- Model Estimate

Considerations

- Capture modeling details
- Differentiate matrix type
- Detailed chemical identification
- Denote estimate type as an extreme range estimate or a central tendency

Modeled Intake/Dose Data Curation

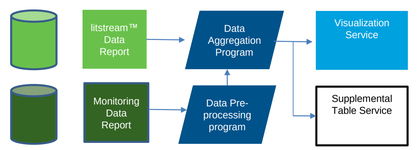


- Study
- Receptor
- Chemical
- Dose Estimate

Considerations

- Capture modeling details
- Differentiate receptors by life stage and country to enable conversion of intakes to dose using exposure factors
- Classify receptor population
  - General population
  - Highly exposed population
  - Occupational exposure
- Detailed chemical identification
- Denote estimate type as an extreme range estimate or a central tendency

Data Aggregation Approach



Flexible data aggregation workflow can accommodate data streams generated both within and outside of litstream™

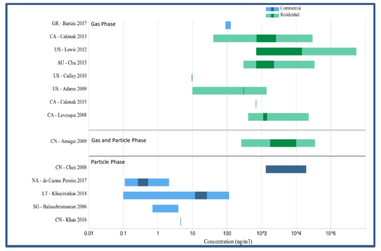
Data Aggregation Workflow

- Combine all data streams
- Convert to common units
- Select aggregation levels
- Calculate
  - Overall range of observations
  - Range of central tendencies
  - Total observations
  - Frequency of detection
- Gather (if applicable)
  - Geographic details
  - Sample details (species, receptor, etc.)
  - Detection limits
  - Sampling date ranges

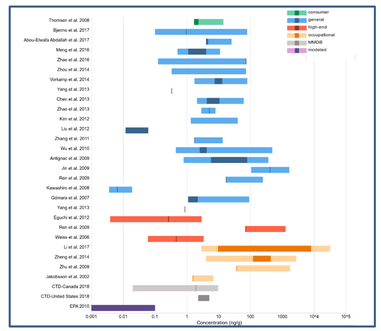
Exposure Data Visualization

- Visualize magnitudes of central tendency and ranges in common units
- Use color to distinguish between exposure population types or data types
- Utilize external services built in node.js/D3

Example Indoor Air Data



Example Blood Concentrations



Exposure Evidence Integration

- Exposure data can be curated and aggregated across a number of fields to explore spatially, temporally, by microenvironment, and more.
- Using this integration approach data are easily summarized and used as model inputs for exposure models.
- The litstream™ flexible form feature is versatile for data extraction and evaluation. We have successfully implemented flexible forms to extract and evaluate data from sources on exposure, toxicology, epidemiology, and economic impacts.

Acknowledgements
Logo Design and Graphics: ***Courtney Skuce***

Technological Tools for Evidence Integration

Shane Thacker, Jennifer Nichols, Ryan Jones, Steven J. Dutton

National Center for Environmental Assessment; Office of Research and Development; US EPA

www.epa.gov/research

Shane Thacker I thacker.samuel@epa.gov I 919-541-5159

Evidence Integration

At the EPA's National Center for Environmental Assessment (NCEA), we work closely with programs throughout the EPA to integrate web-based and desktop computer tools into the assessment process, facilitating evidence integration for science assessment products. By incorporating in-house and third-party tools, both open source and commercial, activities such as the Integrated Risk Information System (IRIS) and the Integrated Science Assessments (ISA) seek to use the best tools for the job, while remaining flexible enough to improve the evidence integration process.

