United States Environmental Protection Agency

Experience with Incorporating GRADE-Based Evidence Integration Frameworks into Environmental Health Evaluations

Kris Thayer, National Center for Environmental Assessment (NCEA) Integrated Risk Information System (IRIS) Division Director

NAS Workshop "Evidence integration in Chemical Assessments: Challenges Faced in Developing and Communicating Human Health Effect Conclusions"
June 3-4, 2019 (Washington DC)

Outline

- Examples
- Overview of areas of method development most pertinent to environmental and occupational health.
  - Epidemiological evidence
  - Animal evidence
  - Mechanistic evidence
  - Evidence Integration

Many methodological issues identified in applying GRADE to evidence integration in environmental health generally fall under GRADE's directness domain

UCSF Navigation Guide: PFOA and Fetal Growth (2014)

REVIEW OCTOBER 2014 | VOLUME 122 | ISSUE 10

Environ Health Perspect; DOI:10.1289/ehp.1307893

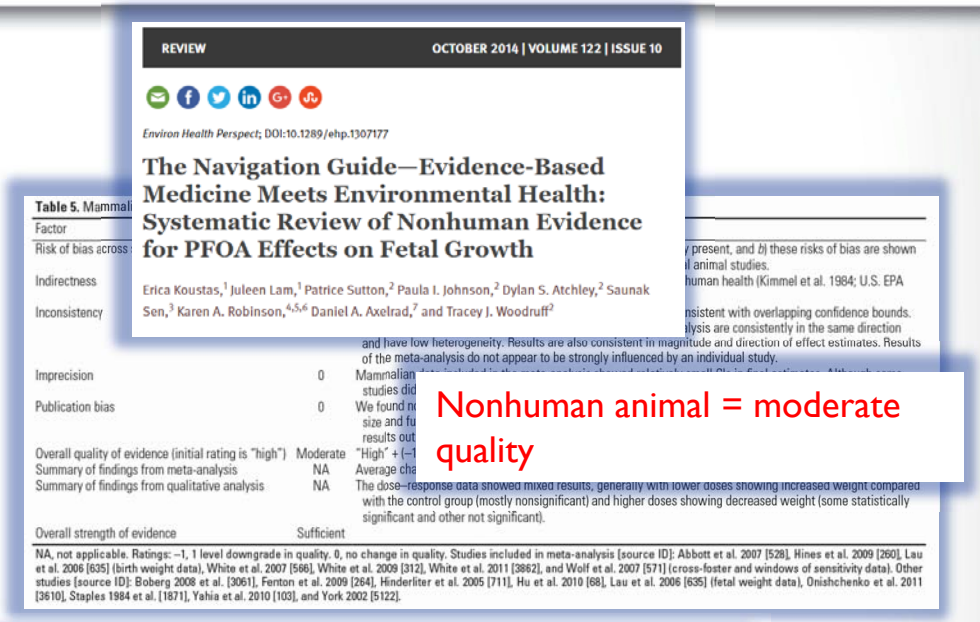
The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth

Paula I. Johnson,[1] Patrice Sutton,[1] Dylan S. Atchley,[1] Erica Koustas,[2] Juleen Lam,[2] Saunak Sen,[3] Karen A. Robinson,[4,5,6] Daniel A. Axelrad,[7] and Tracey J. Woodruff[1]

Imprecision	0	We judged that the CI of the meta-analysis for birth weight is sufficiently narrow.
Publication bias	0	We found no reason to suspect publication bias. The search was comprehensive, and the studies were generally consistent among their findings, regardless of size or funding source.
Upgrade		
Large magnitude of effect	0	We did not consider the estimated effects large.
Dose response	0	Several studies in which association was modeled by categorized incremental exposure showed evidence of a dose–response relationship, but review authors agreed that the evidence was not compelling enough for an upgrade.
Confounding minimizes effect	0	We did not find evidence to suggest that possible residual confounders or biases would reduce effect estimate.
Overall quality of evidence (initial rating is "moderate")	Moderate	Moderate + (0) = moderate. (There were no upgrades or downgrades to change quality from the initial rating).
Summary of findings from meta-analysis	NA	We found decrements in fetal growth associated with PFOA exposure (see results from meta-analyses in Table 5).
Summary of qualitative findings	NA	Studies not included in the meta-analyses presented mixed results, mostly insignificant associations between PFOA and fetal growth (Figure 4; see also Supplemental Material, Figure S1).
Strength considerations		
Quality of body of evidence	NA	Moderate
Direction of effect estimate	NA	Birth weight decreased with increasing exposure to PFOA.
Confidence in effect estimate	NA	It is unlikely that a new study would have an effect estimate that would make the results of the meta-analysis null or insignificant.
Other compelling attributes of the data that may influence certainty	NA	None
Overall strength of evidence	Sufficient	Based on our analysis and interpretation of the evidence, we concluded that there is a positive association between exposure and outcome, and we believe with reasonable confidence that

