Key characteristics of male reproductive toxicants: an approach for screening and sorting mechanistic evidence.

Xabier Arzuaga[1], Martyn T. Smith[2], Catherine Gibbons[1], Niels E. Skakkebæk[3], Erin Yost[1], Brandy Beverly[4], Andrew Hotchkiss[1], Russ Hauser[5], Rodrigo L. Pagani[6], Steve Schrader, Lauren Zeise[7], Gail S. Prins[6,8]

[1]U.S. Environmental Protection Agency, [2]National Center for Environmental Assessment; University of California, Berkeley, School of Public Health; [3]Department of Growth & Reproduction, University of Copenhagen; [4]National Institute of Environmental Health Sciences, National Toxicology Program; [5]Harvard University, T.H. Chan School of Public Health; [6]University of Illinois at Chicago, Department of Urology; [7]California Environmental Protection Agency; Office of Environmental Health Hazard Assessment; [8]University of Illinois at Chicago, School of Public Health.

Xabier Arzuaga | Arzuaga.Xabier@epa.gov | 202-564-1361

Abstract

Since the introduction of ten key characteristics of carcinogens as a basis for organizing mechanistic data on carcinogenesis, the National Academy of Sciences has recommended that key characteristics approaches also be developed for noncancer hazards. The aim of this project was to identify a set of key characteristics that can be used for searching, screening, and sorting mechanistic evidence on chemical-induced toxicological responses in the male reproductive system. An expert workgroup was convened at the University of California-Berkeley in March 2018 to review the key characteristics approach and determine whether it can be applied to endocrine disruptors and male and female reproductive toxicants. For male reproductive toxicants, eight key characteristics were identified based on survey of established mechanisms, and include alterations in: 1) germ cell functions, 2) somatic cell functions, 3) reproductive hormone levels/production, 4) hormone receptors, 5) DNA damage, 6) epigenetic modifications, 7) oxidative stress, and 8) inflammation. As a proof of principle, this set of key characteristics was used to organize mechanistic evidence from in vivo and in vitro studies on the PCB mixture Aroclor 1254 and effects in the male reproductive system. The proposed key characteristics of male reproductive toxicants facilitates the systematic screening and categorization of mechanistic data from diverse research methods, models, and endpoints, as well as from a variety of known pathways for chemical-induced toxicity that can support hazard characterization. Disclaimer: *The views expressed are those of the authors and do not necessarily represent the views or policies of the US EPA.*

Background – Introduction

Hazard identification as part of a human health risk assessment consists of an analysis of the available evidence on chemical-induced adverse health effects that are focused on cancer and non-cancer outcomes such as male reproductive toxicity. Evaluation of epidemiological, toxicological, and mechanistic studies for direct evidence of effects after chemical exposures plays a critical role in the hazard identification process. Analysis of mechanistic events that are precursors to apical endpoints seen in animals and humans support evidence of a hazard, identify potential susceptible populations and lifestages, and inform the human relevance of effects observed in animals, and identify data gaps.

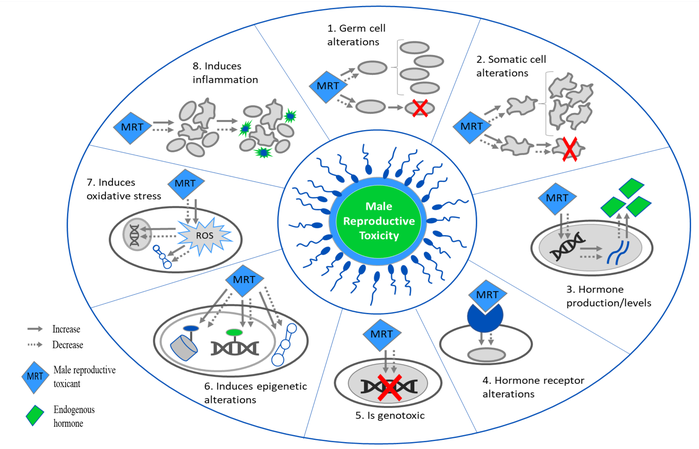
Identification of the eight key characteristics of male reproductive toxicants.

- The ten key characteristics (KCs) of human carcinogens were introduced as the basis of a uniform approach for identifying, organizing, and evaluating mechanistic evidence to support cancer hazard identification (Smith et al. 2016, Guyton et al. 2018).
- In its 2017 report on “Using 21st Century Science to Improve Risk-Related Evaluations”, the National Academy of Sciences, Engineering and Medicine (NASEM) opined that the KCs approach “avoids a narrow focus on specific pathways and hypotheses and provides for a broad, holistic consideration of the mechanistic evidence”
- The 2017 NASEM report also recommended that the KCs approach be expanded to other endpoints, including reproductive effects, endocrine disruption and cardiovascular disease.
- Here, we have attempted to develop a set of key characteristics of male reproductive toxicants based on our current knowledge of the mechanisms by which chemicals cause reproductive toxicity.
  - A workgroup was convened at the University of California-Berkeley on March 2018 to address this topic. The objective was to review the key characteristics approach and determine whether it can also be applied to male reproductive toxicants.
  - Additionally, the review literature was evaluated to identify environmental contaminants known to target the male reproductive system and established cellular/molecular pathways linked to adverse outcomes.

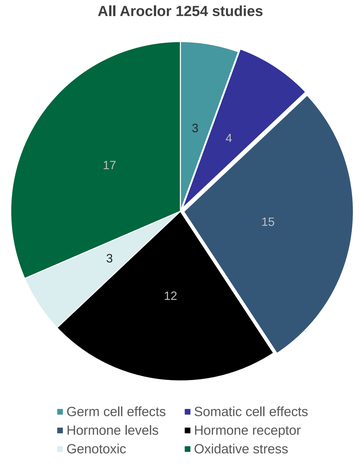
- Each key characteristic is described in the context of mechanisms/pathways by which exposure to male reproductive toxicants can lead to adverse health effects.
- Finally, there is a shift away from whole animal testing for apical endpoints towards high-throughput in vitro testing and toxicogenomic profiling due to cost, timelines, ethical considerations, regulatory constraints on animal testing, and the large volume of chemicals needing evaluation.

Key characteristics of male reproductive toxicants

Characteristic	Examples of relevant evidence
1. Alters germ cell development, functions, or death	Increases in germ cell apoptosis; alterations in sperm acrosome reaction and motility
2. Alters somatic cell development, functions, or death	Increased Sertoli cell apoptosis; alterations in Sertoli cell functions, cytoskeleton, and interactions with germ cells; alterations in Leydig cell development.
3. Alters production and levels of reproductive hormones	Decreased Leydig cell steroidogenic functions; increased hepatic metabolism and excretion of sex hormones.
4. Alters hormone receptor levels/functions	Androgen receptor antagonism, estrogen receptor activation, decreased LH-Receptor expression
5. Is genotoxic	DNA damage, chromosome fragmentation, altered sperm cell chromosome numbers.
6. Induces epigenetic alterations	Altered sperm microRNAs, germ cell DNA methylation patterns, and histone retention sites
7. Induces oxidative stress	Reduced tissue antioxidant levels.
8. Induces inflammation	Increased testicular expression of pro-inflammatory markers and prostaglandin levels
Abbreviations: LH, Luteinizing hormone	



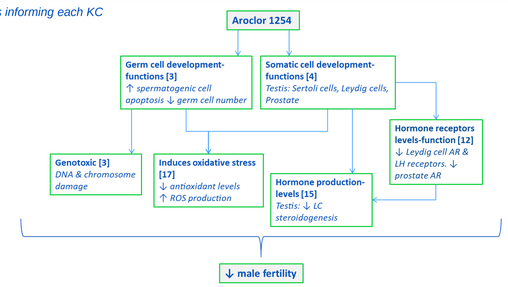
Preliminary evaluation of mechanistic effects caused by the PCB mixture Aroclor 1254



Aroclor 1254 Studies reporting on oxidative stress					
Study & HERO ID	**Age at exposure & exposure duration**	**Exposure route and duration**	**Doses - concentration**	**Species and train**	**Exposure outcomes**
Murugesan et al 2008; 614690	Not reported	Cell culture; 6 or 12 hrs	0, 10^{-10}, 10^{-9}, 10^{-8}, 10^{-7} M	Rat – Wistar	Isolated Leydig cells: ↑ lipid peroxidation, hydrogen peroxide, and hydroxy radical production; ↓ superoxide dismutase, catalase, GPX, GR, GST, and gamma-GT; ↓ cellular levels of Vit C and Vit E
Krishnamoorthy et al 2005; 626034	90 days old	Cell culture; 6, 12 or 24 hrs	0, 10^{-10}, 10^{-9}, 10^{-8}, 10^{-7} M		Isolated Sertoli cells: ↓ activity SOD, CAT, GPx, gamma-GT activities. ↑ lipid peroxidation.
Murugesan et al 2007; 1297571	90 days old	Cell culture; 24 hrs	0, 10^{-10}, 10^{-9}, 10^{-8}, 10^{-7} M		Isolated Leydig cells: ↑ lipid peroxidation, hydrogen peroxide, and hydroxy radical production; ↓ superoxide dismutase, catalase, GPX, GR, GST, and gamma-GT; ↓ cellular levels of Vit C and Vit E (with and without and LH).
Aly 2013; 2149875	Not reported	Cell culture; 3 hrs	0, 10^{-9}, 10^{-8}, 10^{-7} M		↑ reactive oxygen species production and lipid peroxidation in testis mitochondrial fraction. ↓ testicular mitochondria activities of SOD, CAT, GPx and GR, and levels of glutathione and Vit C.
Ateşşahin et al 2010; 1403617	8 weeks old	Injection; 8 weeks	0, 2 mg/kg-day	Rat – Spregue Dawley	↓ SOD activity. No effect on TBARs, GSH, GSH-Px, CAT activities, or testosterone levels.
Krishnamoorthy et al 2007; 626036	90 days old	30 days	0, 2 mg/kg-day	Rat – Wistar	Epididymal sperm: ↑ levels of reactive oxygen species (HO-, H2O2, & lipid peroxides). ↓ levels of alpha tocopherol, GSH, and ascorbic acid. ↓ activity of SOD, CAT, GPx, GR, and GST in epididymal sperm
Selvakumar et al 2011; 759831					In prostate: ↑ lipid peroxidation & levels of H2O2. Effects were ameliorated by co-treatment with Vitamin E.
Murugesan et al 2005; 1416186					↑ Leydig cell lipid peroxidation, hydrogen peroxide, and hydroxyradical levels. ↓ Leydig cell superoxide dismutase, catalase, glutathione-s-transferase, glutathione peroxidase, glutathione reductase, gamma-glutamyl transpeptidase.
Murugesan et al 2005; 1417568					↑ testicular lipid peroxidation, hydrogen peroxide production, hydroxyl radical production. ↓ testicular Vit E and Vit C levels. ↓ testicular superoxide dismutase, catalase, glutathione-s-transferase, glutathione peroxidase, and glutathione reductase.
Murugesan et al 2005; 1417737					Isolated Leydig cells: ↑ lipid peroxidation, hydrogen peroxide and hydroxy radical production; ↓ superoxide dismutase, catalase, GPX, GR, GST.
Venkataraman et al 2004; 1418753					↓ prostate superoxide dismutase, catalase, glutathione-s-transferase, glutathione peroxidase. ↑ prostate lipid peroxidation, hydrogen peroxide production. No effect on prostatic acid phosphatase.
Senthil kuma et al 2004; 1418887					↑ Sertoli cell lipid peroxidation, hydrogen peroxide, and hydroxyradical levels. ↓ Sertoli cell lactate levels. ↓ Sertoli cell superoxide dismutase, catalase, glutathione-s-transferase, glutathione peroxidase, glutathione reductase, gamma-glutamyl transpeptidase.
Sridhar et al 2004; 2162224					↓ prostate superoxide dismutase, catalase, glutathione-s-transferase, glutathione peroxidase, prostatic acid phosphatase. ↑ prostate lipid peroxidation, hydrogen peroxide production.
Raj et al 2014; 2149829	Not reported	Injection; 24 hrs			↑ epididymis H2O2 production, lipid peroxidation (caput, corpus, cauda).

Proposed overview for Aroclor 1254 – induction of six key characteristics of male reproductive toxicants

[ ] number of studies informing each KC



Concluding remarks

- Approaches which facilitate the systematic evaluation of toxicological and mechanistic evidence can improve the transparency and strength of evidence analyses performed as part of a risk assessment.
- The key characteristics of male reproductive toxicants can provide a structure for systematically searching and organizing the relevant literature on mechanistic information in support of an evaluation of a chemical for reproductive toxicity.
- In combination with lifestage and causality applied to human and animal evidence the key characteristics of male reproductive toxicants can be evaluated to identify pathways/networks that are conserved across species and therefore relevant to human health.

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

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable-based ink.

Abstract Sifter: A literature informatics tool for chemical safety assessments

Nancy C. Baker[1] and Thomas Knudsen[2]

[1] Leidos, RTP, NC; [2] NCCT, EPA, RTP, NC

www.epa.gov

Nancy C. Baker | baker.nancy@epa.gov | 0000-0002-8351-9435

Introduction

Identifying environmental chemicals that adversely affect human health or the environment requires assembling information from a wide variety of sources, including the millions of articles in the biomedical literature. Literature informatics approaches can help researchers make use of this literature in more effective ways. We have developed a freely available Excel-based tool called the Abstract Sifter to retrieve and triage articles and to visualize the literature landscape for a set of chemicals.

The tool is supplied with queries that facilitate exploration of mechanistic information by using the language of adverse outcome pathways (AOPs) and Key Characteristics of Carcinogens[2].

Defining mechanisms through PubMed queries

Toxicity type queries

Genetox	(dna/drug effects OR DNA Damage OR chromosome aberrations OR genotoxicity OR micronucleus OR DNA Repair OR mutagenicity te
Cancer	neoplasms or cancer
ReproTox	(reproduction AND (toxicity OR abnormal OR adverse effects))
NeuroTox	(neurotoxicity OR (Nervous system diseases and chemically induced) OR ((neurons OR brain) AND drug effects)
DevTox	(toxicity OR congenital abnormalities OR Prenatal Exposure Delayed Effects) AND (fetus OR embryo OR embryonic development)
Skin sensitization	("allergic" AND "contact" And dermatitis) OR Dermatitis, Allergic Contact[mh]

AOP queries

aop43	1mie	(Vascular Endothelial Growth Factor Receptor-2 OR Vascular Endothelial Growth Factc
aop43	2ke	((angiogenesis OR vasculogenesis) AND (reduce OR inhibit)) OR (Neovascularization, I
aop43	3ke	endothelium OR endothelial AND (damage OR impair OR impairment)
aop44	4ke	Blood circulation AND (drug effects AND toxicity)
aop43	5ao1	birth defects OR Congenital Abnormalities
aop43	6ao2	(bone OR bone development OR calcification OR osteogenesis OR limb bud/drug effe

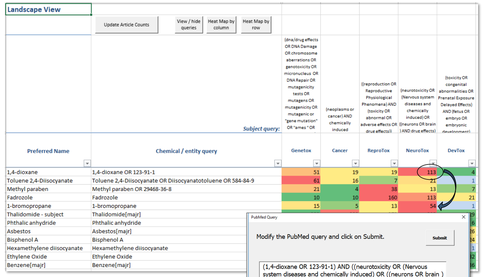
Aopwiki.org

Key Characteristics of Carcinogens queries

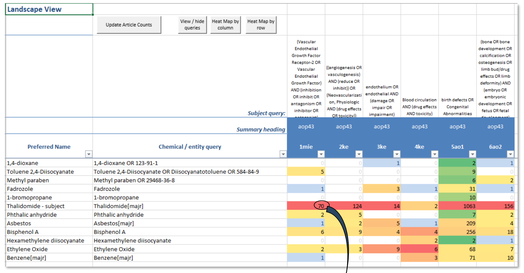
KCC1	Adducts	protein Adducts OR dna adducts OR epoxides OR quinones OR aldehydes
KCC2	DNA	(dna/drug effects OR DNA Damage OR chromosome aberrations OR genotoxicity OR micronucleus OR mutagenicity tests OR mutagens OR mut
KCC3	DNA break/rep	(topoisomerase OR DNA Repair/drug effects OR double-strand break repair)
KCC4	Epigenetics	(DNA methylation OR histone modification OR Histone Code OR Chromatin Assembly and Disassembly OR microRNAs OR Epigenesis, Genetic)
KCC5	Oxidative stres	(oxidative stress OR reactive oxygen species OR mitochondria/drug effects)
KCC6	Inflammation	(inflammation OR macrophages)
KCC7	Immunosuppre	(immunosuppressive OR immune system dysfunction)
KCC8	Receptors	receptors AND (antagonism OR agonism OR drug effects)
KCC9	Cell immortaliz	(cell immortalization OR cell division OR hayflick limit OR Aging/genetics OR Cellular Senescence)
KCC10	Proliferation	(cell proliferation OR cell death OR apoptosis OR Neovascularization, Pathologic)
Cancer	Cancer	(cancer OR neoplasms OR precancerous OR carcinogen OR carcinoma) AND (toxicity OR poisoning OR adverse effects)

Landscape view for three sample mechanistic reviews

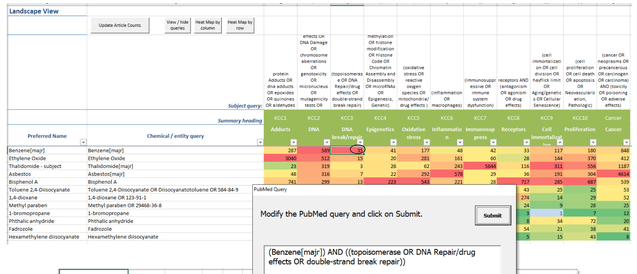
Toxicity Types



AOP 43: Vascular disruption



Key Characteristics of Carcinogens



Landscape Overview

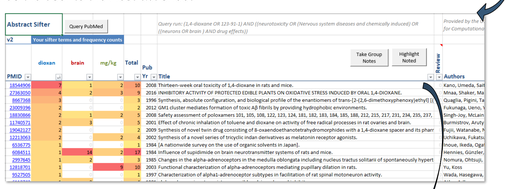
The Landscape sheet lets the scientist build an overview of a literature domain by entering entity and subject matter queries.

The numbers are article counts retrieved by the app… Double-clicking on the article count retrieves the citations and sends them to the Main sheet.

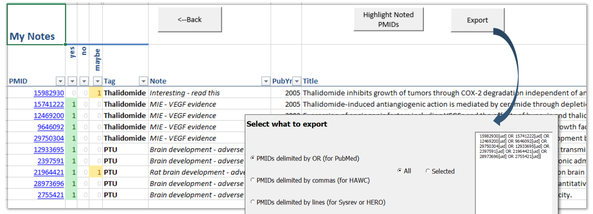
To sift, enter terms or characters of interest here and let the Abstract Sifter count occurrences of the terms in the title and abstract. Then sort. Double-clicking on a row sends the user to the Abstract sheet.

Retrieve and sift the citations

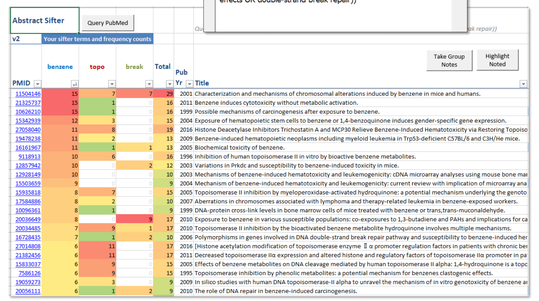
The Main sheet is where the user can browse and use the novel *sifting* technology to find relevant articles.



Citations on the Notes sheet can be exported to other software.



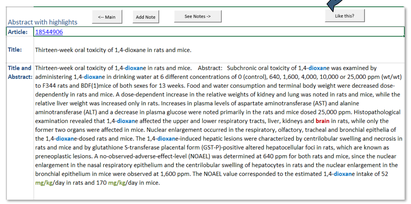
Notes can be taken and noted articles highlighted.



What other things can you do?

Look here to see other useful and fun capabilities.

Here's the Abstract sheet with the sifter terms colorized for easy reading.



Summary

The Abstract Sifter is a free, flexible literature retrieval tool that provides an overview of a complex literature landscape and the capability of focusing on mechanisms of toxicity. Available for download here: ftp://newftp.epa.gov/COMPTOX/Sustainable_Chemistry_Data/Chemistry_Dashboard/Abstract_Sifter/

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

This poster does not necessarily represent U.S. EPA policy.

References: 1) Baker N et al. F1000 Res. 2017 Dec 21;6. pii: Chem Inf Sci-2164.
2) Smith, et al. Environ Health Perspect. 2016 Jun;124(6):713-21.

Implementing machine learning methods in literature searching and screening to identify and categorize mechanistic evidence for the Integrated Science Assessments

Jennifer L. Nichols, Ryan Jones, Michael J. Stewart, and Steven J. Dutton

National Center for Environmental Assessment, US EPA, RTP, Durham, NC, 27711

Jennifer L. Nichols | nichols.jennifer@epa.gov | 919-541-0708

Background

The National Ambient Air Quality Standards (NAAQS) for criteria air pollutants are established by the EPA as mandated by the Clean Air Act to protect public health and welfare. A critical component in reviewing the NAAQS is the development of the Integrated Science Assessments that assess the state of the science informing the relationship between ambient exposures and a range of health and welfare effects. This requires that EPA scientists identify, evaluate, and integrate a broad range of scientific evidence, including mechanistic evidence that is an important aspect to understanding the nature of the relationship between exposure and effect. Because of the vast amount of evidence pertinent to ISAs from both observational and experimental studies, identifying and evaluating the relevant evidence in a comprehensive and efficient manner is a challenge. In the current review of the Ozone NAAQS, recently initiated under an accelerated timeline by directive from the EPA Administrator, methods and approaches for ISA development have been streamlined and modernized to further adopt systematic review methodologies, including for the identification and categorization of mechanistic evidence. As presented in this case study, a combination of machine learning and automated approaches in literature searching through the Health and Environmental Research Online database and the adoption of the SWIFT-ActiveScreener tool, have resulted in a substantially more efficient process to identify and categorize evidence. Ultimately, this streamlined workflow provides an easily adaptable process to effectively search and screen for mechanistic evidence that is a critical component to drawing conclusions in scientific assessments supporting key Agency policy decisions.

Literature Workflow for the Ozone ISA

Keyword search:
- Ozone or o3, 2011/01/01-2018/03/31
- PubMed, Web of Science, TOXLINE
- n = 30,797

References for Automated Topic Classification, n = 29,519
- Experimental seeds, n = 364
- Epidemiology seeds, n = 425

Excluded
- Not peer reviewed, n = 883
- Not English language, n = 329
- Retracted, n = 7

Automated Topic Classified References
- Experimental seeds, n = 3,298
- Epidemiology seeds, n = 1,724

Not Topic Classified, n = 24,497

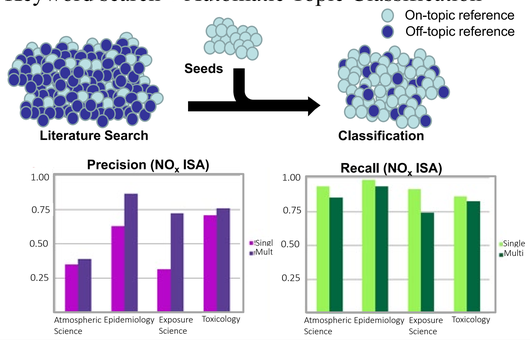
Title/abstract Include
- Experimental, n = 541
- Epidemiology, n = 820

Title/abstract Exclude
- Experimental, n = 701
- Epidemiology, n = 343
- Not screened, n = 2,620

Categorization of References
- Experimental: Health outcome, Exposure duration, Exposure concentration
- Epidemiology: Health outcome, Exposure duration, Location

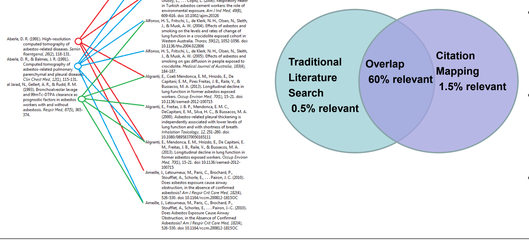
Novel Methods for Literature Identification: Topic Classification and Citation Mapping

Keyword search + Automatic Topic Classification



- Seed items from specific disciplines are used to train an algorithm
- This algorithm is applied to a set of search results and predicts the relevant discipline of each item
- Method is invaluable for large assessments where results must be distributed to specialized teams
- Results from analyses of precision and recall from application in the 2016 Oxides of Nitrogen ISA demonstrate the effectiveness of the automatic topic classification across scientific disciplines.

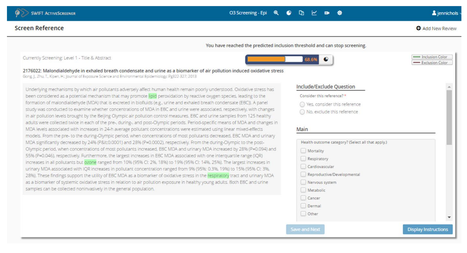
Citation Mapping

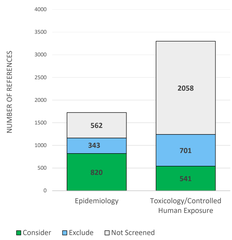


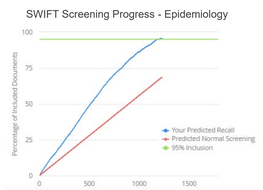
- Starts from a list of known relevant references
- Aggregates expert judgments of relevance, then ranks results by probability of relevance
- Not bound by metadata -- able to return relevant items that would have been missed by keyword searches
- Citations found in both the keyword search and citation mapping have a much higher probability of relevance.
- Citation mapping used to identify ecological references for Ozone ISA, but not health references.

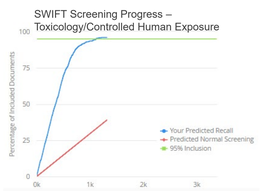
Implementation of SWIFT-ActiveScreener for Title/Abstract Screening

- SWIFT-ActiveScreener (Sciome): title/abstract screening with machine learning in real-time based on exclusion/inclusion decisions
- Allows optional screening questions
- Predicts 95% threshold, comparable to human judgment
- Reduced number of references needed to undergo title/abstract screening
- Added efficiency from SWIFT-AS depends on specificity of the references obtained from the literature search.









Categorization of Evidence and Application of Evidence Maps

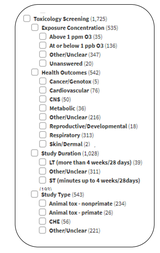
Experimental

Screening Questions

Title/Abstract Screening – Experimental (Animal Toxicology and Controlled Human Exposure

Seed references: n =
Screening questions:
1. Consider this reference (full-text screen)?
   - Yes, consider this reference
   - No, exclude this reference
2. Health outcome category?
   - Mortality
   - Respiratory
   - Cardiovascular
   - Reproductive/Developmental
   - Nervous System
   - Metabolic
   - Cancer
   - Other
   - Unclear
3. Exposure duration?
   - Short-term (up to 28 days)
   - Long-term (more than 28 days)
   - Unclear
4. Study type?
   - Controlled human exposure
   - Animal tox - primate
   - Animal tox – nonprimate
   - Unclear

Tags in HERO Database



Evidence Map

Exposure	Study Type	Health Outcome Category: CV	Resp	Repro/ Dev	Metabolic	Nervous System	Cancer	Dermal	Other	Total
Short-term	CHE	16	40	0	1	4	2	0	9	72
	Tox - Nonprimate	25	9	7	17	23	0	0	17	185
	Tox-Primate	5	8	4	0	0	1	1	0	19
	Unclear	1	8	0	0	0	0	0	0	9
Long-term	CHE	0	0	0	0	0	0	0	0	0
	Tox - Nonprimate	6	23	1	5	12	0	0	7	52
	Tox-Primate	1	5	2	0	0	0	0	0	8
	Unclear	2	0	0	0	1	0	0	0	3
Unclear	CHE	0	0	0	0	0	0	0	0	0
	Tox - Nonprimate	7	62	3	7	9	1	0	23	112
	Tox-Primate	2	15	5	0	0	0	0	2	24
	Unclear	14	69	1	4	7	1	0	44	120
	Total	79	304	23	34	56	5	1	102	604

Refine Scope for ISA

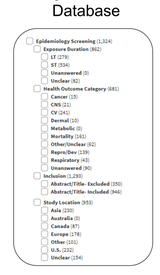
Epidemiology

Screening Questions

Title/Abstract Screening – Epidemiology

Seed references: n = 66
Screening questions:
1. Consider this reference (full-text screen)?
   - Yes, consider this reference
   - No, exclude this reference
2. Health outcome category?
   - Mortality
   - Respiratory
   - Cardiovascular
   - Reproductive/Developmental
   - Nervous System
   - Metabolic
   - Cancer
   - Dermal
   - Other
   - Unclear
3. Exposure duration?
   - Short-term (up to 28 days)
   - Long-term (more than 28 days)
   - Unclear
4. Study locations?
   - U.S.
   - Canada
   - Europe
   - Australia
   - Asia
   - Other
   - Unclear

Tags in HERO Database



Evidence Map

Exposure	Location	Health Outcome Category: Mortality	CV	Resp	Repro/ Dev	Metabolic	Nervous System	Cancer	Dermal	Other	Total
Short-term	US	36	36	38	11	2	0	0	5	1	133
	Canada	10	14	17	1	0	0	0	3	2	48
	Europe	33	17	37	6	2	5	0	1	5	126
	Asia	46	23	58	5	3	8	0	1	16	179
	Other/Unclear	21	8	15	5	1	1	0	0	2	140
Long-term	US	16	8	15	25	4	5	0	2	1	119
	Canada	5	1	4	3	1	0	1	0	1	16
	Europe	5	5	12	16	4	1	2	0	1	49
	Asia	10	17	19	25	4	4	4	1	0	90
	Other/Unclear	6	14	19	40	0	4	1	0	4	88
Unclear	US	9	1	3	1	1	1	1	1	1	19
	Canada	0	0	1	0	0	0	0	0	0	1
	Europe	0	3	4	2	1	0	0	0	2	12
	Asia	0	0	0	2	0	0	0	2	0	4
	Other/Unclear	0	1	2	2	0	4	0	1	1	11
	Total	201	141	324	153	27	33	10	52	37	1028

Refine Scope for ISA

Summary

- The evidence base for ISAs spans many disciplines and is vast. References identified from the literature search for recent ISAs range from approximately 50,000-500,000.
- Automatic topic classification and citation mapping are novel literature identification methods that are modules within the HERO Database. The approaches have successfully reduced the number of off-topic references that disciplinary teams of scientists need to screen for ISAs.
- Recently, SWIFT-AS was implemented in the Ozone ISA development process, adding another layer of machine learning capabilities and effectively reducing the number of references to be screened at the title/abstract level.
- SWIFT-AS optional screening questions were utilized to refine the categorization of references during title/abstract screening.
- For mechanistic evidence, experimental and epidemiologic studies were primarily binned by health outcome category and exposure duration to facilitate a more strategic and systematic consideration of the evidence by the assessment team.
- In future applications, this streamlined workflow could easily be adapted to allow for further refinement and categorization of mechanistic evidence based on assessment needs.

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

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable-based ink.

*The views presented in this poster are those of the authors and do not necessarily represent the views or policy of the U.S. EPA. Mention of trade names or commercial products does not constitute endorsement or recommendation for us.

HAWCPROJECT.ORG: A content management system for human health assessments

Andy Shapiro[1,*], Josh Addington[2], Vickie Walker[1], Kathryn Z. Guyton[3], Andrew Rooney[1], Ivan Rusyn[4]

[1] National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA
* Current affiliation: Infinia ML, Durham,. NC, USA
[2] Kelly Services, Research Triangle Park, NC, USA
[3] International Agency for Research on Cancer (IARC) Monographs Programme, World Health Organization, Lyon, France
[4] Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, USA

Abstract

Decision-makers and researchers frequently conduct literature-based assessments of the potential for chemicals or other exposures to pose a threat to human health. Such assessments typically consist of a critical review of a literature corpus to identify adverse health effects, to extract data for exposure-response relationship modeling, and/or to elucidate toxicity mechanisms. The systematic review methodology increases the transparency and objectivity in an evaluation by using a pre-defined, multistep process to identify, critically assess, and synthesize evidence. In addition to extraction of data, systematic review may also include an assessment of potential bias in a body of literature. A clear and detailed presentation of problem formulation, analysis and outputs, as well as properly documented search strategies and intermediate decisions, are critical to ensure transparency of the process.

We address these challenges by creating a modular, web-based content-management system to synthesize multiple data sources into overall human health assessments of chemicals. This free, open-source web-application, HAWC (Health Assessment Workspace Collaborative, https://hawcproject.org/), integrates and documents the overall workflow from literature search, literature screening, risk of bias assessment, data extraction, dose-response analysis using EPA benchmark dose modeling software (BMDS), and data synthesis by enabling creation of customizable visualizations of evidence and risk of bias.

Each HAWC assessment can be composed of some of all of these steps, based on the goals of the assessment, and at the discretion of assessment owners. User access is assessment-specific; project-managers can create public or private assessments, and can share with their team during development and ultimately release publicly as supplemental information to final reports (e.g., the US National Toxicology Program (NTP) monograph of immunotoxicity associated with PFOA/PFOS exposure, or the National Academy of Science's report on low-dose toxicity from endocrine active chemicals). All data and figures are exportable in user-friendly formats. To date, nearly 500 assessments have been created by users, and has been adopted for use by the NTP, the US EPA, TCEQ, and 34 assessments to date by the WHO IARC Monographs program.

Crucial benefits of such a system include improved integrity of the data and analysis results, greater transparency, standardization and consistency in data collection and presentation.

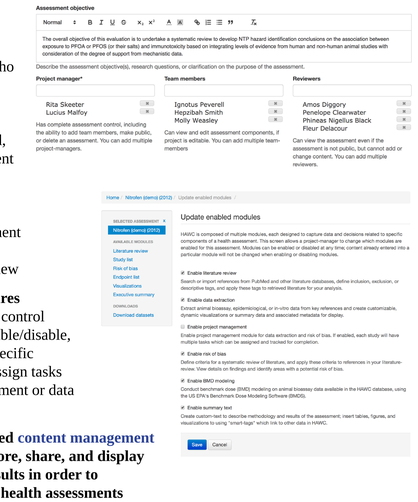
Assessment and permissions

Levels of access:

- **Project managers**: change permissions settings, including who can edit assessment content and which modules are enabled
- **Team members**: add, edit, and delete content
- **Reviewers**: view assessment before assessment is public
- **Public**: if an assessment is made public, the general-public can view

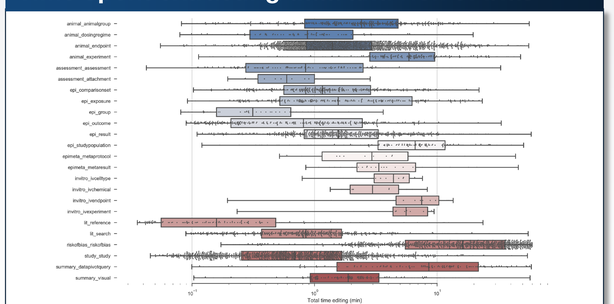
Other assessment features

- Project managers can control which modules to enable/disable, settings are project specific
- Team members can assign tasks for risk of bias assessment or data extraction tasks



HAWC is a web-based content management system to create, store, share, and display data and results in order to conduct human health assessments

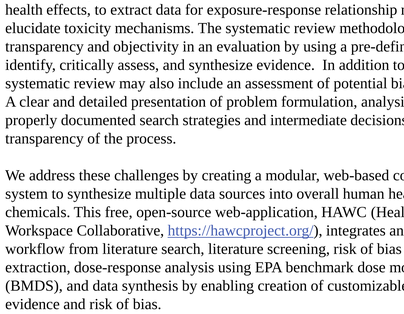
Time spent extracting data and risk of bias



Analytics tracking the amount of time spent editing content in HAWC has been collected since June 21, 2017. While the results are still early, over 8,700 items of content have been edited in HAWC to date. In the future, these data may be informative in estimating the total time/cost of data extraction in systematic reviews, and may help drive cost-benefit analyses for machine learning and semi-automated extraction methods.

Application Modules

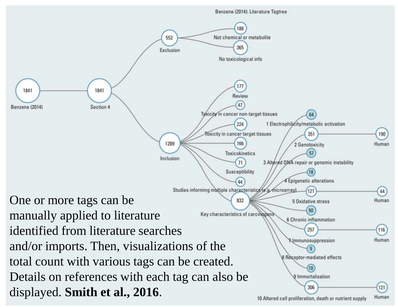
LITERATURE REVIEW



References can be searched in Pubmed, or imported using IDs from PubMed, HERO, or an entire reference database can be imported from Endnote or RefManager.

The example above shows a PubMed query from the IARC Monographs. The query was executed four times, with the final time finding five new references and removing one.

Literature tags are fully customizable for each assessment; they are also nested and counts roll-up in visualizations and other presentations. **Chappell et al., 2016**

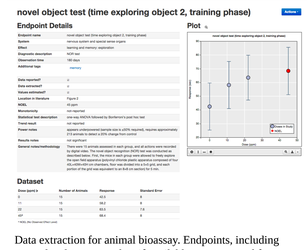


One or more tags can be manually applied to literature identified from literature searches and/or imports. Then, visualizations of the total count with various tags can be created. Details on references with each tag can also be displayed. **Smith et al., 2016**.

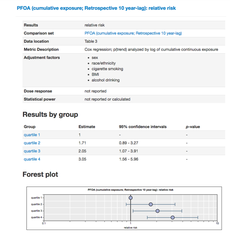
Data in public-HAWC (12/07/2017):

- 882 registered users
- 488 assessments (13 public)
- 216,708 references (52% have been tagged)
- 3,903 references with data extracted (69% have risk-of-bias extracted)
  - 14,316 animal bioassay endpoints (93% include dose-response data)
  - 5,918 epidemiological results (100% include group-level outcomes)
  - 1,658 *in vitro* endpoints (85% include group-level outcomes)
- 1,002 custom visualizations (16% of assessments have visualizations)

ENDPOINT EXTRACTION

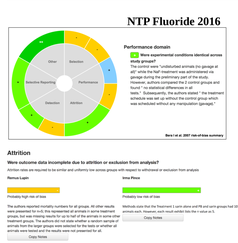


Data extraction for animal bioassay. Endpoints, including group-level summary data if available, are extracted for different animal groups. An animal group is defined as group of related animals for which dose-related response differences are compared. **NTP Fluoride 2016.**



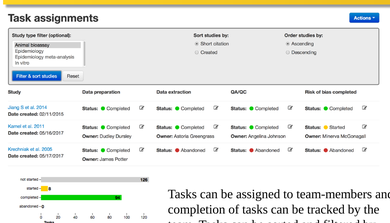
Data extraction for epidemiological output. Outcomes, including group-level summary data if available, are extracted for different groups (case-control or cohort). Adjustment factors can be tracked across outcomes. **NTP PFOS/PFOA 2016.**

RISK OF BIAS

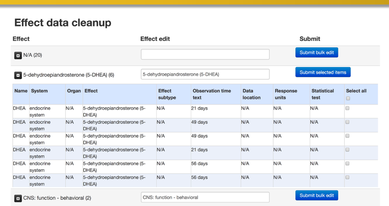


Risk of bias questions can be defined for each assessment and are answered for each study. Multiple reviewers may be required for a response with conflict resolution, or a single reviewer may be used for each study. Heatmaps of responses or strengths/weaknesses across the evidence can be created, to view areas of potential bias across a body of literature.

