EPA
United States
Environmental Protection
Agency

Systematic Review Approaches in Environmental Health Sciences

June 26, 2018

Kristina Thayer

Director, Integrated Risk Information System (IRIS)
U.S. Environmental Protection Agency (US EPA)
Office of Research Development
National Center for Environmental Assessment

Disclaimer: Author's views expressed today do not necessarily reflect the views or policies of the U.S. EPA

EPA
United States
Environmental Protection
Agency

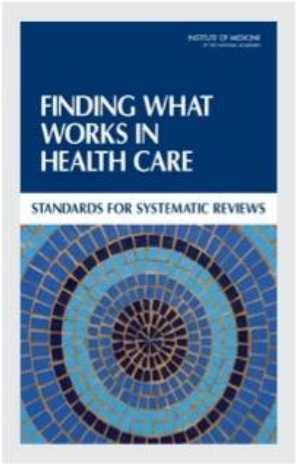
Disclaimer: The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Outline

- **What is systematic review and why should we do it?**
- **Core phases**
  - Frame the question and develop PECO (**P**opulation, **E**xposure, **C**omparator, **O**utcome) criteria
  - Develop a protocol
  - Literature search and screening
  - Individual study quality evaluation ("risk of bias")
  - Approach for assessing confidence in a body of evidence (aka "strength of evidence", "weight of evidence," "evidence synthesis," "evidence integration")

Systematic Review

A structured and documented process for transparent literature review[1]

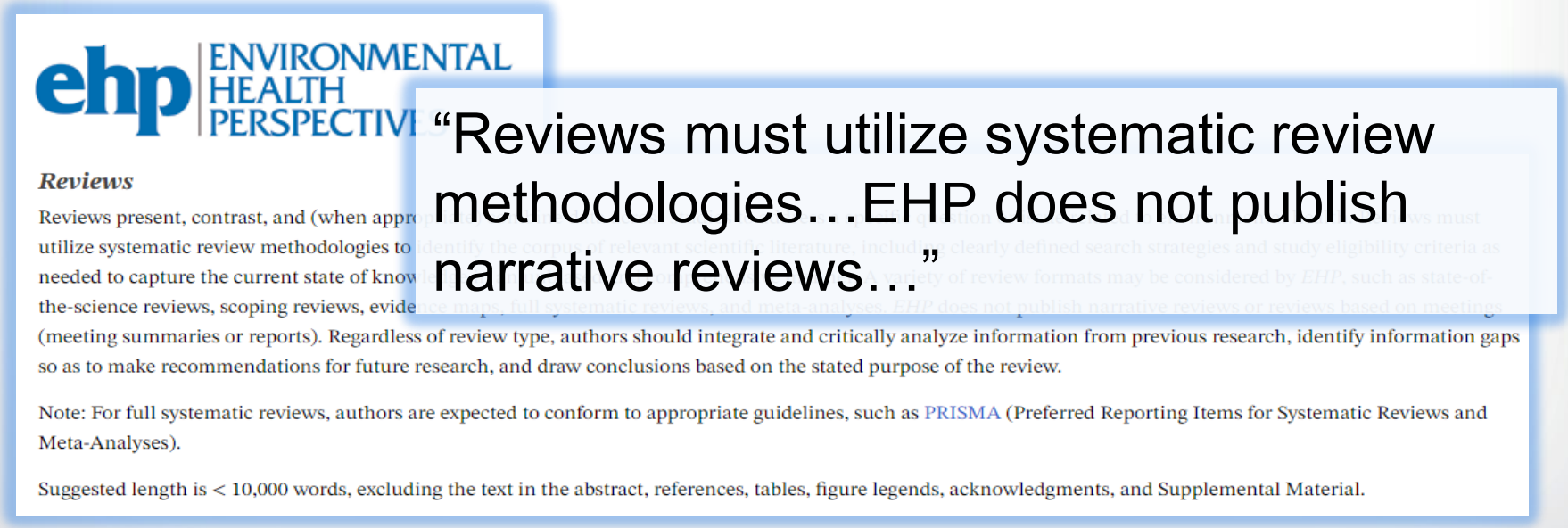


"As defined by IOM [Institute of Medicine], systematic review 'is a scientific investigation that focuses on a specific question and uses explicit, pre-specified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies."

[1] Institute of Medicine. Finding What works in Health Care: Standards for Systematic Reviews. p.13-34. The National Academies Press. Washington, D.C. 2011

Why Systematic Review Matters

- **Enhances transparency and minimizes bias**
- **Can make assessments more "reproducible" BUT not guaranteed - there will be legitimate differences in expert judgements**
- **State of the science – harder and harder to publish narrative reviews**



ehp ENVIRONMENTAL HEALTH PERSPECTIVES

Reviews

Reviews present, contrast, and (when appr
utilize systematic review methodologies to
needed to capture the current state of know
the-science reviews, scoping reviews, evide
(meeting summaries or reports). Regardless of review type, authors should integrate and critically analyze information from previous research, identify information gaps so as to make recommendations for future research, and draw conclusions based on the stated purpose of the review.

Note: For full systematic reviews, authors are expected to conform to appropriate guidelines, such as PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).

Suggested length is < 10,000 words, excluding the text in the abstract, references, tables, figure legends, acknowledgments, and Supplemental Material.

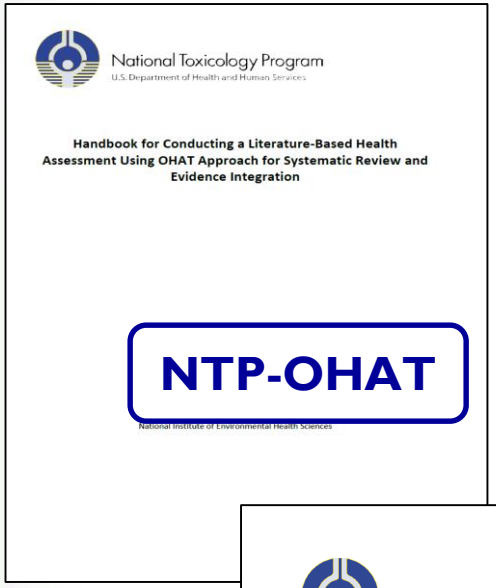
"Reviews must utilize systematic review methodologies…EHP does not publish narrative reviews…"

- **Created in 1985 to foster consistency in the evaluation of chemical toxicity across the Agency.**
- **IRIS assessments contribute to decisions across EPA and other health agencies.**
- **Toxicity values**
  - Noncancer: Reference Doses (RfDs) and Reference Concentrations (RfCs).
  - Cancer: Oral Slope Factors (OSFs) and Inhalation Unit Risks (IURs).
- **IRIS assessments have no direct regulatory impact until they are combined with**
  - Extent of exposure
  - Regulatory options
  - Both of these are t

Congressional budget language for IRIS FY18

the program to do so, while also encouraging the program to ensure that all IRIS methodologies attain the highest scientific rigor. Finally, the Committees urge the expedited completion of the IRIS handbook and direct that the public be afforded an opportunity to provide comment on the handbook before it is placed in use.

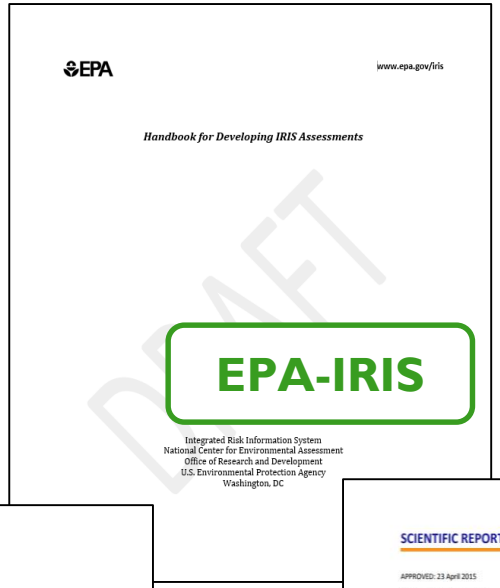
Approaches


National Toxicology Program
U.S. Department of Health and Human Services

Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration

National Institute of Environmental Health Sciences

NTP-OHAT

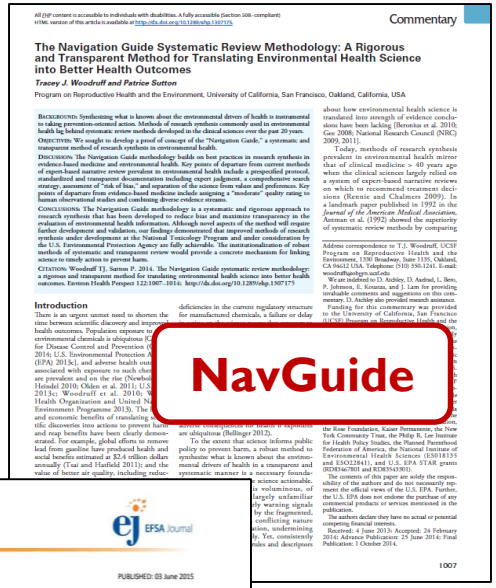

EPA

www.epa.gov/iris

Handbook for Developing IRIS Assessments

Integrated Risk Information System
National Center for Environmental Assessment
Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC

EPA-IRIS

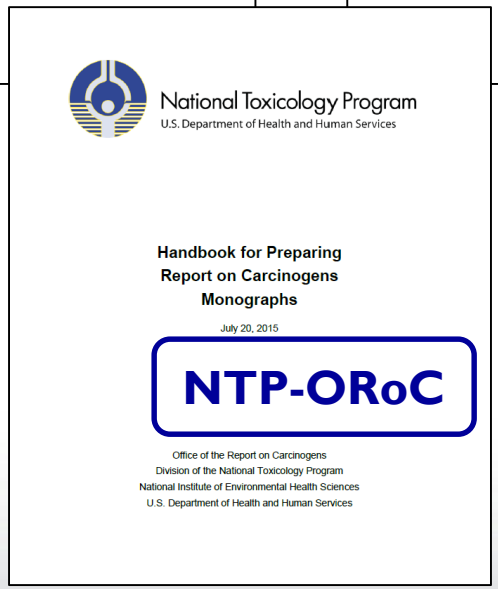

Commentary

The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes

Tracey J. Woodruff and Patrice Sutton

Program on Reproductive Health and the Environment, University of California, San Francisco, Oakland, California, USA