Current Tools

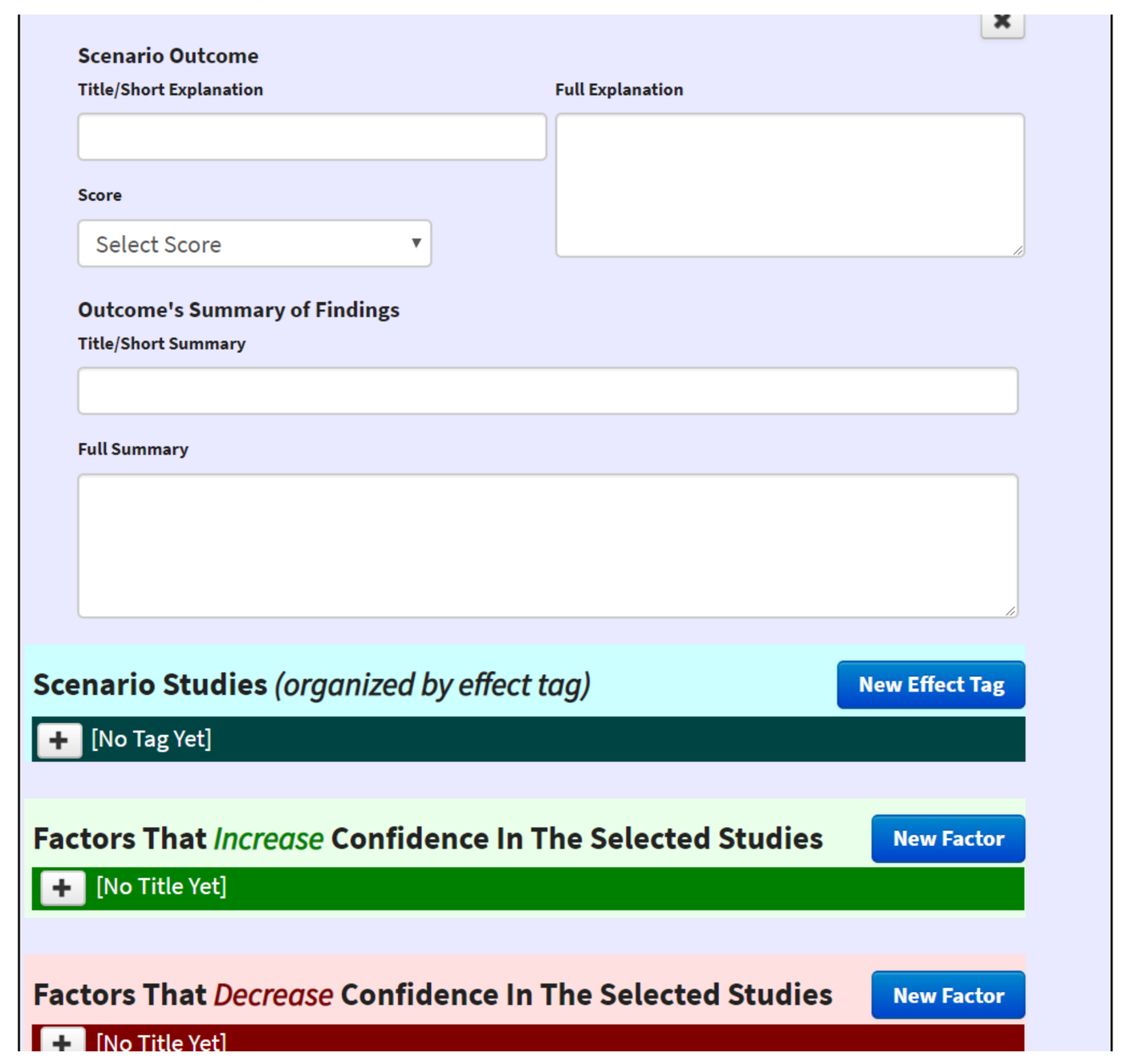
- **Health and Environmental Research Online (HERO)**: Literature search, categorization, acquisition, archiving. Interoperable with HAWC, Distiller, and SWIFT.
- **Health Assessment Workplace Collaborative (HAWC)**: Study evaluation, data extraction, visualization. Interoperable with HERO, BMDS, and Distiller.
- **Benchmark Dose Software (BMDS)**: Dose-response modelling. Interoperable with HAWC.
- **Evidence Partners DistillerSR**: Literature screening, data extraction. Interoperable with HERO and HAWC.
- **Sciome SWIFT-Review and SWIFT-Active Screener**: Literature screening, prioritization, categorization. Interoperable with HERO and HAWC.