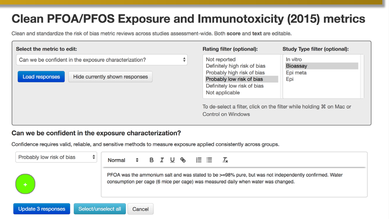
PROJECT MANAGEMENT TOOLS



Tasks can be assigned to team-members and completion of tasks can be tracked by the team. Tasks can be sorted and filtered by study type. Simple task-summary charts and graphs can be presented for project tracking and analytics for completion.

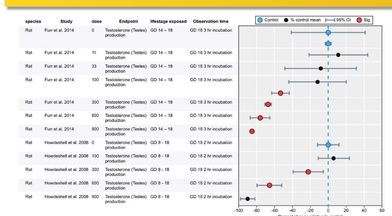


To standardize language across studies within an assessment, HAWC provides tools to "cleanup" extracted terms be more consistent. No formal ontology is used; this allows flexibility in the tool to use assessment-specific vocabulary.

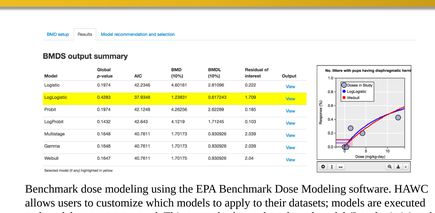


Similar to data cleanup, risk of bias responses can also be cleaned for an assessment. Users have the option to filter responses by metric and score; text can be applied to all responses for consistency across an assessment.

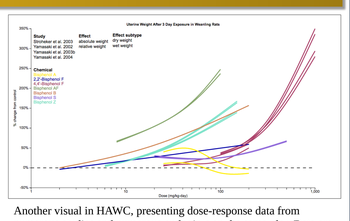
DATA ANALYSIS AND PRESENTATION

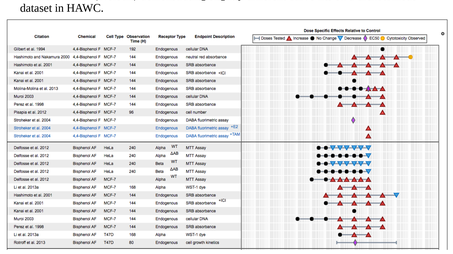


A data pivot created to compare similar endpoints from multiple studies. Data presented is customizable (using sort/filter rules), along with which columns of text to present, and rules for how to apply styles to points can be set manually or with conditional formatting. Here significant response are shown in red. **NAS 2017**.



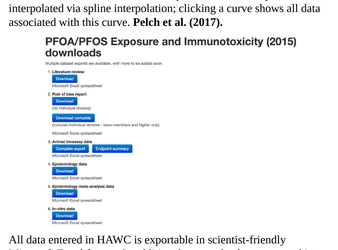
Benchmark dose modeling using the EPA Benchmark Dose Modeling software. HAWC allows users to customize which models to apply to their datasets; models are executed and model outputs are stored. This example shows the selected model (Log-logistic) and a different model (Weibull) demonstrating slightly different curve fits on a dichotomous dataset in HAWC.


Uterine Weight After 3 Day Exposure in Weanling Rats
Study
Effect
Effect subtype
Chemical
Another visual in HAWC, presenting dose-response data from numerous studies and numerous endpoints on the same plot. Data are interpolated via spline interpolation; clicking a curve shows all data associated with this curve. **Pelch et al. (2017).**



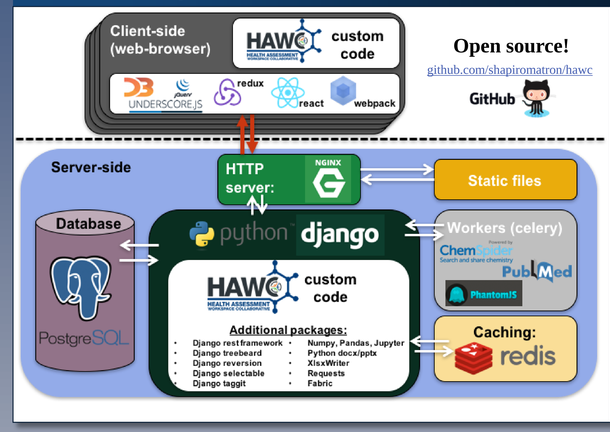
Data pivot barchart with error-bars. Rows are grouped to compare results of the same study. Colors are applied via conditional formatting to separate sexes Finally, a range of interested as added to demonstrate to the user what "normal range" might be for a particular endpoint (example).

A third data pivot showing one row for each *in vitro* endpoint. Instead of presenting mean response values, red triangles represent response significantly higher than control, and blue triangles represented response significantly lower than control. **Pelch et al. (2017).**



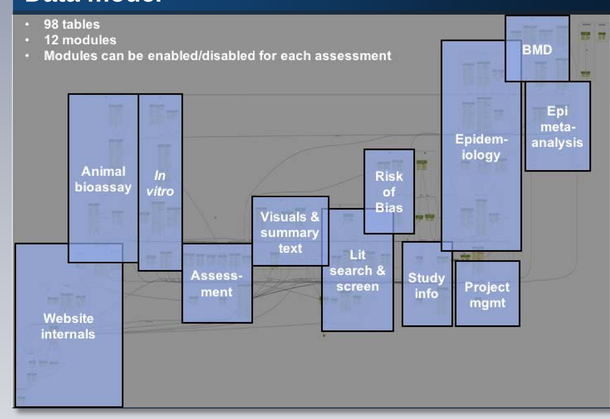
All data entered in HAWC is exportable in scientist-friendly Microsoft Excel format. In addition, because the data are stored in a relational database, numerous other formats are also available. This includes a fully RESTful application programing interface (API), used internally by the HAWC website. **NTP PFOA/PFOS 2016.**

Technical implementation



Data model

- 98 tables
- 12 modules
- Modules can be enabled/disabled for each assessment



Example published projects which used HAWC

Publications citing HAWC include peer-reviewed publications, or final reports from organizations which indicated use of HAWC during analysis or methods guidelines and may not be a complete list:

- Boyles AL, Blain RB, Rochester JR, Avanasi R, Goldhaber SB, McComb S, Holmgren SD, Masten SA, Thayer KA. Systematic review of community health impacts of mountaintop removal mining. Environ Int. 2017;107:163-172. 10.1016/j.envint.2017.07.002
- Chappell G, Pogribny IP, Guyton KZ, Rusyn I. Epigenetic alterations induced by genotoxic occupational and environmental human chemical carcinogens: A systematic literature review. Mutat Res Rev Mutat Res. 2016;768:27-45. 10.1016/j.mrrev.2016.03.004.
- Guha N, Guyton KZ, Loomis D, Barupal DK. Prioritizing Chemicals for Risk Assessment Using Chemoinformatics: Examples from the IARC Monographs on Pesticides. Environ Health Perspect. 2016; 124(12):1823-1829. 10.1289/EHP186.
- IARC Monographs. 2017. Instructions to Authors for the Preparation of Drafts for IARC Monographs. http://monographs.iarc.fr/ENG/Preamble/previous/Instructions_to_Authors.pdf
- Molander L, Hanberg A, Rudén C, Ågerstrand M, Beronius A. Combining web-based tools for transparent evaluation of data for risk assessment: developmental effects of bisphenol A on the mammary gland as a case study. J Appl Toxicol. 2017; 37(3):319-330. 10.1002/jat.3363.
- National Academies of Sciences, Engineering, and Medicine. 2017. Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals. Washington, DC: The National Academies Press. 10.17226/24758.
- National Toxicology Program. 2015. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf
- National Toxicology Program. 2015. Monograph on Identifying Research Needs for Assessing Safe Use of High Intakes of Folic Acid. 2015. Research Triangle Park, NC: National Toxicology Program. https://ntp.niehs.nih.gov/ntp/ohat/folicacid/final_monograph_508.pdf
- National Toxicology Program. 2016. NTP Research Report on Systematic Literature Review on the Effects of Fluoride on Learning and Memory in Animal Studies. Research Triangle Park, NC: National Toxicology Program. https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/01fluoride_508.pdf
- 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. http://ntp.niehs.nih.gov/ntp/ohat/pfoa_pfos/pfoa_pfosmonograph_508.pdf
- Pelch, KE, Wignall, JA, Goldstone, AE, Ross, PK, Blain, RB, Shapiro, AJ, Holmgren, SD, Hsieh, J-H, Svoboda, D, Auerbach, SS, Parham, FM, Masten, SA, Thayer, KA. 2017. NTP Research Report on Biological Activity of Bisphenol A (BPA) Structural Analogues and Functional Alternatives. NTP RR 4. Research Triangle Park, NC: National Toxicology Program. (4): 1-78. https://ntp.niehs.nih.gov/ntp/results/pubs/rr/reports/rr04_508.pdf
- Smith MT, Guyton KZ, Gibbons CF, Fritz JM, Portier CJ, Rusyn I, DeMarini DM, Caldwell JC, Kavlock RJ, Lambert PF, Hecht SS, Bucher JR, Stewart BW, Baan RA, Cogliano VJ, Straif K. Key Characteristics of Carcinogens as a Basis for Organizing Data on Mechanisms of Carcinogenesis. Environ Health Perspect. 2016; 124(6):713-21. 10.1289/ehp.1509912.
- Texas Commission on Environmental Quality. 2016. Ethylene Glycol Development support document. https://www.tceq.texas.gov/assets/public/implementation/tox/dsd/final/feb16/EG.pdf
- Texas Commission on Environmental Quality. 2017. White Paper: TCEQ Guidelines for Systematic Review and Evidence Integration. https://www.tceq.texas.gov/assets/public/implementation/tox/dsd/whitepaper/Proposed%20_Systematic_Review.pdf

Contact information:

Andy Shapiro (shapiromatron@gmail.com)
Ivan Rusyn (irusyn@tamu.edu)

CRAB: Automatic Text Mining of PubMed for Cancer Mechanism/ Mode of Action (MoA)

Amy Wang[1], Ulla Stenius[2], Johan Högberg[2], Imran Ali[2], Simon Baker[3], Ruth Lunn[1], Anna Korhonen[3]

[1]Office of Report on Carcinogens, National Toxicology Program, National Institute of Environmental Health Sciences (NIEHS), [2]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, [3]Language Technology Lab, University of Cambridge, Cambridge, United Kingdom

Abstract

Mechanistic information is often the diverse, challenging, and having-less-developed-approach-and-tool part of systematic review and cancer hazard identification. The CRAB project developed a free public tool http://crab3.lionproject.net enabling users to enter a search term (e.g., a(n) substance, occupation, cancer) and immediately receive PubMed abstracts that are tagged according to scientific evidence, mode of action (MoA) of cancer, and toxicokinetics.

In scientific evidence, study design is tagged for subject (human, animal, cell, subcellular, microorganism), study length, and outcome types (biomarker, tumors, morphological effects, biochemical/cell biological effects etc.)

In MoA, studies are tagged as genotoxic (include event types) or nongenotoxic with events of co-initiation (e.g., DNA repair inhibition), promotion (e.g., specific receptor/pathway activation), progression (e.g., immunosuppression), and multiphase (e.g., transcriptional modification, inflammation). Except electrophilicity and altered nutrient supply, all 10 characteristics of carcinogens are covered.

Additionally, each sentence of the abstract is color-coded as describing the background, objective, method, result, conclusion, related work, or future work of the study.

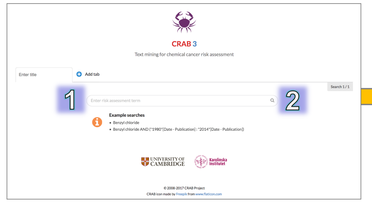
For example, searching "benzo(a)pyrene" lead to nearly 12000 abstracts, among which over 4000 are on genotoxicity (with adducts, strand breaks and mutations being most common, each with over 1300 abstracts), 475 on epigenetics, 0 on angiogenesis, and 50 on toxicokinetic modeling. Users can also search by PubMed IDs to tag specific abstracts.

This tool provides a great coverage of mechanistic information landscape and detailed and thoughtful tagging structure is effective and rapid. In the future CRAB will interface with other databases and software, providing even greater support to existing working practices in systematic review and cancer hazard identification.

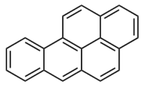
Intended use

As a rapid first step to survey literature landscape of mechanistic information related to carcinogenesis or hazard identification, and to screen the abstracts.

How to use CRAB http://crab3.lionproject.net



1 Type in the search box



1. Chemical name(s), disease, etc.
   - CRAB recognizes benzo(a)pyrene = benzo[a]pyrene

benzo(a)pyrene OR 3,4-benzpyrene OR 3,4-benzopyrene OR 3,4-benz[a]pyrene OR 3,4-benzo[a]

12,066 abstract(s) found

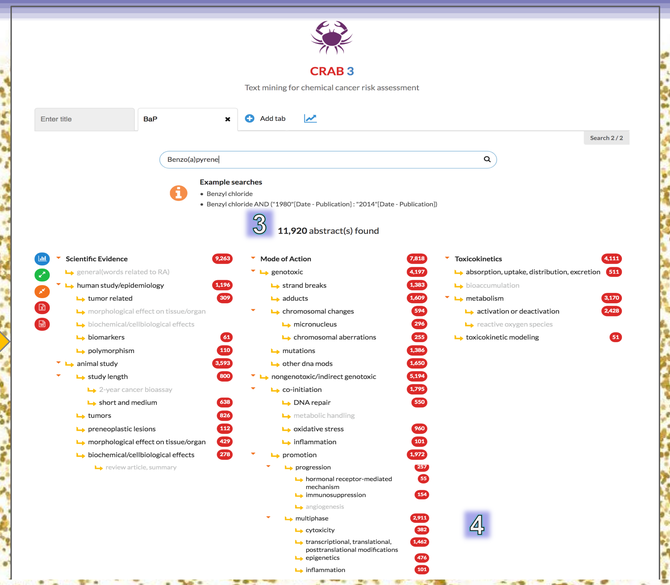
2. PMIDs
   e.g., results from your own search strings
   e.g., manually curated references

3. Interests connected by Bullen logic
   e.g., chemical name AND species AND organ
   e.g., disease AND (juvenile OR infant)

2 Click

3 Results on the landing page

Numbers of references in each "group" (similar to tag)
One abstract might be in multiple "groups" (i.e., multiple tags)
Groups with no abstracts are shown in grey (e.g., 2-year cancer bioassay)



4 Click on the number in a red oval

5 Citation information and abstracts in the selected group are shown

In abstract, each sentence is color-coded for the type of information

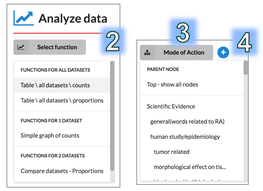
Background | Objective | Method | Result | Conclusion | Related work | Future work

Toxicological effects of benzo[a]pyrene on DNA methylation of whole genome in ICR mice. *2015*

It has been well known that alterations in DNA methylation - an important regulator of gene transcription - lead to cancer. Therefore a change in the level of DNA methylation of whole genome has been considered as a biomarker of carcinogenesis. Previously, a large number of experimental results in genetic toxicology have showed that benzo[a]pyrene could cause DNA mutation and fragmentation. However, there was little to no studies on alterations in DNA methylation of genome directly result from exposure to benzo[a] pyrene. In this paper, possible mechanisms of alterations in whole genomic DNA methylation by benzo[a]pyrene were investigated using ICR mice after benzo[a] pyrene exposure. The blood, liver, pancreas, skin, lung and bladder of ICR mice were removed and checked after a fixed time interval (6 hours) of benzo[a]pyrene exposure, and whole genomic DNA methylation level was determined by high performance liquid chromatography (HPLC). The results exhibited tissue specificity, that is, the level of whole genomic DNA methylation decreases significantly in blood an d liver, rather than pancreas, lung, skin and bladder of ICR mice. This study investigated the direct relationship between aberrant DNA methylation level and benzo[a]pyrene exposure, which might be helpful to clarify the toxicological mechanism of benzo[a] pyrene in epigenetic perspectives.

How to analyze data (compare results) in CRAB

1 Click



2 Click on Select function

Options vary by the number of dataset

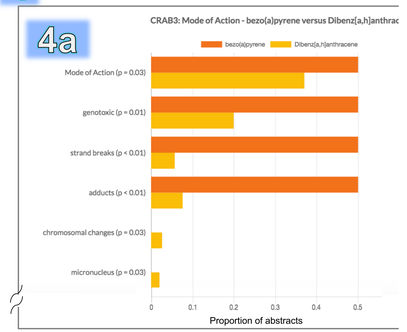
Select the analysis function

A new box shows up on the right

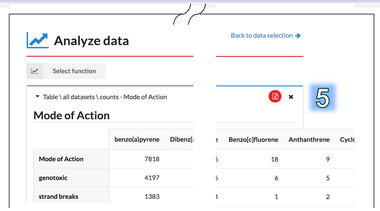
3 Click text next to Select nodes (similar to hieratical tags) to be compared

If the Function chosen is not for all dataset, select dataset(s)

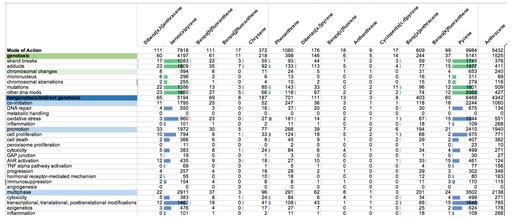
4 Click to see analysis results



4b



5 Click to export analysis results into Excel (color added in Excel)



Strengths

- Great coverage of key characteristics of carcinogens as carcinogenesis mechanism/mode of action
- Easy to use, fast, free and mobile responsive for accessing results on the move
- Practical and detailed grouping (e.g., types of DNA damage and chromosomal changes)

Limitations

- PubMed journal references only
- Search results are from co-occurrence of words, and expert judgment is still needed to identify association (e.g., mutations contributing to carcinogenesis vs. to other function)
- Only the number of references, not the tags associated with each reference, can be exported (exception: the output of a single reference is equivalent to tags of the reference)

Future directions

Develop interfaces with other database and software

Selected references

- Anna Korhonen, Diarmuid O'Séaghdha, Ilona Silins, Lin Sun, Johan Högberg and Ulla Stenius. 2012. Text mining for literature review and knowledge discovery in cancer risk assessment and research. PLoS One. 2012;7(4):e33427.
- Ilona Silins, Anna Korhonen, Ulla Stenius. 2014. Evaluation of carcinogenic modes of action for pesticides in fruit on the Swedish market using a text-mining tool. Front Pharmacol. 2014 Jun 23;5:145. doi: 10.3389/fphar.
- EPA 2010 (External Review draft) Development of a Relative Potency Factor (RPF) Approach for Polycyclic Aromatic Hydrocarbon (PAH) Mixtures. EPA/635/R-08/012A www.epa.gov/iris

A Novel Approach to Evaluate Mechanistic Data for Developmental Neurotoxicity

Mamta Behl[a], Kristen Ryan[a], Jui-Hua Hsieh[b], Frederick Parham[a], Andrew J. Shapiro[c], Bradley J. Collins[a], Nisha S. Sipes[a], Linda S. Birnbaum[a], John R. Bucher[a], Paul M.D. Foster[a], Nigel J. Walker[a], Richard S. Paules[a]

[a] NIH/NIEHS/NTP, RTP, NC, USA | [b] Kelly Government Solutions, RTP, NC, USA | [c] Consultant, Durham NC

Contact: mamta.behl@nih.gov

ABSTRACT

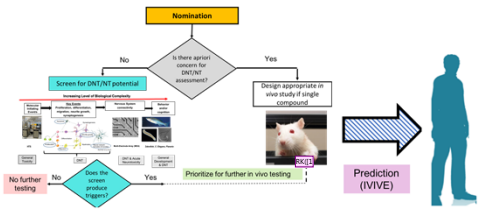
The National Toxicology Program (NTP) receives requests to evaluate chemicals for their potential to cause adverse health effects, including developmental neurotoxicity (DNT). Recent requests have included classes of chemicals such as flame retardants, polycyclic aromatic compounds, perfluoroalkyl substances, and bisphenol A analogs, all with approximately 20 – 50 compounds per class, many of which are commercial mixtures. However, all of the compounds within a class cannot be tested using traditional DNT animal testing guideline studies due to resource limitations, and the desire to be timely. Hence, a biologically relevant screening approach is required to rapidly prioritize compounds for further *in vivo* testing. Since neurodevelopment is a complex process involving multiple distinct cellular processes, it is unlikely that any one assay will be able to address the complexity. Hence, the NTP sought to characterize a battery of *in vitro* and alternative animal assays to quantify chemical effects on a variety of neurodevelopmental processes though a collaborative project. The NTP analyzed data from fourteen assays covering *in vitro* and alternate animal models that capture various aspects of neurodevelopment; results were compared using benchmark concentration (BMC) modeling to determine which assays may provide robust information to prioritize compounds for further studies in mammals. This poster highlights: 1) the overview and goals of the project, 2) strategies involved in compound selection, 3) development of NTP's approach to evaluate utility of these assays for further prioritization, 4) an examination of the validity of mechanistic data generated by these models, and 5) NTP's interactive web-application known as **D**evelopmental **N**euro**T**oxicity **D**ata **I**ntegration *V*isualization **E**nabling **R**esource **(DNT DIVER)**.

BACKGROUND

Introduction

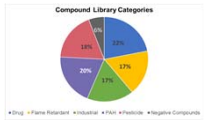
- Recent increases in the rise of neurodevelopmental disorders such as attention deficit hyperactivity disorder (ADHD), dyslexia, autism spectrum disorder and others have prompted further evaluation of the potential contribution of environment toxicants.
- The NTP receives a relatively large set of nominations including those on classes of compounds, chemical spills, and urgent responses with unknown toxicity (e.g. flame retardants, polycyclic aromatic hydrocarbons, Bisphenol A analogs).
- These types of nominations may not necessarily trigger a guideline DNT rodent study and hence may never be identified as DNTs.
- Furthermore, all compounds within the classes cannot be evaluated with stand-alone DNT Guideline studies due to time and resource constraints.
- To develop a strategy to rapidly screen for classes of compounds with potential for DNT, NTP is developing complementary screening approaches using human cell-based, or *in vitro,* assays, as well as alternate animal models such as zebrafish and planaria in combination with the use of *in vivo* testing in rodents.

Decision Tree for DNT Testing at the NTP



Creation of NTP Chemical Library

The NTP created a 91- compound library for distribution to collaborators interested in evaluating a well-characterized set of chemicals in their assay systems. The library consisted of known developmental neurotoxicants (e.g. drugs, pesticides), NTP classes of compounds with unknown neurotoxic potential (flame retardants, Bisphenol A analogs), and negative controls (e.g. saccharin, ascorbic acid). Compounds were added in duplicate to evaluate technical reproducibility.

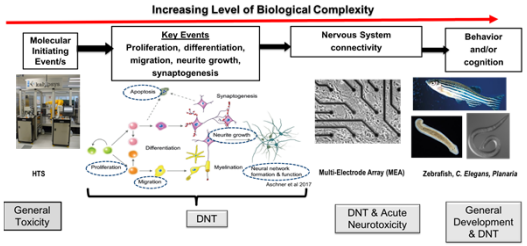


Development of Interactive Web-Application

NTP designed an interactive web-application to analyze, compare, and visualize DNT assays across different models using NTP's Chemical Library. This web-application is called **D**evelopmental **N**euro**T**oxicity **D**ata **I**ntegration and **V**isualization **E**nabling **R**esource **(DNT-DIVER).**

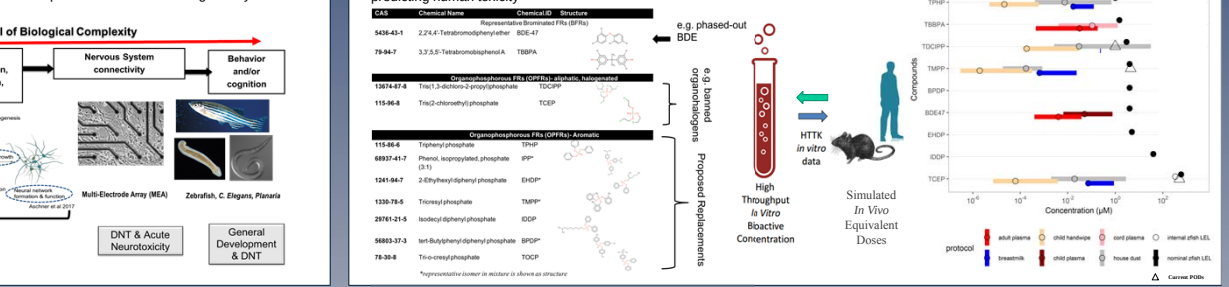
IDENTIFICATION OF ASSAYS FOR DNT BATTERY

- Goal: Evaluate emerging medium- to high- throughput and/or high-content cell-based *in vitro* and alternate animal (e.g., zebrafish, planaria) model systems that represent key neurodevelopmental processes in the development of a DNT screening battery



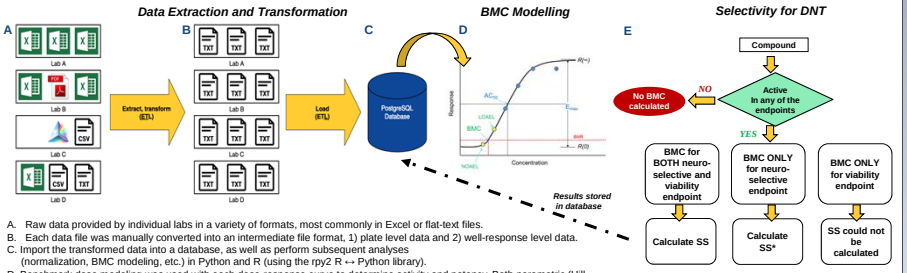
A DEEPER DIVE INTO PRIORITIZING AND PREDICTING POTENTIAL HUMAN TOXICITY

- Goal: Present proof-of-principle using flame retardants as a case example in utilizing *in vitro* and alternate animal models to prioritize compound classes for further testing and predicting human toxicity



DATA FORMATTING AND ANALYSIS

- Goal: Generate a standard database schema for all datasets to allow for comparative benchmark dose modelling (BMD) analyses across datasets. Determine if the compound causes general toxicity (e.g. mortality or loss of viability) or selective neurotoxicity (e.g., a decrease in neurite outgrowth).



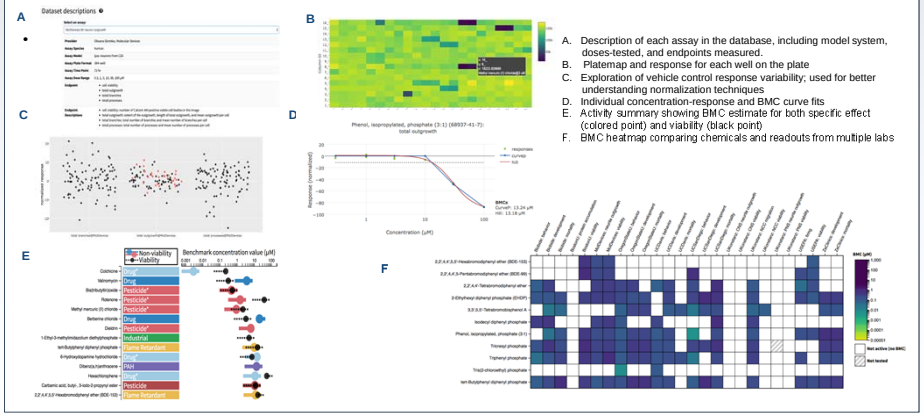
A. Raw data provided by individual labs in a variety of formats, most commonly in Excel or flat-text files.
B. Each data file was manually converted into an intermediate file format, 1) plate level data and 2) well-response level data.
C. Import the transformed data into a database, as well as perform subsequent analyses (normalization, BMC modeling, etc.) in Python and R (using the rpy2 R ↔ Python library).
D. Benchmark dose modeling was used with each dose-response curve to determine activity and potency. Both parametric (Hill model using EPA's TCPL software package) and non-parametric (CurveP) BMD (BMC) approaches were used.
E. SS = selectivity score = log10 (viability BMC/neuro-specific BMC)

* BMC for viability endpoint set at max tested concentration

PRIORITIZING AND VISUALIZING COMPOUNDS USING DNT-DIVER

- Goal: Generate a free, interactive web-based application to compare and prioritize compounds (e.g. flame retardants) known as Developmental NeuroToxicity Integration and Visualization Enabling Resource (DNT-DIVER). This web-application is now publicly available and can be found on the following URL

https://sandbox.ntp.niehs.nih.gov/neurotox/



A. Description of each assay in the database, including model system, doses-tested, and endpoints measured.
B. Platemap and response for each well on the plate
C. Exploration of vehicle control response variability; used for better understanding normalization techniques
D. Individual concentration-response and BMC curve fits
E. Activity summary showing BMC estimate for both specific effect (colored point) and viability (black point)
F. BMC heatmap comparing chemicals and readouts from multiple labs

GLOBAL IMPACT & SUCCESS

NTP's Vision for Global Impact on DNT

- Global public health impact by identifying environmental chemicals that have the greatest potential to affect susceptible populations (pregnant women, infants, children, elderly) and thereby preventing neurodevelopmental diseases (e.g. autism spectrum disorders)

Initial Stepping Stones

- NTP and collaborators are developing rapid screening approaches that use human-based *in vitro* and alternative animal models (e.g. zebrafish, planaria) to prioritize classes of compounds for further *in vivo* testing, and to provide timely information on their DNT potential to risk assessors.
- Since a single assay will not address the complexity of the nervous system, the NTP collated an initial set of assays to form a "battery" as proof-of-principle.
- BMC approach provided a unified data analysis for comparing across data sets (cell-based models and alternative animal models)
- This is the first time that a unified data processing & analysis approach (pipeline) has been applied across a disparate set of assays to screen for DNT.
- A collection of related publications will be made available in a special issue of Toxicological Sciences in January 2019.

DNT-DIVER: A Global Resource

- NTP created a publicly available free interactive web-application known as **D**evelopmental **N**euro**T**oxicity **D**ata **I**ntegration **V**isualization **E**nabling **R**esource **(DNT DIVER)** to visualize comparisons across assays.
- Within this application, comparisons are organized on different tabs including experimental design summary, quality control, chemical-specific concentration response curves, ranking of chemical toxicity by lab/assay, and comparison of results across assays
- Also available on this web-application are downloadable resources including raw and final data, and the NTP Chemical Library that was tested by collaborators in their respective models .

ONGOING DISCUSSIONS & FUTURE DIRECTIONS

- **Cell-based models:** Expansion of biological coverage, inclusion of complex 3-D models with lower throughput, experimental design consensus
- **Alternative Models:** Lack of consensus on experimental design, assessment of internal concentration, metabolites, role of log P
- **Data Processing:** No standards currently in the field, need to understand distribution of data to identify best analytic technique.
- **Chemical Library:** Expansion of chemical library to include more diversity
- **Website maintenance:** Allowance for continual upload of new data (dynamic)
- **Regulatory Perspectives:** Need for a global harmonization effort to ensure that standardized terminology is used across researchers to define and describe findings; linkage to *in vivo* data

ACKNOWLEDGEMENTS

The NTP would like to thank all its collaborators spanning across academia, industry, government and the regulatory community globally for their candid discussions, and their willingness to graciously share data to further the field. This work would not have been possible without their continual support.

The SciRAP Tool for Evaluating *In Vitro* Studies for Use in Hazard and Risk Assessment of Chemicals

Anna Beronius and Johanna Zilliacus, Institute of Environmental Medicine, Karolinska Institutet, Sweden

The Science in Risk Assessment and Policy (SciRAP) initiative

SciRAP was initiated by researchers at Karolinska Institutet and Stockholm University, Sweden.

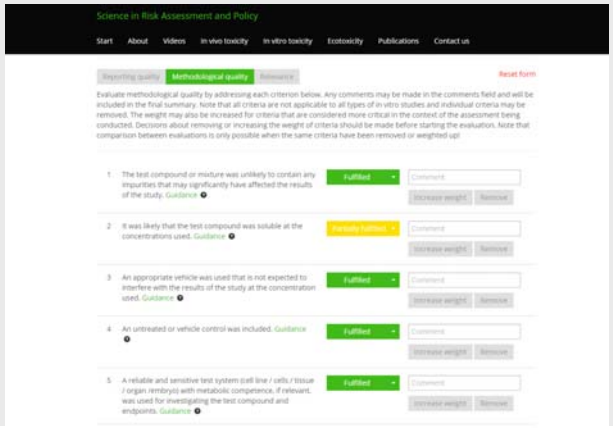
The objective is to provide a systematic and transparent approach for evaluating individual (eco)toxicity studies for chemical hazard and risk assessment.

The SciRAP web-based platform, www.scirap.org

Open access and free of charge:

- Criteria for evaluating reliability and relevance of individual studies.
- An on-line tool for application of the SciRAP criteria.
- Checklists for the reporting of studies.
- Information, guidance and tutorials.

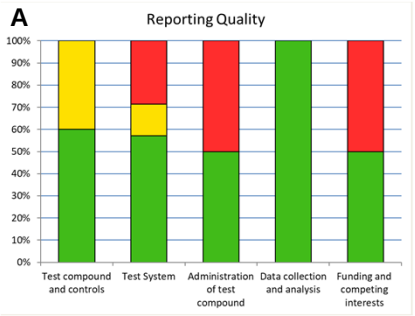
Criteria and tool on the web-based platform

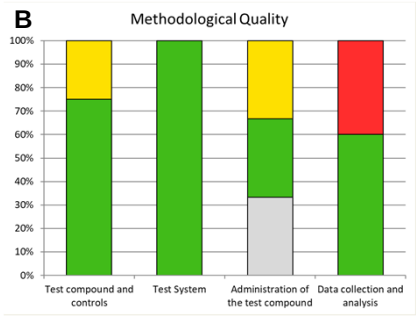


SciRAP evaluation of *in vitro* studies

A tool for evaluating reliability and relevance of *in vitro* studies was developed and added to the SciRAP platform. Evaluation of reliability is divided into evaluation of 23 reporting quality criteria and 15 methodological quality criteria.

The SciRAP criteria were based on requirements and recommendations in OECD test guidelines and guidance documents, such as the OECD Guidance Document on Good In Vitro Method Practices (GIVIMP)





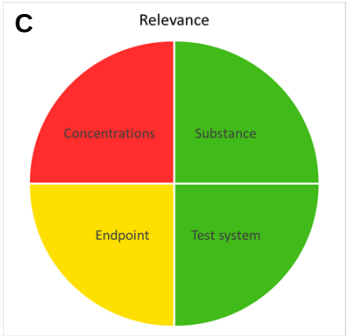


Figure 1. The SciRAP colour profile.

The output of a study evaluation in SciRAP generates a summary, including diagrams showing the distribution of fulfilled (green), partially fulfilled (yellow) and not fulfilled (red) reporting (**A**) and methodological (**B**) quality criteria in different sections. Grey areas indicate criteria that were left as undetermined, e.g. due to poor reporting. The SciRAP tool also generates a numerical score for reporting and methodological quality.

Four relevance items (**C**) are evaluated as directly relevant (green), indirectly relevant (yellow) and not relevant (red).

Summarizing and comparing studies in the same line of evidence

	Test compound and controls				Test System		Adm. of the test compound			Data collection and analysis				
SciRAP criterion:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Study 1														
Study 2														
Study 3														
Study 4														
Study 5														
Study 6														
Study 7														
Study 8														
Study 9														

Figure 2. Example of how evaluation of methodological quality can be summarized and compared for several studies in the same line of evidence.

Application of SciRAP and ongoing studies

- Reliability evaluation of studies for risk assessment under REACH.
- Assessment of Bisphenol AF according to the new EU criteria for endocrine disruptors.
- Assessment of triphenyl phosphate using the SYRINA framework.
- Application of SciRAP in a risk of bias evaluation scheme (together with the US EPA).
- AOP development.

Expert assessment of SciRAP *in vitro*

The SciRAP criteria and tool are currently being assessed for completeness and practical use by experts in the field of *in vitro* testing and health risk assessment.

Please join us!

Anna Beronius, PhD, ERT
Assistant Professor
Institute of Environmental Medicine
Karolinska Institutet
Stockholm, Sweden
E-mail: anna.beronius@ki.se

Karolinska Institutet

OECD Good *In Vitro* Method Practices (GIVIMP) Guidance and Its Implementation

Sandra Coecke[1], Gerard Bowe[1] and Patience Browne[2]

[1] European Commission, Joint Research Centre (JRC), Ispra (VA), Italy; [2] Organisation for Economic Cooperation and Development (OECD), Paris, France

The OECD has published the international guidance document on Good *In Vitro* Method Practices (GIVIMP) to support method developers and end-users working in academic, industry and government laboratories across all 36 OECD member countries and beyond in harmonisation efforts of *in vitro* methods. The development of GIVIMP was coordinated by the JRC's EU Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) within the context of a project of the OECD Test Guidelines Programme. GIVIMP consists of 10 chapters each one addressing key elements of good *in vitro* method practices. Applying GIVIMP during the *in vitro* method development stage will help improve the quality and reliability of generated data needed to support safety decisions.

EURL ECVAM is coordinating a large scale validation study of a set of mechanistically informative alternative methods to detect chemical that disrupt normal thyroid hormone function, in collaboration with the European Union Network of Laboratories for the Validation of Alternative Methods (EU-NETVAL) and the method developers. A total of 17 methods have been identified as candidates by EURL ECVAM taking primarily into account the information reported in an OECD scoping document on *in vitro* and *ex vivo* methods for the identification of modulators of thyroid hormone signalling (OECD No. 207), but also an OECD Detailed Review Paper (OECD No. 178), and feedback received at various Expert Group meetings. Effort has been invested, to assure the quality and long-term availability of the test systems (e.g. cells and tissues), the selection of relevant reference and control items, the completeness of the 17 methods, and to ensure the validation study is performed in compliance with GIVIMP.