NavGuide

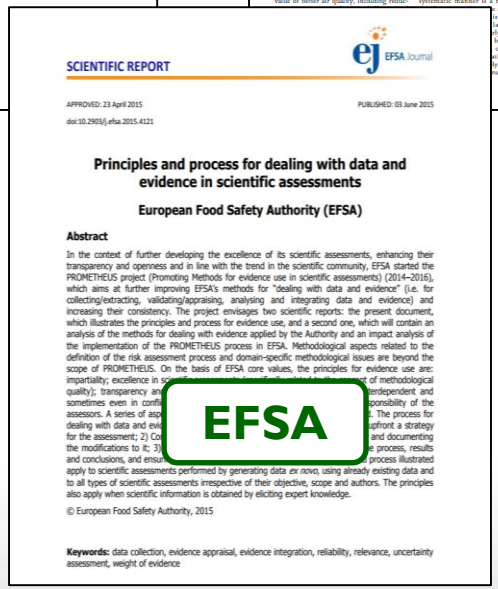

National Toxicology Program
U.S. Department of Health and Human Services

Handbook for Preparing Report on Carcinogens Monographs

July 20, 2015

Office of the Report on Carcinogens
Division of the National Toxicology Program
National Institute of Environmental Health Sciences
U.S. Department of Health and Human Services

NTP-ORoC


SCIENTIFIC REPORT

EFSA Journal

APPROVED: 23 April 2015

PUBLISHED: 03 June 2015

Principles and process for dealing with data and evidence in scientific assessments

European Food Safety Authority (EFSA)

Abstract

© European Food Safety Authority, 2015

Keywords: data collection, evidence appraisal, evidence integration, reliability, relevance, uncertainty assessment, weight of evidence

EFSA

Define the Question(s) and Develop PECO

- Define scope and focus of the review
- Develop PECO criteria (based on PICO used in clinical or health-care based systematic reviews)
  - **P**opulation (or **P**articipants)
  - **E**xposure (modified from **I**nterventions)
  - **C**omparators
  - **O**utcomes
- PECO guides literature search strategy and screening criteria

Example of a Targeted PECO

Step 1. Specify the Study Question

Our objective was to answer the question: "Does fetal developmental exposure to PFOA affect fetal growth in humans?" We developed a PECO (participants, exposure, comparator, and outcomes) statement, which is used as an aid to developing an answerable question (Higgins and Green 2011). Our PECO statement included the following:

Participants: humans who are studied during the reproductive/developmental time period (before and/or during pregnancy or development)

Exposure: exposure to PFOA (CAS# 335-67-1) or its salts during the time before pregnancy and/or during pregnancy for females or directly to fetuses

Comparators: humans exposed to lower levels of PFOA than the more highly exposed humans (i.e., a comparison across a range of exposures)

Outcomes: effects on fetal growth, birth weight, and/or other measures of size, such as length.

Citation: Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, Robinson KA, Axelrad DA, Woodruff TJ. 2014. The Navigation Guide—evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. Environ Health Perspect 122:1028–1039; http://dx.doi.org/10.1289/ehp.1307893

Example of a Broad PECO

PECO element	Evidence
Populations	**Human:** Any population and life-stage (occupational or general population, including children and other sensitive populations). **Animal:** Nonhuman mammalian animal species (whole organism) of any life-stage (including preconception, in utero, lactation, peripubertal, and adult stages).
Exposures	Relevant forms: [chemical x] (*CAS number*) *Other forms of* [chemical x] *that readily dissociate (e.g., list any salts, etc.)* *Metabolites of interest* *Indicate whether mixture studies are included.* **Human:** Any exposure to [chemical X] [via [oral or inhalation] route[s] if applicable]. *Specify if certain exposure assessment methods will NOT be included.* **Animal:** Any exposure to [chemical X] via [oral or inhalation] route[s]. *Specify if certain exposures/study designs will NOT be included, or if a minimum number of dose or concentration levels tested in experimental animal studies is indicated.* Studies involving exposures to mixtures will be included only if they include exposure to [chemical X] alone. Other exposure routes, including [dermal or injection], will be tracked during title and abstract as "potentially relevant supplemental information."
Comparators	**Human:** A comparison or referent population exposed to lower levels (or no exposure/exposure below detection limits) of [chemical X], or exposure to [chemical X] for shorter periods of time. Case reports and case series will be tracked as "potentially relevant supplemental information." **Animal:** A concurrent control group exposed to vehicle-only treatment or untreated control.
Outcomes	All health outcomes (both cancer and noncancer). As discussed above, based on preliminary screening work, EPA anticipates that a systematic review for health effect categories other than those identified (i.e., health effect 1, health effect 2…) will not be undertaken unless a significant amount of new evidence is found upon review of references during the comprehensive literature search.

Source: IRIS Protocol Template

Supplemental Materials

Major categories of "Potentially Relevant Supplemental Material"

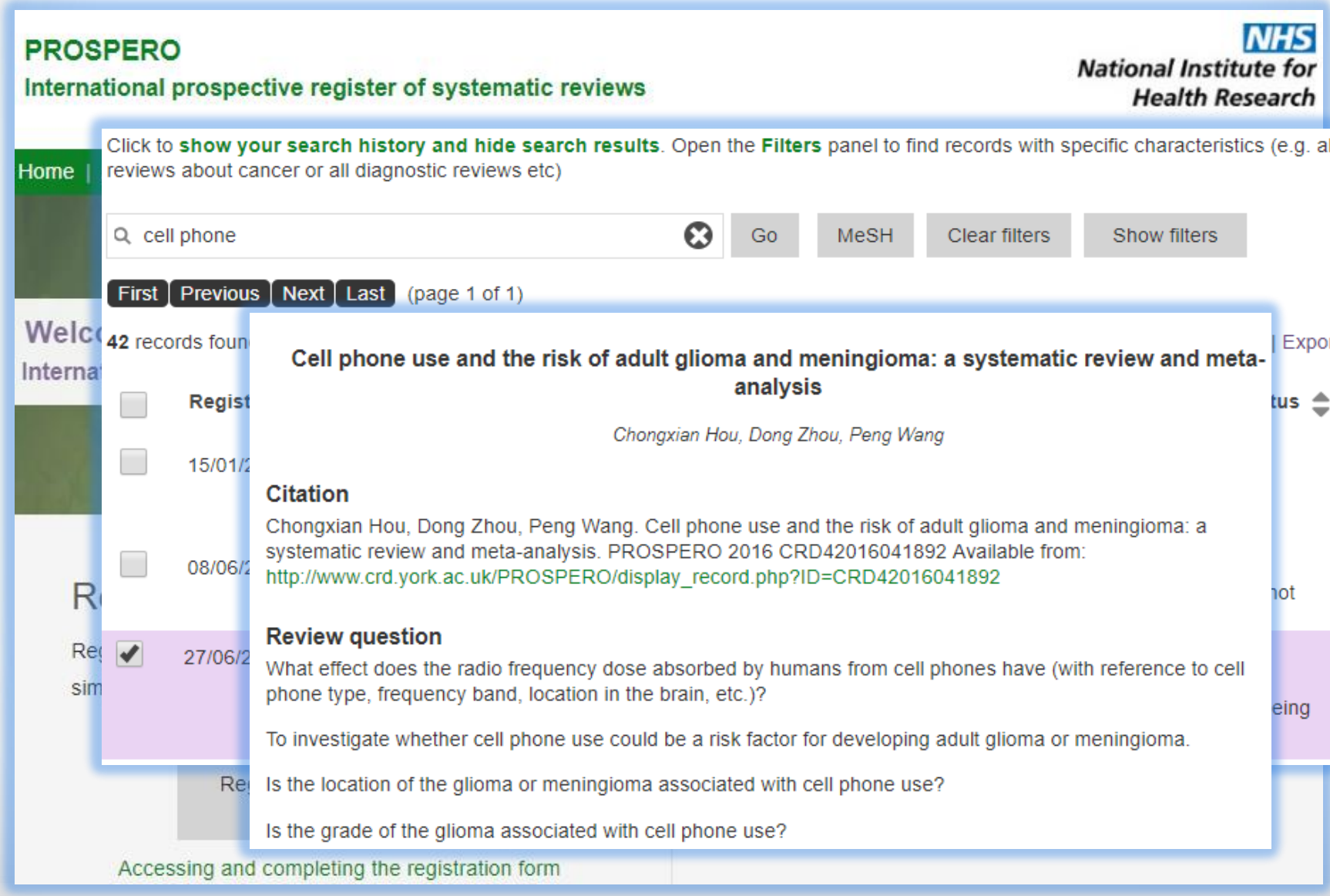
Category	Evidence
Mechanistic	Studies reporting measurements related to a health outcome that inform the biological or chemical events associated with phenotypic effects, in both mammalian and non-mammalian model systems, including in vitro, in vivo (by various routes of exposure), ex vivo, and in silico studies.
ADME and toxicokinetic	Studies designed to capture information regarding absorption, distribution, metabolism, and excretion, including toxicokinetic studies. Such information may be helpful in updating or revising the parameters used in existing PBPK models.
Exposure characteristics	Exposure characteristic studies include data that are unrelated to toxicological endpoints, but which provide information on exposure sources or measurement properties of the environmental agent (e.g., demonstrate a biomarker of exposure).
Susceptible populations	Studies that identify potentially susceptible subgroups; for example, studies that focus on a specific demographic, life-stage, or genotype.
Mixture studies	Mixture studies that are not considered PECO-relevant because they do not contain an exposure or treatment group assessing only the chemical of interest.
Routes of exposure not pertinent to PECO	Studies utilizing routes of exposure that fall outside the PECO scope.
Case studies or case series	In most cases, case reports and case series will be tracked as potentially relevant supplemental information.

Source: IRIS Protocol Template

Systematic Review Reporting Quality Tools

- Journals like to see the protocol as supplemental material and ideally it has been registered before being implemented
- Government-initiated reviews often undergo peer-review and public comment

Prospero


PROSPERO
International prospective register of systematic reviews
NHS
National Institute for Health Research
Home
Click to show your search history and hide search results. Open the Filters panel to find records with specific characteristics (e.g. all reviews about cancer or all diagnostic reviews etc)
cell phone
Go
MeSH
Clear filters
Show filters
First
Previous
Next
Last
(page 1 of 1)
42 records found
Export
Cell phone use and the risk of adult glioma and meningioma: a systematic review and meta-analysis
Chongxian Hou, Dong Zhou, Peng Wang
Citation
Chongxian Hou, Dong Zhou, Peng Wang. Cell phone use and the risk of adult glioma and meningioma: a systematic review and meta-analysis. PROSPERO 2016 CRD42016041892 Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016041892
Review question
What effect does the radio frequency dose absorbed by humans from cell phones have (with reference to cell phone type, frequency band, location in the brain, etc.)?
To investigate whether cell phone use could be a risk factor for developing adult glioma or meningioma.
Is the location of the glioma or meningioma associated with cell phone use?
Is the grade of the glioma associated with cell phone use?
Accessing and completing the registration form

IRIS Protocol Content

3. OVERALL OBJECTIVES, SPECIFIC AIMS, AND POPULATIONS, COMPARATORS, EXPOSURE OUTCOMES (PECO) CRITERIA

The overall objective of this assessment is to identify adverse health effects and characterize exposure-response relationships for these effects of chloroform to support development of toxicity values for this chemical. More specifically, the objective of this is to derive an RfC for chloroform by using inhalation dose-response data from human studies, without the need for route-to-route extrapolation. In addition, the MOA analysis for chloroform derived RfC that methods evaluation