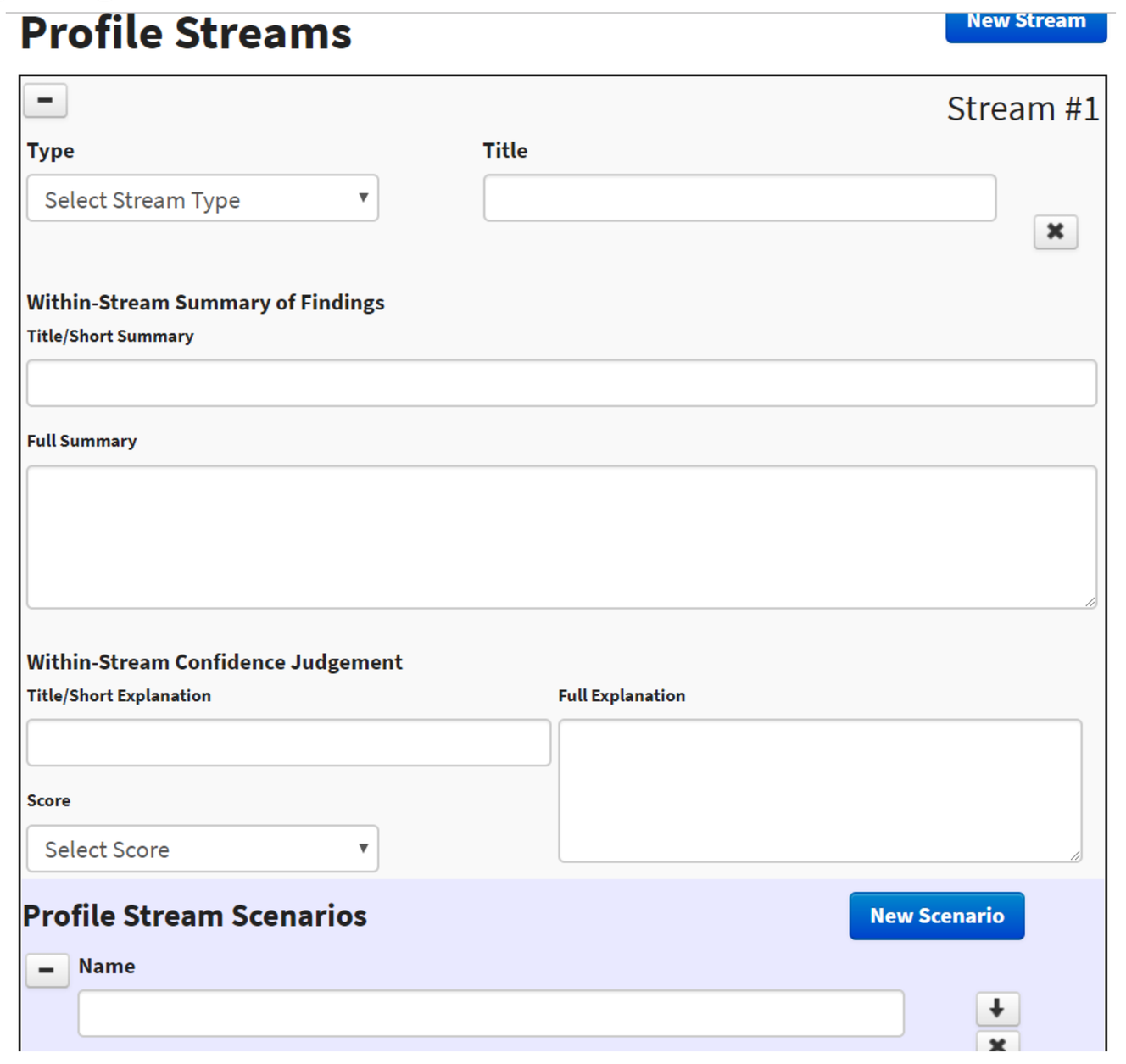
New Tools

Evidence Profile Table

Part of HAWC, the Evidence Profile Table offers a summary explanation of evidence integration in a chemical risk assessment. This view creates greater transparency about the body of evidence by illuminating the rationale behind the assessment findings.