1. Roles and responsibilities

- Describes roles and responsibilities of key actors in the *in vitro* method life cycle
- Targets method developers, test system providers, validation bodies, inter-governmental organisations, suppliers, users, and sponsors
- Provides guidance on documentation requirements (e.g. origins of cells and tissues, identity of the test system)

2. Quality considerations

- Discusses quality assurance versus quality control
- Examines quality risk-based assessment and quality control requirements for development and implementation of *in vitro* methods
- Provides quality considerations regarding the integrity of the data

3. Facilities

- Recommends fit for purpose facilities and a detailed understanding of the work flow
- Indicates that facility design and safety risk assessment and management should be adequate
- Prescribes strategies to avoid cross-contamination

4. Apparatus, materials and reagents

- Highlights the importance of regular maintenance, calibration, and validation
- Instructs on sourcing of materials and reagents (e.g. from well-established suppliers) to ensure the integrity and reliability
- Discusses the use of media in cell culture, including alternatives to animal-sourced serum

5. Test systems

- Elaborates the importance of Good Cell Culture Practice (GCCP)
- Advises the setting of acceptance criteria already at the development stage
- Describes identification and characterisation, sourcing, cell-banking and cryopreservation
- Suggests good practices for contaminants screening, including sterility, mycoplasma, virus testing

6. Test and reference/control items

- Illustrates the preparation and characterisation of test, reference and control items
- Defines the applicability and limitations of the method
- Recommends to identify potential sources of interference with the test system and/or method endpoint

7. Standard operating procedures

- Provides guidance on the development and preparation of *in vitro* method standard operating procedures (SOPs) to ensure consistency and reproducibility of data acquired
- Describes the evolution of a SOP from initial method description to method optimisation and validation

8. Performance of the method

- Analyses development of acceptance criteria for components (e.g.positive and negative controls)
- Reviews the elements of experimental design such as plate layout, the number of replicates, outlier detectin, data analysis
- Examines how to determine the performance of the method, including the assessment of linearity, range, accuracy

9. Reporting of results

- Gives guidance on reporting of data for regulatory purposes including the OECD Mutual Acceptance of Data (MAD)
- Recommends publishing of scientific data to promote more transparency and openness
- Reporting of method validation is also discussed

10. Storage and retention of records and materials

- Discusses requirements relating to the storage and retention of data, records and materials
- Claims the application of data integrity to both paper-based and electronic systems
- Prescribes protection of data, records and materials from deliberate or accidental changes, manipulations or deletions

GIVIMP

OECD Guidance Document on Good *In Vitro* Method Practices

GIVIMP consists of 10 chapters addressing key aspects throughout the *in vitro* method lifecycle from development to regulatory use. It includes definition of roles and responsibilities, facility requirements, characterisation of cells and tissues, selection of control and reference items, checking the performance of the method, development and preparation of standard operating procedures and how to properly report results.

Download GIVIMP

1. Central regulation

1a. CHO-K1 cells thyrotropin-releasing hormone (TRH) receptor activation (beta-galactosidase) measuring agonist and antagonist activities
1b. CHO-K1 cells thyrotropin-stimulating hormone (TSH) receptor activation based on cAMP measurement

2. Thyroid Hormone (TH) synthesis

2a. Thyroperoxidase (TPO) inhibition based on oxidation of Amplex UltraRed
2b. Thyroperoxidase (TPO) inhibition based on oxidation of Luminol
2c. Tyrosine iodination using liquid chromatography
2d. Activation of the sodium iodide symporter (NIS) based on Sandell-Kolthoff reaction

3. Secretion and transport in serum

3a. Thyroxine-binding prealbumin (TTR) / thyroxine-binding globulin (TBG) using fluorescence displacement (ANSA)
3b. Thyroxine-binding prealbumin (TTR) using fluorescence displacement (T4-FITC)

4. Metabolism and excretion

4a. Deiodinase 1 activity based on Sandell-Kolthoff reaction
4b. Glucuronidation of thyroid hormones (THs) using liquid chromatography/mass spectrometry (LC/MS)
4c. Inhibition of thyroid hormones (THs) sulfation using liquid chromatography

5. Local cellular concentrations

5a. Inhibition of monocarboxylate transporter 8 (MCT8) based on Sandell-Kolthoff reaction

6. Cellular responses

6a. Human thyroid hormone receptor alpha (TRα) and Human thyroid hormone receptor beta (TRβ) reporter gene transactivation measuring agonist and antagonist activities
6b. CALUX human thyroid hormone receptor beta (TRβ) reporter gene transactivation measuring agonist and antagonist activities

7. Relevant short term methods integrating multiple MoA

7a. Measurement of intrafollicular thyroxine (T4) using zebrafish eleutheroembryos

8. Integrative cellular *in vitro* methods

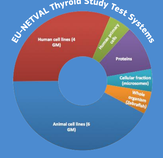
8a. Measurement of proliferation of rat pituitary-derived cell line GH3
8b. Proliferation, migration and oligodendrocyte differentiation of human neural progenitor cells

Legal instruments & agreements

- **Selection of thyroid methods**
  - ✓ 17 methods covering 8 blocks (OECD no 207)
- **Selection of EU-NETVAL 14 facilities**
  - ✓ Training requirements
- **Material transfer agreements (5)**
- **Collaboration agreements (15)**
- **Declarations (15)**

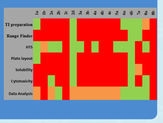
Test systems

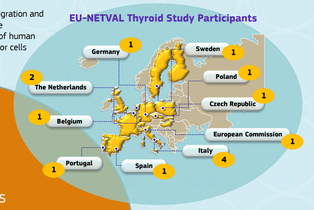
- **Sourcing & documentation**
  - ✓ Commercial, non-commercial, etc.
  - ✓ Information on the GM make-up
- **Authentication**
- **Contamination testing**
- **Cell banking**
  - ✓ Ensuring long term availability



Outline procedures

- **Harmonised approach**
  - ✓ Requirements (materials, equipment, etc.)
  - ✓ Acceptance criteria
- **Identification of reference/control items**
- **Test item preparation**
  - ✓ Range finder, cytotoxicity, solubility, etc.
- **Assessment of completeness**
  - ✓ Identification of missing criteria, etc.





1. Roles and responsibilities
1.1 *In vitro* method developers
1.2 Test system providers
1.3 Validation bodies
1.5 Suppliers of equipment, materials and reagents

2. Quality considerations
2.4 Quality control of test systems
2.5 Quality Control of consumables and reagents

5. Test systems
5.2 Cell and tissue sourcing
5.3 Cell and tissue culture transportation
5.6 Cell line identity and genetic aberrations
5.7 Contaminants screening: sterility, mycoplasma, virus

8. Performance of the method
8.1 Acceptance criteria
8.2 Experimental design
8.2.1 Plate layout
8.2.2 Data analysis

6. Test and reference/control items
6.1 Reference and control items
6.3 Test item preparation
6.4 Concentration range
6.5 Solubility
6.6 Stability
6.7 Solvents

7. Standard operating procedures
7.1 *In vitro* method standard operating procedures development
7.2 Preparing standard operating procedures

© European Union, 2017

The European Commission's science and knowledge service
Joint Research Centre

EU Science Hub: *ec.europa.eu/jrc* @EU_ScienceHub Joint Research Centre EU Science Hub - Joint Research Centre EU Science Hub

Joint Research Centre

Contact: Sandra Coecke
European Commission, Joint Research Centre
Directorate F – Health, Consumers and Reference Materials
Chemical Safety and Alternative Methods
Email: Sandra.COECKE@ec.europa.eu

OECD BETTER POLICIES FOR BETTER LIVES

@EU_ScienceHub
#ECVAM

©European Union, 2018 – icons: Freepik.com

Lessons from using software tools to conduct systematic reviews

Katya Tsaioun[1], Sevcan Akgün Ölmez[1], Daniele Wikoff[2], Robert Wright[3], Hubert Dirven[4], Sebastian Hoffmann[1], EBTC Zebrafish working group, EBTC Tox21 Working Group,
[1] EBTC at Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, [2] ToxStrategies, Asheville, NC, [3] Welch Medical Library, Johns Hopkins University, Baltimore, MD, [4] Norwegian institute of Public Health, Oslo, Norway

1. Introduction

EBTC is an international collaboration of academic, regulatory agencies, non-profits and industry leaders united by the mission to bring evidence-based methodologies into toxicology and risk assessment. EBTC pioneers application of such methodology in toxicology, e.g. by conducting systematic reviews (SR).

We present the processes and lessons learned in using software in two case studies, a SR of the Zebrafish Embryotoxicity Test (ZET) as a predictor of developmental toxicity, and a review of the predictability of the publicly available Tox21/ToxCast data for hepatotoxicity as determined in experimental animals and humans.

2. PECO questions

What is the performance of the Zebrafish Embryo Test (ZET) in predicting the presence and absence of malformations in guidelines studies of prenatal development toxicity in rats and rabbits (OECD TG 414 and equivalents)?

- Index Test
- Outcome
- Comparator Test
- Exposure: chemicals

How well do the ToxCast *in vitro* test predict the liver outcomes in animal tests (rats, dogs, non-human primates) and humans using a data set from 10 drugs?

- Index Test
- Outcome
- Comparator Test
- Exposure

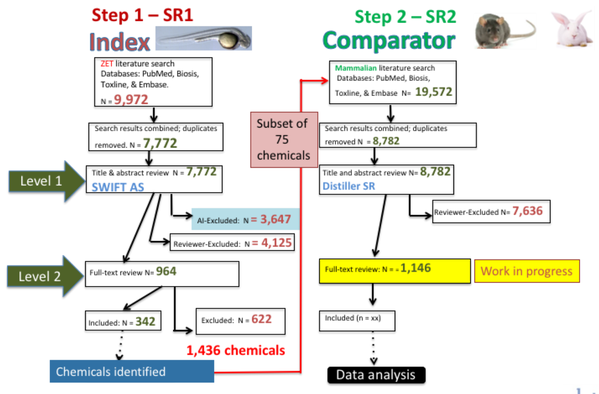
3. SR Software Survey

EBTC has surveyed its Scientific Advisory Council and Board of Trustees about SR software use/experience and features:

- 13 software packages used (in total)
- all offer multiple import and export interfaces
- most offer comprehensive reviewer role management
- most offer screening functionality with key word highlighting
- half are open source and are ranked user-friendly
- 3 offer data extraction
- 2 allow for automatic PRISMA plot generation,
- 1 automatically maps full text papers to title and abstract

4. Methods and ZET SR Process

We developed search strategies for PubMed, Embase, and Web of Science with the expertise of an informationist (RW) to find literature that answers the respective PECO questions. Full search strategies are published in the Zebrafish SR protocol[1] and Tox 21 SR protocol[2].



ebtc 5

Positives:

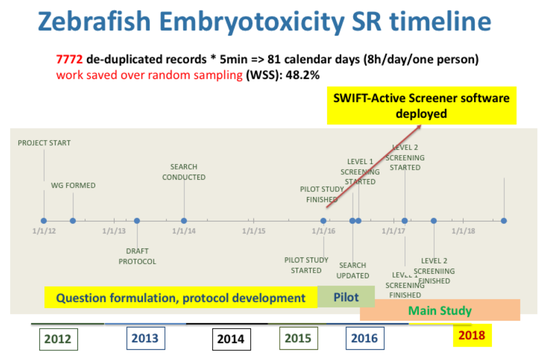
- SWIFT-Active Screener screening TiAb
- SWIFT-Active Screener AI function is validated and works well (saved 50% of time!)
- DistillerSR extraction function

Needs work:

- Support for high-throughput studies (HAWC)
- Text mining for extraction (all)
- Visualization and data sharing (SWIFT-Active Screener)

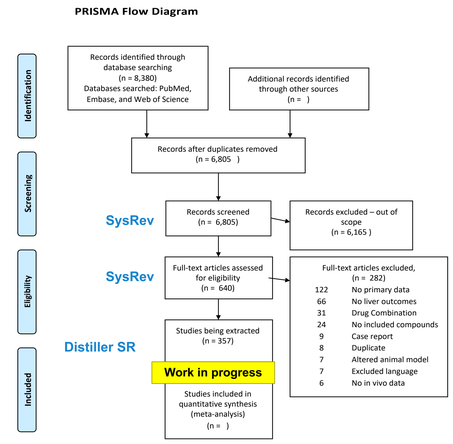
1: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=96120;
2: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=112353

5. Impact of SR software with AI



6. Tox21 SR process

Tox21 SR PRISMA Diagram



Positives:

- Screening TiAb is functional (SysRev)
- Publication-ready reports and PRISMA tables (DistillerSR)

Needs work:

- Connecting Level 1 and Level 2 screening, visualization (SysRev)
- Automatic matching of FT to references (all)
- AI functionality (SysRev, DistillerSR)
- Semi-automation of the hierarchical extraction (all)

7. Conclusions

- Collaborative cloud-based software tools are essential for building capacity, especially with geographically distributed teams
- Piloting the software packages in ZET SR informed the decisions in Tox21 SR, resulting in shorter timelines
- SR software enables engaging multiple reviewers and save time with AI
- No one software fulfills all our SR needs: need to switch mid-project, so files upload and download formats are important
- The areas of most need are:
  - Visualization and sharing
  - Text mining for semi-automated data extraction,
  - Machine learning and AI to improve the prediction of inclusion rate
  - Support for high-throughput studies

Resources: Available tools for SR: http://systematicreviewtools.com/
Contribute to the EBTC survey of SR tools: https://tinyurl.com/y7fj34wb

7. Acknowledgements

EBTC would like to thank Center for Alternatives to Animal Testing and Beagle Freedom Prize for their funding support and to Michelle Angrish, Gunn Vist, Sricharan Bandhakavi, Jyotsna Mehta, Seneca Harvey, Cathalijn Leenaars, Amanda McCormack, Hilda Witters, François Busquet, Burkhard Flick, Yi Zhang, Bianca Marigliani, Nicole Kleinstreuer, Tom Luechtefeld, Brian Howard, Andrew Shapiro, Barbara Livoreil, Emily Sena, Paul Whaley, Andrew Rooney and Rob de Vries for screening, providing software and contributing to the projects and software survey

Systematic literature review of available tools to critically appraise in vitro studies

Katya Tsaioun[1], Robert Wright[2], Rob de Vries[3], Paul Whaley[4], Vickie Walker[5], Caroline Foster[6], Jessica Wignall[6], Andrew Rooney[5], EBTC IV-CAT working group
[1] EBTC at Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, [2] Welch Medical Library, Johns Hopkins University, Baltimore, MD, [3] SYRCLE, Radboud University Medical Center, Nijmegen, The Netherlands, [4] Lancaster University, Lancaster, UK,
[5]National Toxicology Program, NIEHS, Research Triangle Park, NC, [6]ICF (contractor to the Toxicology Program), Durham, NC

Background

In vitro studies are becoming an increasingly important source of evidence in chemical risk assessment. There are concerns, however, about the methodological quality of these studies and this project is designed to systematically collect information on current study quality practices and guidance being utilized on *in vitro* studies. This information can then be used to inform: (1) researchers designing, conducting and reporting primary *in vitro* exposure studies, (2) authors of systematic reviews aiming to assess the risk of bias/study quality of the primary *in vitro* studies included in their review, and
(3) peer reviewers of journals advising on whether to publish a submitted manuscript. The systematic review methods presented here are part of a collaborative project between Evidence-based Toxicology Collaboration at Johns Hopkins Bloomberg School of Public Health (EBTC: www.EBTox.org) and the National Toxicology Program (NTP), with EBTC focusing on the development of guidance and standards to be used in the publishing of *in vitro* studies.

The NTP has pioneered the development and use of systematic review methods to address environmental health questions which requires reaching conclusions by integrating across evidence streams (animal, epidemiological human and mechanistic studies) (Rooney et al, 2014; RoC, 2015; OHAT, 2015). Although there are established methods and ongoing methods development on approaches to assess study quality or risk of bias for human observational studies, and to some extent *in vivo* animal studies, assessing study quality of mechanistic research and particularly studies with in *vitro* exposure regimes without clear established methods a few reviews of current practices. Hence, the aim of this joint NTP and EBTC effort is to conduct a systematic review that will inform critical assessment of *in vitro* studies and development of tools.

The aim of the NTP sub-project is to provide a comprehensive overview of existing critical appraisal tools for *in vitro* studies in order to establish the state of the science. The systematic review will serve as a foundation to consider if there are clear best practices and whether there are common risk of bias items or specific questions used by multiple research groups. However, it is important to note that this review only address current practices in assessing *in vitro* study quality and does not directly assess external validity or applicability of *in vitro* studies for a given research question. When assessing *in vitro* studies in the context of human health questions, the relevance or external validity (e.g., is the dose relevant to human exposure) may be more important than the assessment of study quality.

The aim of the EBTC sub-project is to develop a tool, IV-CAT (*In Vitro* Critical Appraisal Tool), designed to help ensure comprehensive and exacting peer-review of *in vitro* toxicology studies. The first step in the development of IV-CAT is a systematic review of existing critical appraisal tools and reporting standards for *in vitro* research in order to collect potential criteria which could be included in IV-CAT. The extracted criteria will be used as input for a Delphi study to construct the tool.

A multi-stakeholder working group was put together by EBTC and a protocol and search strategy for SR were developed (de Vries & Whaley 2018). The search was designed in such a way as to be suitable for both sub-projects.

Objectives and Scope

NTP and EBTC are collaborating on conducting a systematic review of existing tools which have been developed to either (a) facilitate assessment of the quality of or (b) improve the reporting quality of *in vitro* toxicology studies.

“In vitro toxicology research” is defined as a study in which living parts of animals (including humans) up to and including the level of organization of tissues, but not whole organs, are exposed to chemicals in order to detect adverse effects. This definition excludes: (1) efficacy studies, (2) whole embryo and *ex vivo* studies, and (3) viral toxicity studies. The definition includes studies in cell cultures (including stem cell and organo-typic cultures), and tissue slices. Studies into the safety of drugs (pharmaceutical toxicology) as well as biological effects of exposure to chemical substance (environmental toxicology) are both included.

The search strategy, developed by the EBTC Working Groups, and preliminary results based on the execution by NTP team and the next steps are presented here.

Methods

We developed search strategies for PubMed, Embase, and Web of Science with the expertise of an informationist (RW) to find literature that describes tools used to critically appraise the quality of *in vitro* studies. Each strategy had two parts. Part 1 focused on finding systematic reviews of *in vitro* studies, since systematic reviews typically employ instruments for assessing the methodological quality of included studies. Part 2 searched for studies regardless of publication type which explicitly addressed guidelines for the research design of *in vitro* studies. The two parts of the strategy were combined with the Boolean operator "OR" when searching a database (i.e. "Part 1" OR "Part 2").

We designed the search strategies to achieve a balance of precision and recall in the results. Search terms included keywords applied to relevant search fields (title, abstract, descriptors, etc.) and controlled vocabulary terms (i.e. MeSH and Emtree terms). Part 1 of the search consisted in a systematic review concept and an *in vitro* study concept. Validated systematic review filters were used as the systematic review concepts for PubMed and Embase. The PubMed strategy used the filter from Shojania and Bero (2001), and Embase strategy used the filter (small drop in specificity with a substantive gain in sensitivity) from Wilczynski, Haynes, and the Hedges Team (2007). Part 2 of the search consisted of a guidelines concept, a research design concept, and an *in vitro* study concept. The complete search strategies are available at zenodo.org/record/1493498.

We searched PubMed (1946-present), Embase (Embase.com, 1947-present), and Web of Science (Clarivate Analytics, 1900-present) from the earliest available dates to the dates of the searches. No language limits were placed on the searches.

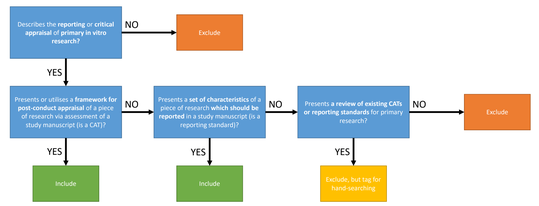


Figure 1. Screening decision tree

Next steps

NTP and EBTC will work in parallel following data extraction from the included studies to publish the review and develop the IV-CAT tool.

The development of IV-CAT will proceed in five steps:

1. Systematically reviewing existing critical appraisal tools and reporting standards for *in vitro* research for criteria to be included in IV-CAT.
2. Consulting with field experts as to the optimal design of an *in vitro* study, to determine a conceptual model for organizing the potential criteria identified from the SR into a provisional appraisal structure.
3. Employing a Delphi process to select the most important from the existing criteria and identify any missing criteria for use in a tool.
4. Creating a draft of the tool based on the results of Step 3.
5. Iteratively test and validate the tool.

Preliminary Results

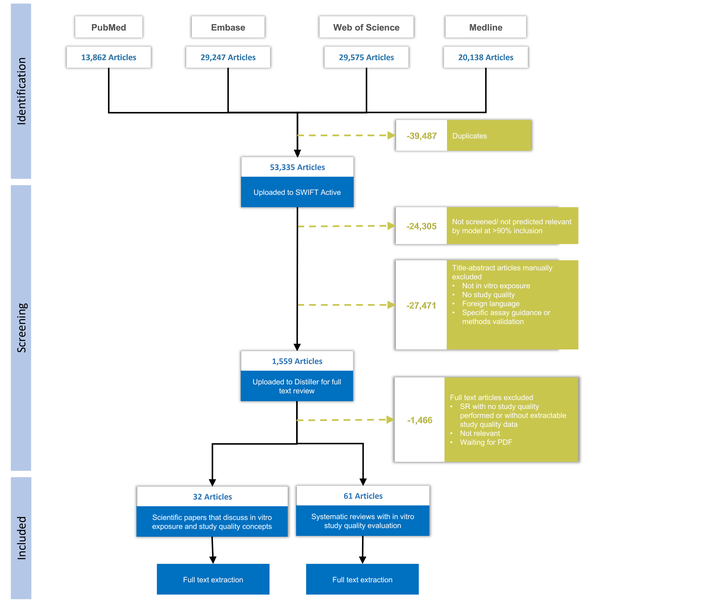


Figure 2. PRISMA plot 11/27/2018

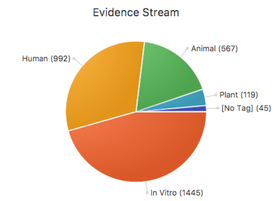


Figure 3. SWIFT-REVIEW analysis of evidence streams
In studies included after Level 1 screen (Title and Abstract)

References

Rooney, A. A., Boyles, A. L., Wolfe, M. S., Bucher, J. R. & Thayer, K. A. Systematic review and evidence integration for literature-based environmental health science assessments. *Environ Health Perspect* **122,** 711–718 (2014).

National Toxicology Program (NTP) Report on Carcinogens. Handbook for Preparing Report on Carcinogens Monographs (2015) [https://ntp.niehs.nih.gov/go/rochandbook]

National Toxicology Program (NTP) Office of Health Assessment and Translation. Handbook for Conducting a Literature-based Assessment Using OHAT Approach for Systematic Review and Evidence Integration (2015) [https://ntp.niehs.nih.gov/go/systematic_review]

de Vries, Rob, & Whaley, Paul. (2018). In Vitro Critical Appraisal Tool (IV-CAT): Tool Development Protocol (Version 1.0.0): zenodo.org/record/1493498

Shojania KG, Bero LA. Taking advantage of the explosion of systematic reviews: an efficient MEDLINE search strategy. Eff Clin Pract. 2001 Jul-Aug;4(4):157-62. [PMID: 11525102]

Wilczynski NL, Haynes RB, Hedges Team. EMBASE search strategies achieved high sensitivity and specificity for retrieving methodologically sound systematic reviews. J Clin Epidemiol. 2007;60(1):29-33

Enhancing Evidence Interpretation and Database Integration via Semantic Matching

Michelle Angrish[1], Sean Watford[2,3], Gail Hodge[4], George Woodall[1], and Anand Mudambi[5]

[1]National Center for Environmental Assessment, US EPA, RTP, Durham, NC, 27711, [2]National Center for Computational Toxicology, US EPA, RTP, Durham, NC, 27711, [3]Oak Ridge Associated Universities, Oak Ridge, TN, 37830, [4]Information International Associates, Oak Ridge, TN, 37830, [5]Office of the Science Advisor, US EPA, Washington DC, 20460

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

Background

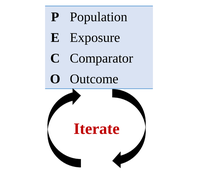


Figure 1▶ Systematic Review

- Through implementation of systematic review the US EPA IRIS program extracts data from ~150 studies per year across 15-20 chemical assessments in the development phase.
- These data are stored in the Health Assessment and Workspace Collaborative (HAWC, https://hawcprd.epa.gov/about/) a free, open-source, and web-based application.

- Data extraction of author reported health findings have introduced a data consistency challenge because words are inconsistently used by authors.
- Inconsistent language may lead to duplication and/or misinterpretation of study findings, make it difficult to efficiently retrieve information from HAWC, and pose a significant barrier to data exchange across different databases.

Objective

Leveraging SR methodologies with transparent and consistent data management practices as evidence addressing key science questions requires accurate curation of author reported data and integration with other databases.

Semantic ontology concept matching can translate disparate author reported terms to:

- **Preserve data in its original context;**
- **Promote consistency within the HAWC database;**
- **Make information more findable by facilitating the process of evidence integration; and**
- **Serve as a point of data integration with other databases of toxicity findings.**

Open Data and Risk Assessments

WHAT IS HAWC AND WHAT IS ITS PURPOSE?

◀Figure 2 Health Assessment Workspace Collaborative (HAWC) is an interactive web-based-content management system for human health assessments that is intended to promote transparency, trackability, structured study evaluation; structured summarization of study design and findings for epidemiological, animal toxicology, and in in vitro studies; and visualization of evidence supporting an assessment.

- HAWC is an interface that allows the data and decisions supporting an assessment to be managed in modules that can be publicly accessed on-line.

◀Figure 3 Word Cloud of 93 HAWC data extraction fields HAWC contains over 275 *in* vitro, animal bioassay, and epidemiology data extraction fields.

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

Term Annotation and Semantic Mapping

SEMANTICS – The study of linguistics, their meaning, relationship, and structure.

- Can aspects of semantics be deployed to drive consistency and interoperability of data (e.g. author reported terms such as assay target, system, organ/tissue, endpoint) used to support risk-based decisions?
  → match author reported terms to ontology concepts that can be used as a point of integration
  → The key element is that there is no need for data producers to use the same language across studies. Data can be retained in its original context and annotated via **ontology class** and semantic matching using resources such as:
    - Bioportal (https://bioportal.bioontology.org/) (Figure 4 below)

ONTOLOGY – "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).

ONTOLOGY CLASS – used to represent the set of all children and persons, respectively, in the application domain,

Figure 4 ▶ Bioportal Ontology Resource

- Bioportal (https://bioportal.bioontology.org/)a comprehensive repository of medical ontologies.

BioPortal Statistics	
Ontologies	734
Classes	9,596,856
Resources Indexed	48
Indexed Records	39,537,360
Direct Annotations	95,468,433,792
Direct Plus Expanded Annotations	144,789,582,932

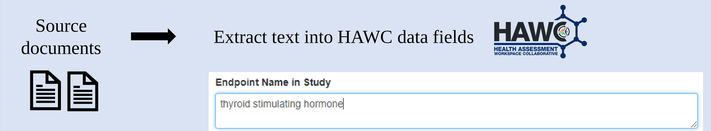
REST API includes a set of ontology resources and endpoint applications liked together.

ANNOTATOR – examines text input and returns relevant ontology classes

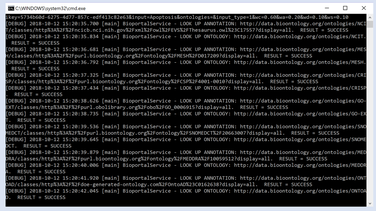
RECOMMENDER – receives text or a list of keywords and suggests appropriate ontologies for them

Figure 5. Data Extraction and Annotation Using Ontology Resources

Source documents → Extract text into HAWC data fields

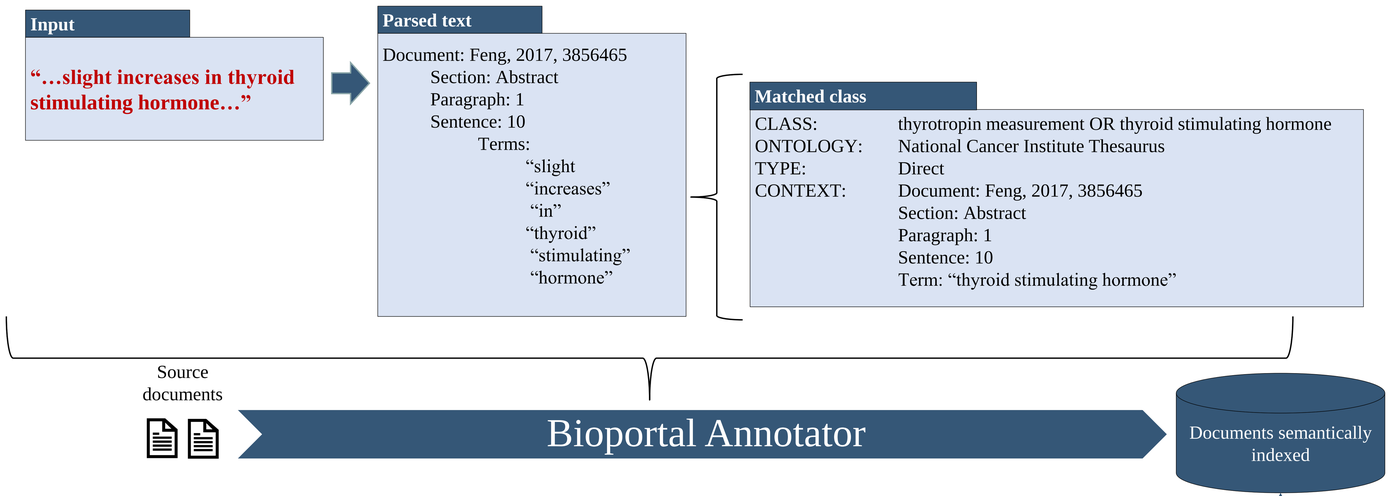


1. Obtain source documents;
2. Extract text into HAWC;
3. Create simple text file as input for the batch search in Bioportal;
4. CSV file output of scored results (scores not shown); and
5. Curate.



Keywords	Matched Class	Matched Ontology	Ont. Acronym	Preferred Name
Thyroid Stimulating Hormone		SNOMED CT	SNOMEDCT	Thyroid stimulating hormone
Thyroid Stimulating Hormone		Read Codes, Clinical Terms Version 3 (CTV3)	RCD	Thyroid stimulating hormone
Thyroid Stimulating Hormone		Medical Dictionary for Regulatory Activities	MEDDRA	Thyroid stimulating hormone
Thyroid Stimulating Hormone		Systematized Nomenclature of Medicine, International	SNMI	Thyroid stimulating hormone
Thyroid Stimulating Hormone	thyroid stimulating hormone	Protein Ontology	PR	thyroid stimulating hormone
Thyroid Stimulating Hormone	Thyrotropin Measurement	National Cancer Institute Thesaurus	NCIT	Thyrotropin Measurement
Thyroid Stimulating Hormone	thyroid stimulating hormone	go-plus	GO-PLUS	thyroid stimulating hormone
Thyroid Stimulating Hormone	thyroid stimulating hormone	Chemical Entities of Biological Interest Ontology	CHEBI	thyroid stimulating hormone
Thyroid Stimulating Hormone	thyroid stimulating hormone	Neuro Behavior Ontology	NBO	thyroid stimulating hormone
Thyroid Stimulating Hormone		National Drug File - Reference Terminology	NDFRT	Thyroid Stimulating Hormone
Thyroid Stimulating Hormone	Thyroid stimulating hormone	Ontology of Biological Attributes	OBA	thyroid stimulating hormone
Thyroid Stimulating Hormone	thyroid stimulating hormone	Children's Health Exposure Analysis Resource	CHEAR	thyroid stimulating hormone
Thyroid Stimulating Hormone	thyroid stimulating hormone	Combined Phenotype Ontology	UPHENO	thyroid stimulating hormone

A



B

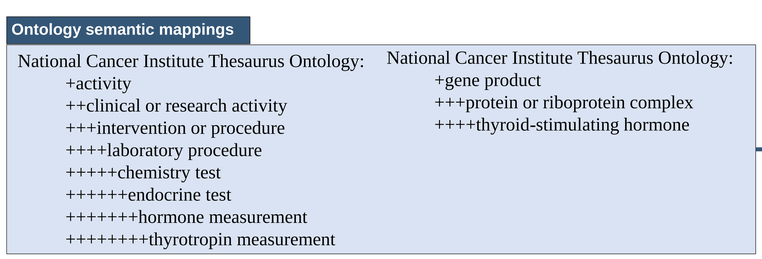


Figure 6 ▶ Semantic matching using the Bioportal Lookup tool (A) followed by semantic mapping (B) and querying against similarly indexed databases.

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable-based ink.

The views expressed are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

Quantitative Weight of Evidence Integration to Compare Confidence in Hypothesized Modes of Action

R.A. Becker,[1] Dellarco V[2], Schoeny R[3]

[1]American Chemistry Council, [2] VL Dellarco Independent Consultant LLC, [3] Rita Schoeny LLC

ABSTRACT

The World Health Organization/International Programme on Chemical Safety mode of action (MOA) framework provides a structure for evaluating evidence in pathways of causally linked key events (KE) leading to cancer. Variability in use of the MOA framework has led to different interpretations of the sufficiency of evidence in support of hypothesized MOAs. To improve incorporation of MOA knowledge in characterizing human health hazards and selecting dose-response extrapolation methods for specific chemicals, we have extended the MOA framework to enable scoring of confidence in the supporting data. This involves selecting hypothesized MOAs, and then, for each MOA, scoring the weight of evidence (WOE) in support of causality for each KE using evolved Bradford Hill causal considerations (essentiality, dose-response concordance, consistency, and analogy). The utility of this method for understanding and communicating the confidence in the evidence supporting selection of the likely operative MOA was documented in Becker et al., 2017 doi10.1016/j.yrtph.2017.02.017). Additional case examples that are under development will be presented: comparing mutagenic to cytotoxic MOAs for chromium VI, 1,4-dioxane and carbon tetrachloride; for dimethylarsinic acid, comparing a cyotoxicity/regenerative proliferation MOA, a mutagenic MOA, and a reactive oxygen species-induced clastogenicity MOA; for diethanolamine comparing a mutagenic MOA to a choline perturbation MOA; and scoring the confidence in the WOE for AfB1, an agent that is considered an archetypal mutagenic carcinogen. This method is intended to promote the full consideration of all underlying science, improve the systematic & transparent evaluation of data & facilitate communication of the relative confidence in the evidence used to identify the likely operative MOA.

Definitions and Terminology

MOA (Mode of Action): A biologically plausible series of chemical-specific Key Events (KEs) starting with exposure and proceeding through the interaction of an agent within a cell, subsequent physiological and tissue or organ changes, resulting in an adverse effect or outcome (Sonich-Mullin et al., 2001, Dellarco & Fenner-Crisp, 2012). In contrast, "mechanism of action" would describe the detailed biochemical and molecular interactions and responses of every step.

KE (Key Event): An empirically observable, reproducible step (which may be measured directly or indirectly, or reasonably inferred based on knowledge of the pathogenesis of the adverse effect) which is a necessary element of the MOA and critical to the outcome (i.e., necessary, but not necessarily sufficient in its own right) (see e.g., Meek et al., 2014a, Meek et al., 2014b).

KER (Key Event Relationship): A scientifically-based relationship that connects one KE to another, defines a direct relationship between the two (i.e., identifies one as upstream and the other as downstream), & facilitates inference or extrapolation of the state of the downstream KE from the known, measured, or predicted state of the upstream KE (OECD, 2016).

MOA Analysis: Evaluation of the extent to which available evidence indicates that the specific chemical acts via a biological pathway to cause an adverse outcome. A KE in MOA analysis must be sufficient "to draw a reasonable working conclusion of the agent's influence on key processes without having to establish the sequence of molecular processes in detail" (Dellarco and Fenner-Crisp, 2012).

Quantitative Confidence Scoring Method

Problem Formulation: To include **identifying postulated MOAs &KEs/KERs** for the adverse outcome(s) of interest & evaluation of data quality/ reliability of each study

Qualitatively evaluate the evidence for each hypothesized MOA (evidence in support of, or inconsistent with, the KEs/KERs) using the evolved Bradford Hill causal considerations

Quantitatively rate each KE/KER using the evolved Bradford Hill causal considerations using a scale of -3 (strong counter evidence) to + 3 (strong evidence convincing evidence of causation)

Composite score for each KE/KER (quantitative rating score x weight assigned for each Bradford Hill causal considerations then adjust for evidentiary value of each KE/KER)

MOA Confidence Score: Integrate evidence of causality by calculating the sum of the scores for all KEs/KERs and dividing by the total maximum score possible

Compare the quantitative confidence scores for the hypothesized MOAs, to identify the MOA(s) for which confidence in the supporting data is highest

Characterize the analysis to include: narrative of comparative confidence scores, uncertainties, identify the KEs/KERs where additional studies would bolster confidence because the existing causal evidence is weak, and implications with respect to potential hazards, and selection of the dose-response extrapolation method(s) for evaluating human health risks

Qualitative & Quantitative Scoring of KEs

Qualitative	Quanti-tative	Description
Strong	3	Multiple studies and/or extensive data provide convincing evidence that the substance causes the key event.
Moderate	2	Some evidence (direct or indirect) indicating the substance causes the key event, but scientific understanding is not yet completely established. There may be some studies that are equivocal.
Weak	1	Very limited evidence (direct or indirect) that the substance causes the key event along this pathway. Scientific understanding of the key event is limited.
No evidence	0	No data available to support or negate causation of this key event by the substance.
Weak Counter	-1	There is very limited contradictory evidence (direct or indirect) that the substance does not cause this key event.
Moderate Counter	-2	Some evidence (direct or indirect) indicating that the key event is not caused by the substance, but scientific understanding is not completely established. There may be some studies that are equivocal.
Strong Counter	-3	Multiple studies and/or extensive data provide convincing evidence that the substance does not cause this key event.