3.1. SI

- state, an
- Registry
- Occupati
- Preventi
- Assessm
- Environ
- Internat
- the Envi
- on Chem
- perform
- Chemica

4.2. LI

the last I
EPA's He
identifie
updated
only on t
in silico)
is preser
range of

4. LITERATURE SEARCH AND SCREENING STRATEGIES

APPENDICES

APPENDIX A. ELECTRONIC DATABASE SEARCH STRATEGIES

Table A-1. Database search strategy

Search	Search Strategy
PUBMED	((("chloroform"[MeSH Terms] OR "1,1,1-trichloromethane"[All Fields]) OR "chloroforme"[All Fields]) OR "trichloromethane"[All Fields]) OR "67-66-3"[EC/RN Number]) AND ("2009"[PDAT] : "3000"[PDAT])
WEB OF SCIENCE	(TS="chloroform" OR TS="1,1,1-trichloromethane" OR TS="chloroforme" OR TS="trichloromethane") AND PY=(2009-2017) NOT (SU="PHYSICS" OR SU="PLANT SCIENCES" OR SU="ENERGY FUELS" OR SU="INSTRUMENTS INSTRUMENTATION" OR SU="COMPUTER SCIENCE" OR SU="LEGAL MEDICINE" OR SU="METALLURGY METALLURGICAL ENGINEERING" OR SU="MECHANICS" OR SU="EDUCATION EDUCATIONAL RESEARCH" OR SU="ACOUSTICS" OR SU="GEOCHEMISTRY GEOPHYSICS" OR SU="MATHEMATICS" OR SU="FORESTRY" OR SU="AUTOMATION CONTROL SYSTEMS" OR SU="MINING MINERAL PROCESSING" OR SU="CONSTRUCTION BUILDING TECHNOLOGY" OR SU="ASTRONOMY ASTROPHYSICS" OR SU="ARCHAEOLOGY" OR SU="OPERATIONS RESEARCH MANAGEMENT SCIENCE" OR SU="ANTHROPOLOGY" OR SU="SPORT SCIENCES" OR SU="ART" OR SU="PALEONTOLOGY" OR SU="TELECOMMUNICATIONS" OR SU="CHEMISTRY" OR SU="POLYMER SCIENCE" OR SU="ENGINEERING" OR SU="ENVIRONMENTAL SCIENCES ECOLOGY" OR SU="FOOD SCIENCE TECHNOLOGY" OR SU="SCIENCE TECHNOLOGY OTHER TOPICS" OR SU="BIOTECHNOLOGY APPLIED MICROBIOLOGY" OR SU="AGRICULTURE" OR SU="SPECTROSCOPY" OR SU="CRYSTALLOGRAPHY" OR SU="INTEGRATIVE COMPLEMENTARY MEDICINE" OR SU="WATER RESOURCES" OR SU="NUTRITION DIETETICS" OR SU="LIFE SCIENCES BIOMEDICINE OTHER TOPICS" OR SU="PARASITOLOGY" OR SU="THERMODYNAMICS" OR SU="OPTICS" OR SU="BIOPHYSICS" OR SU="TROPICAL MEDICINE" OR SU="VETERINARY SCIENCES" OR SU="RESEARCH EXPERIMENTAL MEDICINE" OR SU="MARINE FRESHWATER

6. STUDY EVALUATION (REPORTING, RISK OF BIAS, AND SENSITIVITY) STRATEGY

IRIS assessments evaluate each study's methods using uniform approaches for each group of similar studies s
concerns for the re
that affect the mag
study to detect a tr
animal toxicology s
supplemental mate
prominent role in t

Table 3. S

Epid
Exposure measurem
Outcome ascertainm
Participant selection
Confounding
Analysis
Selective reporting
Sensitivity

Study evalu
The study evaluati
limitations (focusi
result), considerin
null. The study eva
of the results) in th

7. DATA EXTRACTION OF STUDY METHODS AND RESULTS

Data extraction an
elements that may be colle
Choices about what data t
analyses that inform the s
following the identificatio
the data extraction workfl
extraction. Studies evalua
therefore, will not be cons
be less relevant during PE
minimal data extraction.
high confidence studies ar

The data extractio
available for download fr
[NOTE: The following bro
(preferred), Mozilla Foxfi
Internet Explorer.] Data e
independently checked by
by discussion or consultat
verified, they will be "lock
WebPlotDigitizer (http://
information from figures.

8. PHYSIOLOGICALLY BASED PHARMACOKINETIC (PBPK) MODEL IDENTIFICATION, DESCRIPTIVE SUMMARY, AND EVALUATION

PBPK (or classical pharmacokinetic [PK]) models should be used in an assessment when an applicable one exists and no equal or better alternative for dosimetric extrapolation is available. Any models used should represent current scientific knowledge and accurately translate the science into computational code in a reproducible, transparent manner. For a specific target organ/tissue, it may be possible to employ or adapt an existing PBPK model, or develop a new PBPK model or an alternate quantitative approach. Data for PBPK models may come from studies with animals or humans, and may be in vitro or in vivo in design.

8.1. IDENTIFYING PHYSIOLOGICALLY BASED PHARMACOKINETIC (PBPK) MODELS

PBPK modeling is the preferred approach for calculating a human equivalent concentration (HEC) according to the hierarchy of approaches outlined in EPA guidance (U.S. EPA, 2011a). For chloroform, metabolism is a major component of target organ toxicity, and PBPK models are available to account for interspecies differences in metabolism between rats, mice, and humans (Sasso et al., 2013; Corley et al., 1990). Chloroform is metabolized to the reactive metabolites phosgene and dichloromethyl free radical in humans and animals by cytochrome P450-dependent pathways (Gemma et al., 2003; Constan et al., 1999).

Because of the role of metabolism in the production of target organ toxicity, and the reactive

Source: IRIS Chloroform Protocol (2018)
https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=338653

Protocol Content

9. SYNTHESIS WITHIN LINES OF EVIDENCE

For each potential health effect
outcomes; or a broad hazard category),
effect evidence, ar
written to empha
the evidence inte
studies or group c
association, tempo
humans (U.S. EPA
Specificall
first be analyzed a
lack of data withir
the available mec
chloroform, a syn
evaluation of carc

9.1. SYNTHES

To assess

Table 9. Primar
syntheses[a]

Consideration	
Consistency	Repeated exist, the "differing" *Stronger h* *Stronger c*
Biological gradient (dose-response)[b]	Increases concentra or comple necessarily considered
Strength (effect magnitude) and precision	Given wha particular small effe may consi other expl errors and results acr (i.e., low p
Mechanistic evidence related to biological plausibility	Supporting effects; changes in established biol evidence strength. While a lack of strength, it may do so if findings de *Human evidence:* studies in expose *Animal evidence:* studies in exposed animals
Coherence[c]	Findings across the database that fit into a co similarity in results for related effects within a dose-dependent progression of linked effects Conversely, an observed lack of changes that subsequently) with the effect of interest coul informed by the known biological developme toxicokinetic/dynamic understanding of the c
Natural experiments	*Human evidence only*: Reductions in effect thi Although rare, such reductions can provide co
Temporality	*Human evidence only:* The exposure occurs be evaluation of exposure measures for each stu

10. INTEGRATION ACROSS LINES

For the analysis of most health outcomes, IRIS assessme
and mechanistic evidence. Depending on the assessment scope a
animal evidence, conclusions for mechanistic evidence may be b
mechanistic st
are drawn as f

- First, a
chemi
step in
cohere
- In para
the che

WITHIN STREAM CONCLUSIONS

HUMAN EVIDENCE STREAM CONCLUSION

The synthesis of evidence about health effects and mechanisms from human studies is combined (integrated) to draw a conclusion about effects within the stream

Studies and interpretation	Factors that increase confidence	Factors that decrease confidence	Summ
[Health Effect or Outcome Grouping]			
Evidence from Human Studies (Route)			
· References · Study confidence (based on evaluation of risk of bias and sensitivity) and explanation · Study design description	· Consistency · Dose-response gradient · Coherence of observed effects (apical studies) · Effect size (magnitude, severity) · Biological plausibility · Low risk of bias/ high quality · Insensitivity of null/ negative studies · Natural experiments · Temporality	· Unexplained inconsistency · Imprecision · Indirectness/ applicability · Poor study quality/ high risk of bias · Other (e.g., Single/Few Studies; small sample size) · Evidence demonstrating implausibility	Results infor affected/ una Human evid plausibility: d data influenc judgement (precursors in Could be multiple rows (e.g study confidence or popul informs results hetero
Evidence for an Effect in Animals (Route)			
· References · Study confidence (based on evaluation of risk of bias and sensitivity) and explanation · Study design description	· Consistency and Replication · Dose-response gradient · Coherence of observed effects (apical studies) · Effect size (magnitude, severity) · Biological plausibility · Low risk of bias/ high quality · Insensitivity of null/ negative studies	· Unexplained inconsistency · Imprecision · Indirectness/ applicability · Poor study quality/ high risk of bias · Other (e.g., Single/Few Studies; small sample size) · Evidence demonstrating implausibility	Results information (gene affected/ unaffected) acro Evidence informing biolo plausibility for effects in a discuss how mechanistic influenced the within stream judgement (e.g., evidence of coherent molecular changes in animal studies) Could be multiple rows (e.g., by study confidence, species, or exposure duration) if this informs results heterogeneity

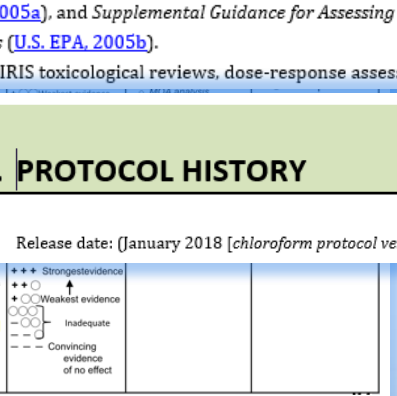


Figure 4. Evidence profile table template.

11. DOSE-RESPONSE ASSESSMENT: STUDY SELECTION AND QUANTITATIVE ANALYSIS

The previous sections of this protocol describe how systematic review principles are applied to support transparent identification of health outcomes (or hazards) associated with exposure to the chemical of interest in conjunction with evaluation of the quality of the studies considered during hazard identification. Selection of specific data for dose-response assessment and performance of the dose-response assessment is conducted after hazard identification is complete, and builds off this step in developing the complete IRIS assessment. The dataset selection process involves database- and chemical-specific biological judgments that are beyond the scope of this protocol, but are discussed in existing EPA guidance and support documents. This section of the protocol provides an overview of points to consider when conducting the dose-response assessment, particularly statistical considerations specific to dose response analysis that support quantitative risk assessment. Importantly, the considerations outlined in this protocol do not supersede existing EPA guidance. Several EPA guidance and support documents provide more detailed considerations for the development of EPA's traditional dose-response values, especially EPA's *Review of the Reference Dose and Reference Concentration Processes* (U.S. EPA, 2002), EPA's *Benchmark Dose Technical Guidance* (U.S. EPA, 2012b), *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 2005a), and *Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens* (U.S. EPA, 2005b).