- Adaptation of GRADEPro Evidence Profiles
- Create multiple rows to cover multiple evidence streams
- Select studies and endpoints added to HAWC
- Streams break down into scenarios
- Endpoints are rated within scenarios
- Confidence judgements build from individual to across-stream
- Findings summaries add to confidence judgments





New Tools, continued

Evidence Profile Table, pictured

Outcome	Studies	Factors That *Increase* Confidence	Factors That *Decrease* Confidence	Confidence Judgement and Summary of Findings for Individual Outcome	Within-Stream Confidence Judgement and Summary of Findings	Inference Across Streams	Across-Stream Confidence Judgement
test animal *(Animal Bioassay)*						•test inference this is another description	⊕⊕○ test total This is a test explanation
Test scenario	neoplastic •*Initial submission: Title of the study (Thacker and Jones 2019)*	•Consistency Relatively Low	•High Risk of Bias This is a factor that would decrease confidence.	⊕⊕○ Moderate test outcome This is another explanation	⊖○○ Slight test judgement This is a full explanation		
test outcome	Overall Odds Ratio •*Title of another study here (Thacker and Nichols 2019)*	•Consistency Relatively Low	•High Risk of Bias This is another factor that would decrease confidence.	⊕⊕⊕ Robust test title test Title of summary And the full summary.	test summary within test summary full		

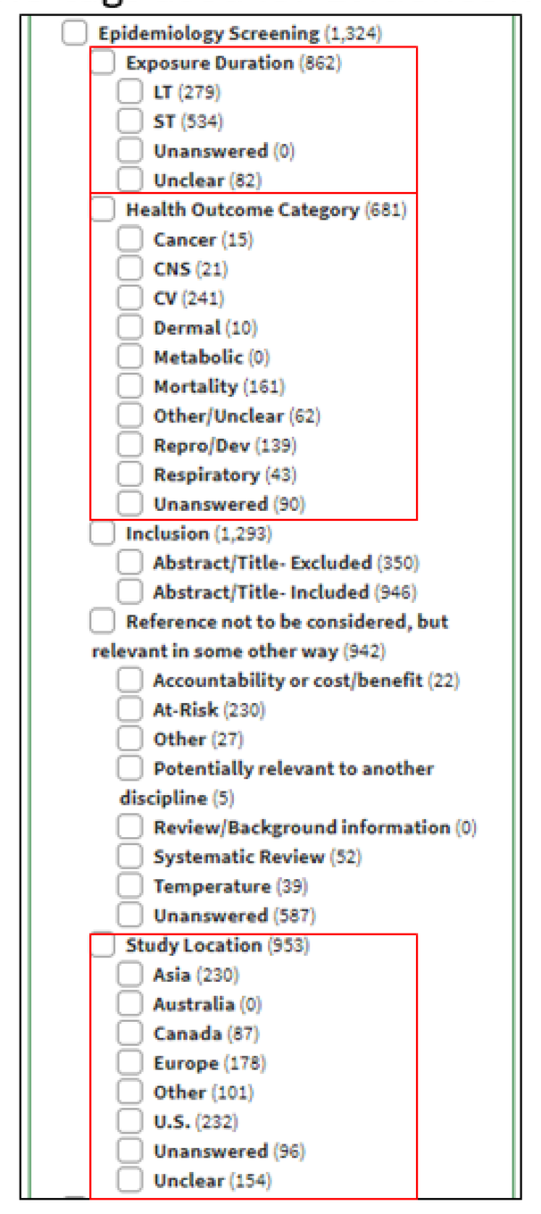
Evidence Mapping

Integrated with the Health and Environmental Research Online (HERO) database, the Evidence Mapping tool allows researchers to create heat maps to visualize and overlay characteristics (e.g., discipline, exposure, concentration, etc.) of the reviewed literature, making it easy to visualize the available evidence.