NA, not applicable.
[a]See the Supplemental Material of Lam et al. (2014) for additional details of rating quality and strength. [b]Criteria for downgrading and upgrading quality are presented in Table 2. [c]A "0" quality rating indicates there were no upgrades or downgrades for each factor being evaluated across the body of evidence.

Human = moderate quality



REVIEW OCTOBER 2014 | VOLUME 122 | ISSUE 10

Environ Health Perspect; DOI:10.1289/ehp.1307177

The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Nonhuman Evidence for PFOA Effects on Fetal Growth

Erica Koustas,[1] Juleen Lam,[1] Patrice Sutton,[2] Paula I. Johnson,[2] Dylan S. Atchley,[2] Saunak Sen,[3] Karen A. Robinson,[4,5,6] Daniel A. Axelrad,[7] and Tracey J. Woodruff[2]

Table 5. Mammali

Factor		
Risk of bias across		
Indirectness		
Inconsistency		
Imprecision	0	
Publication bias	0	
Overall quality of evidence (initial rating is "high")	Moderate	"High" + (−1
Summary of findings from meta-analysis	NA	Average cha
Summary of findings from qualitative analysis	NA	The dose–response data showed mixed results, generally with lower doses showing increased weight compared with the control group (mostly nonsignificant) and higher doses showing decreased weight (some statistically significant and other not significant).
Overall strength of evidence	Sufficient	

NA, not applicable. Ratings: −1, 1 level downgrade in quality. 0, no change in quality. Studies included in meta-analysis [source ID]: Abbott et al. 2007 [528], Hines et al. 2009 [260], Lau et al. 2006 [635] (birth weight data), White et al. 2007 [566], White et al. 2009 [312], White et al. 2011 [3862], and Wolf et al. 2007 [571] (cross-foster and windows of sensitivity data). Other studies [source ID]: Boberg 2008 et al. [3061], Fenton et al. 2009 [264], Hinderliter et al. 2005 [711], Hu et al. 2010 [68], Lau et al. 2006 [635] (fetal weight data), Onishchenko et al. 2011 [3610], Staples 1984 et al. [1871], Yahia et al. 2010 [103], and York 2002 [5122].