For IRIS toxicological reviews, dose-response assessments are typically performed for both

12. PROTOCOL HISTORY

Release date: (January 2018 [*chloroform protocol version 1*])

Source: IRIS Chloroform Protocol (2018)
https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=338653

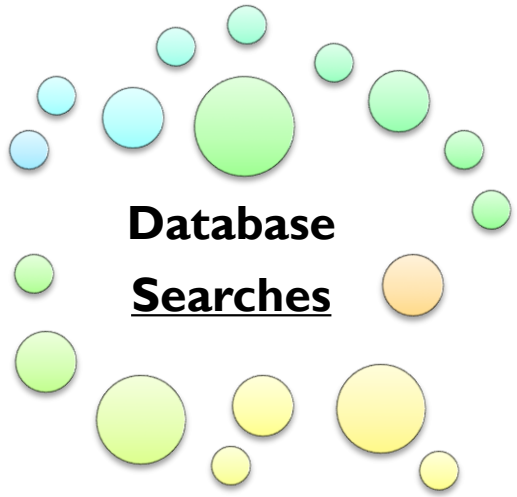
EPA
United States
Environmental Protection
Agency

Literature Searching, Screening, and Inventories*

Office of Research and Development
NCEA, IRIS

*** includes basic methodological details**

Common Literature Searching and Screening Processes



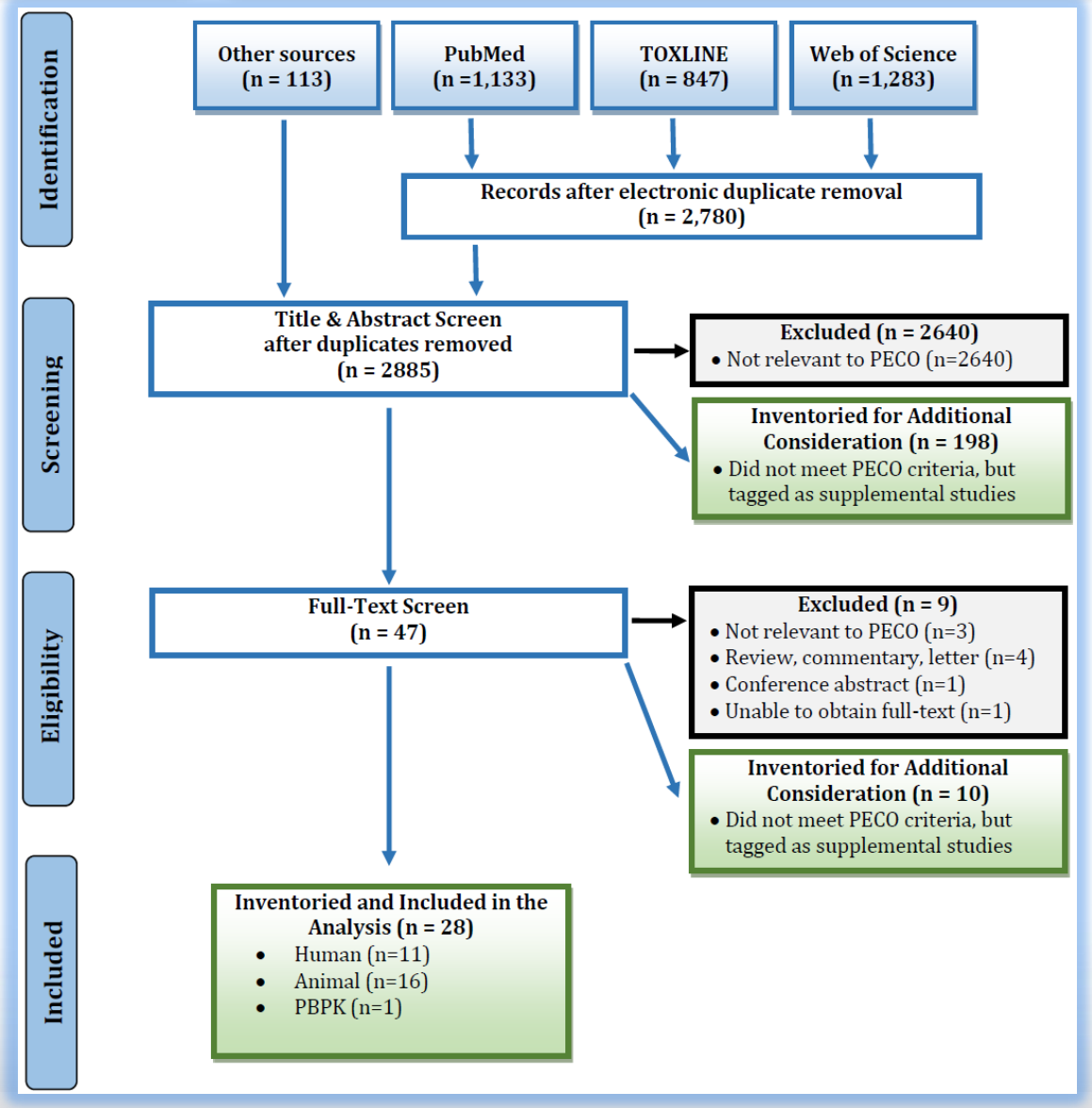
Screening

1. Title/abstract
2. Full text

- Identify peer-reviewed and "gray" (unpublished) literature
- e.g., PubMed, ToxLine, Web of Science, Scopus
- Typically do not apply language-restrictions
- Conduct regular search updates
- Details of search strategy, dates, and retrieved records are presented in protocols and assessments

- Use manual and automated approaches
- 2 independent screeners
- Tag studies as excluded, meeting PECO criteria, or supplemental information
- Review reference list of included studies and relevant reviews to identify studies missed from database searches

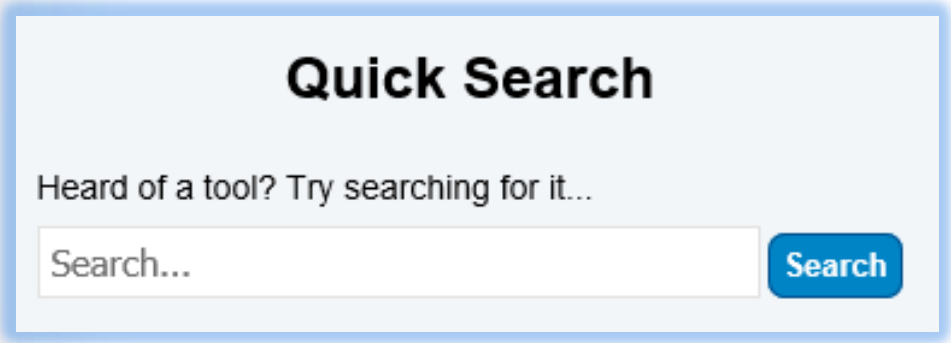
Literature Flow Diagrams

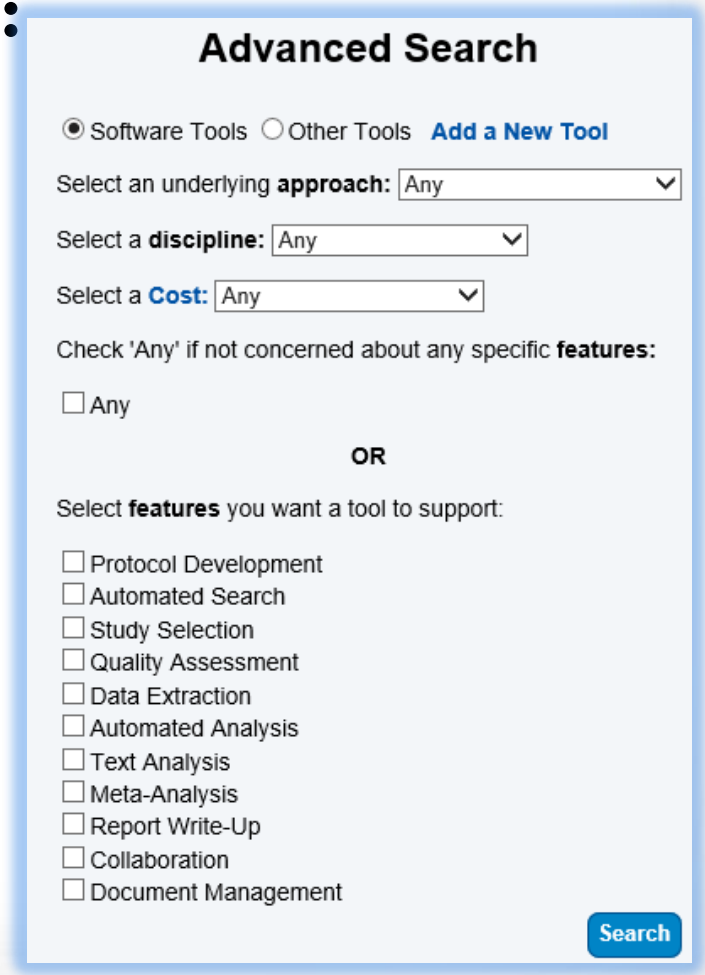


Use of Specialized Tools for Literature Search and Screening

Database of SR software tools:

http://systematicreviewtools.com/





Example Literature Screening Form

SUBMIT FORM and go to or Skip to Next

Forms Independently Entered by 2 Reviewers

1. **Based on Title and Abstract does the article contain relevant human, animal, or in vitro evidence?**

   ◯ Yes ◯ No ◉ No, but has supportive information
   ◯ Unclear (e.g., no abstract) Clear Response

2. **What kind of evidence or supportive information?**

   ☐ human
   ☐ animal
   ☑ in vitro, omics, alternative model systems

3. **What kind of supportive info**

   ☑ MOA/mech (cancer)
   ☐ MOA/mech (non-cancer)
   ☐ case report or poisoning
   ☐ non-inhalation route
   ☐ mixture
   ☐ ADME/PBPK
   ☐ exposure assessment
   ☐ THM, disinfection/chlorinati
   ☐ susceptible population
   ☐ anesthesia/inhalent

P	Human: Any population and life stage (occupational or general population, including children and other sensitive population). The following study designs will be considered most informative: controlled exposure, cohort, case-control, cross-sectional, and ecological. Note: Case reports and case series will be tracked during study screening, but are not the primary focus of this assessment. They may be retrieved for full-text review and subsequent evidence synthesis if no or few informative study designs are available. Case reports can also be used as supportive information to establish biologic plausibility for some target organs and health outcomes. Animal: Nonhuman mammalian animal species (whole organism) of any life stage (including preconception, in utero, lactation, peripubertal, and adult stages).
E	Human: Any exposure to chloroform, including occupational exposures, via inhalation. Exposures quantified by either actual exposure measurements or occupational exposure history are preferred. Studies of chloroform in the context of its use as an anesthetic gas will be excluded Animal: Any exposure to chloroform via inhalation. Studies employing chronic