Evolved Bradford Hill Considerations

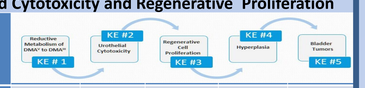
Causal Element	Weight
Essentiality	40%
Empirical Support (dose - incidence concordance and temporal concordance)	40%
Consistency – Different Test Systems	10%
Analogy – Across structurally similar compounds	10%

DIMETHYLARSINIC ACID

Case example: comparing 3 plausible MOAs

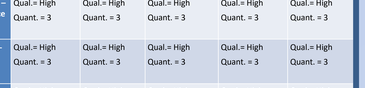
Hypothesized MOA	Quantitative WOE Confidence Score
MOA #1 Sustained Cytotoxicity and Regenerative Proliferation	92
MOA #2- DNA Reactivity and Mutagenesis	-87
MOA #3 - ROS Induced Clastogenicity	-68

Sustained Cytotoxicity and Regenerative Proliferation



Essentiality	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3
Empirical Support – Dose and Incidence Concordance	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3
Empirical Support- Temporal Concordance	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3
Consistency	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3
Analogy	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3	Qual.= High Quant. = 3

DNA Reactivity and Mutagenesis



Essentiality	Qual.= Strong Counter Quant. = -3	Qual.= High Quant. = -2 (inferred)	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3
Empirical Support – Dose and Incidence Concordance	Qual.= Strong Counter Quant. = -3	Qual.= No Evidence Quant. = 0	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3
Empirical Support – Temporal Concordance	Qual.= Strong Counter Quant. = -3	Qual.= No Evidence Quant. = 0	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3
Consistency	Qual.= Strong Counter Quant. = -3	Qual.= High Quant. = -2 (inferred)	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3
Analogy	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3

ROS Induced Clastogenicity

Essentiality	Qual.= Moderate Counter Quant. = -2	Qual.= Moderate Counter Quant. = -2	Qual.= Moderate Counter Quant. = -2	Qual.= Moderate Counter Quant. = -2	Qual.= Moderate Counter Quant. = -2
Empirical Support – Dose and Incidence Concordance	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.=Moderate Counter Quant. = -2	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3
Empirical Support – Temporal Concordance	Qual.= Moderate Counter Quant. = -2	Qual.= Moderate Counter Quant. = -2	Qual.= Moderate Counter Quant. = -2	Qual.= Moderate Counter Quant. = -2	Qual.= Moderate Counter Quant. = -2
Consistency	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3	Qual.= Strong Counter Quant. = -3
Analogy	Qual. = No Data Quant = 0	Qual. = No Data Quant = 0	Qual. = No Data Quant = 0	Qual. = No Data Quant = 0	Qual. = No Data Quant = 0

1,4-Dioxane exposure leading to liver tumors in rodents

Hypothesized MOA	QMOA WOE Score
Cytotoxic regenerative cell proliferation	+57
Mutagenic MOA	-45

Carbon tetrachloride exposure leading to liver tumors in rodents

Hypothesized MOA	QMOA WOE Score
Cytotoxic regenerative cell proliferation	+87
Mutagenic MOA	-36

Induction of liver tumors in rodents by Aflatoxin B1 (AFB1)

Hypothesized MOA	QMOA WOE Score
Mutagenic MOA	+87

Hexavalent chromium exposure leading to small intestinal tumors

Hypothesized MOA	QMOA WOE Score
Sustained small intestine villus cytotoxicity	+87
Mutagenic MOA	-36

Diethanolamine exposure leading to liver tumors in rodents

Hypothesized MOA	QMOA WOE Score
Perturbation of choline homeostasis cytotox.	+87
Endogenous nitrosation of DEA to nitroso-DEA induced mutagenesis	-27

Next Steps

- Improve and refine the quantitative method to address other modes of carcinogenic and non-carcinogenic actions (e.g., reproductive, endocrine, cytotoxic, mutagenic, etc.).
- Building an interactive software program to optimize data evaluation, scoring, engagement of expanding set of experts (expert elicitation /Delphi process) etc.
- Refining evaluation / weighting of the later common KEs? Weigh all such later KEs the same, or should there be a gradation of weights?
- Simplification of the approach (dropping later KEs – those that are common for MOAs; Isn't consistency already included in empirical evaluation?
- Exploring alternate weights for the Bradford Hill considerations
- Determination of the pros and cons of the approach used in the proof of concept for scoring of essentiality? Can this approach be improved?
- Incorporate Human Relevance steps (qualitative human evidence & quantitative human evidence)

Acknowledgements: C. Kirman was the lead author for the diethanolamine and hexavalent chromium case examples; V. Dellarco was lead author for the dimethylarsinic acid case example; M. Manibusan and R. Becker were lead authors for the 1,4-dioxane case example; and R. Schoeny, M. Moore and L.H. Pottenger were lead authors for the AFB1 case example. These investigators also contributed, along with J. Seed , J.M. Kronenberg, B. Meek, J. Foreman, C. Palermo, I. Linkov, and M. Dourson to the development of the method and the publication "Quantitative weight of evidence to assess confidence in potential modes of action" (Becker et al. 2017. Regul Toxicol Pharmacol. 86:205-220 open accaess at https://www.sciencedirect.com/science/article/pii/S0273230017300387?via%3Dihub. M. Manibusan received funding to support this research from the American Chemistry Council (ACC). R. A. Becker is employed by ACC, a trade association of U. S. chemical manufacturers. J. Foreman and C. Palermo are employees of a separately incorporated but wholly owned affiliate of a company that manufactures chemical products. L. H. Pottenger and J.M. Kronenberg are employed by companies that manufacture chemical products. Nominal honoraria and/or travel expenses were provided to some of the authors to support their attendance at meetings of collaborating authors.

References:

Sonich-Mullin et al. 2001. IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regul. Toxicol. Pharmacol., 34 (2001), pp. 146-152, 10.1006/rtph.2001.1493

Dellarco & Fenner-Crisp. 2012. Mode of action: moving toward a more relevant and efficient assessment paradigm. J. Nutr., 142 (2012), pp. 2192S-2198S, 10.3945/jn.111.157396

Meek et al. 2014a. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J. Appl. Toxicol., 34 (2014), pp. 1-18, 10.1002/jat.2949

Meek et al. 2014b. Mode of action human relevance (species concordance) framework: evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. J. Appl. Toxicol., 34 (2014), pp. 595-606, 10.1002/jat.2984

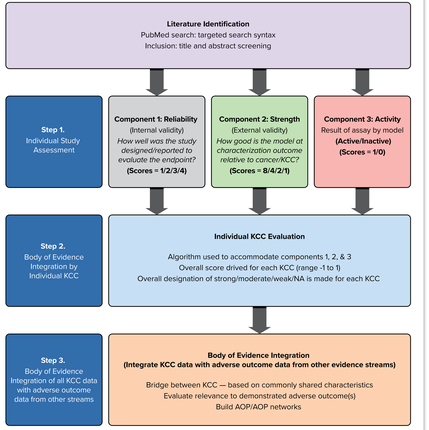
OECD. 2016. Users' Handbook Supplement to the Guidance Document for Developing and Assessing Adverse Outcome Pathways. OECD Series on Adverse Outcome Pathways, No. 1, OECD Publishing, Paris. Available at: http://dx.doi.org/10.1787/5jlv1m9d1g32-en

Systematic Evaluation and Integration of Data Relevant to Mechanisms of Carcinogenesis: Practical Experience in the Identification, Assessment, and Integration of Mechanistic Evidence in Human Health Assessments

G.A. Chappell, S.J. Borghoff, S.H. Fitch, C.L. Doepker, D. Wikoff

ToxStrategies, Inc.

Figure 1. Overview of the method applied to evaluate mechanistic data according to the KCCs, and to integrate the data into cancer evaluations



Background

- Mechanistic data provide valuable information regarding characterization of carcinogenicity.
  - The evaluation and integration of mechanistic data includes complexities:
    - heterogeneity of data types, collection, and reporting
    - understanding the human relevance of endpoints measured in animal and/or cell-based models, or even *in silico* approaches
    - potentially large evidence base with many endpoints to integrate
  - Wikoff et al., (2018) recently published a framework which builds on the key characteristics of carcinogen (KCC) approach by **quantitatively integrating mechanistic** data in a reproducible manner using elements important to risk assessment (e.g., quality, relevance, weight-of-evidence). The output includes an overall conclusion regarding KCC activity as it relates to carcinogenic responses, as well as identification of data gaps and prioritized research needs.

Objective

Apply a systematic approach for the identification, assessment, and integration of mechanistic evidence in human health assessments of carcinogenicity.

Methodology

Literature Identification

- The PubMed database was searched using syntax that included the agent name(s) in addition to terms relating to the ten KCCs (**Table 1**).
- Titles and abstracts from all articles returned via the searches were screened using DistillerSR (Evidence Partners, Ottawa, Canada), and were either included or excluded based on the following criteria (**Table 2**).
- High-throughput screening (HTS) data available through the ToxCast/Tox21 databases were extracted from the U.S. Environmental Protection Agency (U.S. EPA) Summary Files database (https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data; invitrodbv_v2; which was released October 2015 and remains the most up-to-date publicly available version; accessed October 2017 for this study).

Table 1. The Ten Key Characteristics of Carcinogens (from Smith et al. 2016)

Key Characteristic
1. Act as an electrophile either directly or after metabolic activation
2. Be genotoxic
3. Alter DNA repair or cause genomic instability
4. Induce epigenetic alterations
5. Induce oxidative stress
6. Induce chronic inflammation
7. Be immunosuppressive
8. Modulate receptor-mediated effects
9. Cause immortalization
10. Alter cell proliferation, cell death, or nutrient supply

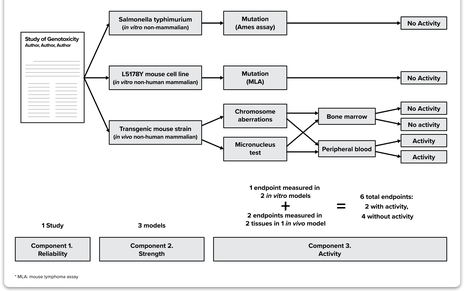
Table 2.

Include:	Exclude:
studies that evaluated the relationship between the agent and endpoints relevant to one or more of the KCC (or lack thereof)	• studies not available in English • studies that did not report original data • studies that did not specify the particular agent tested, and/or multiple agents were evaluated together

Individual Study Assessment (Step 1 of Framework)

- The full text of all included articles were reviewed, and mechanistic data were evaluated and scored for reliability, strength, and activity (**Figure 1**) per Wikoff et al., (2018)..
  - Reliability (i.e. study quality)
    - Reliability was scored on a scale of 1 to 4 (1=reliable without restriction; 2=reliable with restriction; 3=not reliable; 4=not assignable).
  - Strength
    - This ranking represents the confidence in the predictive capability of the model to predict activity of the KCC in humans.
    - The strength of the model was scored highest for human *in vivo* studies (either epidemiological or controlled exposure; score=8), followed by *in vivo* mammalian animal models (score=4), *in vitro* assays in either human or non-human mammalian cells (score=2), and non-mammalian systems (either *in vitro* or *in vivo*) received the lowest score (score=1).
  - Activity
    - For each endpoint, activity scores of 0 or 1 were assigned to indicate a response relative to control or lack thereof, respectively (**Figure 2**).
- ***Across all KCCs, collectively >150 endpoints were evaluated and scored***

Figure 2. Example of characterization of data from a single article, by study, model, endpoint, and activity.



* MLA: mouse lymphoma assay

Body of Evidence Integration: Individual KCCs (Step 2 of Framework)

- These elements were then quantitively integrated to provide a weighted score for each KCC, which facilitates the assessment of the body of evidence in the context of observed tumorigenic responses in animals and humans.

Equation for score (as described in Wikoff et al., 2018):

$$KCC\ Score = \frac{\sum_{i,all}\left[\left(\frac{w_R(-R_i+4)+w_M M_i}{w_R R_{max}+w_M M_{max}}\right)*\frac{E_{Act,i}}{E_{Total,i}}-\left(\frac{w_R(-R_i+4)+w_M M_i}{w_R R_{max}+w_M M_{max}}\right)*\frac{E_{InAct,i}}{E_{Total,i}}\right]}{n_{i,all}}$$

Legend: i = individual study-model; WR =weight of reliability component; Ri =reliability score of the study-model; Rmax =maximum reliability score possible =3; WM =weight of model component; Mi = model score of the study/model; Mmax =maximum model score possible (typically =8); ETotal, i = total number of endpoints in the study-model; EAct,i =number of active endpoints in the study/model; EInAct,i =number of inactive endpoints in the study-model; n,all = number of all study-models in a given KCC. The weights are user-defined, with our case studies typically using WR =1 and WM=2.

- The integrated scores for each KCC are then translated into categorical designations for overall activity of each KCC as related to the agent. This designation is a relative measure that is used to inform the basis of an overall characterization score of activity for the KCC (i.e., not carcinogenic potential) along with expert judgement (which involves consideration of variables such as model, dose, cytotoxicity, etc.) using the criteria in **Table 3**.

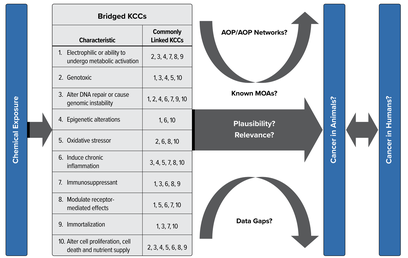
Table 3.

Designation:	Criteria
No Data	No data were available.
No Activity	Overall integrated score is ≤ 0 and no individual endpoints show activity.
Activity Unlikely	Overall AND individual model category scores are ≤ 0, but individual endpoints show activity.
Weak Evidence of Activity	Overall score is ≤ 0, but scores for key model categories (human and/or mammalian models) are > 0.
Moderate Evidence of Activity	Overall score is between 0 and 0.2, with scores for key model categories (human and/or mammalian models) also > 0.
Strong Evidence of Activity	Overall score is > 0.2, with scores for key model categories (human and/or mammalian models) also > 0.

Body of Evidence Integration — KCC Data and Other Evidence Streams (Step 3 of Framework)

- Bridging between KCCs (**Figure 3**):
  - The KCCs are recognized to occur as a set of (commonly linked) events, and are, thus, considered in relation to each other (Smith et al., 2016).
- Anchoring to adverse outcome:
  - KCC data are also then considered relative to any carcinogenic outcomes observed in experimental animals and humans, which allows for the mechanistic KCC to be evaluated in the context of the biological plausibility of a link with observed or potential adverse outcomes (or key events) related to carcinogenicity.

Figure 3. Bridging between KCCs and relating to adverse outcomes.

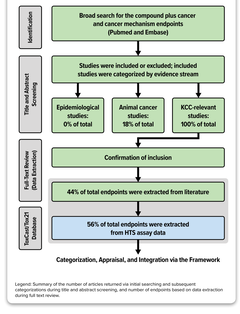


Legend: Integration of mechanistic data according to KCCs to assess biological relevance and plausibility, data gaps, and involvement in potential AOPs.

Results

Categorization of all articles returned via literature searching was conducted based on review of the titles and abstracts. KCC-relevant data were extracted and scored from the literature sources, and data available in the ToxCast/Tox21 screening program was also extracted and incorporated in the review. All extracted data were then scored for reliability, strength, and activity. (see **Figure 4**)

Figure 4. Overview of literature searching and screening results + HTS data.



Legend: Summary of the number of articles returned via initial searching and subsequent categorizations during title and abstract screening, and number of endpoints based on data extraction during full text review.

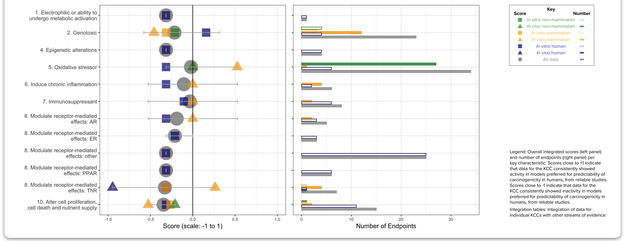
Integration of evidence per KCC

- Overall, nearly all of the integrated scores for each KCC were indicative of a lack of activity, particularly when accounting for study reliability and strength of the model tested.
  - When model types were evaluated individually, the integrated data for some specific model types demonstrated activity (e.g., non-human mammalian in vivo data for KCC #5; see **Figure 5**).
  - Substantial variation was observed in the number of endpoints per KCC.
  - Strength of evidence for activity ranged from "no activity" to "weak evidence of activity" across the KCCs.

Integration of evidence for KCCs with other data streams

- When the KCC-related mechanistic data were integrated with other available data streams related to cancer outcomes, the agent is unlikely to pose a carcinogenic risk to humans (see **Table 4**).

Figure 5. KCC-specific scores and data availability.



Legend: Overall integrated scores (left panel) and number of endpoints (right panel) per key characteristic. Scores close to +1 indicate that the KCC is associated with activity in models preferred for predictability of carcinogenicity in humans, from reliable studies. Scores close to -1 indicate that data for the KCC consistently showed inactivity in models preferred for predictability of carcinogenicity in humans, from reliable studies.
Integration tables: Integration of data for individual KCCs with other streams of evidence.

Table 4. Evidence Integration.

Framework Step 2: Body of Evidence Assessment for Individual KCCs			Framework Step 3: Body of Evidence Integration of KCC with Other Streams					
Characteristic	KCC Break	Rationale	Commonly Linked KCCs	Overall Characterization of KCC Data	INTEGRATION	Key Findings from Other Evidence Streams (as reported by study authors without considerations for study validity)		
						Animal	Human	
1. Electrophilic or ability to undergo metabolic activation	Not Active	All data demonstrated inactivity	NA	169 total endpoints (95 HTS) **Weight of Evidence:** • Most likely not mutagenic, genotoxic nor directly damages DNA • Unlikely to be active or weak for four other KCCs • Lack of activity or data not available for the remaining five KCCs. *Overall lack of KCC activity; no consistent demonstrations of activity for a single KCC.*	**Unlikely to be carcinogenic to humans based on integration of available evidence**	*Limited/inadequate evidence of carcinogenicity*	*Limited/inadequate evidence of carcinogenicity*
2. Genotoxic	Unlikely to be Active/Weak	Integrated score <0; all model category scores are <0 except for human *in vitro*	NA					
3. Alter DNA repair or cause genomic instability	No data	NA	NA					
4. Epigenetic alterations	Not Active	All data demonstrated inactivity	NA					
5. Oxidative stress	Unlikely to be Active/Weak	Integrated score <0; all model category scores are <0 except for mammalian *in vivo*	NA					
6. Induce chronic inflammation	Unlikely to be Active/Weak	Integrated score <0; all model category scores are <0 except for mammalian *in vivo*, for which inflammation was reported in a 2-year cancer bioassay by histopathological exam	NA					
7. Immunosuppressant	Unlikely to be Active	Integrated score <0; individual endpoints show activity	NA					
8. Modulate receptor-mediated effects	Weak for THR only; Unlikely to be Active for AR and ER. Not Active for PPAR and other receptors	Integrated score <0; all model categories scores <0 except for human *in vitro*	NA					
9. Immortalization	No Data	NA	NA					
10. Alter cell proliferation, cell death and nutrient supply	Not Active	All data demonstrated inactivity	NA					

Legend: Integration tables: Integration of data for individual KCCs with other streams of evidence.

Overall Findings

- The methodology, process, and results of this assessment represent a case study of practical experience with the implementation of a systematic review of mechanistic evidence into human health assessments
- In addition to presenting how the KCC data can be systematically integrated with other evidence streams, this case study demonstrates that systematic evaluation of mechanistic data may include substantial complexity, and that consistent evaluation of such data requires a clearly defined process executed by a skilled multidisciplinary team.
- Evaluation of mechanistic data related to the KCC, together with consideration of evidence from experimental animal cancer studies, can be valuable in cases of limited epidemiological data.
- The expertise of the screener impacts which studies will be included in the evidence base and how such will be appraised.
- The Wikoff et al., framework provides a feasible and flexible solution for assessing and integrating mechanistic data in systematic assessments.
  - Rapid incorporation of all data provides a quick overview of activity per KCC.
  - Data visualizations are helpful in facilitating the evidence to decision process.
    - Data source (e.g., literature vs. HTS data) may be especially important for data-poor agents (**Figure 6A**)
    - Using KCC sub-categories can provide additional information (**Figure 6B**)

Figure 6. Data visualization using various parsing strategies.

Legend: A) Example of scoring for all KCC with data separated by model type (left panels) and by data source (right panels, HTS vs. literature) for an agent with minimal KCC data available in the literature. B) Example of scoring for KCC #8 scored as a single KCC and also analyzed by sub-categories for an agent with ER activity.

Conclusion

The Wikoff et al., framework provides a feasible and flexible solution for assessing and integrating mechanistic data in systematic assessments of potential carcinogenicity.

Smith, M.T., et al. "Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis." Environmental Health Perspectives (2016).
Wikoff, D.W. "A framework for systematic evaluation and quantitative integration of mechanistic data in assessments of potential human carcinogens." Toxicological Sciences (2018).

Assimilation of multiple toxicity endpoints to identify and prioritize constituents of concern in oil and gas produced water

C. Danforth,[1] I. Rusyn,[2] W. Chiu,[2] E. Craft[1]
[1]Environmental Defense Fund, New York, NY; [2]Department of Veterinary Medicine & Biomedical Science, Texas A&M University, College Station, TX

Abstract

Produced water (PW) is the largest waste stream associated with oil and gas development. In the United States, onshore operations generate an estimated 900 billion gallons of produced water annually, and its efficient management and disposal is often challenging. While this wastewater is primarily deep-well injected for enhanced oil recovery and disposal, factors including drought and induced seismicity have led to the consideration of its use outside the oilfield. There remain significant gaps in our understanding of potential risks to human health or environmental impacts from such practices. We performed a comprehensive literature search, screening nearly 16,000 articles, to identify and aggregate chemicals detected in produced water. However, the dearth of information on toxicity of these chemicals limits our ability to design effective treatment and monitoring strategies for potential constituents of concern. **This poster will present our method to assimilate mechanistic information on these chemicals and to generate data using *in silico* methods where lacking**. This study not only recognizes and prioritizes which chemicals are of concern currently, but also identifies chemicals with no data that will require further study. This research effort will inform stakeholders and decision-makers on the risks that this complex waste stream may pose, and indicates research and regulatory improvements to reduce those risks.

Background

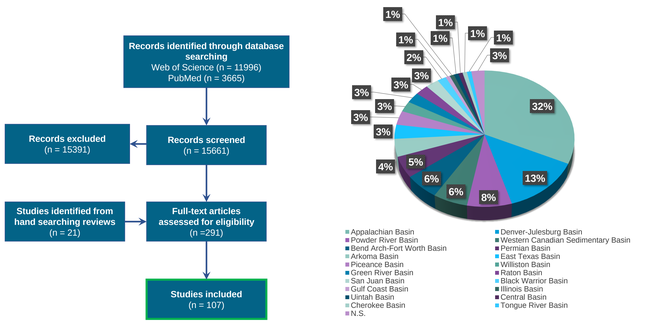
Produced water is a highly complex wastewater and may contain:

- Total dissolved solids (TDS) >10x seawater
- Hydraulic fracturing additives or maintenance chemicals
- Naturally occurring radioactive materials (NORM)
- Petroleum hydrocarbons and other organic compounds

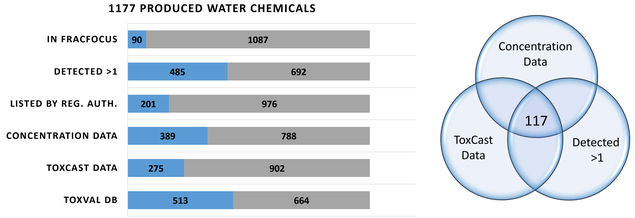
Despite its complexity, variability, and uncharacterized nature, there is increasing demand to consider this wastewater a resource. **This framework will help identify potential contaminants of concern for treatment and environmental monitoring.**

Literature Review to Identify Chemicals in Produced Water

Oil OR Gas **AND** Conventional OR Unconventional **AND** Wastewater **AND** Chemical Characterization

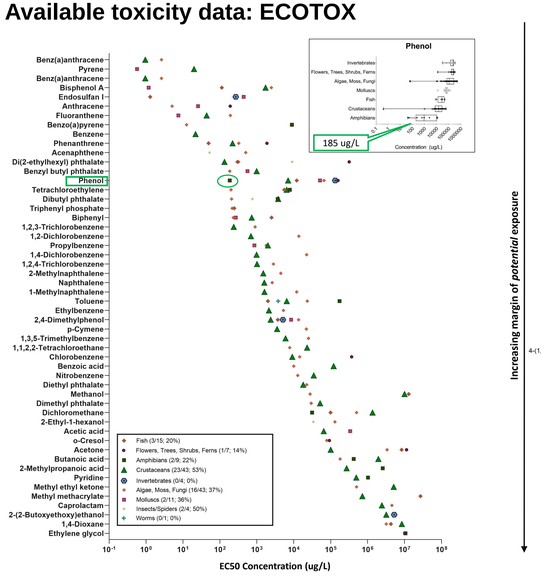


Current literature is skewed to one or two basins, which does not represent areas of current and future produced water generation

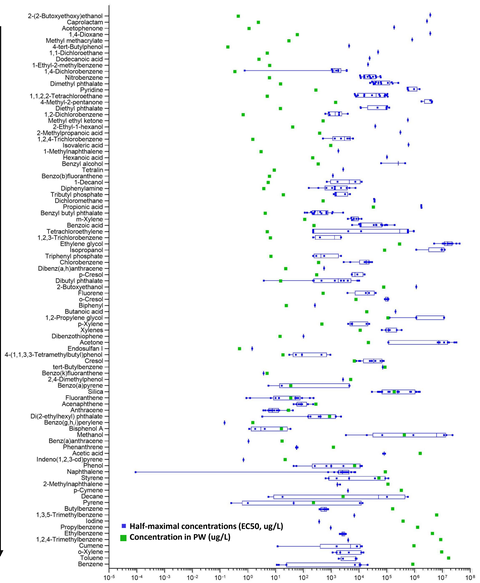


Our effort identified nearly 1200 chemicals in produced water. We selected 117 for further analysis based on available data.

Available toxicity data: ECOTOX

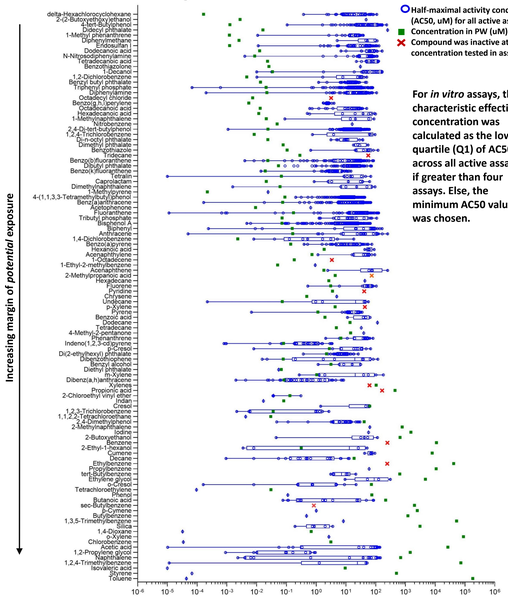


Half-maximal effective concentration (EC50) for aquatic organisms, by species group. For chemicals with more than four studies by species group, Q1 was chosen as the characteristic effective concentration. Else, the minimum value was chosen.



Comparison of half-maximal effective concentrations (EC50, ug/L) concentration of most sensitive species group to reported concentration of each chemical in produced water.

Available toxicity data: ToxCast



For *in vitro* assays, the characteristic effective concentration was calculated as the lower quartile (Q1) of AC50 across all active assays, if greater than four assays. Else, the minimum AC50 value was chosen.

Comparison of half-maximal activity concentrations (AC50, uM) to reported concentration of each chemical in produced water (uM).

Toxicity Values: QSAR or existing values

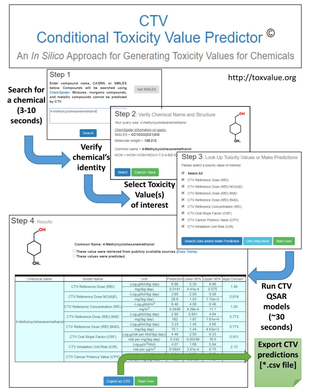
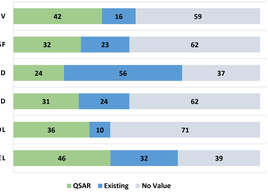


Figure taken from Wignall et al. 2014.



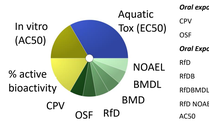
Existing toxicity values were collected by the CTV predictor from the following sources:

- **IRIS**
- **Office of Pesticide Programs**
- **Superfund Regional Screening Levels (CDC/ATSDR, PPRTV, HEAST)**
- **California EPA Office of Environmental Health Hazard Assessment**

ToxPi

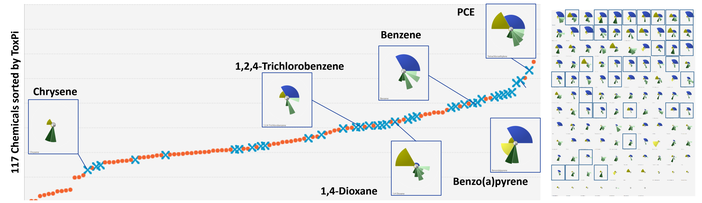
ToxPi = f(ecotox bioassay + *in vitro* bioactivity + toxicity data or values)

$$ToxPi = \sum_{1}^{Eco} w_{Eco} * effect + \sum_{1}^{I} w_i * activity + \sum_{1}^{CTV} w_C * tox\ val$$



Oral exposure, cancer	
CPV	Cancer Potency Value
OSF	Oral Slope Factor
Oral Exposure, non-cancer	
RfD	Reference Dose
RfDB	RfD Benchmark Dose
RfDBMDL	RfD Benchmark Dose, Lower Limit
RfD NOAEL	RfD No observed adverse effect level
AC50	Half-maximal activity concentration
EC50	Half-maximal effective concentration

Example of ToxPi weighting strategy



Distribution of example reference compounds (listed on EPA Drinking Water Standards and Health Advisory Tables) in sorted ToxPi Scores

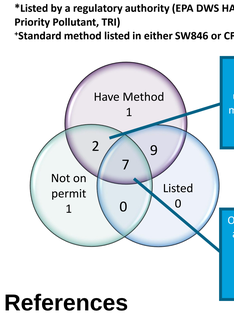
Proposed Ranking

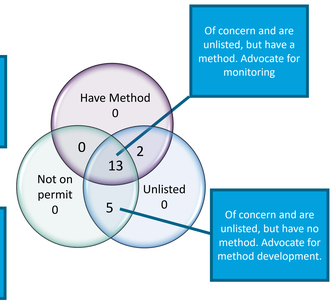
Top twenty ranked chemicals using example weighting strategy, including all chemicals listed by a regulatory authority (L) and excluding those chemicals that have MCLs and/or permitting levels (R)

Rank	CAS #	Name	FracFocus (2013-2017)	Listed*	CWT Permit	Method exists†
1	959-98-8	Endosulfan I				Y
2	127-18-4	Tetrachloroethylene		Y		Y
3	124-18-5	Decane	Y		Y	Y
4	50-32-8	Benzo(a)pyrene		Y	Y	Y
5	56-55-3	Benz(a)anthracene		Y	Y	Y
6	207-08-9	Benzo(k)fluoranthene		Y	Y	Y
7	193-39-5	Indeno(1,2,3-cd)pyrene		Y	Y	Y
8	108-88-3	Toluene	Y	Y	Y	Y
9	79-34-5	1,1,2,2-Tetrachloroethane		Y		Y
10	129-00-0	Pyrene		Y		Y
11	126-73-8	Tributyl phosphate	Y			
12	80-05-7	Bisphenol A	Y	Y		Y
13	191-24-2	Benzo(g,h,i)perylene		Y		Y
14	71-43-2	Benzene	Y	Y	Y	Y
15	108-95-2	Phenol	Y	Y	Y	Y
16	87-61-6	1,2,3-Trichlorobenzene				Y
17	100-42-5	Styrene		Y		Y
18	206-44-0	Fluoranthene		Y	Y	Y
19	108-90-7	Chlorobenzene		Y	Y	Y
20	91-20-3	Naphthalene	Y	Y		Y

Rank	CAS #	Name	FracFocus (2013-2017)	CWT Permit		Method exists†
1	959-98-8	Endosulfan I				Y
3	124-18-5	Decane	Y	Y	Y	Y
11	126-73-8	Tributyl phosphate	Y			
16	87-61-6	1,2,3-Trichlorobenzene				Y
22	104-76-7	2-Ethyl-1-hexanol	Y			Y
23	112-30-1	1-Decanol	Y			
25	503-74-2	Isovaleric acid				
27	140-66-9	4-(1,1,3,3-Tetramethylbutyl)phenol				Y
31	104-51-8	Butylbenzene				Y
32	143-07-7	Dodecanoic acid				
34	64-19-7	Acetic acid	Y			Y
40	107-92-6	Butanoic acid				
43	91-57-6	2-Methylnaphthalene				Y
46	132-65-0	Dibenzothiophene				Y
47	90-12-0	1-Methylnaphthalene				Y
48	57-55-6	1,2-Propylene glycol	Y			
49	115-86-6	Triphenyl phosphate				Y
55	111-76-2	2-Butoxyethanol	Y	Y		Y
56	142-62-1	Hexanoic acid				Y
59	65-85-0	Benzoic acid	Y			Y

***Listed by a regulatory authority (EPA DWS HA or CCL4, RCRA, Priority Pollutant, TRI)**
†Standard method listed in either SW846 or CFR 136





References

Marvel, S. W. *et al.* ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models. *BMC Bioinformatics* **19**, (2018).

Reif, D. M. *et al.* Endocrine Profiling and Prioritization of Environmental Chemicals Using ToxCast Data. *Environmental Health Perspectives* **118**, 1714–1720 (2010).

Wignall, J. A. *et al.* Conditional Toxicity Value (CTV) Predictor: An *In Silico* Approach for Generating Quantitative Risk Estimates for Chemicals. *Environmental Health Perspectives* **126**, 057008 (2018).

Williams, A. J. *et al.* The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *Journal of Cheminformatics* **9**, (2017).

Acknowledgements

We would like to acknowledge the efforts of Carol Kwiatkowski, Kim Schultz, Ashley Bolden of TEDX in performing the literature review and data extraction.

Approaches and tools to search and screen the literature for mechanistic data

Review of Environmental and Human Health Hazards from Alternative Applications of Produced Water

Fabian A. Grimm[1], Karen P. Christensen[1], Karlene S. Lavelle[1], Silvia I. Maberti[2], Melannie S. Alexander[1], Thomas F. Parkerton[2], Dennis J. Devlin[3]

[1]ExxonMobil Biomedical Sciences Inc., Annandale, NJ; [2]ExxonMobil Biomedical Sciences Inc., Spring, TX; [3]ExxonMobil Corporation, Irving, TX

Introduction

Produced water (PW) is naturally occurring water that comes out of a wellbore along with oil and gas. When hydraulic fracturing has been used, some of the water and chemicals in the frac fluid will also return to surface, but they decline over time. PW represents the largest waste stream associated with oil and gas production. Alternative applications of onshore PW are being evaluated as a potentially valuable resource, especially in geographic regions with limited water supply or high demand, including arid and agricultural areas. However, benefits of reintroducing PW into the environment need to be evaluated under consideration of potential environmental and public health impacts. Risk evaluations of PWs are challenging due to the presence of a range of natural and anthropogenic chemicals. Environmental and human health concerns are related to both inorganic and organic chemicals in PW streams including salts, aromatic hydrocarbons, naturally occurring radioactive materials (NORM), and specific classes of chemicals associated with endocrine disrupting properties, e.g. alkylphenols.

Objectives of the Review

- Develop a workflow for software-assisted (scoping and systematic) literature review
- Develop a search strategy to identify peer-reviewed scientific reports on environmental and human health impacts of PW
- Conduct software-assisted screening of literature search results to identify studies containing mechanistic data
- Identify studies reporting on treatment options and related efficiencies in alleviating exposure and health hazards in the peer-reviewed literature

Scoping/ Systematic Literature Review - Approach

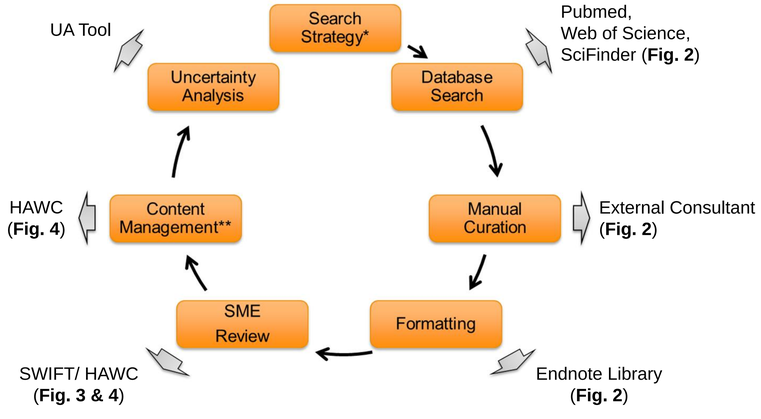


Figure 1: Generalized workflow for software-aided literature review at ExxonMobil Biomedical Sciences Inc., applicable to both scoping and systematic reviews. *For systematic review, define a priori criteria, **For systematic review, include quality assessment and assign independent reviewer. [SWIFT, Sciome Workbench for Interactive computer-Facilitated Text-mining; SME, Subject Matter Expert; HAWC, Health Assessment Workspace Collaborative; UA, Uncertainty Analysis]

Search Strategy & Database Search

Pubmed *higher sensitivity*	("produced water" **OR** "flowback water" **OR** "flow back water" **OR** "formation water" **OR** "reclaimed water" **OR** "degraded water")	1283 hits
Web of Science *higher specificity*	("produced water" **OR** "flowback water" **OR** "flow back water" **OR** "formation water" **OR** "reclaimed water" **OR** "degraded water") **AND** (oil **OR** petroleum **OR** gas) **AND** (exposure **OR** release **OR** discharge **OR** crop **OR** crops **OR** soil **OR** agriculture **OR** aquatic **OR** wildlife **OR** cattle **OR** livestock **OR** fish **OR** invertebrates **OR** tox* **OR** human **OR** health **OR** hazard)	828 hits

RESULTS

MANUAL CURATION & FORMATTING

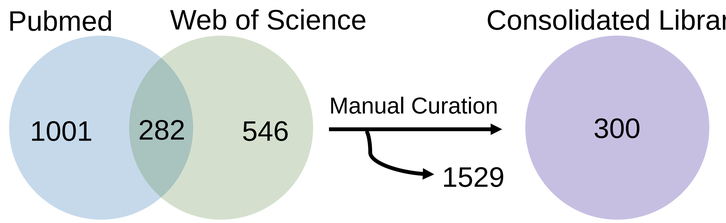


Figure 2: 2111 articles were initially identified and screened. 282 duplicates were removed and articles not relevant to the search topic were excluded. Articles were excluded if they (1) focused on PW derived from other industries, or (2) were unrelated to environmental and/ or human health impacts of PW. Altogether, manual curation revealed 300 peer-reviewed articles relevant to the search topic. A quality assessment of individual studies has not yet been concluded as part of this literature review.