Nonhuman animal = moderate quality

REVIEW OCTOBER 2014 | VOLUME 122 | ISSUE 10

Environ Health Perspect; DOI:10.1289/ehp.1307923

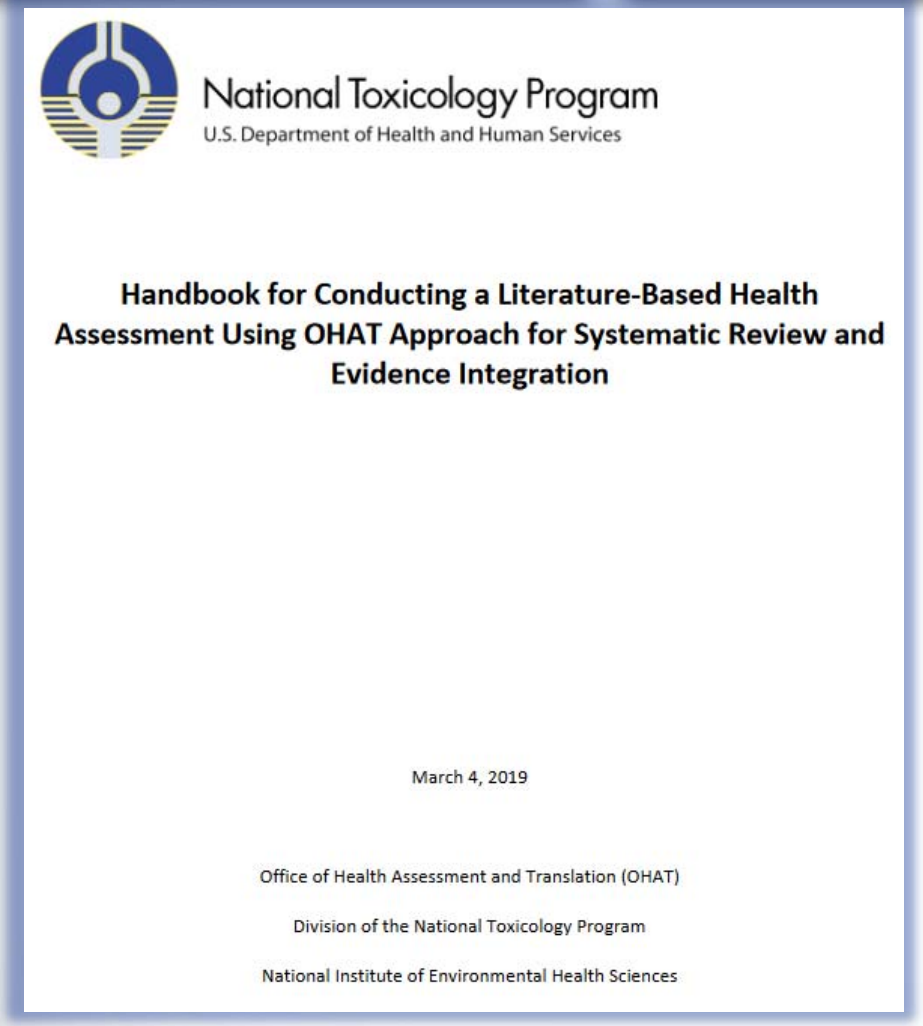
The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Integration of Animal and Human Evidence for PFOA Effects on Fetal Growth

Juleen Lam,[1] Erica Ko

Sen,[3] Karen A. Robin

Integration of human and animal evidence led to conclusion of "known" to be toxic to human reproduction and development

NTP-OHAT Handbook (2015, Updated 2019)



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

March 4, 2019

Office of Health Assessment and Translation (OHAT)

Division of the National Toxicology Program

National Institute of Environmental Health Sciences

NTP-OHAT Handbook (2015, Updated 2019)

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

Handbook for Conducting a
Assessment Using OHAT Approa
Evidence In

March 4,

Office of Health Assessment

Division of the National

National Institute of Environ

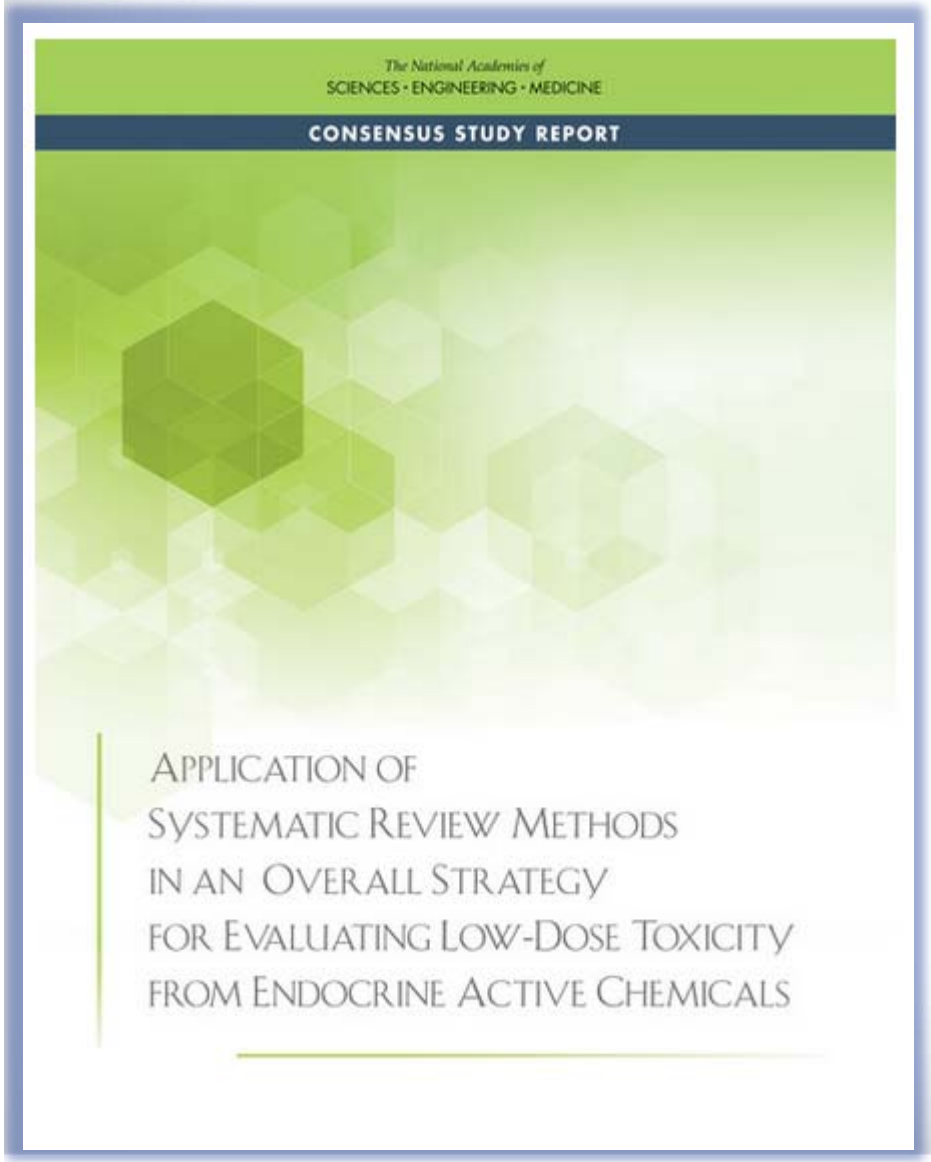
Table 6. Evidence Profile of the Main Findings for PFOA Immunotoxicity