Example Evidence Map

Exposure	Location	Health Outcome Category: Mortality	CV	Resp	Repro/Dev	Metabolic	Nervous System	Cancer	Dermal	Other	Total
Short-term	US	16	36	58	11	2	0	0	0	5	128
	Canada	10	14	12	1	1	0	0	3	7	48
	Europe	23	36	38	6	2	5	0	1	5	116
	Asia	39	41	70	5	3	8	1	1	10	178
	Other	11	52	59	2	1	1	0	0	2	128
Long-term	US	15	8	18	44	5	4	5	0	7	106
	Canada	5	1	4	3	1	1	1	0	1	17
	Europe	5	5	11	16	6	0	2	0	1	46
	Asia	10	16	19	25	4	4	4	1	6	89
	Other	3	12	3	15	0	4	1	0	4	42
Unclear	US	3	1	3	1	1	1	1	1	1	13
	Canada	0	0	1	0	0	0	0	0	0	1
	Europe	0	3	4	2	1	0	0	0	2	12
	Asia	0	1	8	2	0	0	0	2	0	13
	Other	0	1	2	2	0	4	0	1	1	11
Total		140	227	310	135	27	32	15	10	52	948

HERO tags used for Evidence Map



- In HERO, scientists use tags to categorize literature for possible use in chemical risk assessment projects
- Using the tool, scientists create crosstabs between sets of tags that code literature by characteristics
- The results are color-coded, creating heat maps for easy visualization of the intersection totals
- The result is a map showing the amount of possible evidence between characteristics, such as location and exposure
- Researchers can layer the characteristics into sets and subsets, adding visual organization
- Interoperable with Distiller

New Tools, continued

Evidence Inventory

The Evidence Inventory tool, hosted within HERO, facilitates data extraction and portrayal by providing researchers a template to collect and categorize data from the relevant literature and then create summary tables of the extracted information. The summary tables are then ready for export into assessment documents, allowing readers to review the evidence behind the chemical risk assessment.