SEARCH HIT CATEGORIZATION IN SWIFT REVIEW

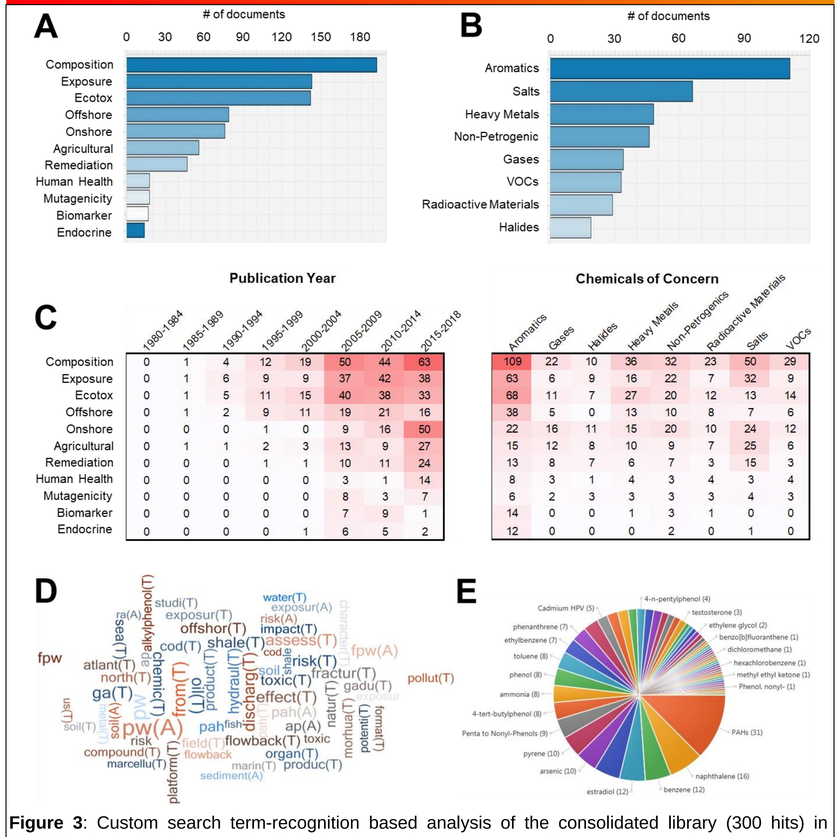


Figure 3: Custom search term-recognition based analysis of the consolidated library (300 hits) in SWIFT indicated that a majority of studies focused on PW composition, exposure and ecotoxicological studies (Figure 3A-D). While similar numbers of studies focus on onshore and offshore PW, there is a proportionally much stronger recent increase in the onshore PW literature, whereas articles investigating primarily offshore PW remain comparably stagnant (Figure 3A). Consistent with increasing interest in alternative applications for onshore PW, there is a rising number of publications related to agricultural uses, treatment and remediation, biomarker discovery to track exposures, and human health impacts in recent years. Search term-recognition based publication hits related to endocrine-disrupting chemcials are shown in Figure 3E. [Abbreviations: T=title, A=abstract, EDC=endocrine disrupting chemical, VOC=volatile organic compound]

LITERATURE SCREENING IN HAWC

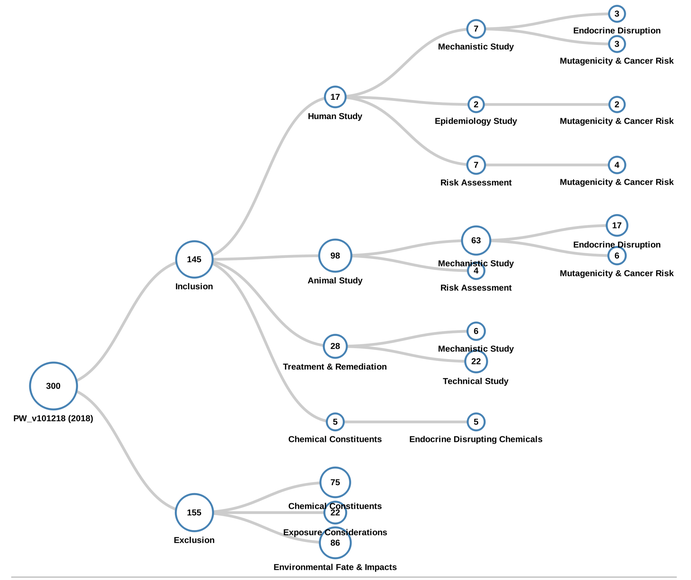


Figure 4: In order to identify mechanistic and supporting studies related to endocrine activity and mutagenicity, the curated search results were imported into "Health Assessment Workspace Collaborative" (HAWC) for subject matter expert review. Altogether, twenty manuscripts relevant to endocrine disrupting potential and fifteen papers evaluating mechanisms of mutagenicity and mutagenic potential of PW were identified. In total, six studies reported results related to PW treatment-related alleviation of toxicity using mechanistic data.

A

Reference	Organism	Exposure	Summary
Hull *et al.* (2018) Sci Total Environ. 644: 183-192	Bacteria	Untreated PW (220 day collection), 0.45 µm filtered PW, and SPE-concentrated PW	Mutagenicity (AMES test) was observed in SPE-concentrated samples
Yao *et al.* (2015) Toxicol Appl Pharmacol. 288:121-130	Human bronchial epithelial cells	6 week exposure to filtered (0.22 µm) PW from Marcellus Shale. Dilutions ranged from 0.13% to 8% in cell culture medium	Cells exhibited altered morphology, induced colony formation and exhibited enhanced cell migration in a concentration-dependent manner.
Durant *et al.* (2016) J Nat Gas Sci Eng. 34: 969-978	Human	Dermal exposure modeling for various chemical constituents of PW, assuming full hand exposure for three hours, once per week for four years	No increased lifetime cancer risk indicated for most constituents except for benzo(a)pyrene and heptachlor
Bloomdahl *et al.* (2014) J Nat Gas Sci Eng. 21: 348-356	Human	Exposure modeling for 12 VOCs	No increased lifetime cancer risk indicated

B

Reference	Organism	Exposure	Summary
Li *et al.* (2007) Environ Toxicol Chem. 26: 416-423	Bacteria	Raw oilfield PW effluent & PW treated by photocatalysis, electro oxidation, and photoelectrocatalysis	Mutagenicity (AMES test) was observed upon treatment with untreated PW at extract concentrations of 4-5 µg/ plate. All three treatment technologies were able to effectively detoxify direct-acting mutagenic organic constituents, but residual mutagenicity was observable after 60 min of treatment
Casini *et al.* (2006) Mar Environ Res. 62: S347-S351	Mosquito Fish	High concentrations (25-50%) of PW, PW treated conventionally, and innovatively by zeolites	Increased, but insignificant micronuclei formation was observed following exposure to untreated PW. Treatment was effective in detoxification.

Table 1: Summary of exemplary studies reporting mutagenicity and evaluating human lifetime cancer risk following exposure to PW or PW constituents (**A**), and PW treatment-related detoxicification effieciency of mutagenic activity (**B**)

CONCLUSIONS

- The developed workflow for software-assisted literature reviews was efficient in identifying and categorizing 300 relevant peer-reviewed publications
- PW is a rapidly increasing subject of scientific inquiry and is undergoing a shift from offshore to onshore focus with ecotoxicological and exposure science studies representing the major focus areas at this point
- The human health effects literature is lagging and so are risk assessments at this point

OUTLOOK

- Conduct a study quality assessment of identified mechanistic and supporting studies and evaluate the potential for risk evaluation

Acknowledgements

The authors would like to thank Dr. R. Jeffrey Lewis and Dr. Jennifer Foreman of ExxonMobil Biomedical Sciences Inc. for technical support. Funding for this literature review was provided by ExxonMobil Corporation.

Incorporating Mechanistic Evidence and Systematic Review Tools to Assess the Biological Plausibility of Cardiovascular Effects in Integrated Science Assessments

Michael J. Stewart, Ellen Kirrane, Thomas Luben, Michelle Angrish, Jennifer Nichols

National Center for Environmental Assessment, US EPA, RTP, Durham, NC, 27711

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

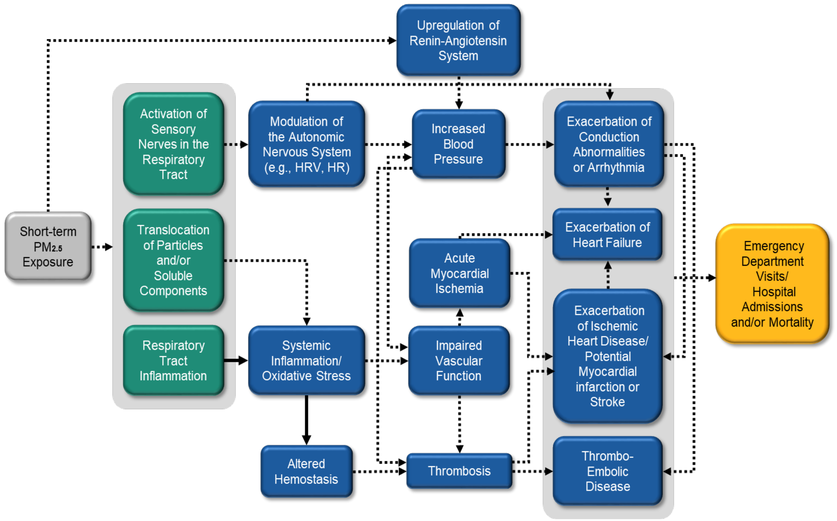
Background

National Ambient Air Quality Standards (NAAQS) are set for the six criteria pollutants: particulate matter (PM) ozone (O_3), oxides of sulfur, oxides of nitrogen, lead, and carbon monoxide. Primary NAAQS are set to protect public health- including sensitive populations such as children, older adults and people with chronic diseases. The Integrated Science Assessments (ISAs) identify, evaluate, and synthesize the best available and most policy-relevant exposure and health evidence, and communicate critical science judgments regarding the extent to which a specific health effect is related to exposure to a specific criteria pollutant. In making causality determinations, it is important to provide evidence that can plausibly link the inhalation of a criteria pollutant to downstream health effects that are systemic in nature. In the External Review Draft of the 2018 ISA for PM, a new and innovative approach was taken to systematically assess the biological plausibility for epidemiologic results indicating positive associations between ambient $PM_{2.5}$ concentrations and serious health outcomes such as ischemic heart disease (IHD), heart failure, and mortality. This approach leveraged mechanistic animal toxicology evidence along with human health endpoint data to identify biologically plausible pathways by which inhalation exposure to $PM_{2.5}$ could lead to these health outcomes. Here, we describe this approach and these biologically plausible pathways, placing emphasis on the role of mechanistic data in their construction. In addition, using the currently being developed 2018 draft O_3 ISA as an example, we briefly describe how this approach will be improved upon in future ISAs through the expanded use of systematic review tools and techniques.

Approach

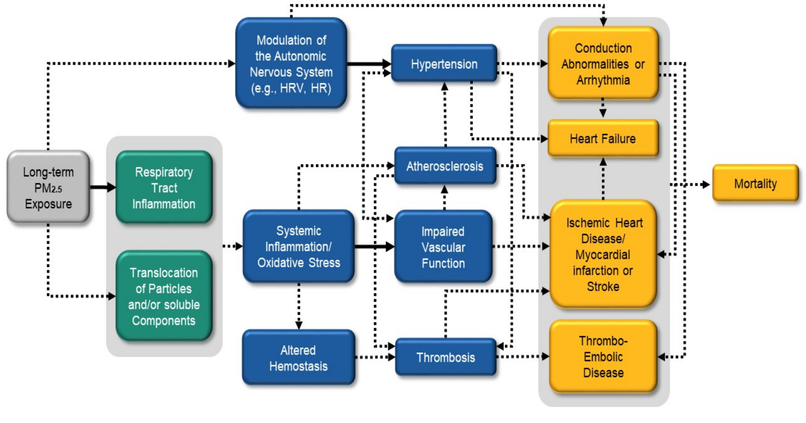
- Review the controlled human exposure, animal toxicological and epidemiologic literature examining the relationship between exposure to PM and endpoints ranging from changes in biomarker expression to associations with serious cardiovascular outcomes such as stroke, myocardial infarction and mortality
- Identify mechanistic studies in animals that used techniques such as pharmacological inhibitors, or knockout mice to elucidate the linkage between two or more endpoints
- Using the information above, construct a plausibility figure that depicts potential pathways leading from inhalation exposure of PM to an apical endpoint such as an emergency department visit for a cardiovascular-related event
- Separate figures are created for long- and short-term exposure to $PM_{2.5}$, $PM_{10-2.5}$ and ultrafine particles (see examples 1 and 2)
- Boxes are color coded:
  - Gray represents an exposure to a particular size fraction of PM
  - Green represent an initial event- often evidence of biomarker expression (e.g., cytokine expression in the respiratory tract),
  - Blue represents an intermediate event- may be evidence of biomarker expression (e.g., increased markers of coagulation) or be clinical in nature
  - Orange depicts apical associations reported in epidemiologic studies
- Figures are located at the beginning of each health chapter to serve as a roadmap for the health effects that are going to be discussed in more detail later in that chapter
- Figures are *not* developed using on a weight of evidence approach, as long as one study demonstrates an effect, it is included as a "box" on the figure. Dotted lines are proposed pathways between two boxes, while solid lines specifically link two boxes based on mechanistic studies using a pharmacological inhibitor or genetic knockout model following exposure to PM

Example 1: Potential biological pathways for cardiovascular effects following *short-term* exposure to $PM_{2.5}$



NOTE: The boxes above represent the effects for which there is experimental or epidemiologic evidence, and the dotted arrows indicate a proposed relationship between those effects. Solid arrows denote mechanistic evidence of the relationship as provided by an inhibitor of the pathway or a genetic knock out model. Progression of effects is depicted from left to right and color coded (grey, exposure; green, initial event; blue, intermediate event; orange, apical event). Here, apical events generally reflect results of epidemiologic studies

Example 2: Potential biological pathways for cardiovascular effects following *long-term* exposure to $PM_{2.5}$



*See NOTE under Example 1

Incorporating additional systematic review tools into future ISAs

- The 2018 O_3 ISA is under development and a draft has not been completed yet. However, the literature screening process has begun with the use additional systematic review tools
- Title abstract screening was done using SWIFT-ActiveScreener(AS), a software application employing machine learning in real-time based on inclusion/exclusion decisions to predict relevance of references
- Screening questions were used during title/abstract screening to "tag" references to disciplines/topics
- A "tag" for mechanistic studies was added within SWIFT-active to ensure these studies were reflected in the appropriate biological plausibility section
- Full text screening (conducted outside of SWIFT-AS), and decisions were recorded in Evidence Inventories (see below)

Sample Evidence Inventory

Study Details				Health Outcome Tags								Exp Dur Tags			Exp Conc Tags		Study Type Tags				Other Tags								
HERO ID	Author	Year	Title	Resp	CV	Reprod/Dev	Nervous	Met	Cancer	Dermal	Other	ST	LT	Unclear	Above	At or Below 1ppm	Other	Tox Human	Tox Animal	Chem	Other/Unclear	REVIEW/BACKGR	QUAL	SR	PotentiallyRelevs	Mechanistic	Dosimetry	ATRISK	Other
4256321	Uludag, M	2017	Effects of ozone treatment in endotoxin induced shock model in rats	N	Y	N	N	Y	N	N	N	Y	N	N	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
4256695	Mason, R	2013	Biomarkers Of Oxidative Stress Study	Y	Y	N	N	N	N	N	N	Y	N	N	N	N	N	N	Y	N	N	N	N	N	Y	N	N	N	N
4256698	Mason, R	2014	Biomarkers of oxidative stress study																										
4251532	Kuz'mina,	2012	[The effect of ozone therapy on the level of blood cholesterol in cerebrovascular	N	Y	N	Y	Y	N	N	N	N	Y	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	Y
4256780	Mason, R	2015	Biomarkers Of Oxidative Stress Study	Y	Y	N	N	N	N	N	Y	Y	N	N	N	N	N	N	Y	N	N	N	N	N	Y	N	N	N	N
4251943	Wang, L; C	2014	Ozone oxidative preconditioning inhibits renal fibrosis induced by ischemia and reperfusion	N	Y	N	N	N	N	N	N	Y	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	Y
4252568	Wagner, J(	2013	channels in ozone-induced decreases in blood pressure and heart rate in rats on a high fructose diet	N	Y	N	N	N	N	N	N	N	N	Y	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
4252569	Wagner, J(	2014	exposure to a mixture of ozone and rural ambient fine particles (PM2.5) in rats on a high fructose diet	Y	Y	N	N	N	N	N	N	N	N	Y	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
4253769	Lynch, H; S	2013	Weight-of-evidence evaluation of the cardiovascular effects of ozone exposure	N	Y	N	N	N	N	N	N	Y	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
4247395	Meng, W;	2017	reperfusion injury: A role for oxidative preconditioning in attenuating mitochondrial injury	N	Y	N	N	N	N	N	N	N	Y	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N

Summary

- The PM ISA incorporated a new and innovative approach for using mechanistic evidence to provide biological plausibility for the associations reported in epidemiologic studies
- This systematic approach results in a biological plausibility figure, as well as accompanying text that depicts the potential pathways leading from inhalation exposure of PM to an apical endpoint such as an emergency department visit for IHD or heart failure
- This approach will be improved upon in future ISAs by incorporating additional systematic review tools. This includes the use of software employing machine learning for literature screening and a "tag" for identifying mechanistic studies relevant to specific biological plausibility sections
- In addition to increasing transparency, the use of systematic review tools increases the efficiency of ISAs

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

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable-based ink.

*The views presented in this poster are those of the authors and do not necessarily represent the views or policy of the U.S. EPA. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Development of a Novel Evidence Integration Framework that Incorporates Human, Animal and Mechanistic Data

AUTHORS

Sandra Sulsky, Tracy Greene, Greg Mariano, Farah Chowdhury, Allison Franzen, P. Robinan Gentry, Ramboll; Donna C. Smith, Willie J. McKinney, Jr., Altria Client Services

ABSTRACT

To accurately characterize human health hazards, human, animal and mechanistic data must be integrated and the relevance to the research question of all 3 lines of evidence must be considered. Mechanistic data are often critical in fully integrating animal and human data and characterizing relevance and uncertainty. This novel evidence integration framework (EIF) provides a method for synthesizing data from comprehensive, systematic, quality-based assessments of the epidemiological and toxicological literature, including *in vivo* and *in vitro* mechanistic studies. The data are organized using both a disease-based and mechanism-based scheme, providing a method for assimilating and using mechanistic information to support evidence synthesis. The disease-based scheme uses the evidence of human health outcomes studied in the best quality epidemiological literature to organize the toxicological data according to authors' stated purpose, with the pathophysiology of the disease determining the potential relevance of the toxicological data. The mechanism-based scheme organizes the data based on the proposed mechanisms of effect and mechanistic data supporting key events leading to each human health endpoint, with the epidemiological data providing corroboration or no corroboration of causality. The EIF includes a method to cross-classify and describe the concordance of the data, and to characterize its uncertainty. A case study with nicotine is presented focused on the integration of evidence related to non-acute exposure and cancer. The results of the case study highlight knowledge gaps, demonstrate how different conclusions may be drawn depending on the organiz ation of the data, and show the impact of uncertainties on the strength of causal inference.

INTRODUCTION

Scientifically justifiable decisions regarding the potential for health effects resulting from chemical exposure require integration of multiple streams of evidence (i.e., epidemiological, animal and mechanistic), as well as consideration of the breadth and quality of that evidence.

METHODS

The EIF is applied following a comprehensive, systematic, critical review of the epidemiological and toxicological literature. This case study demonstrates the EIF by applying it to assess the relationship between non-acute nicotine exposure and cancer. The epidemiological literature focused on comparisons between users of smokeless tobacco – as a surrogate for nicotine exposure – compared with non-users of tobacco. The toxicological data were for chronic or repeated *in vivo* or *in vitro* exposure to nicotine, *per se*.

The best quality epidemiological and toxicological literature was identified and documented using standard techniques (Table 1 and Table 2), with modifications as necessary to allow for the evaluation of quality for mechanistic data.

Table 1. Epidemiology study attributes and their contributions to assessment of quality[1]

Study attribute	Quality criteria
Study objectives	Clearly stated Relevant to research questions
Study methods	Adequately described Appropriate for objectives Minimize selection and information bias Reasonable statistical power
Outcome measurement	Well-defined, reasonably specific Accurate measurement or diagnosis Proper time frame for risk of outcome
Exposure measurement	Well-defined, specific Verified (e.g., with biomarkers) Accounts for changes over time
Control of confounding	Known risk factors considered and measured Reasonable analysis method(s) used (stratification, multivariate statistical models)

Table 2. Toxicology study attributes and their contributions to assessment of quality[2]

Study quality category	Quality score	Requirements
Adequate	1	Studies or data from the literature or reports that are carried out or generated according to generally valid and/or internationally accepted testing guidelines (preferably performed according to good laboratory practices [GLP]), or studies in which the test parameters documented are based on a specific (national) testing guideline (preferably performed under GLP), or studies in which all parameters described are closely related/comparable to a guideline method.
	2	Studies that are not performed according to GLP or specific testing guidelines but are well documented and scientifically acceptable. For *in vivo* animal studies this includes the following criteria: • Data on the test animals including species, sex and strain • Purity/composition/origin of the test substance • Number of animals evaluated • Scope of the investigation per animal (e.g., clinical chemistry, hematology, organ weights, pathology or histopathology) and description of the methods • Description of the changes observed • Control group or historical control data of the laboratory • Description of the test conditions • Description of the route and doses of administration • Dose/concentration relationship is possible For *in vitro* studies this includes the following criteria: • Description of the test system and test method in details • Purity/composition/origin of the test substance (for *in vitro* studies only, if the suppler was noted it was assumed that the purity was acceptable) • Data on the dose/concentration differentiated according to the toxicity of the test substance on the test system; information on volatility • Data on secondary effects which may influence a result (solubility, impurities, pH shifts, influence on the osmolarity, etc) • Appropriate negative/positive controls as integral parts of the test References on adequacy of the method should be given or generally known.
Fair	3	Studies or data from the literature which do not meet the criteria for a quality score of 2. These include studies for which the methods were not clearly defined or specified in the study, in which only one dose was administered, or small numbers of animals per group were tested. The appropriate number of animals per group was determined based on recommendations from OECD guidelines for comparable study types.
	4	Studies or data from the literature which do not provide sufficient experiment details and which are only listed in short abstracts or secondary literature (books, reviews, etc).
Not useful	5	Criteria for *in vivo* animal studies: • Exposure not specific to chemical of interest • Single or acute exposure • Inappropriate or irrelevant species tested • Route of exposure not relevant to humans • If the endpoint being considered was related to addiction or reproductive/developmental toxicity • If the animals tested underwent any type of alteration or injury prior to testing Criteria for *in vitro* studies: • Exposure not specific to the chemical of interest • The endpoint was not related to the human health outcome(s) of interest, according to the author's purpose

After synthesizing each line of evidence independently, the data are integrated in stages and a gap analysis can be performed.

Stage 1: Disease-based integration

After identifying the human health endpoints that have been evaluated in the best-quality epidemiological studies, the toxicological evidence pertinent to each endpoint is organized in 3 ways, based on the proposed pathophysiology (i.e., the data must be demonstrated to be relevant to a key event in the development of the disease).

- The data are grouped to correspond to the human health outcomes from the epidemiological literature and divided into those providing evidence of a statistically significant change in response to exposure (evidence of an effect) and those that do not (no evidence of an effect)
- The relevance of the toxicological data to a human health outcome is classified as likely or uncertain based on what is known regarding the modes of action of the chemical relative to the development and/or progression of the specific human disease
- The toxicological data are classified into 1 of 6 major and minor domains (i.e., pathology, gross, functional, biochemical, genotoxicity and gene expression)[3]
  - Major evidence: endpoints clearly related to or associated with the development of the human disease, such as the apical endpoint; gross physiological or histopathological changes that are clearly analogous to health effects observed in humans with relevant exposures; or that have been identified as key events in the development of specific human health effects
  - Minor evidence includes endpoints that are less directly analogous or are less clearly related to the human disease in question
- Conclusions based on the integrated toxicological data are shown in Table 3

Table 3. Disease-based integration of overall toxicological conclusions:[a]

		If the majority of evidence in minor domains is:[a]								
		Insufficient		**Balanced**[b]		**Suggestive**		**Not suggestive**		
If the majority of the evidence in major domains is:[a]	**Insufficient**	a. Insufficient	b. Balanced[c]	c. Insufficient[d]	d. Suggestive[c]	e. Insufficient[d]	f. Not suggestive[c]	g. Insufficient[d]		
	Balanced[b]	h. Balanced	i. Balanced		j. Suggestive[c]	k. Balanced[d]	l. Not suggestive[c]	m. Balanced[d]		
	Suggestive	n. Suggestive	o. Suggestive		p. Suggestive		q. Balanced[c]	r. Suggestive[d]		
	Not suggestive	s. Not suggestive	t. Not suggestive		u. Suggestive[c]	v. Not suggestive[d]	w. Not suggestive			

[a] The major and minor domains are: biochemical, pathological, functional, gross or genotoxic. The gene expression domain is minor, only (no corresponding major domain)

[b] To account for uncertainties, "balanced" means 40–60% of the endpoints suggest evidence of an effect from nicotine exposure

[c] Greater than 60% of the minor evidence is from a primary cell line, or from *in vivo* gross data related to precancerous lesions

[d] Greater than 60% of the minor evidence is from a cancer cell line, or from *in vivo* gross data related to the apical endpoint but not of relevant route or duration to observe endpoint, or from an animal model with questionable relevance to human health (e.g., AJ mouse model)

Stage 2: Mechanism-based integration

The mechanism-based component of the EIF is driven by the toxicological data. For this cancer analysis, mechanisms related to direct measurements of cancer hallmarks or enabling characteristics.[4] Endpoints are categorized as being of likely versus uncertain relevance based on their specificity. Epidemiological data corroborate toxicological findings based on similarity to apical endpoints or tissue-type concordance.

Stage 3: Overall synthesis

The overall conclusions from the epidemiological and toxicological data are integrated as shown in Table 4.

Table 4. Integrated summary conclusions possible based on the integration of the epidemiological and toxicological conclusions

		If the toxicology evidence is:								
		Insufficient		**Balanced**		**Suggestive**		**Not suggestive**		
If the epidemiology evidence is:	**Insufficient**[a]	a. Insufficient	b. Balanced[b]	c. Insufficient[b]	d. Suggestive		e. Insufficient[c]	f. Not suggestive[d]		
	Balanced	g. Balanced	h. Balanced		i. Suggestive		j. Not suggestive[e]	k. Balanced[e]		
	Suggestive[f]	l. Suggestive	m. Suggestive		n. Suggestive		o. Balanced[e]	p. Suggestive[e]		
	Not suggestive	q. Not suggestive	r. Not suggestive		s. Suggestive[g]	t. Not suggestive[e]	u. Not suggestive			

[a] If there are no relevant endpoints available in the epidemiology data of adequate quality, include that statement and indicate the conclusions are based on toxicology data only

[b] If the toxicological evidence is balanced and is from minor domains and/or the majority of the data (i.e., > 60%) is from cancer a cell line, the overall evidence is insufficient. If the toxicological evidence is balanced and is from major domains and/ or the majority of the data (i.e., > 60%) is from primary cell lines, the overall evidence is balanced

[c] Insufficient amount of epidemiological data that is of adequate quality

[d] Insufficient amount of epidemiological data that is of adequate quality and toxicological data related to the human apical outcome is not suggestive of an effect

[e] If the toxicological evidence is not suggestive and is only from minor domains and/or the majority of the data (i.e., > 60%) is from cancer cell lines, the overall evidence is balanced or suggestive (depending on the overall epidemiological conclusion); if the toxicological evidence is not suggestive and is from major domains and primary cell line data, the overall conclusion is that the evidence is not suggestive

[f] If epidemiology data corroborate the toxicology data, then the conclusion is pushed up one level (e.g., from balanced to suggestive)

[g] If the toxicological evidence is suggestive and is from minor domains and/or the majority of the data (i.e., > 60%) is from a cancer cell line, the overall evidence is not suggestive. If the toxicological evidence is suggestive and is from major domains and/or the majority of the data (i.e., 60%) is from primary cell lines, the overall evidence is suggestive

RESULTS

Searches for studies of any type of cancer in smokeless tobacco users versus non-users and for cancer-related endpoints studied in connection with repeated exposure to nicotine returned 32,247 potentially relevant articles.

- 44 epidemiology studies of 48 types of cancer
- 314 toxicology studies; 100/314 relevant to the disease-based component; 312/314 relevant to the mechanism-based component (some studies provide data that are relevant to either or both components)

Disease-based component of the EIF: Two examples illustrating dependence of results on outcome definitions

The evidence for 22 specific cancers was evaluated based on signals indicating potential associations between smokeless tobacco use and various cancers available in the epidemiological evidence. There were limited epidemiology data for 26 other specific cancers, yielding insufficient evidence to apply the EIF. Examples of how the evidence was organized is provided in Figures 1 and 2, which assisted in drawing conclusions and integrating evidence using Table 6.

Figure 1. Stomach cancer

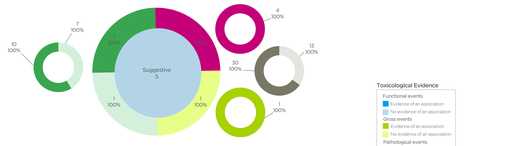


Figure 2. Gastrointestinal/digestive system cancer

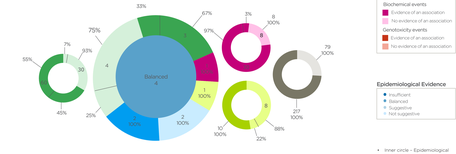
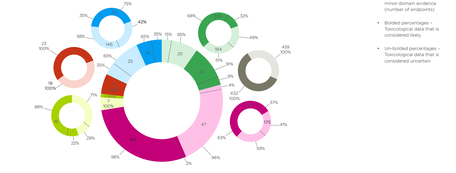


Figure 3. Mechanism-based approach



The conclusions drawn from the available data depend on the organizational scheme and outcome definitions. Table 5 summarizes the data for digestive systems cancers.

Table 5. Summary of epidemiological, toxicological, and overall integrated results for the disease-based approach to evidence integration

Cancer type	Epidemiological[a]	Toxicological[b]	Overall[c]
Digestive system			
Stomach cancer	Suggestive	Balanced	Suggestive
Colon cancer	Suggestive	Not suggestive	Not suggestive
Pancreatic cancer	Balanced	Balanced	Balanced
Rectal cancer	Balanced	Not suggestive	Not suggestive
Gastrointestinal/digestive system cancers combined	Balanced	Balanced	Balanced
Overall conclusion: The integrated toxicological and epidemiological evidence suggests that **nicotine may not play a role** in the development of **digestive system cancers.**			

[a] Integrated epidemiological evidence relating smokeless tobacco use to specified diseases shown in the table rows

[b] Integrated toxicological evidence relating nicotine exposure to the development of endpoints (not necessarily the apical endpoint) potentially associated with the development of the diseases shown in the table row

[c] The overall conclusions reflect the integration of the epidemiological and toxicological evidence according to the system summarized in Table 6

Mechanism-based component of the EIF

312 studies reported data potentially relevant to the hallmarks and enabling characteristics of cancer.[4] They reported 2,420 relevant endpoints which were categorized into the domains as providing either major or minor evidence (Figure 5).

Table 6. Summary and integration of evidence from the mechanism-based component of the EIF

Domain	Epidemiological	Toxicological	Overall
Gross	Balanced	Not suggestive	Not suggestive
Functional	Insufficient	Balanced	Balanced
Pathological	Suggestive	Balanced	Suggestive
Biochemical	Balanced	Balanced	Balanced
Genotoxicity	Balanced	Suggestive	Suggestive
Expression change	Insufficient	Balanced	Balanced

Gap analysis

- In the human domain there are no relevant studies of nicotine *per se*. Studies of smokeless tobacco users tend to be subject to exposure misclassification and to provide incomplete information on key details of exposure, including duration and amount of smokeless tobacco used prior to disease detection
- The toxicological data provide information about exposure to nicotine *per se*, but very few studies directly measured the incidence of cancer in animals *in vivo*. Such studies would provide the strongest evidence of a potential relationship between nicotine exposure and cancer risks in humans
- The limited number of repeated dose studies that focused on tumor incidence in animals *in vivo* could be viewed as a data gap, or as a consequence of the general movement away from animal testing. A more pertinent and general question to be addressed is whether *in vitro* gene expression studies that are being substituted for animal studies can predict the development of the disease of interest

Observations based on the examples

- Specificity in both the outcome and exposure measures is required to link the epidemiological with the toxicological data, and to draw appropriate discipline-specific conclusions
- The examples for nicotine illustrate how conclusions can be affected by the ways the endpoints are defined and combined
- The EIF provided a sensitive system for detecting signals potentially indicating a nicotine effect; these signals should be interpreted with full appreciation of the amount and quality of the available evidence and any uncertainties related to drawing conclusions (e.g., using smokeless tobacco as a surrogate for nicotine per se exposure) that can be used as guideposts for future, focused research and surveillance

CONCLUSIONS

- There are inherent differences in the data available from the human observational studies (e.g., the epidemiological evidence relates to tobacco, a mixed exposure) and the data available from experiments (e.g., the toxicological evidence that relates to nicotine exposure)
- The EIF accounts for quality and quantity of the underlying data, and uncertainties in the conclusions the EIF can support
- The EIF uses information about specific disease etiology to link the epidemiological and toxicological evidence
- The disease-based component identifies toxicological data relevant to human health endpoints via pathophysiological processes that lead to the disease
- The mechanism-based component directly assesses the evidence in major and minor domains that address the potential for exposure to affect the progression to a disease state or health endpoint
- The EIF is flexible and can incorporate new information at any level (epidemiological, *in vivo* and *in vitro*), as it becomes available
- The EIF systematizes integration of epidemiological and toxicological data derived from comprehensive, systematic, and critical reviews of the relevant literature. Applying the EIF can identify previously undetected signals of potential causal associations between chemical exposures and human health outcomes and identifies knowledge gaps. These results provide a direction for future, targeted investigations to address the uncertainties in the available existing evidence

REFERENCES

1. Sulsky, S. I., F. H. Hooven, et al. (2002). "Critical review of the epidemiological literature on the potential cardiovascular effects of occupational carbon disulfide exposure." Int Arch Occup Environ Health 75(6): 365-380.

2. Criteria modified from Klimisch, H. J., M. Andreae, et al. (1997). "A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data." Regul Toxicol Pharmacol 25(1): 1-5.3. R Core Team. R: A language and environment for statistical computing, 2015.

3. Building on Dekant, W. and J. Bridges (2016). "A quantitative weight of evidence methodology for the assessment of reproductive and developmental toxicity and its application for classification and labeling of chemicals." Regul Toxicol Pharmacol 82: 173-185.

4. Hanahan D, Weinberg RA. (2011) Hallmarks of cancer: the next generation. Cell 144(5):646-74

Building and Evaluating the Utility of an Adverse Outcome Pathway Network for Arsenic-Induced Diabetes

Ingrid L. Druwe[1], J. Allen Davis[2], Jeff Gift[1], Ila Cote[1], Janice S. Lee[1]

EPA, Office or Research and Development, National Center for Environmental Assessment – Research Triangle Park, [2] EPA, Office or Research and Development, National Center for Environmental Assessment – Cincinnati;

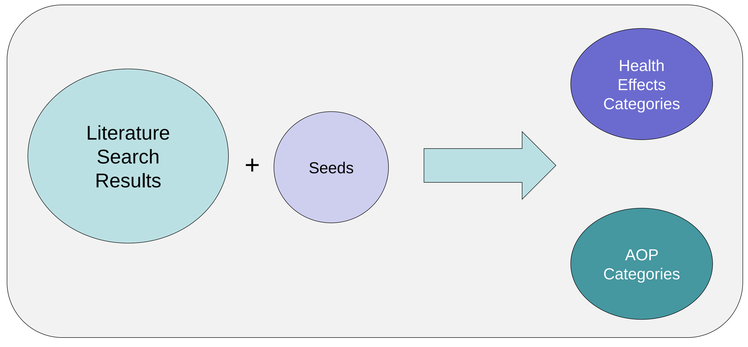
Ingrid L. Druwe I druwe.ingrid@epa.gov **I** 919-541-2452

Identifying Arsenic-Specific Evidence

During the 2015 Arsenic Development Plan meeting with the National Academy of Science the National Research Council (NRC) recommended that EPA conduct mode of action (MOA) analyses to facilitate understanding of exposure-response relationships to below the observed range. The MOA approach was thought worthwhile to pursue as a complement to the analysis of epidemiological evidence (+1000 studies on a multitude of health outcomes). Importantly, the NRC was not clear whether an MOA analysis would be feasible for the stated purpose of understanding dose-response relationships below the observable range. As a case example, an MOA analysis of diabetes was pursued. Diabetes was one outcome the NRC asked EPA to prioritize for consideration in the assessment.

Our goal was to perform MOA analyses to aid in informing low dose extrapolation. Knowing that the arsenic database is very large we performed a targeted literature screening for MOA coupled with a clustering approach. We used machine learning to prioritize studies for screening. References were clustered into groups based on language similarity (i.e., natural language processing) using OmniViz reference visualization software (Instem, Staffordshire, United Kingdom).