	Factors decreasing confidence "---" if no concern; "↓" if serious concern to downgrade confidence					Factors increasing confidence "---" if not present; "↑" if sufficient to upgrade confidence				
INITIAL CONFIDENCE for each body of evidence (# of studies)	Risk of Bias	Unexplained Inconsistency	Indirectness	Imprecision	Publication Bias	Large Magnitude	Dose Response	Residual Confounding	Consistency Species/Model	FINAL CONFIDENCE RATING
Immunotoxicity Based on Evidence for Suppression of the Antibody Response										
Human										
Initial Moderate (4 prospective studies)[a]	---	---	---	---	---	---	---	---	---	**Moderate**
Initial Low (2 cross-sectional studies)[b]	---	---	---	---	---	---	---	---	---	**Low**
Confidence Across Human Bodies of Evidence	No change for considering across study designs									**Moderate**
Animal										
Initial High (7 mammal studies)	↓	---	---	---	---	---	↑	---	---	**High**

References:
Human: Granum (2013)[a], Grandjean (2012)[a], Kielsen (2016)[b], Looker (2014)[a], Mogensen (2015)[a], Stein (2016)[b]
Animal: DeWitt (2008, 2009a, 2016), Hu (2010), Loveless (2008), Vetvicka (2013), Yang (2002a)

NTP (National Toxicology Program). 2016. *Monograph on Immunotoxicity Associated with Exposure to Perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS)*. Research Triangle Park, NC: National Toxicology Program.
http://ntp.niehs.nih.gov/ntp/ohat/pfoa_pfos/pfoa_pfosmonograph_508.pdf