Reference Status

Show 10 entries

	Unreviewed	Some Reviews	Included	Excluded	Conflict	Fully Reviewed
Level 1 - Title Abstract	420	1137	24	513	43	580
	Unreviewed	Some Reviews	Included	Excluded	Conflict	Fully Reviewed

- Use of machine-learning/natural language processing approached can reduce the screening burden by at least 50%

SUBMIT FORM and go to Skip to Next

Draft example based on chloroform using Distiller

EPA
United States
Environmental Protection
Agency

Evaluating Quality of Individual Studies

Office of Research and Development
NCEA, IRIS

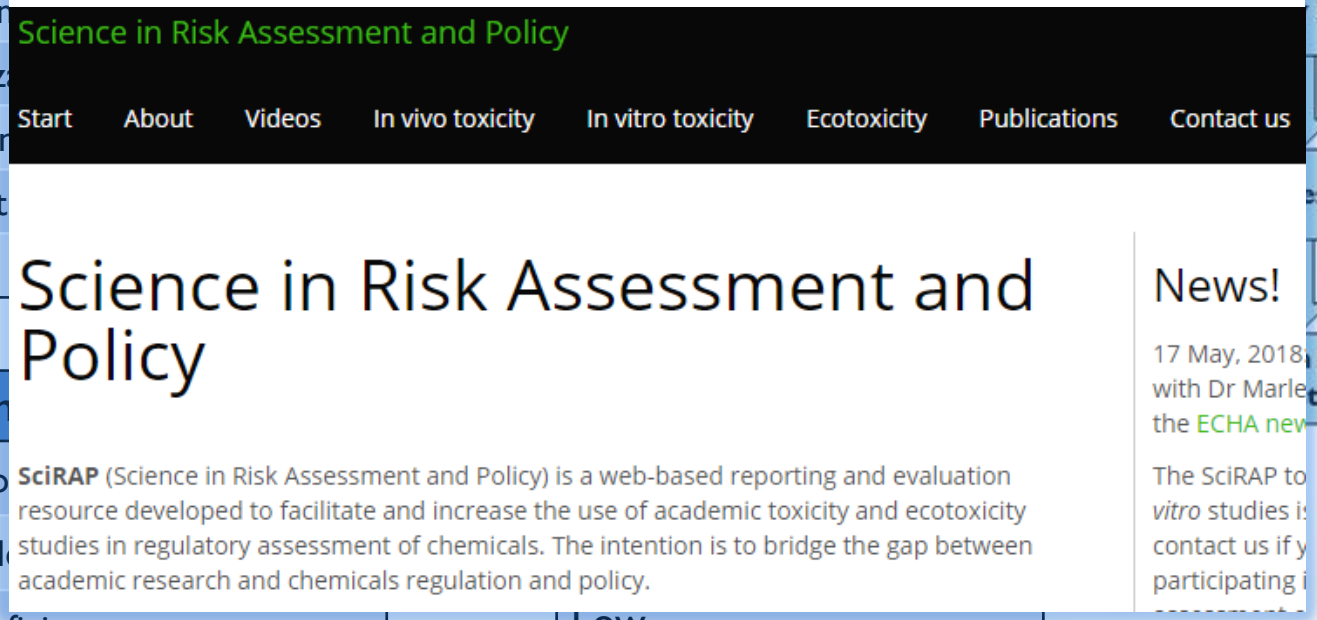
Aspects of Study Quality

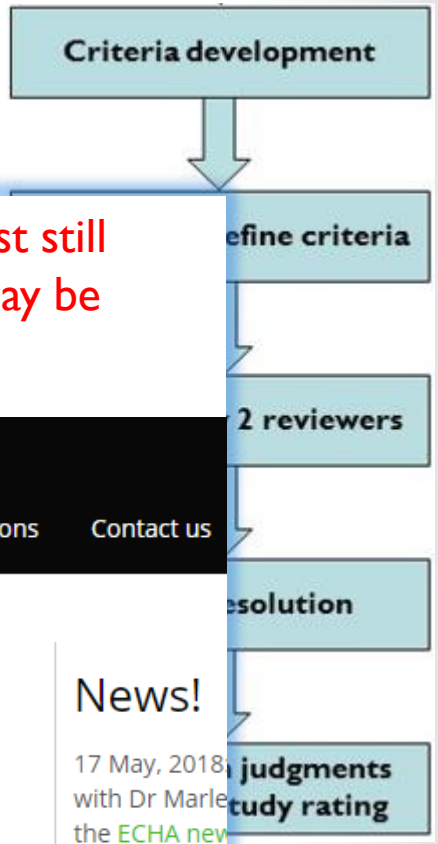
- Reporting quality
- Internal validity ("risk of bias")
- Applicability ("directness") to the topic

Example: EPA IRIS Approach

Individual study level domains	
Animal	Epidemiological
Reporting Quality	Exposure measurement
Allocation	
Blinding	
Variable Control	
Selective Reporting an	
Exposure Characteriz	
Utility of Study Design	
Outcome Assessment	
Results Presentation	

- Tools are under-developed for in vitro studies. Most still focus on reporting quality (e.g., ToxRTool). SciRap may be promising http://www.scirap.org/