From HERO				Level 2 Screening									
HERO ID	Author	Year	Title	ISA Relevant	PECO Relevant	Reliability/Study Quality	Section	Health Endpoint	Study Design	Cohort or Study Name	Study Population Details	Sample Size	Cou
123456	Thacker, S; Joı	2016	Chemical pollutants and thei	Yes	Yes	High Quality	Infl/Ox Stress/Inju	FeNO	Panel		Adults with type II diabetes mellitus.	69	U.S.
123456	Thacker, S; Joı	2016	Chemical pollutants and thei	Yes	Yes	High Quality	Infl/Ox Stress/Inju	FeNO	Panel		Adults with type II diabetes mellitus.	69	U.S.
123456	Thacker, S; Joı	2016	Chemical pollutants and thei	Yes	Yes	High Quality	Infl/Ox Stress/Inju	FeNO	Panel		Adults with type II diabetes mellitus.	69	U.S.
123456	Thacker, S; Joı	2016	Chemical pollutants and thei	Yes	Yes	High Quality	ED Visit - AsEx		Time-series		All ages	611,970	U.S.
123456	Thacker, S; Joı	2016	Chemical pollutants and thei	Yes	Yes	High Quality	ED Visit - AsEx		Time-series		All ages	611,970	U.S.
123456	Thacker, S; Joı	2016	Chemical pollutants and thei	Yes	Yes	High Quality	ED Visit - AsEx		Time-series		All ages	611,970	U.S.
123456	Thacker, S; Joı	2016	Chemical pollutants and thei	Yes	Yes	High Quality	ED Visit - AsEx		Time-series		All ages	611,970	U.S.
123456	Thacker, S; Joı	2016	Chemical pollutants and thei	Yes	Yes	High Quality	ED Visit - AsEx		Time-series		All ages	611,970	U.S.
123457	Thacker, S; Joı	2015	Air pollution and quality of li	Yes	Yes	Adequate	ED Visit - AsEx		Time-series		Ages 5 and older	165,056	U.S.
123457	Thacker, S; Joı	2015	Air pollution and quality of li	Yes	Yes	Adequate	ED Visit - AsEx		Time-series		Ages 5 and older	165,056	U.S.
123457	Thacker, S; Joı	2015	Air pollution and quality of li	Yes	Yes	Adequate	ED Visit - AsEx		Time-series		Ages 5 and older	165,056	U.S.
123457	Thacker, S; Joı	2015	Air pollution and quality of li	Yes	Yes	Adequate	ED Visit - AsEx		Time-series		Ages 5 and older	165,056	U.S.
123458	Nichols, J; Jon	2015	Solar flares and leaded gasoli	Yes	Yes	Adequate	ED Visit - AsEx	Asthma/Wheeze	Time-series		All ages	34,086	U.S.
123458	Nichols, J; Jon	2015	Solar flares and leaded gasoli	Yes	Yes	Adequate	ED Visit - AsEx	Asthma/Wheeze	Time-series		All ages	34,086	U.S.
123458	Nichols, J; Jon	2015	Solar flares and leaded gasoli	Yes	Yes	Adequate	ED Visit - AsEx	Asthma/Wheeze	Time-series		All ages	34,086	U.S.
123458	Nichols, J; Jon	2015	Solar flares and leaded gasoli	Yes	Yes	Adequate	ED Visit - AsEx	Asthma/Wheeze	Time-series		All ages	34,086	U.S.
123458	Nichols, J; Jon	2015	Solar flares and leaded gasoli	Yes	Yes	Adequate	ED Visit - AsEx	Asthma/Wheeze	Time-series		All ages	34,086	U.S.
123458	Nichols, J; Jon	2015	Solar flares and leaded gasoli	Yes	Yes	Adequate	ED Visit - AsEx	Asthma/Wheeze	Time-series		All ages	34,086	U.S.
123458	Nichols, J; Jon	2015	Solar flares and leaded gasoli	Yes	Yes	Adequate	ED Visit - AsEx		Time-series		All ages	10,377	U.S.
123458	Nichols, J; Jon	2015	Solar flares and leaded gasoli	Yes	Yes	Adequate	ED Visit - AsEx	Pneumonia	Time-series		All ages	32,166	U.S.
123458	Nichols, J; Jon	2015	Solar flares and leaded gasoli	Yes	Yes	Adequate	ED Visit - AsEx		Time-series		All ages	186,449	U.S.
123459	Jones, R; Dutt	2014	Transhuman influences on ai	Yes	Yes	High Quality	ED Visit - AsEx		Case-crossover		All ages	122,607	U.S.
123459	Jones, R; Dutt	2014	Transhuman influences on ai	Yes	Yes	High Quality	ED Visit - AsEx		Case-crossover		All ages	122,607	U.S.
123459	Jones, R; Dutt	2014	Transhuman influences on ai	Yes	Yes	High Quality	ED Visit - AsEx		Case-crossover		All ages	122,607	U.S.
123459	Jones, R; Dutt	2014	Transhuman influences on ai	Yes	Yes	High Quality	ED Visit - AsEx		Case-crossover		All ages	122,607	U.S.

- Producing a chemical risk assessment document requires extracting and reformatting the data in cited studies into tables
- Previously, this was largely a manual task without required standardization
- This new tool uses spreadsheets formatted for different disciplines to help standardize data extraction
- Once the data is extracted, the spreadsheets are transformed into sets of tables useful in the document production process

Future Development

- Store data in HERO for repeated use in assessments
- Develop search and reporting capabilities for extracted data
- Through text and concept mining tools, automate the first pass at categorization and tagging
- Visualize the results of automated categorization in Evidence Maps
- Create tighter integration between HERO and HAWC
- Create, improve, and utilize web service APIs for HERO and HAWC to ease integration with third-party software
- Allow web-based data entry for Evidence Inventory
- Integrate Tableau visualization software with HAWC, Distiller, and Evidence Inventory tools
- Integrate Evidence Prime's Pupil automated data extraction software with Distiller, HAWC, and Evidence Inventory tools
- Investigate possible standards for extracted data formats to ease data migrations
- Investigate and implement tools for automated table and graph data extraction
- Work on ontologies for data extraction to make the data more easily searchable
- Employ agile development processes to test and incorporate new and useful tools into the assessment process

U.S. Environmental Protection Agency
Office of Research and Development