NAS (2017) Low Dose Toxicity From Endocrine Active Chemicals



NAS (2017) Low Dose Toxicity From Endocrine Active Chemicals

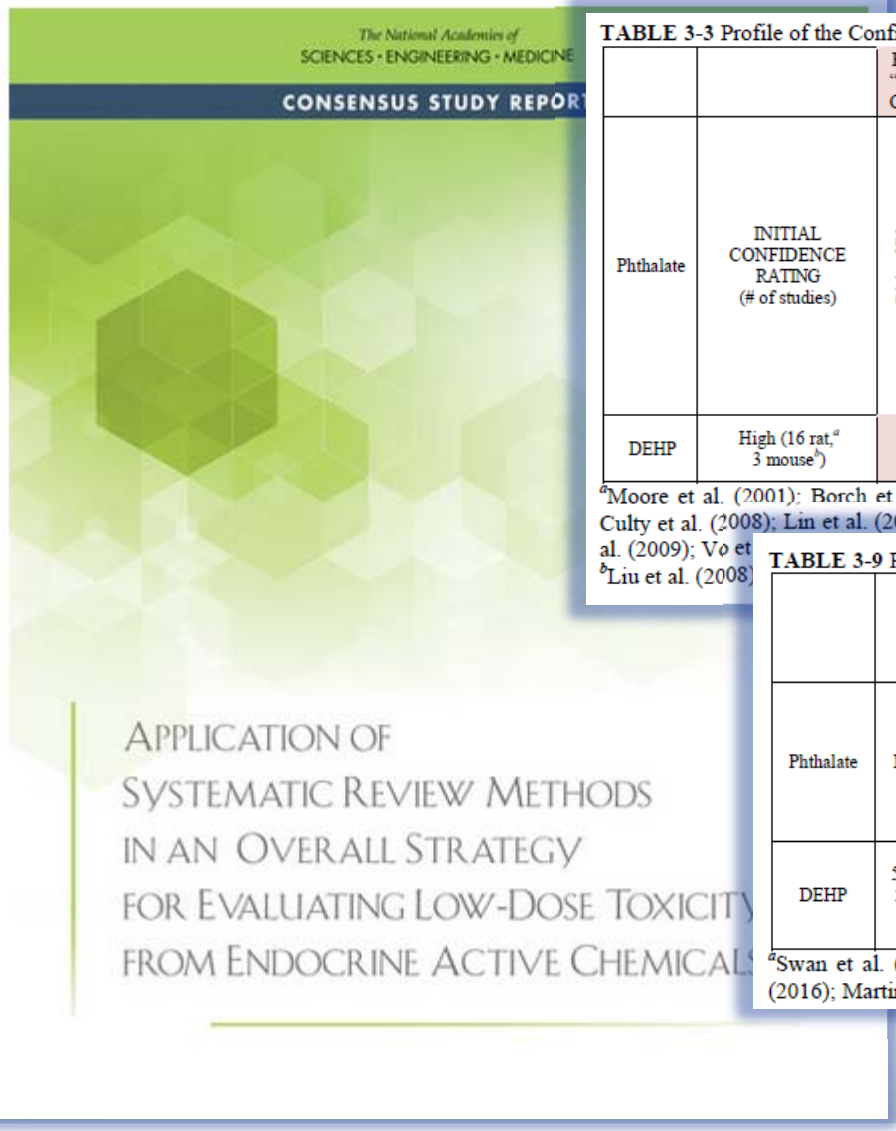


TABLE 3-3 Profile of the Confidence in the Body of Evidence on DEHP and AGD in Animals

		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
[b]Liu et al. (2008)