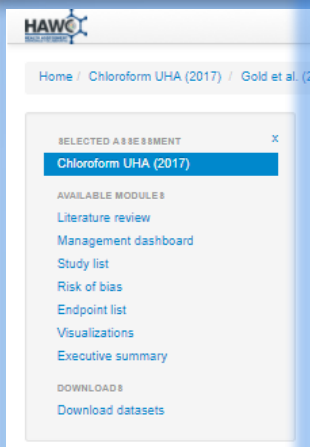


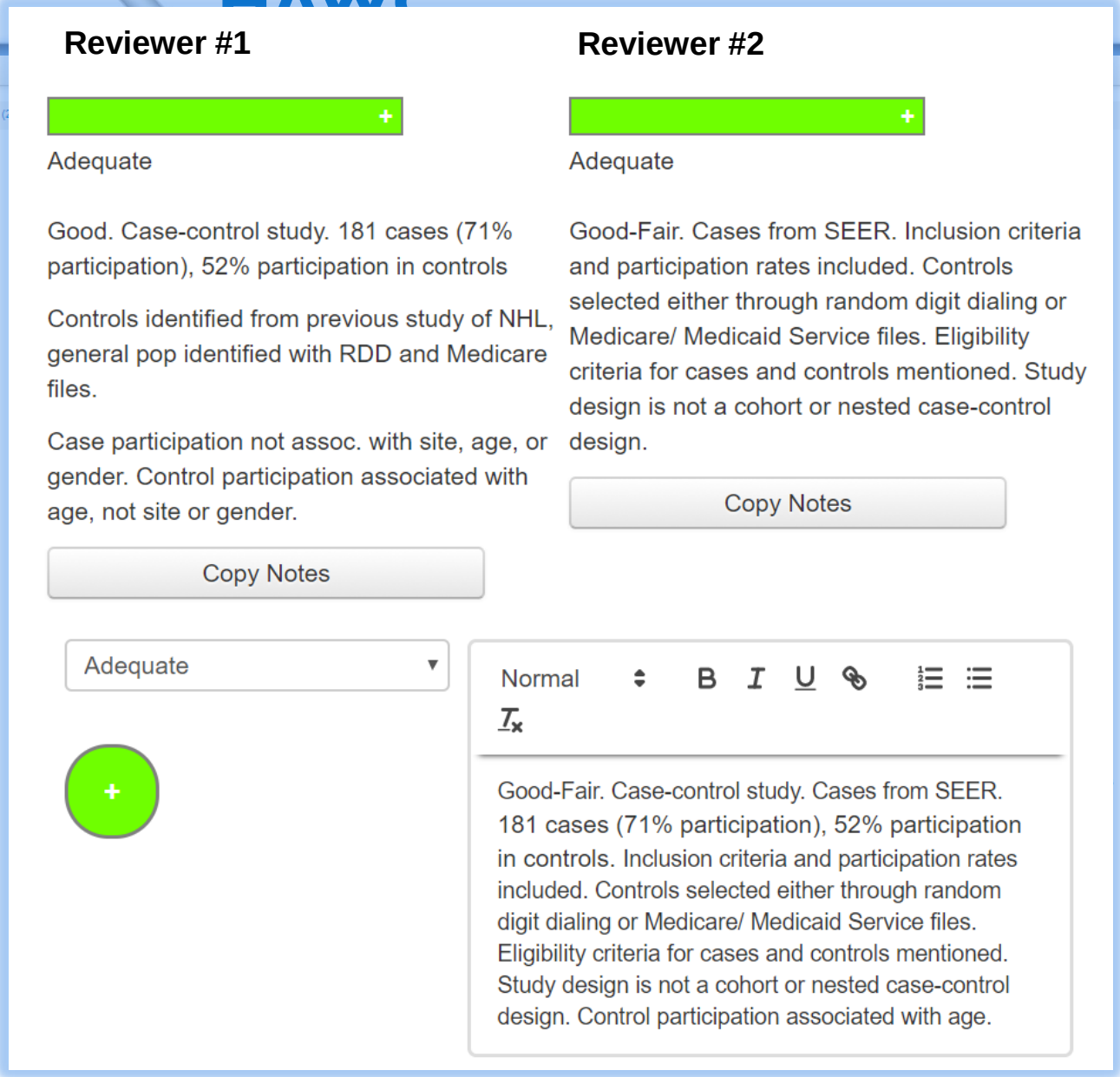


Dor	
++	Go
+	Ad
-	Deficient
--	Critically Deficient

Low
Uninformative

Study Evaluation Workflow in HAWC



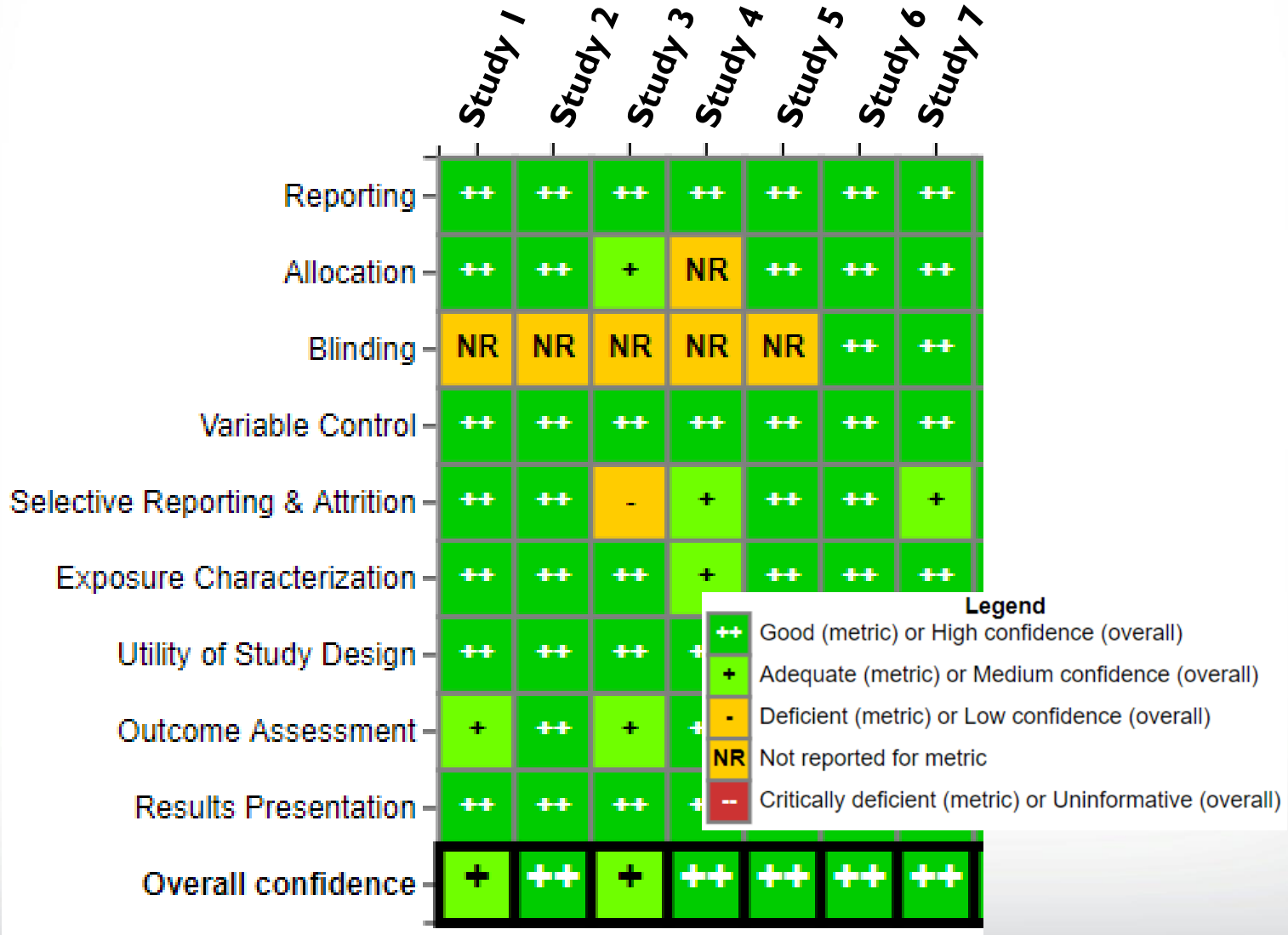


Individual Studies in HAWC

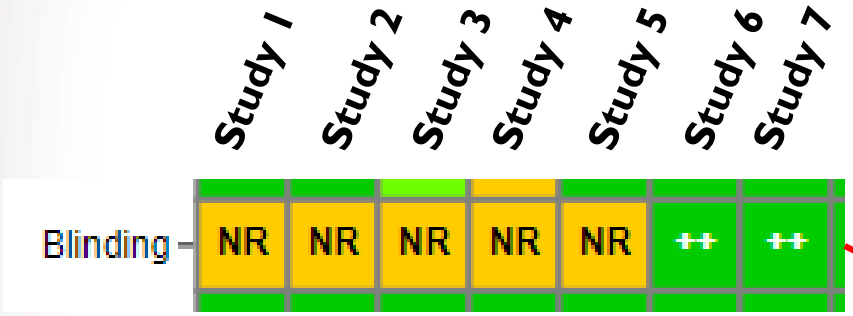
Medium confidence

Uninformative

Study Evaluation Summary in HAWC (Animal Studies)



Example Study Evaluation for Blinding



++

Good

Not fully blinded (interpreted as good)

Body and organ weights, clinical chemistry, hormone measurements: NTP standard pre-chronic or chronic studies are not blinded during in life portions for technical reasons associated with running large animal bioassays (i.e., blinding increases risk for dosing errors and hinders cage side recognition of abnormal behaviors). Potential concern for bias was mitigated for these endpoints which were measured using automated/computer driven systems, standard laboratory kits, relatively simple, objective measures (e.g., body or tissue weight). Additional details provided during personal communication (6/17/2018) clarified that data collection for recent NTP studies is heavily automated, including use of bar coded animal ID chips with wand chip readers, use of scales that automatically calculate daily dose based on body weight, use of scales for organ weights that are electronically captured into a Provantis system for data recording. Data are also analyzed with automated statistical packages. Outliers are flagged by the programs and a human asked to decide whether to censor or include.

Good

Histopathology: Blinding during the initial evaluation of tissues is generally not recommended as masked evaluation can make the task of separating treatment-related changes from normal variation more difficult and may result in subtle lesions being overlooked (Crissman, 2004). A blinded pathology working group (PWG) review was carried out on coded pathology slides to minimize the potential for observational bias.

EPA
United States
Environmental Protection
Agency

Assessing Confidence in a Body of Evidence (aka “strength of evidence”, “weight of evidence,” “evidence synthesis,” “evidence integration”)

Trends in Evidence Synthesis and Integration

- Recommended element in systematic review protocols

Section and topic	Item number	Checklist item
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)
FROM: Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1. doi: 10.1186/2046-4053-4-1		

- Integrating evidence across streams can be qualitative or quantitative, but qualitative is far more common
- Typically, conclusions are reached within evidence stream prior to integrating across streams

Develop Within Evidence Stream Conclusions Prior to Integrating Across

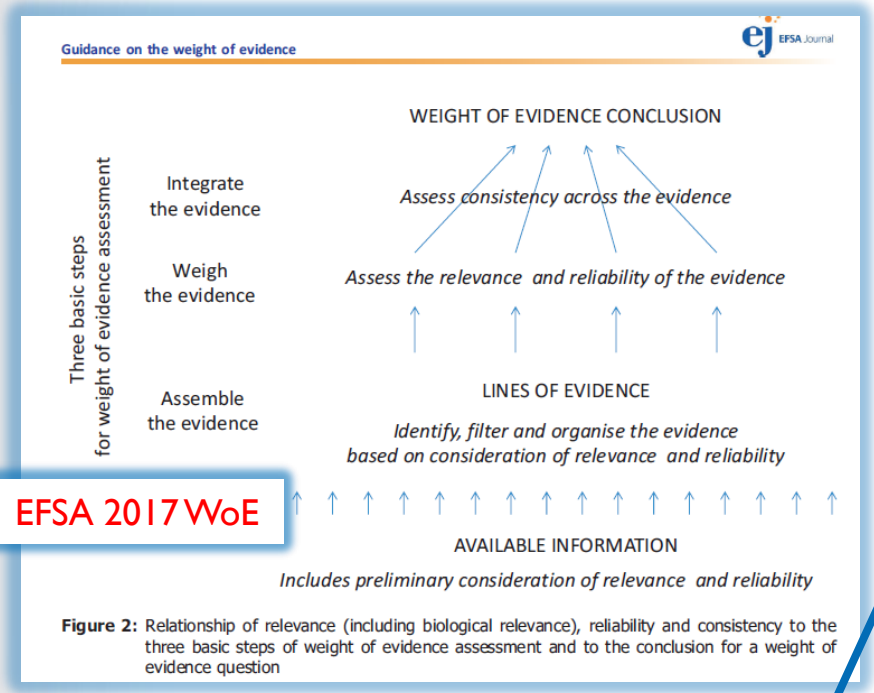
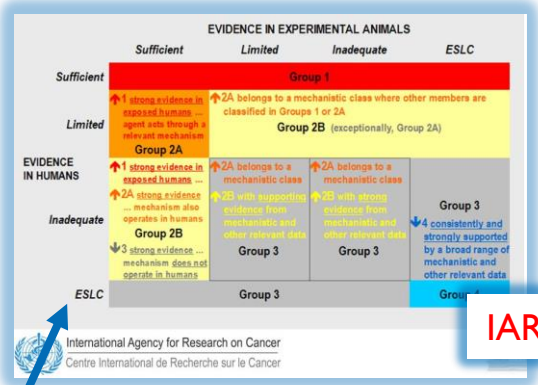
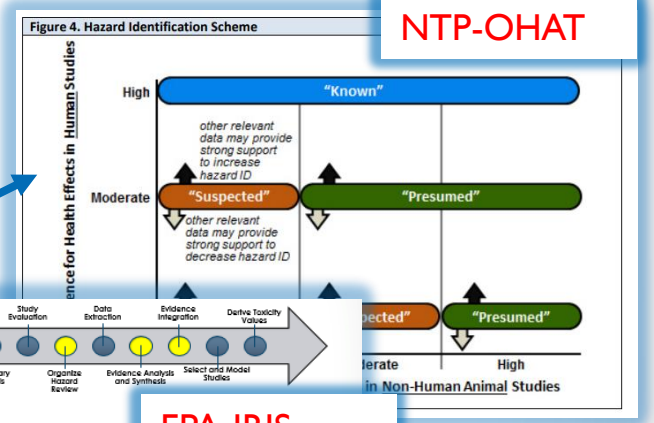


Figure 2: Relationship of relevance (including biological relevance), reliability and consistency to the three basic steps of weight of evidence assessment and to the conclusion for a weight of evidence question

EFSA 2017 WoE



IARC



NTP-OHAT

EPA-IRIS

mechanistic information used to increase/decrease integrated conclusions from human and nonhuman animal evidence