- The initial literature search was designed to be comprehensive and not miss potentially relevant studies; clustering helped more efficiently identify those references most likely to contain data relevant to hazard identification.
- Approximately 900 additional references were used as “seeds;”
- The “seeds” and literature search results were combined, and, using natural language processing, the titles and abstracts of the references were grouped based on similarity.
- Reference clusters containing one or more of the “seed” references were selected to create the health effects cluster of 3,715 references that were then manually screened for relevance. Approximately 196 studies clustered to the diabetes outcome tag.
- No particular individual study evaluation tool was used due to the large number of studies, diversity in the type of mechanistic considered in the AOP analysis, and lack of accepted study evaluation tool at the time of the analysis

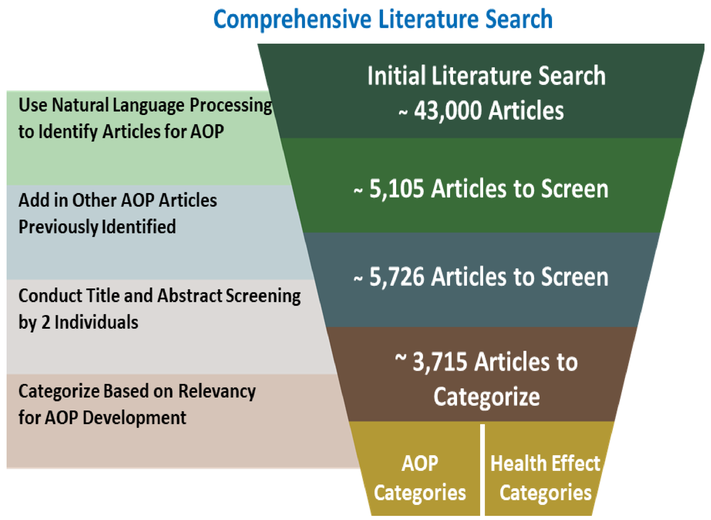


~ 3,715 Papers into 2 Broad Categories

Health Effect Specific		AOP Specific	
102	Bladder	218	ADME
116	Cardiovascular	426	Cell viability, Proliferation, Cycle Changes
81	Developmental	86	Cellular Differentiation, Malignant Transformation
21	Digestive System	354	Gene Expression Changes
70	Endocrine System / Diabetes	96	Immune Mechanisms
88	Hematopoietic System	75	Endocrine Mechanisms
152	Immune System/Lymphatic	100	Epigenetic Mechanisms
266	Liver	241	Non-Oxidative DNA /Chromosomal Damage
122	Nervous System	29	Specific Proteotoxicity
93	Renal	653	Oxidative Stress Effects
84	Reproductive & Pregnancy	17	Regenerative Proliferation
110	Respiratory	95	Vascular Mechanisms
171	Skin	240	Other AOP
504	Cancer		
	Categories not mutually exclusive		

Step 1 Natural Language processing combined literature search results and 900 seeds and grouped the tiles and abstracts based on similarity; (Categories were not mutually exclusive).

We decided to use the AOP framework to: organize and synthesize the refences & examine mechanistic commonalities among diseases and response modifiers, including susceptibility factors; help inform quantitative data. Conceptual models built and arsenic data overlay on top.



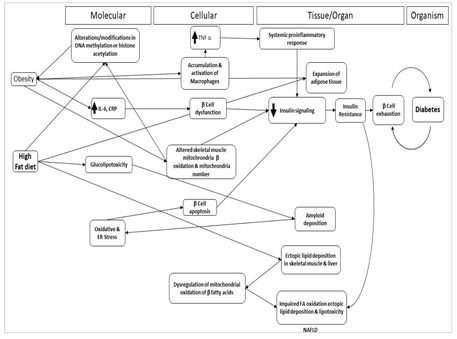


Figure 1: Postulated adverse outcome network for idiopathic diabetes in humans

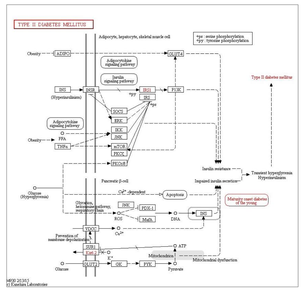


Figure 2: Kyoto Encyclopedia of Genes and Genomes (KEGG) database for diabetes mellitus in humans

Lessons Learned

- We performed a targeted literature search for arsenic MOA and used clustering to identify and tag AOs using studies from the previous IRIS arsenic assessments as seeds. We took the information under the diabetes tag and overlaid the information onto the AOP for idiopathic diabetes disease. This allowed us to identify key events in the progression of iAs-induced diabetes. **(Figure 3; Table 1).**
- The clustering approach was useful in identifying mechanistic relevant literature from this very large database.
- The AOP approach was useful in organizing references for evaluating the MOA.
- The MOA evaluation provided additional support by identifying arsenic-specific mechanisms and risk modifier likely to increase risk of diabetes,
- The challenges presented in this analysis included the doses required to observe diabetogenic effects in rodent models were great and not physiologically relevant to humans.
- **This mode of action analysis was conducted for diabetes but ultimately determined to be insufficient to dictate dose-response shape, especially given the availability of epidemiological data that could be used for dose-response analysis. Therefore additional analyses were not pursued.**

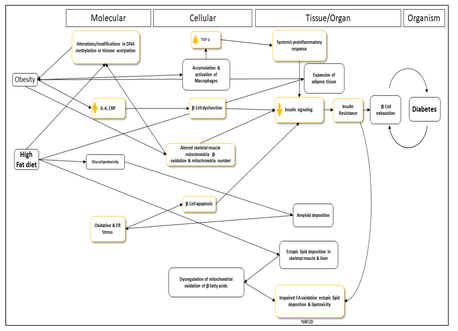


Figure 3: Postulated adverse outcome network for arsenic-induced diabetes in humans.

Key Event	Evidence	References
Molecular		
Alterations/modifications in DNA methylation or histone acetylation		Bailey KA et al, 2013 Hong et al., 2012, Xie Y et al., 2007
Increases in IL-6, CRP	Increased IL-6 transcription Increased CRP	Chakraborty et al 2012 Druwe et al, 2012 Wu et al 2003
Oxidative and ER Stress	ROS production in primary pancreatic β cells and hepatocytes derived from swiss albino mice; Increased pancreatic GSH, oxidative damage	Chakraborty et al 2012 Izquierdo-Vega et al, 2004
Cellular		
Increased TNF alpha	Increased TNF alpha transcription	Chakraborty et al, 2012 Yu et al, 2002
Altered skeletal muscle β oxidation and mitochondrial number		
Beta cell dysfunction	Decreased insulin secretion via decrease proteolysis of SNAP-25 Impaired insulin secretion & insulin mRNA expression	Diaz-Señor et al, 2008 Diaz-Señor et al, 2006
Beta cell apoptosis	Decreased β cell viability in response to 10 μM NaAsO3 in vivo (rat, oral)	Diaz-Señor et al, 2006
Tissue/Organ		
Systemic proinflammatory response	Elevated expression of IL-8 and TGF-β Increased miRNAs involved in diabetes and inflammation	Escobar-Garcia et al, 2012 Rager et al, 2014
Decreased insulin signaling	Inhibition of glucose-stimulated insulin secretion	Douillet et al, 2013
Insulin resistance	Inhibition of insulin-activated signal transduction in adipocytes Impaired glucose tolerance	Douillet et al, 2013 Wauson et al, 2002 Paul et al, 2008
Impaired fatty acid oxidation, ectopic lipid deposition and lipotoxicity	Increased pancreatic LPO	Izquierdo-Vega et al, 2004

Table 1: Representative evidence supporting adverse outcome pathway (AOP) network for arsenic-induced diabetes (Figure 3)

Step 2: Establishing the Disease-Base AOP:

- In response to NRC’s tiering of health outcomes, MOA analyses was conducted for diabetes
- To delineate a postulated mode of action for arsenic-induced diabetes, the molecular basis for idiopathic diabetes disease, irrespective of a specific chemical insult, was first established.
- The AOP framework (Villenueve et al., 2014) was used to organize and identify important key events and data gaps in the arsenic-induced diabetes MOA.
- To identify the key events (KE) leading to the adverse outcome (AO), we performed a literature search in PubMed and identified peer reviewed medical reviews of idiopathic diabetes disease. We screened the results and included publications that described mechanisms and or molecular events in the onset of idiopathic diabetes mellitus disease. We assembled the AOP for idiopathic diabetes by binning the results into key events in the disease process. **(Figure 1)**
- The adverse outcome network was compared to the KEGG (Kyoto Encyclopedia of Genes and Genomes) database for diabetes mellitus in humans to ensure concordance **(see Figure 2).**

Step 3 Identifying Arsenic-specific Modification in the Diabetes Network

- After establishing a general disease-based network for diabetes, we performed a targeted literature search for arsenic MOA and used clustering to identify and tag AOs using studies from the previous IRIS arsenic assessments as seeds.
- We took the information under the diabetes tag and overlaid the information onto the AOP for idiopathic diabetes disease. This allowed us to identify key events in the progression of iAs-induced diabetes.

The findings and conclusions in this presentation have not been formally disseminated by the U.S. EPA and should not be construed to represent any agency determination or policy.

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

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable-based ink.

The Key Characteristics of Carcinogens as an Organizing Principle for Mechanistic Evidence: Ethylene Oxide as a Case Study

Jason M. Fritz[1], Nagu Keshava[2], Suryanarayana V. Vulimiri[2], Catherine Gibbons[3]

U.S. EPA; [1]Office of Enforcement and Compliance Assurance/Office of Criminal Enforcement, Forensics and Training/National Enforcement Investigations Center; [2]Office of Research and Development/National Center for Environmental Assessment, Washington Division; 3Office of Research and Development/National Center for Environmental Assessment, Integrated Risk Information Systems (IRIS) Division

Jason M. Fritz, Ph.D. | Fritz.Jason@epa.gov | 303-462-9091

Structured Overview

Background – Ethylene Oxide (EtO) Assessment

In 2016, the U.S. Environmental Protection Agency's Integrated Risk Information System Program finalized a cancer assessment of ethylene oxide (EtO), characterizing it as "carcinogenic to humans" following inhalation exposure. EtO induces lymphoid and breast cancers in both humans and rodents, as well as other tumors in rats and mice (U.S. EPA, 2016).

Role of Mechanistic Data

While strong epidemiological evidence was instrumental in the human health hazard characterization process, evaluation of the animal and mechanistic data was also critically important (Jinot et al., 2018). Core concepts from the key characteristics of carcinogens (KCCs) (Smith et al., 2016), a pragmatic means of categorizing and evaluating the weight of evidence for mechanisms of carcinogenesis, were adopted in the organization of the mechanistic data summary sections supporting the mode of action analysis.

Evolving Approach to Evaluating Mechanistic Data

In subsequent work, the mechanistic evidence identified in the comprehensive literature search included in the IRIS assessment has been further reviewed and organized in a systematic manner using the KCCs as an organizing principle; this information was reviewed using a weight of evidence approach and integrated into adverse outcome pathways.

Conclusions Based Upon Key Characteristics of Carcinogens (KCCs)

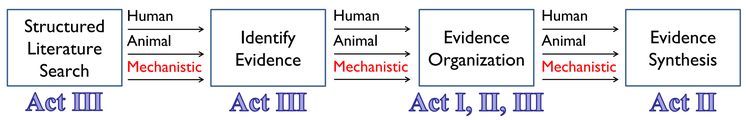
Strong and consistent evidence indicates that EtO is both electrophilic and mutagenic, representing two of the 10 KCCs; conversely, evidence for oxidative stress, another KCC, was neither strong nor consistent. Evidence of coherence in genetic or genomic damage in similar tissues across rodents and humans provides further support, linking relevant associations across data streams. The evaluation and discussion of cancer mechanisms was facilitated by using the KCCs as a central organizing principle to evaluate mechanistic data.

Challenges and Future Directions

One significant challenge in the systematic evaluation of cancer mechanisms was a paucity of mechanistic data identified from the EtO assessment literature search to evaluate 7/10 of the KCCs; specific supplemental literature searches using a digital content management system (i.e., Health Assessment Workplace Collaborative, or HAWC) have since been performed to locate published information pertinent to each KCC, based upon the methods recently described by Guyton et al. (2018). KCCs and Evidence Profile Tables (EPTs) are used in the structured evaluation of mechanistic data relevant to cancer hazard assessment, along with the database assembled and published in the EtO Assessment (U.S. EPA 2016), and described subsequently by Jinot et al. (2018).

Evolving Mechanistic Data Evaluation – in Three Acts

Hazard Identification Process (Adapted from NRC 2014, Review of EPA's IRIS Process)



Act I – US EPA IRIS Assessment of EtO (U.S. EPA, 2016)

- Mechanistic evidence sections organized by KCCs
- Mechanistic evidence separately synthesized according to 2 mode of actions (MOAs)

Act II – Discussion of Cancer Mechanisms (Jinot et al., 2018)

- Mechanistic evidence discussion utilized structured tables for high-level summary of primary observations by data stream and key event
- Mechanistic evidence synthesis of 2 MOAs followed structured summary

Act III – KCCs to Facilitate Structured Evaluation (*ongoing*…)

- Structured literature searches designed specifically to capture KCCs
- Evidence organization and summary of observations facilitated by EPTs

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

Act I – US EPA IRIS EtO Assessment

Relevant KCC	Possible Key Event
1) Is Electrophilic or Can Be Metabolically Activated	Protein adducts
	DNA adducts
2) Is Genotoxic	Point mutations in reporter genes or surrogate markers
	Point mutations in proto-oncogenes or tumor suppressor genes
	Chromosomal effects
5) Induces Oxidative Stress	ROS or lipid peroxidation products
	DNA oxidation
	Glutathione species levels

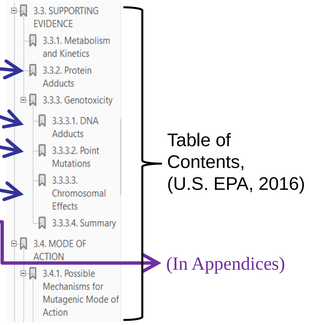


Figure 1. Table of contents from the IRIS EtO assessment (U.S. EPA, 2016), illustrating mechanistic data summary and organization by KCC, supporting MOA analysis.

Act II – Discussion of Cancer Mechanisms

Directly Mutagenic MOA

Relevant KCC	Possible Key Event	Weight-of-Evidence: Exposed Humans	Weight-of-Evidence: Experimental Systems (*In Vivo*)	Evidence Integration and Conclusions
1) Is Electrophilic or Can Be Metabolically Activated	**DNA adducts**	**N7-HEG** adducts were non-significantly elevated in white blood cells in two studies.	**N7-HEG** adducts increased in multiple rat and mouse tissues following repeated exposures to ≥ 3 ppm. **O^6-HEG** and **N3-HEA** adducts were also detected in rats exposed to 300 ppm; **O^6-HEG** and **N1-** and **N^6-HEA** adducts increased in an exposure-related manner in mice exposed to 100 or 200 ppm.	EtO is a direct-acting agent capable of alkylating DNA, proteins and non-protein sulfhydryls (e.g. GSH). **Strong and consistent evidence** demonstrating protein adduction in humans and rodents, as well as DNA adduction in rodents, along with **weak but not inconsistent evidence** from the limited human database of DNA adduct studies.
	Protein adducts	Exposure-response relationships have been reported with **hemoglobin adducts** in several studies; such adducts can be used as biomarkers of recent human exposure to EtO.	**Hemoglobin adducts** are linearly correlated with exposures up to at least 33 ppm in rats and mice; **adduction increased** at exposures ≥ 100 ppm, consistent with decreased GSH availability for detoxification. EtO-conjugated glutathione (**HESG**) increased in an exposure-related manner in mouse lungs following subchronic exposures ≤ 200 ppm.	
2) Is Genotoxic	**Genetic mutations**	***HPRT*** mutant frequency was significantly increased in peripheral blood leukocytes in one study following high exposures.	***Hprt*** or ***Lacl*** mutation incidences increased in several tissues in rats and mice, and dominant lethal effects were observed in germ cells.	**Strong and consistent evidence** supporting the induction of point mutations and chromosomal effects in rodent tissues, and the induction of chromosomal effects in humans. Most tumor mutations in genes regulating repair, proliferation, or survival in EtO-exposed mice occurred at purine nucleotides, consistent with evidence of EtO DNA adduction.
	Mutations in DNA repair, proliferation, or survival genes	*No evidence available*	***Kras*** mutation incidence was higher and the mutational spectra differed in lung and Harderian gland tumors of EtO-exposed mice, compared with tumors from control mice. The ***Trp53*** and ***Hras*** mutational spectra were likewise different in mammary gland carcinomas from EtO-exposed mice; these genes were frequently mutated together, while ***Trp53*** mutations and protein expression were induced in an exposure-dependent manner.	
	Chromosomal effects	**Chromosome aberrations** and **sister chromatid exchanges** were elevated in peripheral blood leukocytes from populations with prolonged and/or intense exposures.	**Sister chromatid exchanges** were induced in lymphohematopoietic tissues in non-human primates, rats, and rabbits; **chromosome aberration** incidence increased in similar tissues in mice, as well as in germ cells.	

Oxidative Stress MOA

Relevant KCC	Possible Key Event	Weight-of-evidence: Exposed Humans	Weight-of-evidence: Experimental Systems (*In Vivo*)	Evidence Integration and Conclusions
5) Induces Oxidative Stress	**↑ ROS, lipids oxidation or lipid-dNTP products**	*No evidence available*	**Croton-dG** increased in mouse lungs in an exposure-related manner with subchronic exposures ≤ 200 ppm; other lipid-nucleotide adduct species evaluated were unaffected.	**No evidence** supports DNA or GSH oxidation following subchronic EtO exposure. The decreased GSH and GSSG levels likely resulted from EtO-GSH conjugation, and not oxidation. No direct measures of ROS were reported, and **limited evidence** supports increased levels of a single oxidized lipid-DNA adduct (CrotondG).
	↑ 8-OHdG levels or direct DNA oxidation	In human lung epithelial cells, keratinocytes and PBLs exposed *in vitro*, **oxidative DNA damage** (Fpg-dependent comet assay) was not increased.	**8-OHdG** levels did not increase in mouse lungs following subchronic exposures ≤ 200 ppm.	
	↑ GSH oxidation (↓GSH/GSSG)	*No evidence available*	Both **GSH** and **GSSG** decreased in the lungs of mice an exposure-related manner following subchronic exposures ≤ 200 ppm; however, the **GSH:GSSG ratio** and **total glutathione content** were not affected.	

Table 1. Summarized support for mechanistic key events from the IRIS EtO assessment (Adapted from U.S. EPA, 2016 and Jinot et al., 2018), organized by KCC, key event, MOA and evidence stream.

Act III – KCCs in Lit. Searching and Evidence Identification

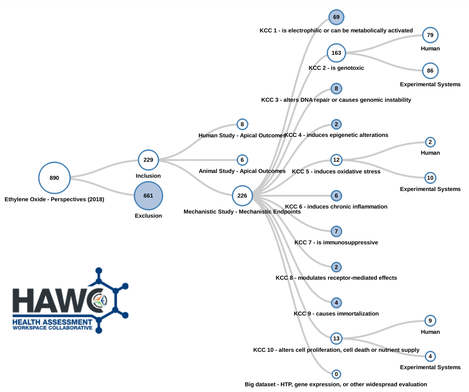


Figure 2. LitFlow tagtree from HAWC. Independent mechanistic literature search strings were constructed (Guyton et al., 2018), for KCCs1-3, KCC4, KCC5, KCCs6-7, KCC8, and KCCs9-10. Circles indicate number of studies categorized following title/abstract screen; a significant number of identified studies were excluded for lack of relevance, and there was a paucity of data relevant to many KCCs.

Act III – KCCs/AOPs/MOAs in Evidence Organization

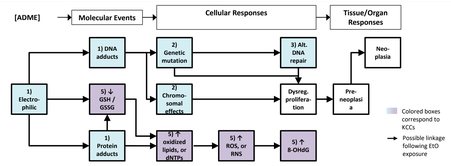


Figure 3. Plausible pathways relating key events (including relation to KCCs) leading from exposure to carcinogenesis, relevant to genotoxicity and oxidative stress activity.

Outcomes (KCCs and key events)		Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings	Within stream evidence judgement	Overall conclusion
Directly Mutagenic MOA							⊕⊕⊕ STRONG – Mutagenic MOA Based upon: • Strong and consistent evidence of direct alkylating activity, genotoxicity and mutagenicity in humans and a variety of experimental systems • Evidence of temporality and an exposure-response gradient between • Widespread coherence across outcomes and species • Biological plausibility between supported key events and carcinogenesis Relevance to humans: • Strong evidence that the key precursor events are anticipated to occur in humans and progress to tumors, including human evidence of chromosome damage, such as chromosomal aberrations, SCEs, and micronuclei.
1) Electrophilic	DNA adducts	*Many* – Table 3-6; e.g. Zhang et al., 2015; Marsden et al., 2009, 2007; Young et al., 2007; Ruoyn et al., 2005, etc.	• Exposure-response gradient • Coherence (species) • Pro-mutagenic	• Few studies available (human) • Effect size (human)	⊕⊕⊕ STRONG, consistent and coherent in rodent studies	⊕⊕⊕ STRONG and consistent evidence of protein adduction in humans and rodents, and DNA adduction in rodents; while human evidence of DNA adducts are limited.	
	Protein adducts	*Many* – Section 3.3.2; e.g. IARC 2008; Walker et al, 1993, 1992, etc.	• Exposure-response gradient • Coherence (species) • Coherence (molecular targets)		⊕⊕⊕ STRONG, consistent and coherent		
2) Genotoxic	Genetic mutations	*Many* – Table 3-6; e.g Major et al., 2001; Walker et al., 2000; Tates et al., 1999, etc.	• Coherence (species)	• Few studies available (human) • Effect size (human)	⊕⊕⊕ STRONG, consistent and coherent in rodent studies	⊕⊕⊕ STRONG and consistent evidence for induction of point mutations and chromosomal effects in rodent tissues, consistent with strong and consistent evidence of chromosomal effects in humans.	
	Mutations in DNA repair, proliferation, or survival genes	Hong et al., 2007; Houle et a., 2006 (from NTP, 1987); Parsons et al., 2013	• Coherence (tissues) • Mutational fingerprint • Biological plausibility (*ras*, *Trp53* genes)	• No data (human) • Coherence (temporal)	⊕⊕○ MODERATE, coherence in multiple tissues from rodent studies, and plausible fingerprint		
	Chromosomal effects	*Many* – Tables 3-7 and 3-8; e.g. Donner et al., 2010; Major et al., 1996, etc.	• Coherence (species, tissues) • Exposure-response gradient	• Few studies available for some endpoints (e.g. micronuclei incidence)	⊕⊕⊕ STRONG, consistent and coherent		
Oxidative Stress MOA							
5) Oxidative stress	ROS, lipid oxidation	Zhang et al., 2015		• Few studies available • Effect size	○○○ INADEQUATE	⊕○○ SLIGHT evidence for increased levels of an oxidized lipid-DNA adduct, and no evidence of direct DNA or GSH oxidation, or ROS generation.	
	DNA oxidation	Zhang et al., 2015; Nagy et al., 2013	• Exposure-response gradient in CrotondG (mice)	• Few studies available • Effect size (human) • Coherence (within, between species)	⊕○○ SLIGHT Increases in one, but not other markers, in one study/sex/species		
	GSH oxidation	Zhang et al., 2015; Brown et al., 1998		• Few studies available • Effect size • Changes attributed to alkylation	○○○ INADEQUATE		

Table 2. Preliminary EtO cancer mechanism EPT, summarizing conclusions on the strength of available evidence relevant to KCCs and key events, for both mutagenic and oxidative stress MOAs (note that glutathione can be oxidized, alkylated, and/or depleted via metabolism).

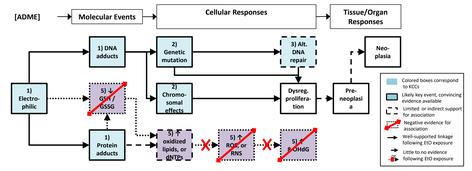


Figure 4. Illustration of observations summarized in EPT above, overlaid on relevant mechanistic pathways, as part of evidentiary support for MOA analyses and cancer synthesis.

Observations and Future Directions

- KCCs provide a topical basis for focused, transparent and reproducible literature searches
- Summary tables and EPTs are useful as structural tools for organizing large volumes of mechanistic data and highlighting results of evaluation by subject-matter experts in what could be a transparent, reproducible, and understandable manner
- Data relevance, organization, weight of evidence, and pathway impact determinations are currently highly expert judgement-driven, and subjective; more work is needed to develop objective, pragmatic, and reproducible *a priori* criteria

Selected References

- Guyton KZ, Rusyn I, Chiu WA, Corpet DE, van den Berg M, et al. 2018. Application of the key characteristics of carcinogens in cancer hazard identification. Carcinogenesis. Apr 5;39(4):614-622. DOI: 10.1093/carcin/bgy031
- Jinot J, Fritz JM, Vulimiri SV & Keshava N. 2018. Carcinogenicity of ethylene oxide: key findings and scientific issues, Toxicol Mech Methods. 28:5, 386-396, DOI: 10.1080/15376516.2017.1414343
- National Research Council. 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press. DOI: 10.17226/18764
- Smith MT, Guyton KZ, Gibbons CF, Fritz JM, Portier CJ, et al. 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ Health Perspect. 124:713–721
- U.S. EPA. 2005. Guidelines for carcinogen risk assessment. EPA/630/P-03/001F. Washington, DC, U.S. Environmental Protection Agency, Risk Assessment Forum
- U.S. EPA. 2016. Evaluation of the inhalation carcinogenicity of ethylene oxide (CASRN 75-21-8): In support of summary information on the Integrated Risk Information System (IRIS). EPA/635/R-16/350Fa. Washington, DC, U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment

The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable-based ink.

Using Mechanistic Information to Support Evidence Integration and Synthesis: A Case Study with Formaldehyde

AUTHORS

Robinan Gentry, Tracy Greene, Allison Franzen, Ramboll; Chad M. Thompson, ToxStrategies

Formaldehyde is one of the most comprehensively studied chemicals, with over 30 years of research focused on understanding the development of cancer following inhalation. The extensive database has been reviewed by multiple authoritative bodies, focusing on upper respiratory tract cancer and leukemia to characterize the potential for carcinogenicity. Modes of action (MOA) have been proposed for each endpoint, with mechanistic data developed to help understand if these MOAs are operable in humans. When consistent with the problem formulation, proposed MOAs help to narrow the focus of literature search and review of mechanistic data. Formaldehyde is also produced endogenously, and therefore requires mechanistic considerations to understand the potential for health effects that can be attributed to exogenous exposure. An approach is presented to organize and integrate mechanistic data to evaluate the key events (KE) postulated for both upper respiratory tract cancer and leukemia. The approach utilizes pharmacokinetic data and existing MOA frameworks to assess evidence around the key events in proposed MOAs, considering modified Bradford Hill criteria, as well as evaluation of the quality and consistency of the mechanistic studies. Output from available tools, (eg biologically based dose-response models, are also considered in evaluating proposed cancer MOAs for formaldehyde). This case study demonstrates the utility of a mechanism driven approach for a chemical with extensive data to support an MOA for one endpoint and limited/no data to support another endpoint. These two contrasting examples show the role and value of mechanistic data in drawing conclusions from evidence integration and synthesis; reducing uncertainty and improving confidence in the hazard assessment conclusions.

INTRODUCTION

- Epidemiological data for both upper respiratory cancers and leukemia are weak and inconsistent
- Animal evidence provides support for upper respiratory tract cancers; however, no animal evidence is available to support leukemia
- Multiple authoritative reviews have been conducted, some finalized and others in draft form, highlighting the mechanistic data that have been relied upon for regulatory decision making for both endpoints; however, different conclusions have been drawn based on the data available at the time of the review and whether or not mechanistic data were fully integrated
- Multiple MOAs have been proposed for each endpoint
- Mechanistic data demonstrate some proposed MOAs are biologically implausible
- The World Health Organization (WHO) International Programme on Chemical Safety (IPCS) framework provides a method for integration of mechanistic data and the application of modified Bradford Hill criteria to evaluate consistency, concordance of dose-response relationships across key events, coherence of the database and biological plausibility

METHODS

Formaldehyde case study

- MOAs have been proposed for both upper respiratory tract cancer and leukemia
  - Proposed MOAs evaluated were taken from regulatory assessments and open literature
- The IPCS framework was selected since it affords an approach to systematically review studies that have been relied upon in regulatory decision making and integration of multiple streams of evidence to draw conclusions regarding biological plausibility of health effects (Figure 1)
- Study quality considerations
  - Methods that allow differentiation between endogenous and exogenous formaldehyde (critical as formaldehyde is endogenously present) were deemed more reliable than those that could not differentiate between the two
  - *In vivo* studies in uncompromised animals were deemed more reliable than those performed with compromised systems (eg rats receiving radiation to destroy bone marrow; lung transplant models in mice with thrombocytopenia)
  - Gaseous formaldehyde studies were deemed more reliable than formalin based
- Most of the proposed MOAs for each cancer type have some common key events; as such, the common key events were grouped for evaluation (Table 1)

Figure 1. Integration of multiple streams of evidence to draw conclusions regarding biological plausibility of health effects

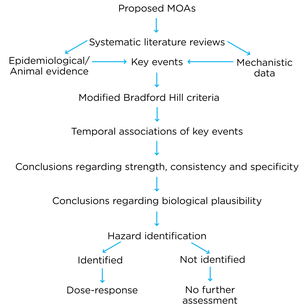


Table 1. Key events associated with proposed MOAs

Upper respiratory tract cancer	Leukemia
Increase of formaldehyde in nasal tissue	Conversion of formaldehyde to methanediol for transport through portal of entry
Cytotoxicity	Formaldehyde systemic circulation and migration through the blood
Chronic increase in cell proliferation	Macromolecular binding to hematopoietic stem cells and progenitor blood cells in the blood
Spontaneous mutations	Formaldehyde induced genotoxicity in the hematopoietic stem cells and progenitor blood cells derived from bone marrow or in the bone marrow
Nasal tumor formation	Damaged stem cell release into systemic circulation
	Incorporation of initiated stem cell into the bone marrow
	Initiation of leukemia

RESULTS

Upper respiratory tract cancer

In 2006, the MOA for formaldehyde-induced nasal tumors in rodents and the human relevance of these tumors was formally analyzed using IPCS frameworks (top row of Table 2a; McGregor et al 2006).

- Note that McGregor et al characterized KE3 "genotoxicity" as an increase in DNA-protein crosslinks (DPX)-mediated replication errors occurring under high cell proliferation conditions; McGregor et al characterized this as a hypothesis
- McGregor et al characterized the MOA as supportive of a non-linear MOA and thus non-linear risk assessment approaches

Research conducted after McGregor et al (2006) have better characterized all key events in the proposed MOA and better explain why there is a highly non-linear dose response for rat nasal tumors and support conclusions about non-linear risk assessment approaches (bottom row of Table 2a)

- Analytical techniques that are capable of distinguishing between endogenous and exogenous DNA adducts has demonstrated that exogenous adducts only add to background adduct levels when concentrations exceed 1 ppm and the adducts do not demonstrate dose-response concordance with the tumor response
- Despite exposure to concentrations of formaldehyde that increase cell proliferation and induce DPX, neither micronuclei (MN) nor gene mutations have been observed in target nasal tissue under these favorable conditions
- A study in transgenic mice susceptible to carcinogenesis by genotoxicants and exposed to ≤ 15 ppm revealed no evidence of neoplastic lesions in the nasal cavity; however, there was evidence of non-neoplastic lesions, which demonstrates delivery of formaldehyde to the target tissue
- New unpublished data indicate the absence of detectable DPX in nasal tissue of rats exposed to ≤ 300 ppm for 28 days

Results from biologically based dose-response (BBDR) models (Conolly et al 2003, 2004), which integrated the available mechanistic data at that time, demonstrate a non-linear dose-response curve. They also provide additional support for sustained cytotoxicity being a significant contributor to the observed tumorigenicity following exogenous formaldehyde exposure and support DPX as a biomarker of exposure.

The integration of formaldehyde-specific dose-response and temporal data published through 2018 (Table 2a, b), in combination with the results of BBDR modeling, indicate that the totality of data support the biological plausibility of a cytotoxicity-regenerative cell proliferation MOA inconsistent with assumptions of low-dose linearity for respiratory tract tumors.

Table 2a. Key events relevant for MOA of upper respiratory tract cancer (adapted from McGregor et al 2006)

	Cytotoxicity	Cell proliferation	Genotoxicity[a]	Mutations	Tumors
Up to 2006	Kerns et al (1983) Monticello et al (1991) Monticello et al (1996)	Monticello et al (1991) Monticello et al (1996)	DPX in target tissue (although DPX is not genotoxicity *per se*)	No target tissue studies measuring mutation	Kerns et al (1983) Monticello et al (1996)
2006 and later	Andersen et al (2008) Andersen et al (2010) Example data:	Andersen et al (2008) Andersen et al (2010) Meng et al (2010) Speit et al (2011) Example data:	Meng et al (2010) Speit et al (2011) Example data:	Meng et al (2010) Morgan et al (2017) Example data:	Morgan et al (2017) Example data: No nasal tumors in p53+/- mice No neoplastic lesions in p53+/- mice

[a] McGregor et al includes DPX as evidence for this key event. Since published data indicate DPX at all levels of exposure, we do not include this in the bottom row of the table. Instead, we present data that focus on genotoxicity endpoints (mutations, MN).

Table 2b. Formaldehyde-specific dose-response and temporal concordance data for upper respiratory tract cancer

Dose (ppm)	Cytotoxicity	Cell proliferation	DPX/DNA adducts	Genotoxicity (including mutations)	Tumors
≤0.03			-[a]		
0.3			+/-		
0.7	-	-	+	-	-
1	-[a]	-[a]	+[a]	-	-[a]
2	+	+/-	+	-	-
6	+	+	+	-	+/-
10	+	+	+	-	+
15	+	+	+	-	+

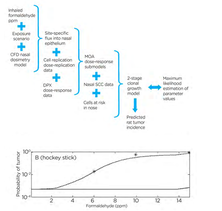
[a] Effect observed; - Effect not observed experimentally; +/- Effect variably observed (for tumors incidence was not significant by one-sided Fisher's Exact Test [FET]); [a] inferred based on lower dose; blank, effect not studied at specific dose

Dark grey shading indicates region where effect is negative or where effect has not been studied empirically

[a] Unpublished data presented at EUROTOX 2018

Results of BBDR modeling

Biologically Motivated Computational Modeling of Formaldehyde Carcinogenicity in the F344 Rat



Model indicates little/no contribution from mutagenic component of DPX

Table 3a. Formaldehyde-specific dose-response and temporal concordance data for leukemia

Dose (ppm)	Formaldehyde reaches systemic circulation[a]	Macromolecular binding to bone marrow or blood	*In vivo* genotoxicity[a] in HSCs/HPCs or primitive pluripotent cells	Damaged stem cell released into circulation	Incorporation of initiated cell back into bone marrow
0.2	No data	No data	- [c]	No data	No data
0.3	-	No data	- [c]	No data	No data
0.4	No data	- [b] - [c]	+/- [b] - [c]	No data	No data
0.5	-	- [c]	- [b] - [c]	No data	No data
0.6	No data	No data	- [c]	No data	No data
0.7	-	- [b]	- [c]	No data	No data
0.8	No data	+ [b] - [c]	+ [b] - [c]	No data	No data
1	-	- [c]	- [c]	No data	No data
2	-	- [b] - [c]	- [c]	No data	No data
2.4	No data	+ [b] + [c]	+ [b] + [c]	No data	No data
3	No data	No data	- [b]	No data	No data
5	No data	No data	- [c]	No data	No data
6	-	- [b] - [c]	- [c]	No data	No data
9	-	- [b]	No data	No data	No data
10	-	- [b] - [c]	- [c]	No data	No data
13.5	No data	No data	- [c]	No data	No data
14	No data	- [c]	No data	No data	No data
15	-	- [b] - [c]	- [b] - [c]	No data	No data

[a] Effect observed; - Effect not observed experimentally; +/- Effect variably observed

[a] Genotoxicity for supporting evidence is defined by the study authors to be DNA protein crosslinks formation or biomarkers of oxidative stress (ie reactive oxygen species (ROS), malondialdehyde (MDA), measured glutathione (GSH) and cytochrome P450 1A1 and glutathione s-transferase theta 1 expression)

[b] Bone marrow data; [c] Blood data

[a] Exogenous formaldehyde was reported to not be present in multiple tissues, including the lung, liver, bone marrow, blood, spleen, thymus

- Computational fluid dynamic (CFD) models estimate formaldehyde flux (dosimetry) to regions of nasal tissue
- DPX formation was estimated as a function of flux (and assumed to be pro-mutagenic)
- Cytolethality-regenerative cellular proliferation (CRCP) was estimated as a function of flux
  - Assumed to be a proxy for cytotoxicity
  - Cell division was assigned probability of mutation
  - CRCP was independent of DPX (ie related to CFD-predicted flux)
- Model included time-to-tumor component
- Modeled predictions of tumor formation indicate minimal role for DPX (as a pro-mutagenic event)
- Model predictions are consistent with data published after the BBDR model and MOA analysis were published (Table 2a)

Leukemia

- Much scientific debate surrounds conclusions that formaldehyde exposure is associated with leukemia; for example, a review of the draft USEPA (2010) IRIS conducted by the NAS (2011) concluded that associations between formaldehyde exposure and leukemias or related hematopoietic cancers were not supported, appeared to be subjective in nature, and that no clear scientific framework had been applied in reaching that conclusion
- As with upper respiratory tract tumors, the IPCS framework provides a method for integration and evaluation of the biological plausibility of a causal association
- In contrast to the upper respiratory tract tumor data, the application of the framework clearly highlighted the lack of formaldehyde-specific dose-response information for multiple (common or unique) key events in any of the proposed MOAs (Table 3a, b)

Table 3b. Integration of evidence for biological plausibility for proposed MOA

Reference	Formaldehyde reaches systemic circulation	Macromolecular binding to bone marrow or blood	*In vivo* genotoxicity[a] in HSCs/ HPCs or primitive pluripotent cells	Damaged stem cell released into circulation	Incorporation of initiated cell back into bone marrow
Evidence to support proposed MOA	No data	Ye et al (2013)	Wei et al (2016) Ye et al (2013) Zhang et al (2013)	No data	No data
Evidence inconsistent with proposed MOA	Edrissi et al (2013) Kleinnijenhuis et al (2013) Lai et al (2016) Lenq et al (2018) Lu et al (2010, 2011) Moeller et al (2011) Speit et al (2009) Swenberg et al (2011) Yu et al (2015)	Casanova and Heck (1987) Casanova et al (1984) Casanova et al (1988) Edrissi et al (2013) Heck et al (1985, 1989) Heck and Casanova (2004) Kleinnijenhuis et al (2013) Lai et al (2016) Lu et al (2010, 2011) Moeller et al (2011) Speit et al (2009) Swenberg et al (2011) Yu et al (2015)	Dallas et al (1992) Fleig et al (1982) Kligerman (1984) Speit et al (2009) Vieges et al (2010) Zeller et al (2011)	No data	No data

[a] Genotoxicity for supporting evidence is defined by the study authors to be DNA protein crosslinks formation or biomarkers of oxidative stress (ie reactive oxygen species (ROS), malondialdehyde (MDA), measured glutathione (GSH), and cytochrome P450 1A1 and glutathione s-transferase theta 1 expression)

Blue citations indicate studies that report a lack of the event observed experimentally

Green citations indicate studies that report the event observed experimentally

- Evidence of dose-response was limited to studies conducted in mice in one laboratory or studies conducted in animals with compromised systems that were not specifically exposed to formaldehyde. With the absence of supporting data for multiple key events, in conjunction with the absence of evidence of leukemia or hematotoxicity in animals exposed to formaldehyde for chronic durations, the integration of evidence and application of the modified Bradford Hill criteria indicated a lack of biological plausibility of any of the proposed MOAs for leukemia

EVIDENCE INTEGRATION TABLES FOR EACH ENDPOINT

Upper respiratory tract cancer

Table 4. Evidence integration for a cytotoxicity-regenerative cell proliferation MOA

Modified Bradford Hill consideration	Supporting evidence	Potentially inconsistent evidence
Dose-response concordance	Cytotoxicity and cell proliferation precede tumor formation. Dose and temporal concordance for all key events.	None.
Consistency, specificity	High degree of consistency across studies with regard to doses that induce cytotoxicity, increase unit length labeling index (ULLI) and increase tumor incidence.	None.
Biological plausibility	Consistency with known risk of increased cell replications.	None.