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

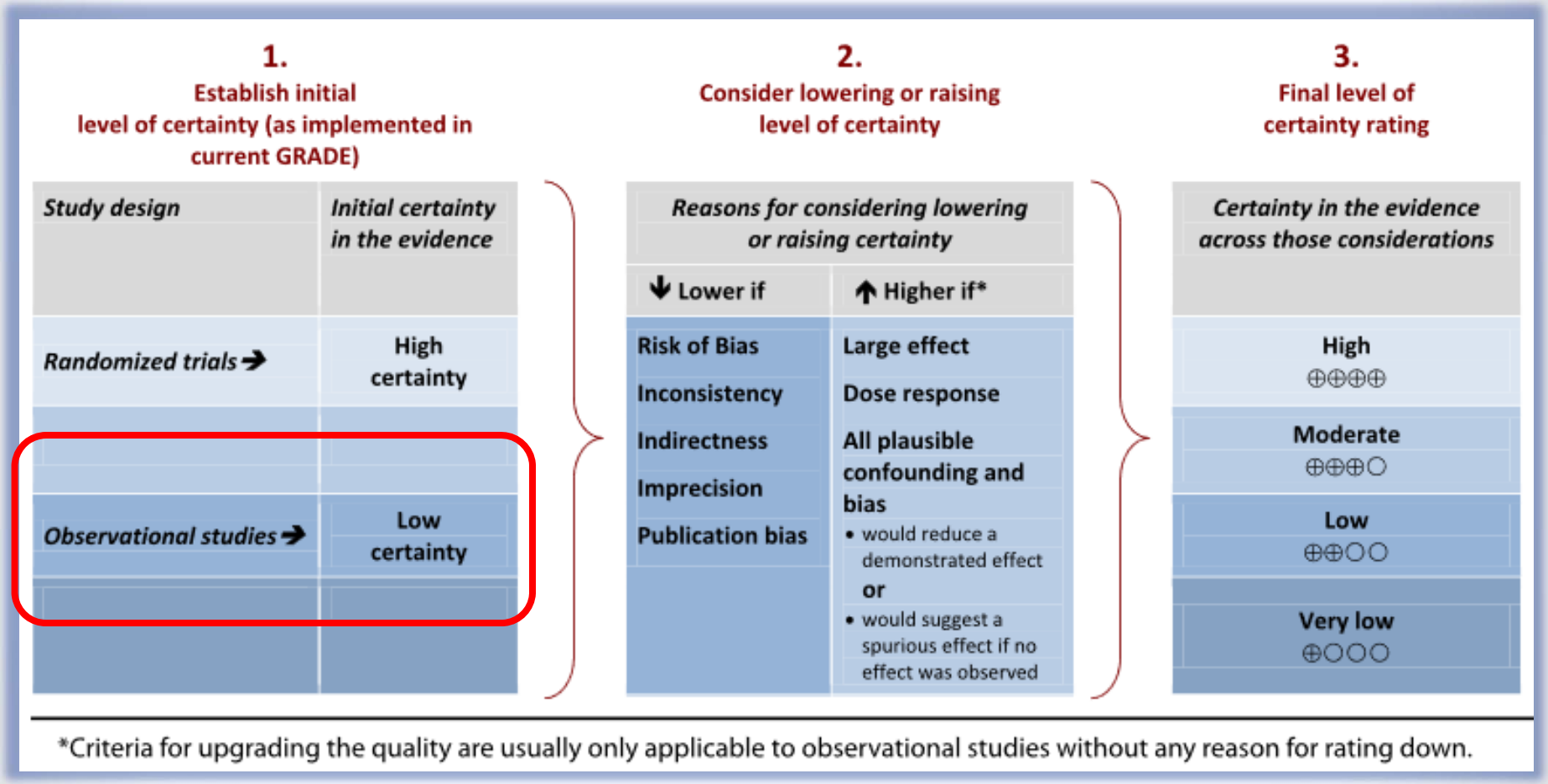
			Factors Decreasing Confidence "—" If No Concern; "↓" If Serious Concern to Downgrade Confidence					Factors Increasing Confidence "—" If Not Present; "↑" If Sufficient to Upgrade Confidence			
Phthalate	Metabolite(s)	INITIAL CONFIDENCE RATING (# of studies)	Risk of Bias	Unexplained Inconsistency	Indirectness	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).