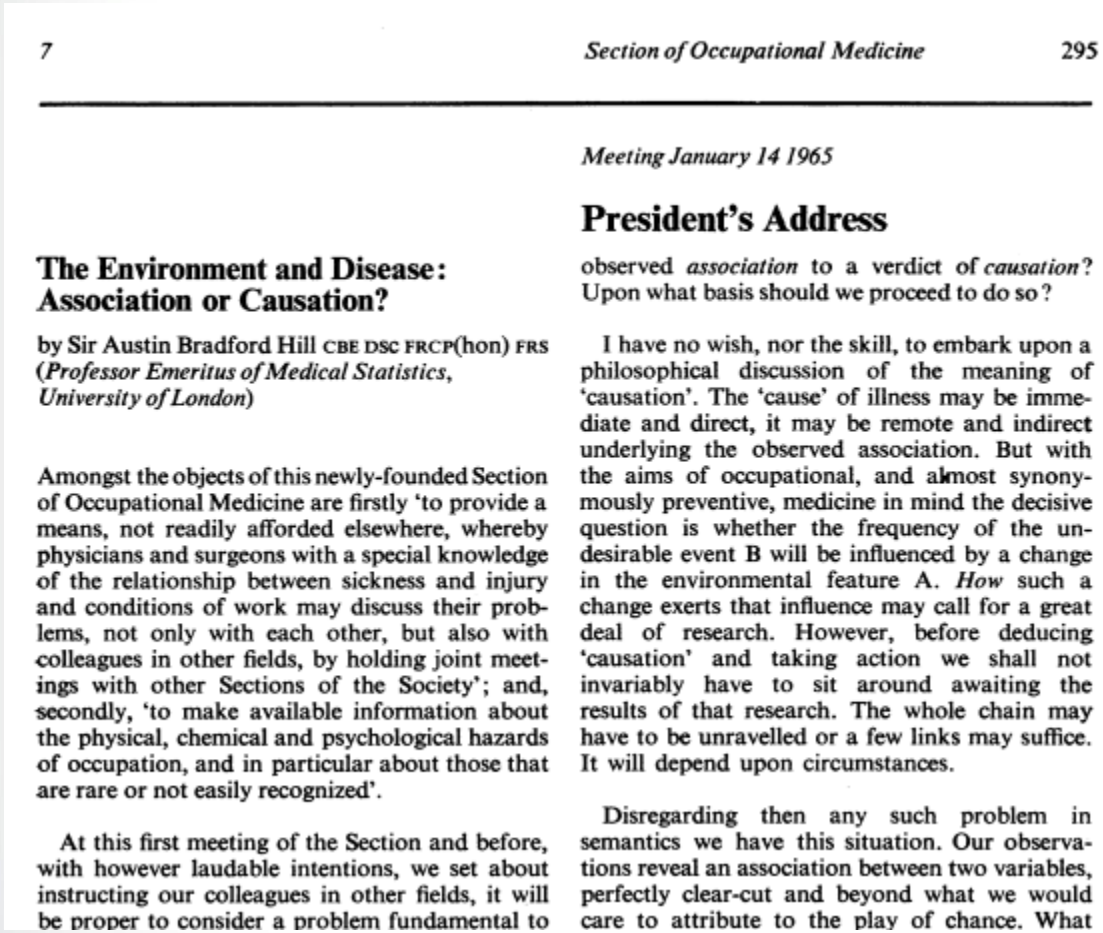
Organize and analyze evidence

- Synthesis of each line of evidence (human, animal and mechanistic evidence) - to identify important health effects potentially linked to exposure, and to analyze results to inform strength of evidence

Develop judgements regarding strength of evidence

- Integration within evidence streams – to develop judgements about the strength of evidence for health effects in each human and animal evidence stream incorporating mechanistic information
- Integration across evidence streams - to develop a conclusion about whether exposure to a substance may cause a health effect in humans

Hill Considerations



7 *Section of Occupational Medicine* 295

Meeting January 14 1965

The Environment and Disease: Association or Causation?

President's Address

by Sir Austin Bradford Hill CBE DSC FRCP(hon) FRS
(*Professor Emeritus of Medical Statistics, University of London*)

Amongst the objects of this newly-founded Section of Occupational Medicine are firstly 'to provide a means, not readily afforded elsewhere, whereby physicians and surgeons with a special knowledge of the relationship between sickness and injury and conditions of work may discuss their problems, not only with each other, but also with colleagues in other fields, by holding joint meetings with other Sections of the Society'; and, secondly, 'to make available information about the physical, chemical and psychological hazards of occupation, and in particular about those that are rare or not easily recognized'.

At this first meeting of the Section and before, with however laudable intentions, we set about instructing our colleagues in other fields, it will be proper to consider a problem fundamental to

observed *association* to a verdict of *causation*? Upon what basis should we proceed to do so?

I have no wish, nor the skill, to embark upon a philosophical discussion of the meaning of 'causation'. The 'cause' of illness may be immediate and direct, it may be remote and indirect underlying the observed association. But with the aims of occupational, and almost synonymously preventive, medicine in mind the decisive question is whether the frequency of the undesirable event B will be influenced by a change in the environmental feature A. *How* such a change exerts that influence may call for a great deal of research. However, before deducing 'causation' and taking action we shall not invariably have to sit around awaiting the results of that research. The whole chain may have to be unravelled or a few links may suffice. It will depend upon circumstances.

Disregarding then any such problem in semantics we have this situation. Our observations reveal an association between two variables, perfectly clear-cut and beyond what we would care to attribute to the play of chance. What

- Strength
- Consistency
- Specificity
- Temporality
- Biological gradient
- Plausibility
- Coherence
- Experiment
- Analogy

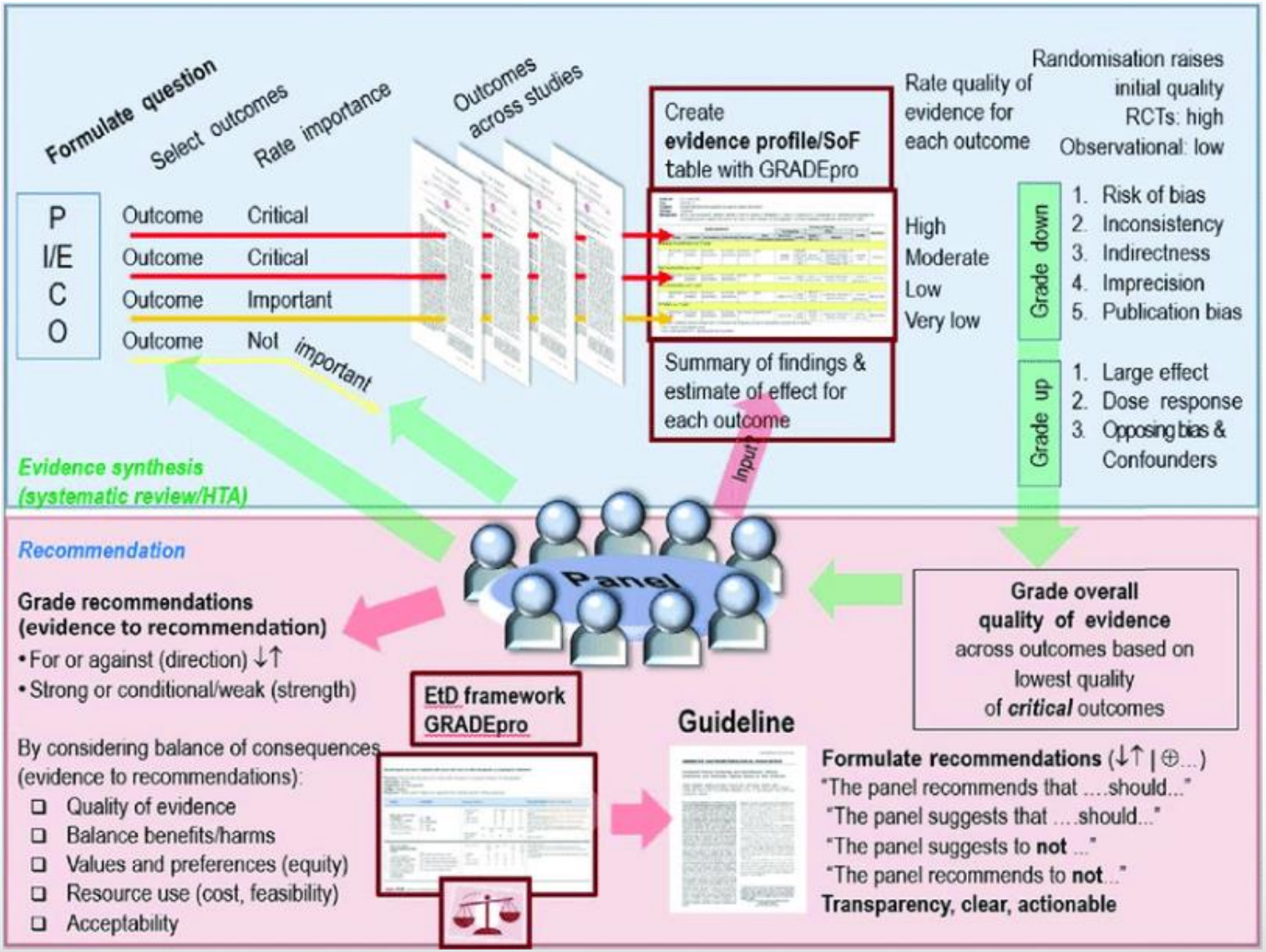
…..but Hill does not discuss how to operationalize these concepts

Hill, Austin Bradford. "The Environment and Disease: Association or Causation?" *Proceedings of the Royal Society of Medicine* 58.5 (1965): 295–300. Print.

GRADE

- Widely used (100+ organizations)
- GRADE Certainty in the Evidence (CiE) framework
  - Are the research studies well done? Risk of bias
  - Are the results consistent across studies ? Inconsistency
  - How directly do the results relate to the question? Indirectness
  - Is the association precise - due to random error? Imprecision
  - Are these all of the studies that have been conducted? Pub. Bias
  - Is there anything else that makes us particularly certain? Large associations, worst case scenario predictors still allows strong conclusions, exposure-effect relation
- GRADE conducts research and develops guidance
  - Publications, handbook, software application (GRADEpro/GDT), bi-annual meetings, use of case examples to address methodological challenges
  - GRADE Working Group has open and free membership www.gradeworkingroup.org
- GRADE is adaptable, e.g., GRADE frameworks for interventions, prognostic factors, values and preferences, etc.

GRADE Evidence to Decision Making



NAS (2017) Low Dose Toxicity From Endocrine Active

TABLE 3-9 Profile of the Confidence in the Body of Evidence on DEHP and AGD in Humans

			Factors Decreasing Confidence "—" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence			
Phthalate	Metabolite(s)	INITIAL CONFIDENCE RATING (# of studies)	Risk of Bias	Unexplained Inconsistency	Indirectness	Imprecision	Publication Bias	Large Magnitude	Dose Response	Residual Confounding	FINAL CONFIDENCE RATING
DEHP	MEHP; 5-oxo-MEHP; 5OH-MEHP; sumDEHP metabolites	Moderate (6 prospective)[a]	—	—	—	—	—	—	—	—	Moderate

[a]Swan et al. (2008); Bustamante-Montes et al. (2013); Bornehag et al. (2015); Swan et al. (2015); Jensen et al. (2016); Martino-Andrade et al. (2016).

TABLE 3-3 Profile of the Confidence in the Body of Evidence on DEHP and AGD in Animals

		Factors Decreasing Confidence "—" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence					
Phthalate	INITIAL CONFIDENCE RATING (# of studies)	Risk of Bias	Unexplained Inconsistency	Indirectness	Imprecision	Publication Bias	Large Magnitude	Dose Response	Residual Confounding	Consistency Across Species/Models	Rare Outcome	FINAL CONFIDENCE RATING
DEHP	High (16 rat,[a] 3 mouse[b])	↓	—	—	—	—	↑	↑	—	—	—	High

[a]Moore et al. (2001); Borch et al. (2004); Jarfelt et al. (2005); Wolfe and Layton (2005); Andrade et al. (2006); Culty et al. (2008); Lin et al. (2008, 2009); Christiansen et al. (2009, 2010); Gray et al. (2009); Martino-Andrade et al. (2009); Vo et al. (2009); Li et al. (2013); Zhang et al. (2013); Jones et al. (2015).
[b]Liu et al. (2008); Do et al. (2012); Pocar et al. (2012).