Leukemia

Table 5. Evidence for proposed MOA for initiation of leukemia by direct damage to hematopoietic stem cells in the bone marrow

Modified Bradford Hill consideration	Supporting evidence	Potentially inconsistent evidence
Dose-response temporal concordance	Integration of the available evidence provides inadequate support of a dose-response relationship for any of the key events in the proposed MOA.	Multiple studies in multiple species provide evidence indicating macromolecular binding and genotoxicity in the bone marrow do not occur in animals or humans. Also, the maximum amount of formaldehyde inhaled at the WHO International Academy of Quality and Safety (IAQS) and available for systemic distribution is >1,000x less than endogenously produced. As such, there is no dose-response concordance between multiple key events proposed to be associated with leukemia and inhaled formaldehyde.
Consistency, specificity	Integration of the available evidence indicates a lack of consistency across studies.	High degree of consistency across studies to support the lack of evidence of macromolecular binding, cytotoxicity or genotoxicity occurring in the bone marrow.
Biological plausibility	Integration of the available evidence indicates a lack of supporting evidence for biological plausibility.	Studies in multiple species demonstrate that formaldehyde does not reach the bone marrow in significant quantities and the available data indicate that direct mutagenic effects of formaldehyde in the bone marrow do not occur *in vivo*.

CONCLUSIONS

- The IPCS framework provides a method for systematically and critically reviewing mechanistic data and integrating these data with other streams of evidence (epidemiological, animal) (Tables 4 and 5)
- The framework allows for consideration of study quality, as well as considerations that may be specific to the chemical of interest; in this case study, endogenous production of formaldehyde and whether methods can determine exogenous versus endogenous formaldehyde
- This case study also demonstrates that the framework will work in determining biological plausibility for endpoints for which data from multiple streams of evidence are available to provide supporting evidence for key events in a proposed MOA, as well as those for which the evidence is inconsistent

Table Builder: A content management system for carcinogenicity health assessments for the IARC Monographs and the NTP Report on Carcinogens

Andy Shapiro[1,*], Ruth Lunn[1], Gloria Jahnke[1], Pam Schwingl[2], Kate Guyton[3], Dana Loomis[3], Neela Guha[3,†]

[1] National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA
* Current affiliation: Infinia ML, Durham, NC, USA
[2] Integrated Laboratory Services, Durham, NC, USA
[3] International Agency for Research on Cancer (IARC) Monographs Programme, World Health Organization, Lyon, France
† Current affiliation: School of Community Health Sciences, University of Nevada, Reno, Nevada, USA

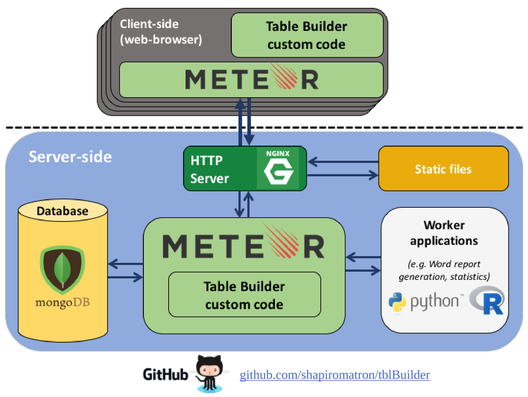
Abstract

The International Agency for Research on Cancer (IARC) Monographs Program and the NTP Report on Carcinogens (RoC) are tasked with evaluating evidence to determine whether agents, exposure scenarios, or mixtures pose a cancer hazard to humans. This process requires an extensive literature search and systematic review of the evidence. Evidence synthesis requires extracting and interpreting data in multiple domains, including 1) human exposure, 2) epidemiologic evidence, 3) animal evidence in test model systems, and 4) mechanistic evidence on key characteristics of carcinogenicity, such as genotoxicity. Standardized tables are necessary to synthesize and evaluate evidence in a transparent and systematic manner; they are included in final reports to summarize the strength of these findings.

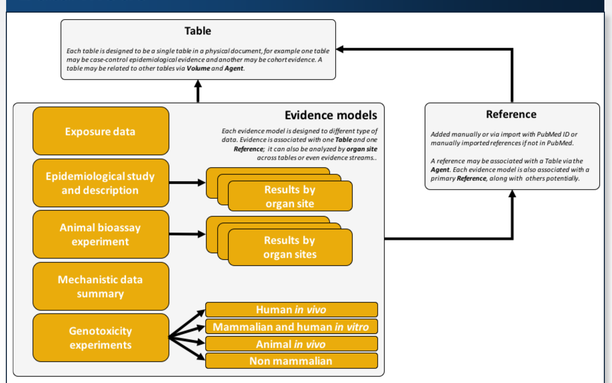
The Table Builder is a web-based content management system that was designed to capture data and facilitate analyses for human health assessments. This software allows collaborators to extract reported evidence, list potential covariates and confounders, and indicate study strengths and limitations. The data extraction fields are standardized for each evidence stream; this allows multiple collaborators to enter data in a consistent format in parallel. The software is reactive; whenever a user changes any data in the system, it is updated for all other users of the system in real-time (a feature that is indispensable during IARC Monograph meetings). Statistical analysis can be performed in the software (such as Cochrane Armitage trend test or pairwise tests for animal bioassay data). Data visualizations can be created (e.g. forest plots) and filtering of data by cancer site, which facilitate the interpretation and synthesis of data for report writing, especially when the number of extracted elements are large. Further, data can be managed in the software system and quality control (QA/QC) of data entry is integrated into the software. Finally, reports can be downloaded in Microsoft Word format, or data can be downloaded in Microsoft Excel.

The Table Builder software was designed using the Meteor Javascript web framework and uses a Mongo database, and is open-source and publicly available at https://github.com/shapiromatron/tblBuilder. To date, the Table Builder software has been used for 11 IARC and 4 NTP RoC Monographs (including monographs that are finalized and currently under development).

System architecture



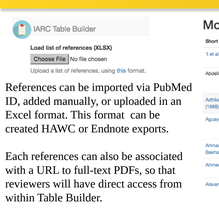
Data model



The Table builder data model has additional flexibility (and complexity) compared to more traditional SQL databases as it uses a document-object model. Each evidence element is in an evidence-specific collection; it is associated with a table and one primary reference. Tables have user-specific access/control permissions.

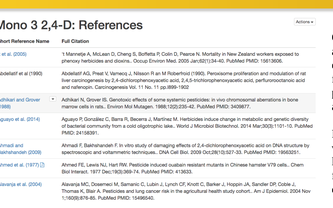
Application overview

REFERENCE UPLOAD AND TRACKING



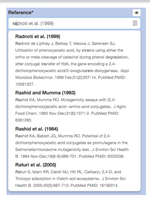
References can be imported via PubMed ID, added manually, or uploaded in an Excel format. This format can be created HAWC or Endnote exports.

Each references can also be associated with a URL to full-text PDFs, so that reviewers will have direct access from within Table Builder.



Once a reference is added, it automatically appears when entering data on data-entry forms. Typing a letter auto-populates the list of available references.

References are associated with **Agents**; the selection list only includes references for a single Agent to limit choices.



The IARC and NTP Table builder are two separate websites and databases, but they both share common application source-code.

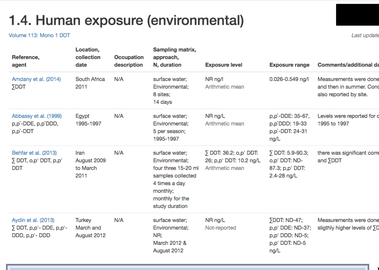
The software was created in the summer of 2014, and was first used on IARC 112 and the 14th Report on Carcinogens.

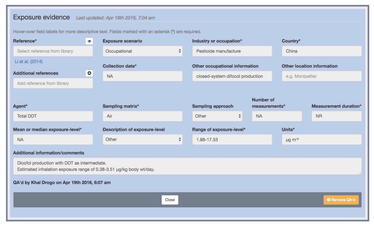
Collection	IARC table builder	NTP table builder
Users	238	23
Tables	277	61
References	8,755	387
Exposure evidence	509	146
Epidemiology descriptions	1,918	169
Epidemiology results	4,661	656
Epidemiology confounders	N/A	165
Animal evidence	442	195
Animal endpoints	1,195	522
Mechanistic evidence	100	13
Genotoxicity evidence	640	4

Data in table builder website for each collection, last updated 08/21/17.

DATA EXTRACTION

Exposure

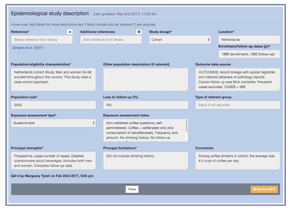




Exposure data can be added for each reference and may include a particular value or a range of values.

After adding data, staff can mark it QA'd, and it can no longer be edited.

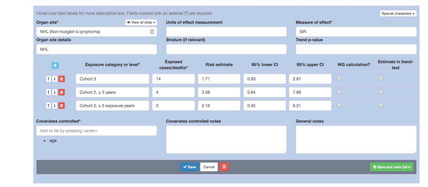
Epidemiology

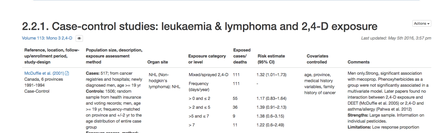


Epidemiology data is entered at the study description level to capture details on key study characteristics including study design, location, population details, exposure assessment methods, outcome source and methods, and overall study strengths and limitations.

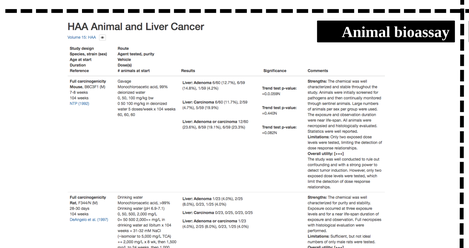
With each study description, results are entered which are associated with an organ site (from a controlled drop-down list); one or more groups can be entered, including the number of exposed cases, covariates, risk estimates and confidence intervals. Simple forest plots can be created with these data, but they can also be exported for more advanced analyses and graphical displays.

Using this nested structure makes it easier to add multiple results for the same epidemiological study; these descriptions can also be cloned on the same table and across tables for more rapid and consistent data entry.





Animal bioassay



Animal bioassay data is extracted similarly to epidemiological data; animal evidence captures the experimental protocol for a collection of related animals (e.g., male F344 rats), and animal evidence results captures tumor counts at an organ site.

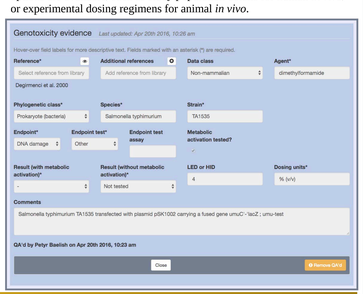
Genotoxicity

Reference	Test system	Endpoint test	Results	Results (with metabolic activation)	Agent, LED/HID dose	Comments
Chen, 2004 Human in vivo	blood, peripheral blood cells Workers having long-term DMF exposure history	DNA damage/ DNA strand breaks	(+)	NA	dimethylformamide	Comet assay, tail extent, olive tail moment and tail DNA/total DNA (%) significantly different from controls in workers with 6 – 7 years of exposure to DMF.
Xuan et al. (2006) Mammalian and human in vitro	Human Liver/hepatocytes	DNA damage/ DNA strand breaks	+	Not tested	Dimethylformamide, 1.58 mM [114 µg/mL]	Phosphorylated histone gamma H2AX was used as a biomarker of double strand breaks. Immunofluorescence assay
Xing et al. (2014) Animal in vivo	Rat Wistar Female Lymphocytes	DNA damage/ DNA strand breaks	+	NA	dimethylformamide, 50 mg/kg bw, p.o, 1x / day, 14 days	Comet assay, average tail length and tail moment significantly increased in all exposed groups ($p<0.05$ or $p<0.01$)
Midorikawa et al. (2000) Non-mammalian	Acellular systems HP-labelled DNA fragments/Cu(II)	DNA damage/ DNA strand breaks	+	Not tested	degraded dimethylformamide containg H_2O_2, 2 % (v/v)	In the presence of Cu(II) degraded DMF caused site specific damage at thymine and cytosine residues. Purified DMF enhanced the damage caused by H_2O_2.

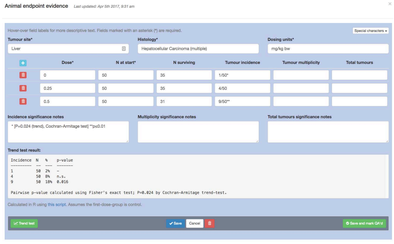
Example endpoint and endpoint tests (data-class specific):

- Chromosomal damage: aberrations, aneuploidy, sister chromatic exchange, etc.
- Cell transformation
- DNA damage: adducts, strand breaks, cross-links, unscheduled synthesis, oxidation, etc.
- DNA repair
- Mutation: forward, reverse, gene conversion, oncogene, tumor suppressor, etc.

This module enables data capture from diverse evidence streams concerning whether an agent is genotoxic, and exhibits related key characteristics of carcinogens as identified by Smith et al., 2016. Data are extracted from studies of four data classes: human *in vivo*, mammalian and human *in vitro*, animal *in vivo*, and non mammalian. Each data class has a specific set of relevant endpoints and required questions for data extraction; for example, in the example below, a species and strain can be specified for non-mammalian assays, and results can be reported both with and without metabolic activation. For other data classes, other options are available, such as study population details for human *in vivo*, or experimental dosing regimens for animal *in vivo*.



ADDITIONAL ANALYSIS



For some animal bioassay references, statistics across endpoint-groups may not be reported, or the approaches used may not be clear to the extractor.

Table Builder reports tumor pairwise significance using Fisher's Exact test and trend-significance using the Cochran Armitage test.

Results from the analysis are saved with the dataset; extractors can add results to the summary findings, if relevant. *(IARC table builder only)*

Colon & rectum

Reference	Alcohol consumption	BMI	age	physical activity	physical activity, BMI	red meat consumption
Reed et al. (2014) Germany 2000-2009 Cohort	no	no	yes	yes	-	no
Saldvirth et al. (2007) Sweden 1977-1981 Cohort	No	-	yes	-	yes	no
Yi'Qhai et al. (2003) 11 U.S. states Enrolled 1976, followed June 1998-May 1998 Cohort	yes	yes	yes	yes	-	yes
Bolton et al. (2017) Spain 2008-2013 Case-Control	yes	yes	yes	yes	-	yes
Aarnio et al. (2015) 11 U.S. states Enrolled 1976, followed June 1998-June 2008 Cohort	yes	yes	yes	yes	-	yes

When determining the strength of evidence across multiple epidemiology studies, it is important to evaluate the potential for confounding. The variables of concern explorer allows users to identify potential confounders, evaluate whether they are addressed in each study, and allows for easy comparisons across studies.

In the colon and rectum cancer example, yes and no refer to whether confounding can be reasonably "ruled out"; i.e., the findings cannot be entirely explained by the confounder. The rationale for this decision is captured in the variables of concern analysis for each study. *(NTP table builder only)*

Environmental exposure and prostate cancer

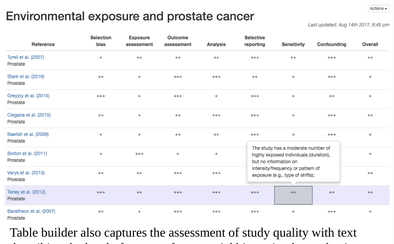


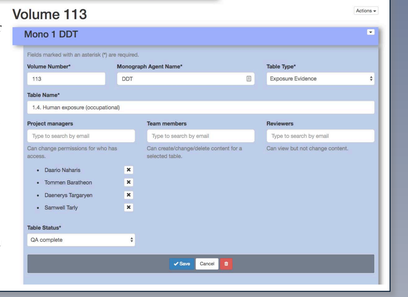
Table builder also captures the assessment of study quality with text describing the level of concern for potential biases (such as selection, information, and confounding) as well as an assessment of study sensitivity (the ability to detect a true effect). The example above presents this assessment in an interactive web view which shows the rationale for the assessment.

The data can also be exported into Microsoft Excel for heatmap creation or Microsoft Word tables. *(NTP table builder only)*

Volumes, agents, and table permissions

Multiple tables can be created with each Agent; each table contains data from a single evidence stream (e.g. exposure, epidemiology, animal bioassay).

Indicators show the status in the project lifecycle for each each table. This makes it easy to view a quick snapshot of an agent's evaluation.



Each table has its own set of rules for access-control.

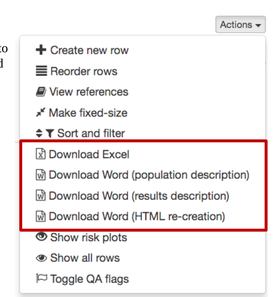
A table must have one or more project managers; they have full control of the project including the ability to modify access to mark data as QA/QC.

Team-members can extract data.

Reviewers have a read-only view of the data.

Tables in Table Builder do not have a public view; they are expected to be internal-only.

Scientist-friendly data exports

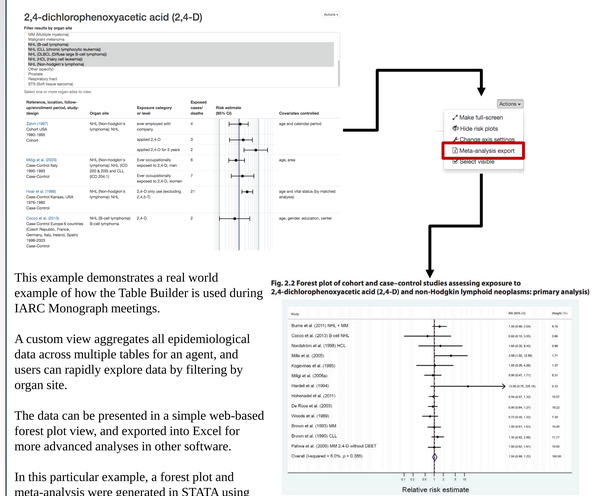


All data from Table Builder is designed to be exported into formats which are useful for both subsequent analysis and data presentation.

Each table has an Excel export, which is designed to capture all data in a long-format for import into other software for data-analysis or archival purposes.

Exports can also be created for Microsoft Word reports and are expected to be basis for final tables presented in main reports (such as the IARC Monographs) or appendices (such as the RoCs).

From data exploration to final publication



This example demonstrates a real world example of how the Table Builder is used during IARC Monograph meetings.

A custom view aggregates all epidemiological data across multiple tables for an agent, and users can rapidly explore data by filtering by organ site.

The data can be presented in a simple web-based forest plot view, and exported into Excel for more advanced analyses in other software.

In this particular example, a forest plot and meta-analysis were generated in STATA using data from Table Builder.

Fig. 2.2 Forest plot of cohort and case–control studies assessing exposure to 2,4-dichlorophenoxyacetic acid (2,4-D) and non-Hodgkin lymphoid neoplasms: primary analysis)

http://monographs.iarc.fr/ENG/Monographs/vol113/mono113-03.pdf

Contact information:

Andy Shapiro (shapiromatron@gmail.com)
Kate Guyton (Guytonk@iarc.fr)
Ruth Lunn (lunn@niehs.nih.gov)

Use of the Key Characteristics of Carcinogens and Read Across-Like Approaches for Cancer Hazard Evaluation: A Case Study with Haloacetic Acids

Stanley T. Atwood,[1] Ruth M. Lunn,[2] Sanford C. Garner, [1] Gloria D. Jahnke[2]

[1]Contractor in support of the National Institute of Environmental Health Sciences (NIEHS) Report on Carcinogens, Integrated Laboratory Systems, Inc. (ILS), Research Triangle Park, NC, United States

[2]Office of the Report on Carcinogens, Division of the National Toxicology Program, NIEHS, Research Triangle Park, NC, United States

Abstract

The Report on Carcinogens (RoC) is a congressionally mandated public health report that identifies cancer hazards. It identifies agents, substances, mixtures, or exposure circumstances known or reasonably anticipated to be human carcinogens based on systematic evaluation of the scientific literature and application of specific listing criteria. These criteria include listings based on mechanistic evidence indicating that the agent either acts through relevant cancer mechanisms in humans or belongs to a structurally related class of chemicals previously listed in the RoC. Thus, these criteria allow for the use of mechanistic data and read across-like approaches to evaluate chemical agents.

This RoC cancer hazard evaluation focuses on 13 haloacetic acids (HAAs) found as water disinfection by-products. HAAs are formed during the disinfection process when chlorine-based disinfectants react with organic matter naturally present in the source water resulting in by-products potentially affecting public health. Of the 13 HAAs, 6 have been tested for carcinogenicity in rodents and 5 were positive for cancer. A primary objective of this evaluation was to determine if read across-like techniques would enable listing HAAs as a class, subclass or individually. Mechanistic data based on the 10 key characteristics of carcinogens (KCs) along with disposition and toxicokinetic data were organized and clear trends were identified related to the halogen substitution pattern. Although these data were insufficient to propose listing haloacetic acids as a class, they supported NTP listing recommendations of *reasonably anticipated to be a human carcinogen* for two haloacetic acids with no animal cancer data plus four haloacetic acids with animal cancer data.

Objective and approach

1) To determine whether haloacetic acids found in drinking water should be evaluated for carcinogenicity individually, as a subclass, or as a class for possible listing in the RoC,
2) Using key characteristics of carcinogens (KCs) and read across-like approaches.

Methods and process

- Thirteen HAAs found in disinfected drinking water were reviewed (**Figure 1**). (See RoC protocol and RoC Monograph on Haloacetic Acids Found as Water Disinfection By-Products for literature search strings and detailed methods.)
- Studies (5681) were uploaded into Health Assessment Workspace Collaborative (HAWC) system and 21 animal cancer studies were identified. Six of the HAAs had animal cancer studies.
- Animal cancer data were uploaded into systematic review web-based software and study quality assessed. Cancer endpoints for animal studies are listed in **Table 1**.
- A comparative read-across approach was used to identify relative chemical and mechanistic HAA potencies and compare to observed cancer endpoints (**Tables 2 and 3**).
- Metabolism and toxicokinetic (TK) data were used to determine the nature and extent of metabolism of tri-HAAs to di-HAAs (**Figures 2 and 3).**

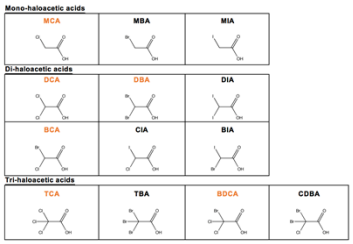
Abbreviations for haloacetic acids (HAAs)

MONO- HAAS: monochloroacetic acid (MCA, CA), monobromoacetic acid (MBA, BA), monoiodoacetic acid (MIA, IA)

DI-HAAs: dichloroacetic acid (DCA), dibromoacetic acid (DBA), diiodoacetic acid (DIA), bromochloroacetic acid (BCA), chloroiodoacetic acid (CIA), bromoiodoacetic acid (BIA)

TRI-HAAs: trichloroacetic acid (TCA), tribromoacetic acid (TBA), bromodichloroacetic acid (BDCA), chlorodibromoacetic acid (CDBA)

Figure 1 Six of thirteen HAAs reviewed have animal cancer studies



HAAs in red print have animal cancer studies

Table 1 Cancer endpoints from animal studies

Neoplasm or tissue	DCA				DBA				BCA				TCA				BDCA			
	Rats		Mice		Rats		Mice		Rats		Mice		Rats		Mice		Rats		Mice	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Liver	X		X	X			X	X			X	X			X	X			X	X
Mononuclear-cell leukemia						X														
Malignant mesothelioma					X				X								X			
Mammary gland										X								X		
Lung							X													
Skin																	X			
Harderian gland																			X	
Large intestine									X	X										

- **Four HAAs met RoC criteria for two species or multiple cancer sites and were sufficient to propose a RoC listing for 4 HAAs.**
- Monochloroacetic acid (MCA) exposure did not produce tumors in experimental animals. Trichloroacetic acid (TCA) had one cancer site in one species and did not meet RoC criteria.

Table 2 Read-across approaches included chemical and mechanistic data related to cancer

13 HAAs	mono HAA	di HAA	tri HAA
Properties	Electrophilicity		
TK Toxicokinetics	Comparative data		
Key events	Potencies Oxidative stress, Genetic toxicity, Animal Cancer		

Table 3 Relative HAA potencies for mechanistic and chemical endpoints

Endpoint	Monohaloacetic acids			Dihaloacetic acids						Trihaloacetic acids			
	CA	BA	IA	DCA	DBA	DIA	BCA	CIA	BIA	TCA	TBA	BDCA	CDBA
Toxicokinetics[r]													
Oral bioavailability (%)	100			81	30		47			100	62	96	100
Unbound fraction (%)	27			94	89		93			53	18	49	55
Total clearance (mL/kg/h)	262			267	491		1037			92.5	754	286	486
Renal (% of total)	59			1.1	2.6		3.6			45.5	22.7	31.1	37.4
Non-renal (% of total)	41			98.9	97.4		96.4			54.5	77.2	68.9	62.6
Chemical properties[s]													
pK_a	2.97	2.96	2.95	1.41	1.39		1.4	1.47	1.67	0.66	0.03	0.05	0.04
E_{LUMO} (deprotonated)	0.43	8.68	7.18	8.44	7.51		7.78	6.40	6.46	7.13	6.12	6.65	6.42
TR_{OSH}	0.71	2.18	2.15	1.68	1.48		1.93	1.35	0.62	1.01	2.26	1.56	1.1
TR_{NSA}	0.92	0.84	0.81	0.5	1.9		3.18	1.17	1.47	0.99	2.61	0.73	0.95
Oxidative stress in vitro[t]													
ARE-bla	3.7	192	278	0.17	8.3		7.1	45.5	38.5	N	2.3	0.5	0.2
ARE-bla	4	90.9	196	0.06	4		2.2	10	18.9	N	1.5	0.25	0.46
Oxidative stress *in vivo*													
8-OHdG				1.4	2.9		2.9			1.2		1.7	
TBARS				129	250		290			67		240	
Genetox *in vitro*													
SOS-umu[u,v]	60	2400	15400	180	2564		2941	5263	9091	60	142860	9091	9091
Ames TA100 (-S9)[f]	44	6588	14129	36	183					N	N		
Ames TA100 (-S9)[g]				5.2	61.9		60.6			N	1.2	31.6	1.7
Comet CHO cells[h]	2439	58820	114900	N	556	500	333		313	N	400	N	71
HGPRT CHO cells[i]	8.7	14.6	536	2.8	66.2					N			
FSS-blu[j]	5882	105260	212730	N	3846		4348	9091	9091	N	N	N	N
Genetox. *in vivo*[l]				0.5	1		N			N	N		
PPARα *in vitro*[k]	2	20	100	1	1					2	1		
PPARα *in vivo*[l]	N			2	3.5		N			4.3		N	
GST-τ inhibition[m]				45	83		81			N			
Animal carcinogenicity													
Species/tumor site[n]	N			2	3		3			1		3	

Darker to lighter shade = high to low value or potency, white = negative (N), gray = no data.

Results for Table 3

- Electrophilicity, genetic toxicity, and oxidative stress were major key carcinogenicity characteristics of the HAAs. Most of the HAAs were associated with most of the characteristics of carcinogens (KCs).
- Brominated tri-haloacetic acids considered as a subclass (**circled, Table 3**).

Several factors lead to major limitations in read-across data confidence for some of the HAAs:
- Lack of cancer data for iodoacetic acids, and monoacetic acids.
- Lack of in vivo potency data for oxidative stress and genetic toxicity.
- Possible dissimilar mode of action of carcinogenicity of dichloroacetic acid and trichloroacetic acid.

Figure 2 Metabolism of tri-HAAs to di-HAAs

- Tri-HAAs are metabolized by P450 reductive dehalogenation, forming dihaloacetic acids via a dihaloacetic acid radical intermediate.

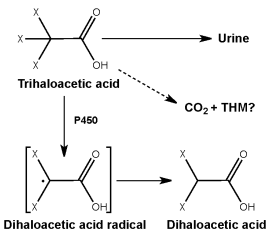


Figure 3 TBA and CDBA are metabolized to animal carcinogens

Metabolism of tri-brominated HAAs (Figure 2) and supporting mechanistic data led to cancer predictions for two brominated HAAs (CDBA, TBA) without cancer data.

- Tri-HAAs with both Cl and Br always lose a Br
- Br loss from Tri-HAA corresponds 1:1 to Di-HAA formation
- Br substitution for Cl enhances metabolism
- **TBA and CDBA: 1) no animal cancer data but are metabolized to animal carcinogens (BCA and DBA) 2) have similar physicochemical, toxicokinetic, and biological effects (Table 3).**

No. of Bromines	Parent	Relative extent of metabolism	Metabolite
0	TCA	→	DCA
1	BDCA	➡	DCA
2	CDBA	➡	BCA
3	TBA	➡	DBA

Conclusions: NTP listing recommendation

Six HAAs are *reasonably anticipated to be human carcinogens:*

- **Dichloroacetic acid (DCA), dibromoacetic acid (DBA), bromochloroacetic acid (BCA) and bromodichloroacetic acid (BDCA)**
  - Sufficient evidence from studies in experimental animals and supporting mechanistic data that demonstrate biological plausibility
- **Chlorodibromoacetic acid (CDBA) and tribromoacetic acid (TBA)**
  - CDBA and TBA are directly metabolized to BCA and DBA, respectively, which are two HAAs that caused cancer in experimental animals and supporting mechanistic data that demonstrate biological plausibility

References

RoC Handbook is available at: https://ntp.niehs.nih.gov/go/rochandbook

Final Report on Carcinogens Monograph on Haloacetic Acids Found as Water Disinfection By-Products and related materials are available at: https://ntp.niehs.nih.gov/go/790113

Protocol for Monograph: https://ntp.niehs.nih.gov/go/790113

RoC Process/Listing Criteria: https://ntp.niehs.nih.gov/go/rocprocess

Acknowledgments

Integrated Laboratory Systems, Inc (Support provided through NIEHS contract number HHSN273201100004C)

Additional Scientific Support:

Andrew Ewens, PhD, DABT
Alton Peters, MS
Jessica Geter, MSLS
Whitney Arroyave, PhD

Administrative Support:

Ella Darden, BS
Tracy Saunders, BS

An Approach to Integrating Mechanistic Information in Human Health Risk Assessment: A Case Study with Hexavalent Chromium

Authors: Chad M. Thompson, Mina Suh, Deborah M. Proctor, Laurie C. Haws, and Mark A. Harris

Introduction

Mode of action (MOA) analysis is inherently an exercise in data integration. Two important challenges to MOA analysis for carcinogens with large databases include focusing the analysis on the data most relevant for the risk assessment problem, and succinctly summarizing large amounts of literature into a coherent series of key events. Rather than develop an altogether new framework, we propose to combine two existing data integration methods that together provide an approach for assimilating mechanistic information for the goals of risk assessment. In this two-phased approach, we first adapt the 10 factors that Eastmond (2012) identified as having historically been used by regulators when determining whether a carcinogen acts through a mutagenic or non-mutagenic MOA. Several of the 10 factors require summarizing highly relevant information for risk assessment irrespective of the existence of a MOA, including: the nature of the tumors of interest, properties of the chemical of interest, pharmacokinetics, and *in vivo* genotoxicity. As such, these factors (or summaries) provide important information that can subsequently feed into a formal MOA analysis. Eastmond (2012) identified *evidence for an alternative non-mutagenic MOA* as a factor; however, this presupposes that a MOA analysis already exists. Herein, we propose that several of the 10 factors can be used to summarize critical data *before* formal MOA analysis begins. The second phase of our approach is to conduct a formal MOA analysis using a framework such as that proposed by IPCS or U.S. EPA. In this poster, we demonstrate this two-phase approach using hexavalent chromium as an example (**Figure 1**).

Methods and Results

Eastmond (2012) reviewed over 30 MOA determinations by regulators and identified 10 factors that are important for consideration of whether a carcinogen acts by a mutagenic MOA (see Table in Figure 1). Eastmond concluded that these factors should "improve the quality and consistency of mutagenicity MOA decisions made by both governmental and non-governmental entities", as well as "facilitate the refinement of previous mutagenic MOA determinations as new information becomes available".

Phase 1

Six of the 10 factors are particularly relevant for informing MOA analyses by allowing for systematic identification and organization of mechanistic data. A case example with hexavalent chromium (Cr(VI)) is shown in **Table 1**. Relevant data are summarized in **Figures 2-4** and **Tables 2&3**.

Phase 2

Mechanistic data from Phase 1 can then be synthesized and integrated to develop and evaluate one or more MOAs. This poster presents an example of data integration demonstrating a non-mutagenic, cytotoxicity-regenerative cell proliferation MOA (**Figures 5&6**). The organizational structure from Phase 1 could also be used to explore other MOAs. Note: Evidence for an "alternative" (i.e. non-mutagenic) MOA is one of the 10 factors identified in Eastmond (2012) (**Figure 1**).

Risk Assessment

- Results from Phases 1 & 2 feed into the overall evidence integration and can inform the most appropriate approach(es) for quantitative dose-response analysis (**Figure 7**).
- Exposure information can, in some instances, provide further insight into which MOA is operational and which risk assessment approach is supported by the data (**Figure 7**).

Figure 1. An Approach to Integrating Mechanistic Information in Human Health Risk Assessment

Phase 1. Data Collections/Organization

Ten Factors for Influencing Mutagenic MOA Determinations

Eastmond Factor	Brief Description
Nature of tumors	Understanding the distribution of tumors (e.g., site-specific, multi-organ), time of onset, and underlying biology of tissue(s) can inform MOA
Mutations in rodent tumors	Commonly mutated genes can inform MOA by identifying possible early targets; the types of mutations can also inform MOA
Chemical properties	Chemical properties, metabolites, and/or degradation products
Metabolism and toxicokinetics	Understanding the absorption, distribution, metabolism, and excretion of the chemical informs an understanding and interpretation of both in vitro and in vivo genotoxicity test results
Structural similarity to known carcinogens	Comparison of results for the chemical of interest with those of chemicals with similar structure, metabolites, or chemicals that act through similar mechanisms
Understanding the assay system	Understanding the principles, strengths, and weaknesses of short-term genotoxicity assays
In vivo genotoxicity (especially in target organs)	Positive results in the in vivo test systems have almost always been necessary for expert panels to conclude that the agent induces cancer through a mutagenic mechanism
Origin or mechanism of observed effects	Understanding the origins of the genotoxic lesion or damage (e.g., oxidative vs. direct binding) can play an important role in mutagenic MOA determinations
Quality, quantity, and reproducibility of data	Consideration of data quality—including study design (e.g., OECD), conditions (e.g., GLP), and expertise (e.g., student vs. professional) can help sort through conflicting or contradictory findings
Evidence for alternative MOAs	Evidence supporting an alternative or complementary MOA could influence the interpretation of the mixed or ambiguous genotoxicity results.

adapted from Eastmond (2012)

Collect Data

Integration

Phase 2. MOA Analysis

Problem Formulation (Purpose-oriented)
What is the decision context (e.g., priority setting, quantitative risk assessment)?
Can mode of action help inform the decision?

Mode of Action Framework
- Hypothesis based
- Evidence in support of key events based on modified Bradford Hill considerations
- Qualitative and quantitative species concordance

Mode of Action Knowledge Informs

- **Risk Assessment**: Human relevance; Human variation; Species extrapolations; Life stage effects; Dose-response extrapolation; Combined exposures
- **Assessment-specific Data Generation**: Targeted testing (in vivo and in vitro); Non-test methods (QSAR, read-across, modeling)
- **Research**: Diagnostic biomarkers; Expert systems; New test methods; Non-test methods (QSAR, read-across, modeling); Therapeutic intervention to treat intoxication

Adapted from Meek et al. (2014)

SR Process

Scoping

Problem Formulation

Broad Literature Search

Develop Protocols for Systematic Reviews

Identify Evidence

Evaluate Studies

Integrate Evidence

Dose-Response Assessment and Derivation of Toxicity Values

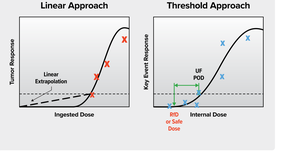
Hazard Evidence

Human / Animal / Mechanistic

Systematic Reviews

Adapted from NRC (2014)

Risk Assessment



Phase 1: Data Collection & Organization

Table 1. Use of 6 Factors Most Relevant for MOA Analysis

Factor	Summary
Nature of tumors	adenomas (with lower incidence of carcinomas) in duodenum and proximal jejunum of B6C3F1 mice; oral cavity tumors in F344 rats (see **Figure 2**)
Mutations in tumors	unknown (no data from tumors)
Chemical properties	Cr(VI) is reduced to Cr(III) by bodily fluids Cr(VI) that escapes extracellular reduction enters cells through anion transporters, due to the structural similarity of chromate, sulfate, and phosphate anions Cr(III) is taken up by pinocytosis and passive diffusion Inside cells, Cr(VI) is reduced to Cr(III) by low-molecular-weight thiols (e.g., GSH, cysteine), high-molecular- weight thiols (e.g., proteins), and antioxidants (e.g., ascorbate). Depending on the specific reducing partner, different levels of reactive intermediates (Cr(V) and Cr(IV)) are formed, including free radicals Cr(III) conjugates may bind to DNA, and are potentially mutagenic
Metabolism and toxicokinetics	little/no enzymatic metabolism; non-enzymatic extracellular reduction of Cr(VI) to inert Cr(III) **Figure 3** shows the disposition of Cr(VI) within the alimentary canal (inset shows Cr localization)
***In vivo* genotoxicity**	**Table 2** and **Figure 4** summarize *in vivo* genotoxicity data in target tissues
Structural similarity	Adapted to include carcinogens that induce similar phenotype effects in target tissue(s) (**Table 3**).