United States Environmental Protection Agency

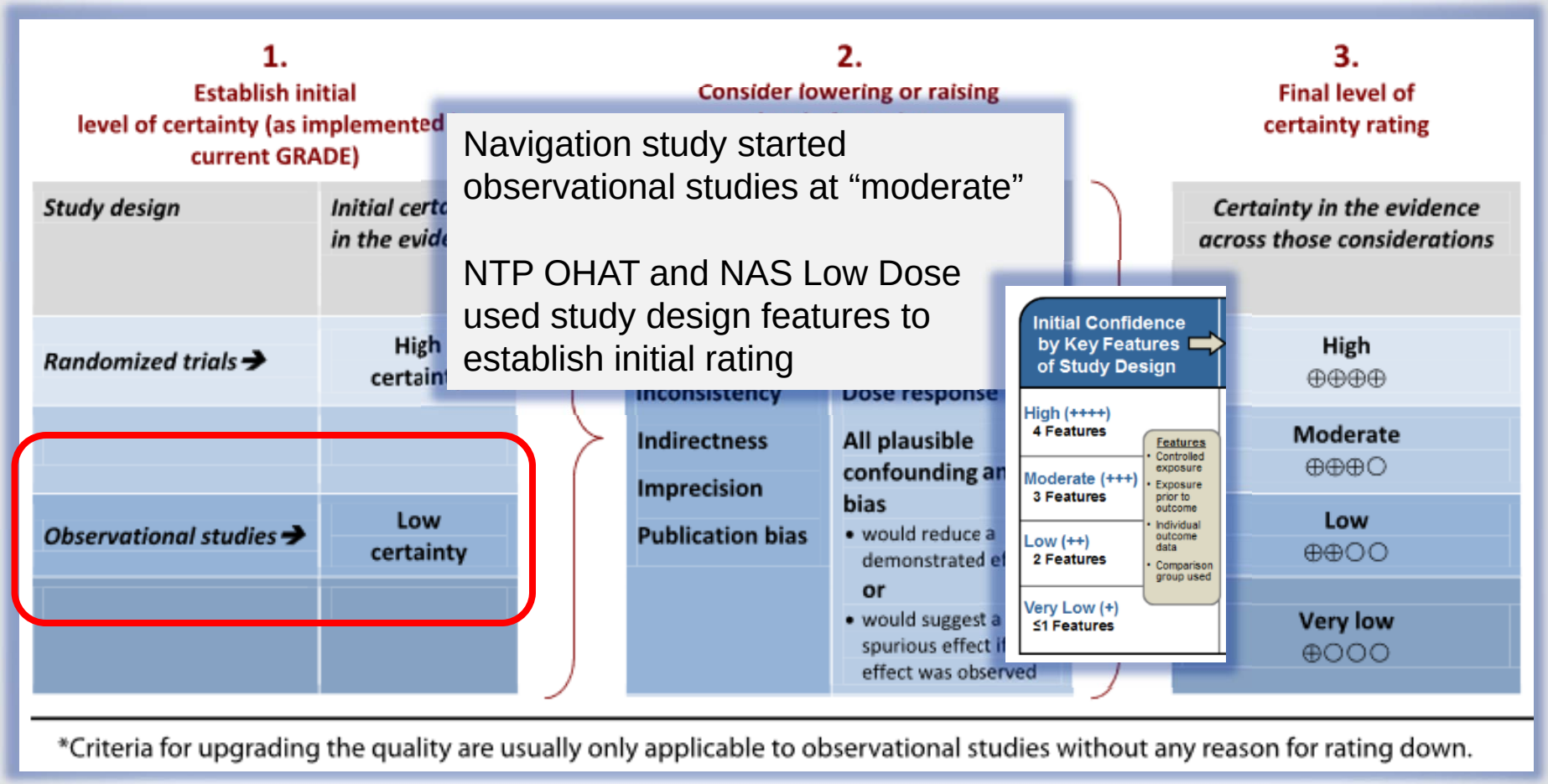
Initial Level of Certainty Rating of Low for Observational Studies
&
Impact of ROBINS-I Risk of Bias Tool for Non-Randomized Studies of Interventions

Initial Certainty for Observational Studies



Schunemann, H. J., et al. (2018). "GRADE Guidelines: 18. How ROBINS-I and other tools to assess risk of bias in non-randomized studies should be used to rate the certainty of a body of evidence." J Clin Epidemiol.

Initial Certainty for Observational Studies



Schunemann, H. J., et al. (2018). "GRADE Guidelines: 18. How ROBINS-I and other tools to assess risk of bias in non-randomized studies should be used to rate the certainty of a body of evidence." J Clin Epidemiol.

United States Environmental Protection Agency

RESEARCH METHODS AND REPORTING

OPEN ACCESS

ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions

Jonathan AC Sterne,[1] Miguel A Hernán,[2] Barnaby C Reeves,[3] Jelena Savović,[1,4] Nancy D Berkman,[5] Meera Viswanathan,[6] David Henry,[7] Douglas G Altman,[8] Mohammed T Ansari,[9] Isabelle Boutron,[10] James R Carpenter,[11] An-Wen Chan,[12] Rachel Churchill,[13] Jonathan J Deeks,[14] Asbjørn Hróbjartsson,[15] Jamie Kirkham,[16] Peter Jüni,[17] Yoon K Loke,[18] Theresa D Pigott,[19] Craig R Ramsay,[20] Deborah Regidor,[21] Hannah R Rothstein,[22] Lakhbir Sandhu,[23] Pasqualina L Santaguida,[24] Holger J Schünemann,[25] Beverly Shea,[26] Ian Shrier,[27] Peter Tugwell,[28] Lucy Turner,[29] Jeffrey C Valentine,[30] Hugh Waddington,[31] Elizabeth Waters,[32] George A Wells,[33] Penny F Whiting,[34] Julian PT Higgins[35]

For numbered affiliations see end of article.