Mechanistic evidence: "The mechanistic data developed in vitro and in animal models provide evidence that the DEHP effects on AGD in humans identified by the committee's systematic review are biologically plausible….but were not sufficient to result in an upgrade in the committee's final hazard identification."

Final Hazard Conclusion on AGD

On the basis of the committee's evidence integration of the animal and the human evidence on DEHP and effects on AGD and consideration of relevant mechanistic data, the committee concluded that DEHP is presumed to be a reproductive hazard to humans.

IRIS Within-Stream (Human; Animal Stream) Evidence Judgment Considerations

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

Light blue rows highlight mechanistic inferences; "temporality" and "natural experiments" not shown

IRIS Evidence Profile Table

Studies	Factors that increase strength	Factors that decrease strength	Summary of findings	Strength of the evidence judgement	Inference across lines of evidence	Integrated Evidence Conclusion
[Health Effect or Outcome Grouping]						
Evidence from Human Studies (Route)					*Examples:* • *Human relevance of findings in animals* • *Cross-stream coherence*	*Describe conclusion for the integration of all available evidence*
• *References* – *Study design*	*Examples:* • *Consistency*	*Examples:* • *Unexplained*	• *Results across studies* • *Human mechanistic evidence informing*	*Describe strength of the evidence from human*		
Evidence for an Effect in Animals (Route)						
• *References* – *Study design description* – *Study confidence*	*Examples:* • *Consistency* • *Effect size* • *Dose-response gradient* • *Coherence of observed effects* • *Low risk of bias*	*Examples:* • *Unexplained inconsistency* • *Imprecision* • *High risk of bias*	• *Results across studies* • *Animal mechanistic evidence informing biological plausibility for effects in anima*	*Describe strength of the evidence from animal studies* + + + Strongest evidence + + ○ + ○○ Weakest evidence ○○○ Inadequate		

Step 1 – Evidence Integratic of Human or Animal Evidenc

Step 2 – Evidence Integration Across All Lines of Evidence

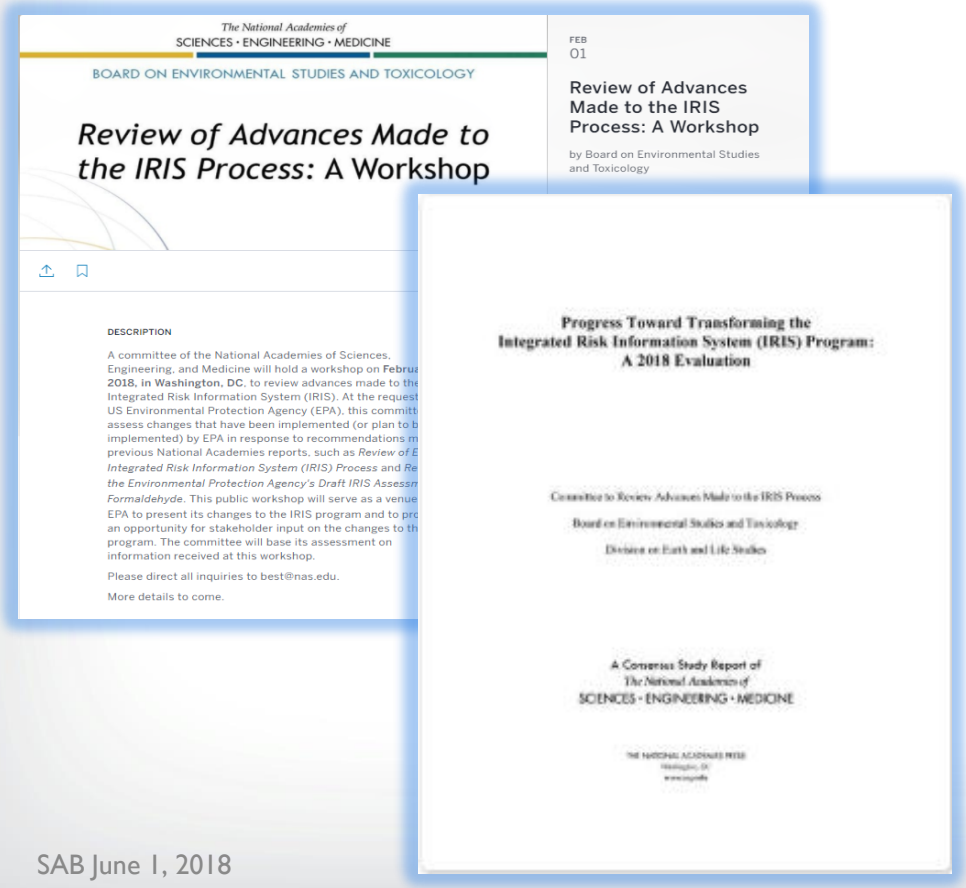
Evidence Profile Table for Diisobutyl Phthalate (DIBP) and Male Reproductive Toxicity

Outcome		Studies	Factors that increase confidence	Factors that decrease confidence	Summary of findings and confidence judgement for individual outcome	Within-stream confidence judgement for male repro	Inference across streams	Across-stream confidence judgement
HUMAN STUDIES							**Relevance of animal data to humans** •Role of testosterone-dependent and –independent pathways in male reproductive system development, maturation, and function is conserved across mammalian species. **Cross-stream coherence** •Testosterone is reduced with phthalate exposure in both humans and animals during different lifestages. **Susceptibility** •Developmental stages are particularly susceptible to perturbation by phthalates **Other relevant information** •Evidence from DBP, a structurally similar phthalate, indicates male reproductive toxicity with stronger evidence in humans, likely due to higher exposure levels and a larger number of studies	⊕⊕⊕ High confidence that DIBP causes male reproductive toxicity, based on robust animal evidence, moderate evidence in humans, and supportive mechanistic evidence. Evidence from animals is presumed relevant to humans. Lower level of evidence in humans can be explained by low sensitivity and few available studies. Mechanistic evidence suggests effects are conserved across species.
Testosterone (adult)		All cross sectional studies **Medium confidence** Meeker and Ferguson (2014) Pan et al., 2015 **Low confidence** Chang et al. (2015) Den Hond et al. (2015)	• Consistency • Minimal risk of bias in medium confidence studies	• Few studies available	⊕⊕◯ MODERATE Inverse associations between DIBP exposure and testosterone levels in 3/4 studies (Meeker and Ferguson et al., 2014, Pan et al., 2015, Chang et al., 2015), 2 of which were statistically significant. No studies examined exposure-response gradient.	⊕⊕◯ MODERATE Based on data for testosterone in adults, supported by slight evidence in other outcomes with low sensitivity and few available studies explaining lack of clear associations.		
Anogenital distance (AGD), semen parameters, pubertal development, time to pregnancy, hypospadias/cryptorchidism					⊕◯◯ SLIGHT			
ANIMAL STUDIES								
Gestational exposure	Testosterone	**High confidence** Borch et al. 2006 Furr et al. 2014 Hannas et al. 2011 Hannas et al. 2012 Howdeshell et al. 2008 Saillenfait et al. 2017 **Medium confidence** Wang et al. 2017	• Consistency • Exposure-response gradient • Effect size • Biological plausibility • Minimal risk of bias		⊕⊕⊕ ROBUST A dose-related decrease in testicular androgen levels or production was observed in all studies in rats and mice that evaluated this endpoint. Several of these studies also demonstrated decreased testicular expression of genes in the steroidogenesis pathway.	⊕⊕⊕ ROBUST Supported by consistency and coherence across outcomes. The greatest weight of evidence came from gestational exposure studies, whereas postnatal exposure studies were limited by risk of bias concerns.		
	Male morphological development	**High confidence** Borch et al. 2006 Saillenfait et al. 2006 Saillenfait et al. 2008 Saillenfait et al. 2017 **Medium confidence** Wang et al. 2017	• Exposure-response gradient • Effect size • Minimal risk of bias • Biological plausibility • Inconsistency may be explained by differences in species		⊕⊕⊕ ROBUST All rat studies observed a dose-related increase in effects consistent with decreased testosterone and INSL-3, including increased time to puberty, decreased AGD, nipple retention, cryptorchidism, hypospadias, exposed os penis, and cleft prepuce. No effects on AGD were observed in mice (Wang et al. 2017).			
	Sperm evaluation and histopathological effects in testis or epididymis	**High confidence** Saillenfait et al. 2008 **Medium confidence** Borch et al. 2006 Wang et al. 2017	• Consistency • Exposure-response gradient • Effect size • Biological plausibility		⊕⊕⊕ ROBUST Adverse effects on the testis and/or sperm were observed in rats and mice, including a dose-related increased incidence of pathological lesions of the testis (Borch et al. 2006, Saillenfait et al., 2008), epididymal oligo- or azoopermia (Saillenfait et al. 2008), and decreased sperm concentration and motility (Wang et al. 2017).			
	Reproductive organ weight	**High confidence** Saillenfait et al. 2008 **Medium confidence** Wang et al. 2017	• Biological plausibility • Exposure-response gradient • Minimal risk of bias • Inconsistency may be explained by differences in species or dose	• Few studies	⊕⊕◯ MODERATE Decreased reproductive organ weights were observed in rats (Saillenfait et al. 2008), whereas a consistent trend in testis weight was not observed in mice (Wang et al. 2017).			
Postnatal exposure	Testosterone				◯◯◯ INDETERMINATE			
	Sperm evaluation and histopathological effects in testis or epididymis	**Low confidence** Oishi and Hiraga 1980c Foster et al. 1981	• Consistency • Biological plausibility	• High risk of bias	⊕⊕◯ MODERATE Rats were found to have increase testicular atrophy (Foster et al. 1981) and decreased spermatocytes and spermatogonia (Oishi and Hiraga 1980c).			
	Reproductive organ weight	**Medium confidence** Oishi and Hiraga 1980a Oishi and Hiraga 1980b Oishi and Hiraga 1980c Oishi and Hiraga 1980d **Low confidence** Foster et al. 1981 U. Rochester 1954 Zhu et al. 2010	• Biological plausibility	• High risk of bias • Unexplained inconsistency	⊕⊕◯ MODERATE In rats, a dose-related decrease in absolute testis weight was consistently observed (Oishi and Hiraga 1980c-d, Foster at al. 1981, University of Rochester 1954). In mice, Zhu et al. (2010) observed decreased testis weight in the highest dose group, whereas Oishi and Hiraga (1980a-b) observed increased testis weight.			

Outcomes with slight or indeterminate evidence received a full systematic review, but were not significant contributors to the overall conclusion, so the details of the evidence are not provided here.

NAS IRIS Workshop Report

- **A consensus report by the National Academy of Sciences on progress made in the IRIS Program (based on a February 1-2, 2018 workshop) is now available**



[Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation](#) (released April 11, 2018)

EPA
United States
Environmental Protection
Agency

Questions?

thayer.kris@epa.gov