Table 2. Summary of *In Vivo* Genotoxicity Results in Target Tissue

Genotoxicity Assay	+/-
Duodenal Micronucleus Assays	
B6C3F1; f; ≤180 ppm; 7 days; intact crypts; Swiss roll	-
B6C3F1; f; ≤180 ppm; 7 days; full sections; transverse	-
B6C3F1; f; ≤180 ppm; 90 days; intact crypts; transverse	-
B6C3F1; f; ≤180 ppm; 90 days; full sections; transverse	-
Crypt γ-H2AX Immunostaining	
B6C3F1; f; ≤180 ppm; 7 days; Swiss roll	-
B6C3F1; f; 180 ppm; 90 days; transverse	-
Mutation Assays	
B6C3F1; f; ≤180 ppm; 90 days; duodenum; *kras* codon 12 GAT	-
Big Blue TgF344 rats; m; 180 ppm; 28 days; gingival mucosa; *cII*	-
Big Blue TgF344 rats; m; 180 ppm; 28 days; duodenum; *cII*	-
gpt delta mice; m; 30 ppm; 28 & 90 days; duodenum	-
DNA Modification	
SKH-1; f; 5 and 20 ppm; 9 months; forestomach, stomach, duodenum; 8-oxo-dG	-
SKH-1; f; 5 and 20 ppm; 9 months; forestomach, stomach, duodenum; DPX	-
B6C3F1; f; ≤180 ppm; 90 days; duodenum & oral cavity; 8-oxo-dG	-
F344; f; ≤180 ppm; 90 days; duodenum & oral cavity; 8-oxo-dG	-
B6C3F1; f; ≤180 ppm; 7 and 90 days; duodenum & oral cavity; Cr-DNA	±
F344; f; ≤180 ppm; 7 and 90 days; duodenum & oral cavity; Cr-DNA	±

All exposures were via drinking water; DPX, DNA-protein crosslinks; +, positive; -, negative; ±, equivocal

Table 3. Comparison of Intestinal Key Events and Critical Effects Between Cr(VI) and Captan/Folpet

Effect	Captan and Folpet	Cr(VI)
Dosimetry (duodenum is primary target)	• hydrolytically stable in the lower pH of the stomach • hydrolyzed in higher pH of SI • duodenum receives bulk of exposure	• readily reduced to inert Cr(III) in low pH of stomach • unreduced Cr(VI) enters the higher pH of SI • Cr(VI) absorbed into villi via ion mimicry • tissue Cr levels highest in the duodenum (see **Figure 3**)
Molecular interactions	• can react with thiols directly and form thiophosgene, which also reacts with thiols	• reacts with thiols (GSH, cysteine) • reactive intermediates can be formed
Villous effects	• cytotoxicity leading to crypt hyperplasia and villus atrophy and blunting • healthy proliferation with blunted villi indicate villi are site of damage	• cytotoxicity leading to crypt hyperplasia and villus atrophy and blunting • healthy proliferation with blunted villi indicate villi are site of damage — XRF maps show Cr is localized to villi (see **Figure 3**)
Crypt hyperplasia	• expanded crypt compartment via H&E stain • proliferation evident at 1, 2, 4, and 104 weeks of exposure	• expanded crypt compartment via H&E stain • increased cells/crypt and crypt length • increased Ki67 and other transcript changes • proliferation evident at 1, 13, and 104 weeks of exposure
Tumorigenesis	• increase in stem-cell proliferation increases chance of tumor due to probability of DNA replication error	• increase in stem-cell proliferation increases chance of tumor due to probability of DNA replication error • background kras mutation frequency may be high in B6C3F1 mice
In vivo genotoxicity in target tissues	• captan and folpet were negative in duodenal MN assays • folpet was negative in duodenal comet assay (captan not tested) • captan was negative in DNA binding (folpet not tested)	• negative duodenal MN assays • negative for aberrant γ-H2AX immunostaining in crypts • negative for aberrant proliferative foci on villi • negative for increased kras codon 12 GAT MF • negative for mutation in gpt delta mice • negative increase in cII mutants in oral mucosa & duodenum of TgF344 rats • XRF maps indicate Cr-DNA binding is confined to villi • toxicogenomic analyses indicate minimal DNA damage signaling at high doses
In vitro genotoxicity	• captan and folpet mutagenicity in Ames assays are mitigated by S9	• Cr(VI) mutagenicity in Ames assays is mitigated by S9

Figure 2. Tumor incidence in 2-year cancer bioassay. (Left) Oral mucosa tumor incidence in F344 male and female rats (these were the only treatment-related tumors in rats). (Right) Incidence of duodenal diffuse epithelial hyperplasia (DEH) (dotted lines) and small intestine tumors (solid lines) in B6C3F1 male and female mice (these were the only treatment-related tumors in mice).

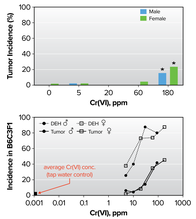
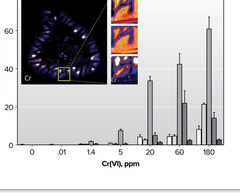


Figure 3. Location of Cr in the mouse duodenum following 13 weeks of exposure. Measures of total chromium (Cr) in the oral mucosa, glandular stomach, duodenum, jejunum, and ileum following 13 weeks of exposure to Cr(VI) in drinking water (bars darken as progress down the GI tract (e.g., oral=white; ileum=darkest). Data plotted are mean and SEM. The inset shows a representative XRF map of a duodenal section from a mouse exposed to 180 ppm Cr(VI) for 13 weeks. The main map is for Cr signal. The region in the gold box was magnified, and XRF maps were generated for calcium (top), sulfur (middle), and Cr (bottom). The crypt regions are marked with a white asterisk.



Phase 2: MOA Analysis

Figure 5. Temporal and dose concordance for data supporting a cytotoxic/regenerative hyperplasia MOA. Bars represent incidence in H&E stained duodenal sections; blue lines represents mean ± sd for enterocytes per crypt in separate assays and/or studies (note: the short bars indicate empirical observations with 0% incidence). Dose concordance of intestinal diffuse epithelial hyperplasia (DEH), adenomas, and carcinomas in mice, with dose expressed as Cr(VI) intestinal flux estimated by a PBPK model.

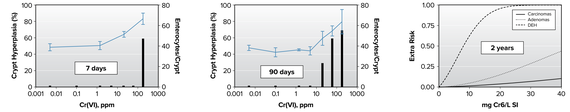


Figure 6. Schematic of MOA for Cr(VI)-induced intestinal cancer with select data supporting key events. This MOA is very similar to other intestinal carcinogens that have been concluded to act via chronic cytotoxicity and regenerative hyperplasia (captan, folpet).

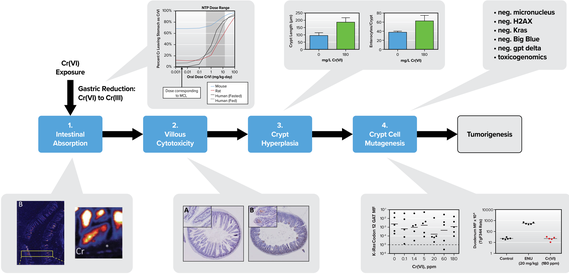
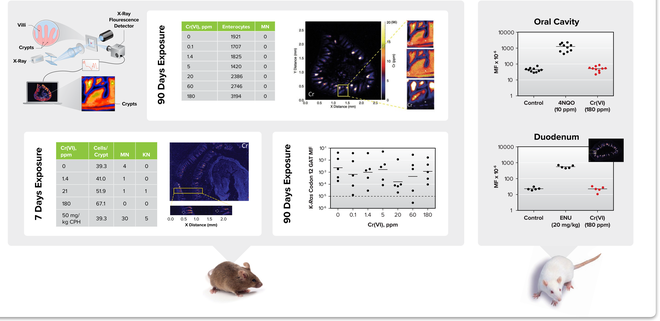
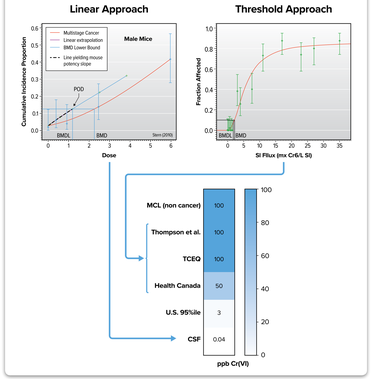


Figure 4. In vivo genotoxicity assays. **(Left, mice)** Tables show the results of duodenal micronucleus (MN) assays in B6C3F1 mice exposed ≤180 ppm Cr(VI) for 7 or 90 days. Near each table is a representative X-ray fluorescence (XRF) microscopy image of a duodenum section from a mouse exposed to 180 ppm Cr(VI). The 7-day duodena were prepared by Swiss roll technique, whereas the 90-day samples are transverse sections. In both studies, chromium is detected in villi but not crypts. Also shown are results for analysis of kras codon 12 GAT mutant frequency (MF) in scraped duodenum (crypts and villi). **(Right, rats)** Results from transgenic rodent (TGR) mutation assay in Big Blue® TgF344 rats exposed to 180 ppm Cr(VI) for 28 days. Results for ENU and 4NQO are significantly different from respective controls. Also shown is an XRF map from a transverse duodenal section from a F344 rat exposed to Cr(VI) for 90 days—indicating similar localization of chromium in the duodenum of both mice and rats.



Risk Assessment

Figure 7 . Systematic assessment of mechanistic data from Phases 1 & 2 inform which risk assessment approach (linear or nonlinear) is most consistent with the available data. Furthermore, a linear approach results in toxicity standards much lower than most U.S. exposures (95th-%ile = 3 ppb; median ~1 ppb). Health Canada, TCEQ, and Thompson et al. have proposed similar non-linear approaches based on recent MOA data.



Conclusions

- In practice, Phases 1 and 2 were completed in reverse order for Cr(VI), with a MOA analysis published in 2013 and evaluation of 10 Factors published in 2017. However, our experience with these methods and frameworks led us to appreciate the value of considering many of the 10 factors as organizing principles *prior* to conducting a MOA analysis.
- Phase 1 helps organize the most relevant data for conducting a MOA analysis. This phase can also help determine whether a MOA analysis should be performed. For example, a MOA for Cr(VI) was published in 2010 despite the lack of target tissue genotoxicity data and no apparent consideration of similarly acting intestinal carcinogens. Using the approach described herein and data available in 2010, the 2010 MOA analysis might have concluded that the MOA was unknown as opposed to mutagenic.
- Phase 2 helps assess how strongly the data, collected in Phase 1, support a given proposed MOA.
- Together, Phases 1 & 2 help inform the most appropriate approach(es) for conducting quantitative dose-response analysis in a Risk Assessment/Evaluation.
- It is recognized that this case example does not fully employ systematic review methods such as appraising studies for validity; however, such methods could be incorporated into the general approach described herein.

Select References (all references available upon request)

Eastmond, D. A. (2012). Factors influencing mutagenic mode of action determinations of regulatory and advisory agencies. *Mutat Res, 751*, 49-63

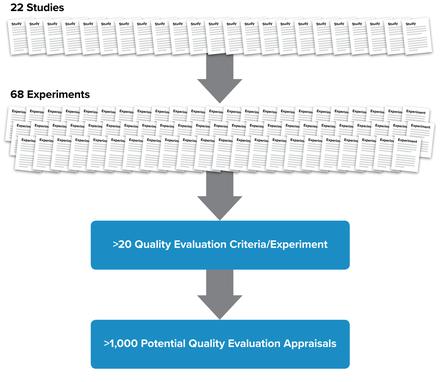
Thompson, C. M., Proctor D. M., Suh M., Haws L. C., Kirman C. R., Harris M. A. (2013). Assessment of the Mode of Action Underlying Development of Rodent Small Intestinal Tumors Following Oral Exposure to Hexavalent Chromium and Relevance to Humans. *Critical Reviews in Toxicology*, 43, 244-274.

Thompson, C. M., Suh M., Proctor D. M., Haws L. C., Harris M. A. (2017). Ten factors for considering the mode of action of Cr(VI)-induced gastrointestinal tumors in rodents. *Mut Res*, 823, 45-57

Application of Mechanistic Data Quality Criteria in Assessment of the Relationship Between Congenital Heart Defects and TCE Exposure—A Case Study

Authors: Wikoff, D[1]; Urban J[2]; Chappell G[2]; Haws, L[2]

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



Introduction

- Tools for evaluating the quality of mechanistic studies — especially *in vitro* experiments — within the systematic review paradigm are not well established. Heterogenous study designs preclude simple adoption of existing tools and require development of new tools for appraising validity.
- One recently proposed tool is that from the USEPA-OPPT's 2018 Application of Systematic Review (SR) in TSCA Risk Evaluations[1] (i.e., the TSCA study quality tool).
- The evidence base for trichloroethylene (TCE) and congenital heart defects (CHDs) in offspring presents an opportunity to explore study quality tools for mechanistic data. Previously, a Risk of Bias (RoB) assessment was published on the human and animal toxicity studies (Wikoff et al., 2018)[2]. Findings indicated the limited supporting human and animal data suffered from high risk of bias; however, tools to evaluate the quality of the mechanistic dataset were not readily available.

Objective

To evaluate the utility of the TSCA study quality tool for mechanistic data using the TCE-CHD mechanistic evidence base as a case study.

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[3]; Wikoff et al., 2018). Search results conducted by Wikoff et al. (2018) were re-reviewed for identification of relevant mechanistic literature. 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.
- Any studies evaluating TCE exposure (*in vivo* or *in vitro*) relative to any mechanistic aspect of biomolecular and/or physiological pathways of cardiac development were included.
- Studies were categorized based on the assay type(s): in vivo, in vitro (cell culture, *in ovo, ex ovo, ex vivo*)

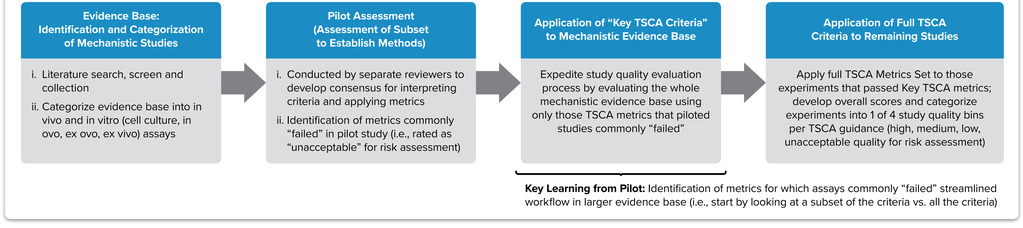
Selection of Screening Criteria:

- Primary Tool: TSCA study quality tools
  - The TSCA study quality tool is comprised of 7 domains, defined by 25 metrics (*in vitro*; per Table G-16 of TSCA SR guidance) or 24 metrics (*in vivo*; per Table G-14 of TSCA SR guidance).
  - Studies/experiments with any metrics scoring a "4" are characterized as "unacceptable" for risk assessment and excluded from further evaluation. For the remaining experiments, overall quality was determined per the weighted scoring calculations and categorizations based on Tables G-8 (*in vitro*) and G-7 (*in vivo*).
- Secondary Tools: two additional tools were selected for comparison purposes and applied to a subset of data:
  - SciRAP[4] — a scoring method developed for quantifying reporting and methodological quality evaluations; the developers recommend using scores to rank study quality. Each evidence base is evaluated on a case-by-case basis, and no default score cutoffs are provided for characterizing overall quality. There are 3 evaluation criteria for each metric, none of which included a mechanism for excluding low quality studies from the evidence base.
  - ToxRTool (Toxicological data Reliability Assessment Tool)[5] — developed to evaluate toxicology study quality based on the Klimisch reliability scoring system, in which toxicology studies are assigned to one of four scores for reliability. Only 2 evaluation criteria are considered for each metric ("0" or "1"); a "0" for any of the 6 "Red Criteria" results in an automatic lowering of evaluation quality category.

Study Quality Assessment (Figure 1)

- All quality assessments were carried out by two PhD scientists (GC, JU) with experience conducting and/or assessing *in vivo* and *in vitro* studies. In cases of conflict, a third scientist (DW) was consulted to facilitate a consensus solution.
- Study quality evaluation was conducted at the experimental level; many studies reported multiple experiments.
- A pilot assessment was conducted in which the two analysts independently reviewed a subset of 10 *in vitro* experiments. All decisions, interpretations, and refinements were documented. A similar pilot was conducted for mechanistic data reported in *in vivo* studies.
- Based on findings of the pilot assessment, a tiered approach was used to assess *in vitro* study quality:
  - Key Metrics Evaluation: The five Key Metrics were assessed; those studies that scored "4" on any of these metrics were removed from further study quality evaluation.
  - Complete Study Quality Evaluation: Those studies that scored 1-3 on all of the Key Metrics were then further evaluated using all the quality metrics.
- A subset of the data were also subjected to quality evaluation using SciRAP and ToxRTool methods for comparison purposes.

Figure 1. Summary of Study Quality Evaluation Process Developed for TCE-CHD Mechanistic Evidence Base



Results

Mechanistic Evidence Base: 22 studies were identified that tested TCE in an experimental model designed to examine a mechanistic aspect of cardiac development. These studies comprised a total of 68 unique experiments (**Table 1**).

Table 1. TCE-CHD Mechanistic Studies: Breakdown of Experiment Types

Experiment Type	Number of Experiments	Tools Used to Evaluate Quality of Experiments Relevant to Potential TCE-CHD Mechanism
In vivo	5	USEPA OPPT TSCA Animal Toxicity Study Metrics (Table G-14)
In vitro (cell culture)	26	USEPA OPPT TSCA In Vitro Toxicity Study Metrics (all 26); SciRAP Tool for Evaluating In Vitro Toxicity Studies (2 of 26); ECVAM ToxRTool for Assessing Reliability of In Vitro Toxicity Studies (2 of 26)
In vitro (in ovo)	21	USEPA OPPT TSCA In Vitro Toxicity Study Metrics (all 21); SciRAP Tool for Evaluating In Vitro Toxicity Studies (1 of 21); ECVAM ToxRTool for Assessing Reliability of In Vitro Toxicity Studies (1 of 21)
In vitro (ex ovo)	3	USEPA OPPT TSCA In Vitro Toxicity Study Metrics (all 3)
In vitro (ex vivo)	12	USEPA OPPT TSCA In Vitro Toxicity Study Metrics (all 12)
Unknown model	1	USEPA OPPT TSCA In Vitro Toxicity Study Metrics (1)
Total	68	-

Pilot Assessment:

- All 10 *in vitro* experiments scored a "4" (unacceptable for use in risk assessment) in at least one Metric
- Five TSCA study metrics that the pilot experiments most commonly scored "4's" were identified as "Key Metrics" (**Table 2**). ***The list of commonly failed metrics was used to expedite subsequent evaluations of the mechanistic evidence base by focusing on these five (vs. all 25 for each study) as a means of improving efficiency and workflow.***

Table 2. Metrics Identified as Critical per Pilot Evaluation using TSCA Study Quality Metrics for In Vitro Experiments

TSCA Metric No.	Metric Title	Metric Description	TSCA Definition of "Unacceptable" for Risk Assessment (=4)
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)	• Information on preparation and storage was not reported, OR • Serious flaws reported with test substance preparation and/or storage conditions will have critical impacts on dose/concentration estimates and make the study unusable (e.g., instability of test substance in exposure media, test substance volatilized rapidly from the open containers that were used as test vessels).
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?	• No information on exposure duration(s) was reported, OR • The exposure duration was not appropriate for the study type and/or outcome of interest (e.g., 5 hours for reconstructed epidermis in skin irritation test, 24 hours exposure for bacterial reverse mutation test).
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)?	• The outcome assessment methodology was not reported, OR • The assessment methodology was not appropriate for the outcome(s) of interest (e.g., cells were evaluated for chromosomal aberrations immediately after exposure to the test substance instead of after post-exposure incubation period).
22	Data Analysis	• Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)?	• Statistical methods were not appropriate (e.g., Student's t-test used to compare 2 groups in a multi-group study, parametric test for non-normally distributed data), OR • Statistical analysis was not conducted AND data were not provided, preventing an independent statistical analysis.
24	Cytotoxicity	• Were cytotoxicity endpoints defined, if necessitated by study type, and were methods for measuring cytotoxicity described and commonly used for assessments?	• Cytotoxicity endpoints were not defined, methods were not described, and it could not be determined that cytotoxicity was accounted for in the interpretation of study results.

Assessment of Study Quality Across the Evidence Base:

- 33% of the evidence base was of acceptable quality for risk assessment (**Figure 2**). In all cases except one, all experiments within a given study were similarly categorized as acceptable or unacceptable (**Figure 3**).
- 44 of the 68 experiments received at least one unacceptable score in a Key Metric. Of those remaining after Key Metric assessment, all experiments were "acceptable" for risk assessment following assessment with all TSCA metrics
- Of the 24 experiments considered acceptable (no 4's), 14 scored as "high" quality and 10 scored as "medium" quality. No studies scored as "low quality."
- Very few of the experiments involving *in vivo* or cell culture *in vitro* models were of acceptable study quality for risk assessment. By contrast, a larger portion of the *in ovo* and *ex ovo* experimental models were scored to be of sufficient quality to be retained for further consideration in risk assessment (**Figure 4**).
- The most commonly "failed" metrics were those associated with 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 5**). More than 1/3 of the experiments scored "4" on multiple metrics.

Comparison of Quality Using Secondary Tools:

- Across the three tools, the ToxRTool outcomes for the selected *in vitro* studies more consistently reflect the TSCA results than SciRAP, although the general dichotomy in study quality between the cell culture and *in ovo* experiments was consistent among all three approaches (**Table 3**).
- Two of the three tools (TSCA and ToxRTool) allowed for direct comparisons of overall study designations.
  - Comparison of the overall quality designations were similar with respect general categorization, though two of the three would have been designated in the lowest of all categories under TSCA relative to the second lowest in ToxRTool.
- Comparison of findings to the output from SciRAP are less direct, requiring the analysts to interpret the output relative to the overall scores. Visual inspection generally shows concordance (e.g., more red in Study 19 vs. Study 2 across graphics).

Figure 2. TCE-CHD Mechanistic Studies by Study Quality Category Based on TSCA Systematic Review Guidelines

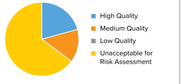
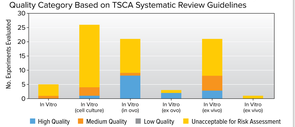


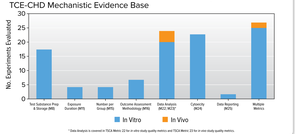
Figure 3. TCE-CHD Evidence Base: Study Quality Evaluation of Mechanistic Studies Using TSCA Systematic Review Guidelines

Figure 4. TCE-CHD Mechanistic Studies by Model Type* and Study Quality Category Based on TSCA Systematic Review Guidelines



* The TSCA in vitro study quality metrics were used to evaluate experimentally designed assays of cell culture, chicken egg ("in ovo"), and whole embryo culture (chicken "ex ovo" and rodent "ex vivo") models.

Figure 5. TSCA Study Quality Metrics Scored "Unacceptable" Across TCE-CHD Mechanistic Evidence Base



* 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. Comparison of In Vitro Study Quality Evaluation Tools

Study number	Experiment Model & Endpoint	Study Quality Evaluation Outcomes		
		TSCA	ToxRTool	SciRAP
2	In ovo: Level of apoptosis in atrioventricular cushion and outflow tract sections of HH24 staged chick embryo hearts	• Score: 1.3 • Interpretation: High Quality Study • Consequence: Retain in TCE-CHD evidence base for risk assessment	• Score: 17 (No "Red" zeros) • Interpretation: Reliable w/o restrictions (k1) • Consequence: Useful; check relevance for intended purpose	• Reporting Quality Score: 78 out of 100 • Methodological Quality Score: 85 out of 100 • Consequence: None defined
14	In vitro: Protein activity in transformed human liver cell line	• Score: Not calculated; multiple metrics scored =4 • Interpretation: Unacceptable for risk assessment • Consequence: Remove from evidence base for TCE-CHD risk assessment	• Score: 12 (1 "Red" zero) • Interpretation: Not reliable (k3) • Consequence: Generally not to be used as a key study; may be useful in WoE approach or as supportive information	• Reporting Quality Score: 50 out of 100 • Methodological Quality Score: 50 out of 100 • Consequence: None defined
19	In vitro: Gene transcription levels in transformed rat heart cell line	• Score: Not calculated; multiple metrics scored =4 • Interpretation: Unacceptable for risk assessment • Consequence: Remove from evidence base for TCE-CHD risk assessment	• Score: 12 (2 "Red" zeros) • Interpretation: Not reliable (k3) • Consequence: Generally not to be used as a key study; may be useful in WoE approach or as supportive information	• Reporting Quality Score: 43 out of 100 • Methodological Quality Score: 46 out of 100 • Consequence: None defined

Overall Findings and Lessons

- The TSCA study quality metrics provide a relatively thorough method for assessing study quality in mechanistic studies.
  - The metrics and metric scoring provide a transparent means of assessing quality as well as a tool to facilitate decisions regarding reliance or inclusion of a study based on quality.
- The heterogeneity of mechanistic study designs provides significant challenges in context of a "one size fits all" quality evaluation approach. Particular challenge was met when using the animal toxicity study metrics for *in vivo* studies of mechanistic parameters.
- Evaluation of the quality of mechanistic/in vitro studies has many challenges related to implementation and workflow:
  - Resource-intensive process with 25 metrics to consider.
  - Difficult to achieve consistent evaluation output between reviewers inter- and intra- assessment.
  - Pilot exercises are critical to determine consensus/agreement on metric criteria refinement, interpretation, and application.
  - The breadth and depth of evidence analyst expertise plays a critical role in making quality determinations across metrics.
- It is difficult to assess study quality without specific consideration of the PECO.
  - The context of the research question is critical for establishing chemical-specific parameters and defining/interpreting the study quality metrics and criteria used to score/rate evidence base.
- The TSCA study quality metrics, as well as SciRAP and ToxRTool, *generally* provide a measure of internal validity though have significant focus on reporting elements. These tools *do not* provide explicit measures of relevance as it relates to external or construct validity in a manner that is consistent with established considerations of such (e.g., mammalian models are more relevant to human risk assessment than non-mammalian models).

Conclusions

- The TSCA study quality metrics provide a relatively thorough method for assessing study quality in mechanistic studies as well as providing a transparent process for determining use of such data in the context of risk assessment. Topic-specific refinements and development of operating procedures improved the utility of this tool.
- The rationale and utility of assessing study quality of mechanistic data should be addressed during problem formulation and protocol development. The specific use of mechanistic data in a risk assessment (e.g., hazard identification, MoA evaluation) should be clarified *a priori* as such context is important to determining the approach for assessing study quality as well as to ensuring that resources are used to provide meaningful output to inform risk assessment and management decisions. In the case of TCE-CHD, it is unlikely that evaluation of the mechanistic data would have changed the earlier risk of bias study conclusions related to development of toxicity values (i.e., CHDs are not a suitable endpoint upon which to base a quantitative risk assessment for TCE).

1 USEPA. 2018. Application of Systematic Review in TSCA Risk Evaluations. Office of Chemical Safety and Pollution Prevention. EPA Document#740-P1-8001. May 2018.

2 Wikoff D, Urban JD, Harvey S, Haws LC. Role of Risk of Bias in Systematic Review for Chemical Risk Assessment: A Case Study in Understanding the Relationship Between Congenital Heart Defects and Exposures to Trichloroethylene. Int J Toxicol. 2018 Mar/Apr;37(2):125-143.

3 Makris SL, Scott CS, Fox J, Knudsen TB, Hotchkiss AK, Arzuaga X, Euling SY, Powers CM, Jinot J, Hogan KA, Abbott BD, Hunter ES 3rd, Narotsky MG. A systematic evaluation of the potential effects of trichloroethylene exposure on cardiac development. Reprod Toxicol. 2016 Oct;65:321-358.

4 http://www.scirap.org/

5 https://ec.europa.eu/jrc/en/publication/articles-journals/toxrtool-new-tool-assess-reliability-toxicological-data

A case study approach to systematically identify and organize mechanistic information informing thyroid hormone adverse outcome pathways.

Andrea Kirk[1], Kristan Markey[2], Isabelle Lee[3], Pamela Noyes[4], Nancy Baker[5], Seema Schappelle[2], Stanley Barone[2].

[1]Office of Land and Emergency Management, Environmental Protection Agency
[2]Office of Chemical Safety and Pollution Prevention, Environmental Protection Agency
[3]University of Pennsylvania, Perelman School of Medicine
[4]Office of Research and Development, Environmental Protection Agency
[5]Leidos, Inc.

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

Abstract

The US Environmental Protection Agency's (EPA) uses systematic review (SR) and systematic evidence mapping for multiple purposes including elucidation of biological pathways and identification of reference chemicals for model and test guideline evaluation. SRs use well-defined, transparent, consistently applied approaches to evaluate available research on a topic. We are adapting and building upon standard systematic review frameworks (such as IOM, OHAT, and TSCA) to support the development of high throughput screening strategies to interrogate thyroid perturbations by:

- Identifying molecular mechanisms of xenobiotic disruption of thyroid hormone-related networks;
- Identifying new potential high and low throughput test methods for further development and evaluation of molecular initiating events of adverse outcome pathways; and
- Identifying diverse reference chemicals with a range of potencies.

Here, we outline our approach for interrogating thyroid hormone networks beginning with the rate-limiting mechanism in thyroid hormone synthesis, the Sodium/Iodide Symporter (NIS). NIS, which serves to concentrate iodide from the bloodstream into the thyroid, is often described as the "starting point" thyroid hormone-related networks. Disruption of these networks by xenobiotics, disease or poor nutrition may lead to diverse, lifelong adverse outcomes. Elucidation network function, molecular initiating events, and chemicals that induce them, will aid in the development of high throughput testing strategies.

The proposed approach is expected to facilitate and further refine efforts to design SR approaches for interrogation of mechanistic lines of evidence.

Introduction

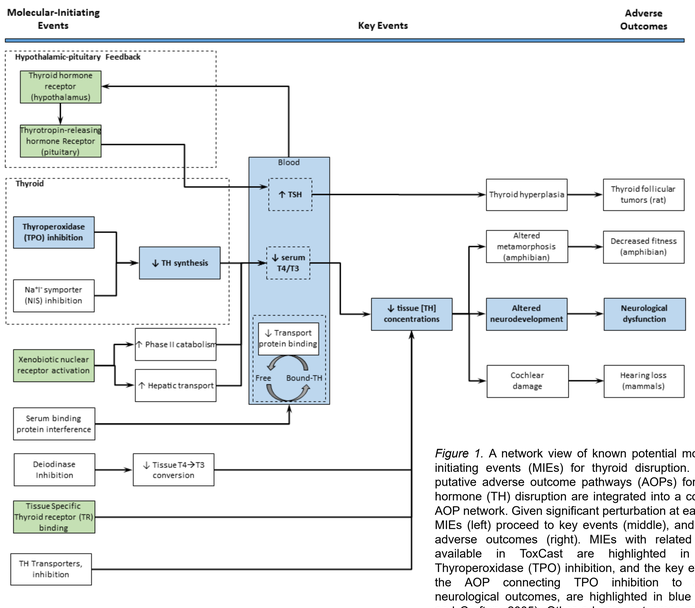


Figure 1. A network view of known potential molecular-initiating events (MIEs) for thyroid disruption. Several putative adverse outcome pathways (AOPs) for thyroid hormone (TH) disruption are integrated into a combined AOP network. Given significant perturbation at each step, MIEs (left) proceed to key events (middle), and then to adverse outcomes (right). MIEs with related assays available in ToxCast are highlighted in green. Thyroperoxidase (TPO) inhibition, and the key events of the AOP connecting TPO inhibition to adverse neurological outcomes, are highlighted in blue (Zoeller and Crofton, 2005). Other adverse outcomes related to TH perturbation area are shown, including hearing loss, altered amphibian metamorphosis, and rodent thyroid tumors (Crofton and Zoeller, 2005; Pickford, 2010; McClain, 1989).

Adapted from Paul Friedman, et al (2016) - DOI: 10.1093/toxsci/kfw034

Automation approaches

Early search strategies of the thyroid literature identified 1+ million potentially relevant articles across a broad range of biological complexity. These results prompted the need to adapt and develop workflows that integrate elements of evidence mapping together with data extraction, machine learning systems and natural language processing, and harmonization of ontologies for management and optimization of large scale pathway-based systematic reviews. Given the breadth and complexity of the literature space, key innovations include modularizing the workflows, inventorying the study elements and experiments rather than papers, and tracking data extraction against inventories. It is unlikely that the complete literature space could ever be surveyed by a single research group.

Organizational Approach

The EPA's Endocrine Disruptor Screening Program has identified the following key targets for molecular initiating events (MIEs). These include:

- Iodide uptake through the sodium iodide symporter (NIS);
- Apical transport of iodide across the basolateral membrane
- Incorporation of iodide into thyroglobulin
- TH synthesis
- TSH-Receptor interactions
- TRH-Receptor function
- Serum transport of Thyroid Hormone
- Membrane interactions and cross-membrane transport
- TH-TH-Receptor interactions
- Hepatic clearance

See Chapter 4 in the November 2017 FIFRA SAP White Paper available at:
https://www.regulations.gov/document?D=EPA-HQ-OPP-2017-0214-0012

Simplified Workflow

Literature search strategies and MIE modularization

- To guide the exploration, we created a thyroid pathway literature map (250k articles) by text mining expert-generated key thyroid pathway terms against chemicals for PubMed citations. MeSH terms will be extended using ontological approaches to mine other standard databases such as Web of Science, ECOTOX KB, Agricola, TOXNET, etc.
- Modules of the literature map are refined using 34 candidate thyroid reference chemicals identified in Wegner, et al (2016) DOI: 10.1016/j.reprotox.2016.08.016.

Thyroid pathway literature map
Uses AOP stages/categories heatmap

Human_Binding Proteins	Human_Deiodinases	Human_Hepatic Catabolism	Human_Hormones	Human_Sodium iodide symporter	Human_Synthesis	Human_Thyroid Hormone Receptors	Human_Thyroid stimulating hormone receptor	Human_Thyrotropin releasing hormone receptor	Human_TR Controlled Genes	Human_Transporters	Human_Energy	Human_Thyroid Gland	Human_Cognition_IQ	Human_Clinical Conditions	Chemical Name
133	234	0	3217	59	438	5	23	0	0	12	1	5935	143	3446	Iodine
4	36	4	678	1	17	17	1	0	0	19	0	528	5	1293	Amiodarone
1	30	4	2500	0	37	2	38	1	0	0	2	216	6	799	Methimazole
1	30	0	2096	0	20	2	8	0	0	1	0	188	7	727	Propylthiouracil
2	2	0	363	0	12	0	2	0	0	0	45	453	169	565	Lithium

Figure 2. Excerpt of the thyroid literature map

- After MIE module refinement, we use Natural Language Processing (NLP) to topic model based on bag-of-words approaches to exclude clinical studies (target in vitro & in vivo mechanistic studies) for the remainder of studies in a particular MIE module.
- Following manual inventorying of an MIE Module, the entire literature space is "rescored" using NLP approaches to identify other articles that might fit within an MIE Module.
- We are also building correlation models between titles/abstracts and the full-texts to improve targeting of screening questions and understand biases in title/abstract writing.
- MIE modules will be explored based on FIFRA SAP Whitepaper and future developments.

Expert manual curation

The iterative approach uses SR Review Software to query thyroid experts at two distinct stages:

Evidence Mapping (current)	Data extraction (future)
1. Gather basic study information 2. Identify mechanistic studies relevant to Thyroid-related Molecular Initiating Events (MIEs) 3. Sort studies by MIE 4. Determine if there is original mechanistic data for possible extraction 5. Evaluate paper quality based on high-level "TSCA-like"/"IRIS-like" criteria	1. Evaluate experiment quality based on high-level "TSCA-like"/"IRIS-like" criteria 2. Extract data on dose-response for each interrogated chemical. 3. Evaluate experiment extracted data based on "SciRAP-like" criteria

Evidence integration approaches to support the 3 primary goals (molecular mechanisms; test methods; and reference chemicals) will be developed as the evidence mapping supports continuation of work.

Pilot Case Study

Selected MIE: Iodide Uptake Inhibition through Sodium Iodide Symporter
Adverse Outcomes: Neurological and other developmental abnormalities . . .
Reasons for selection:

- Considered to be an early MIE in the thyroid pathways
- Rate limiting process for thyroid hormone synthesis
- Specifically relevant to EDSP and broadly relevant to human health and clinical outcomes
- Relatively well-understood and simple mechanism with many publications will enable us to readily determine effectiveness, efficiency and completeness of the review framework.

Evaluating Quality of Mechanistic Studies

Reliability (internal validity) of the data will be evaluated at 3 stages using domain approaches (evaluation based on overall considerations in each area):

- 1st tier: Screening of overall paper quality based on high-level transparent domain approach, adapted for type of study -> do we review experiments to extract from paper?
- 2nd tier: Screening of overall initial experiment quality based on high-level transparent domain approach, adapted for type of study -> do we extract from experiments?
- 3rd tier: Specific reporting and methodological reliability nominally based on SCIRAP approach, adapted for type of study -> what is the reliability of the experiment?

Evaluation Domain	Definition
Test Substance	Whether information provided in the study provides a reliable confirmation that the test substance used in a study is characterized for identity, purity, and properties.
Exposure Characterization	Assess the validity and reliability of methods used to measure or characterize exposure. Whether exposure to the test substance was characterized using a method(s) that provides valid and reliable results, whether the exposure remained consistent over the duration of the experiment, and whether the exposure levels were appropriate to the outcome of interest.
Test Design	Whether experimental design enables the study to distinguish the effect of exposure from other factors. This domain includes metrics related to the use of control groups and randomization in allocation to ensure that the effect of exposure is isolated.
Test Organism/ Test Model	Assess the appropriateness of the population or organism(s), group sizes used in the study (i.e., number of organisms and/or number of replicates per exposure group), and the organism conditions to assess the outcome of interest associated with the exposure of interest.
Outcome Assessment	Assess the validity and reliability of methods, including sensitivity of methods, that are used to measure or otherwise characterize the outcome(s) of interest.
Data Presentation and Analysis	Assess whether appropriate statistical methods were used and if data for all outcomes are presented.

Table 1. 1st tier evaluation guidelines based on overall paper quality. Reviewers give a single overall answer.

Iterative Pathway EM & SR Conceptual Model

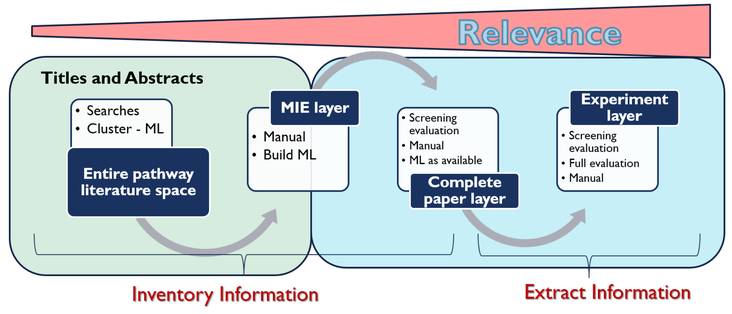


Figure 3. Conceptual model of the entire evidence mapping and systematic review framework

U.S. Environmental Protection Agency
Office of Chemical Safety and Pollution Prevention

Printed on 100% recycled/recyclable paper with a minimum 50% post-consumer fiber using vegetable-based ink.