Correspondence to: J A C Sterne jonathan.sterne@bristol.ac.uk

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;355:i4919 http://dx.doi.org/10.1136/bmj.i4919

Non-randomised studies of the effects of interventions are critical to many areas of healthcare evaluation, but their
therefor
and app
weaknes
("Risk Of
Studies
tool for
estimate
effective
interven
not use
units (in
individu
The tool will be particularly useful to those undertaking systematic reviews that include non-randomised studies.

such as cohort studies and case-control studies in which intervention groups are allocated during the course of usual treatment decisions, and quasi-randomised studies in which the method of allocation

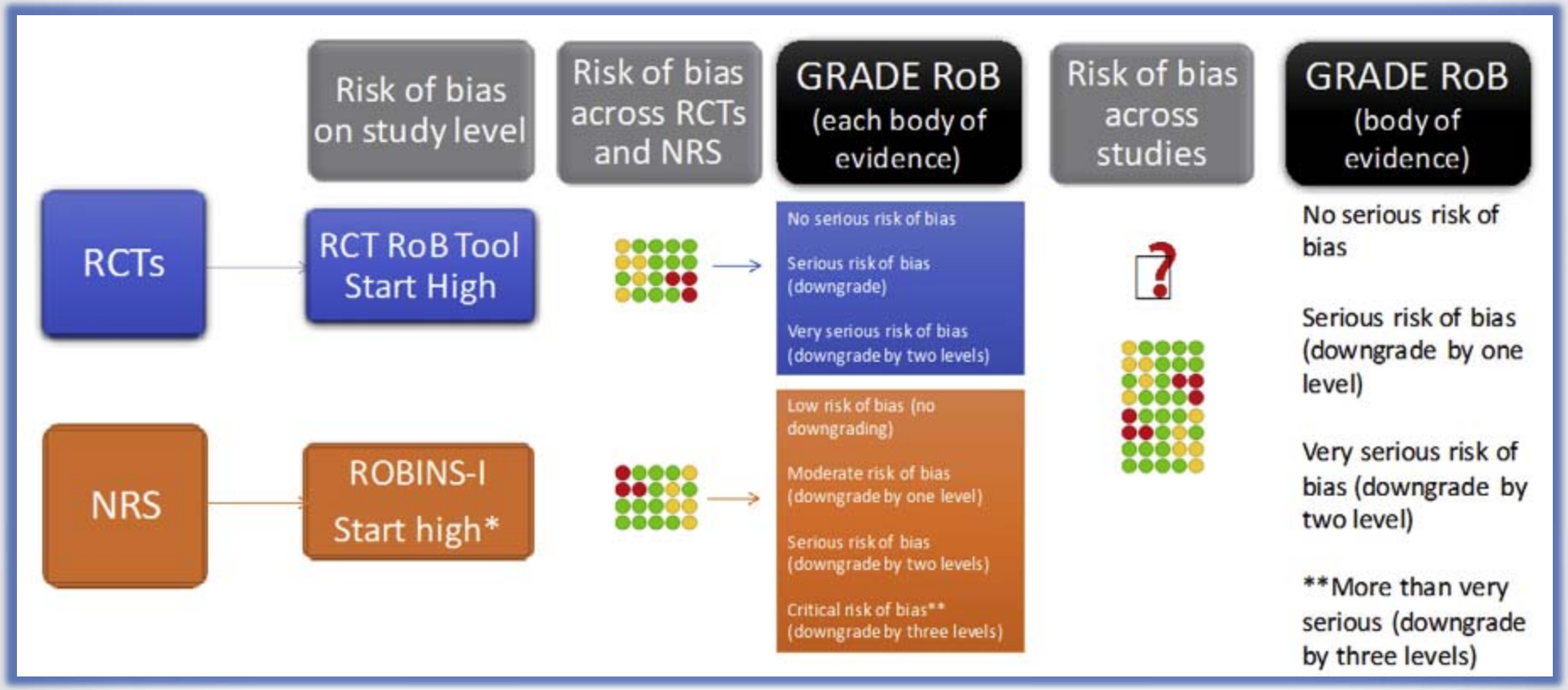
understanding of the strengths and weaknesses of the evidence.

Many tools to assess the methodological quality of observational studies in the context of a systematic review have been proposed.[4,5] The Newcastle-Ottawa[6] and Downs-Black[7] tools have been two of the most pop-

"Evaluations of risk of bias in the results of NRSI are facilitated by considering each NRSI as an attempt to emulate (mimic) a "target" trial. This is the hypothetical pragmatic randomized trial, conducted on the same participant group and without features putting it at risk of bias, whose results would answer the question addressed by the NRSI. Such a "target" trial need not be feasible or ethical…"

Initial Certainty for Observational Studies

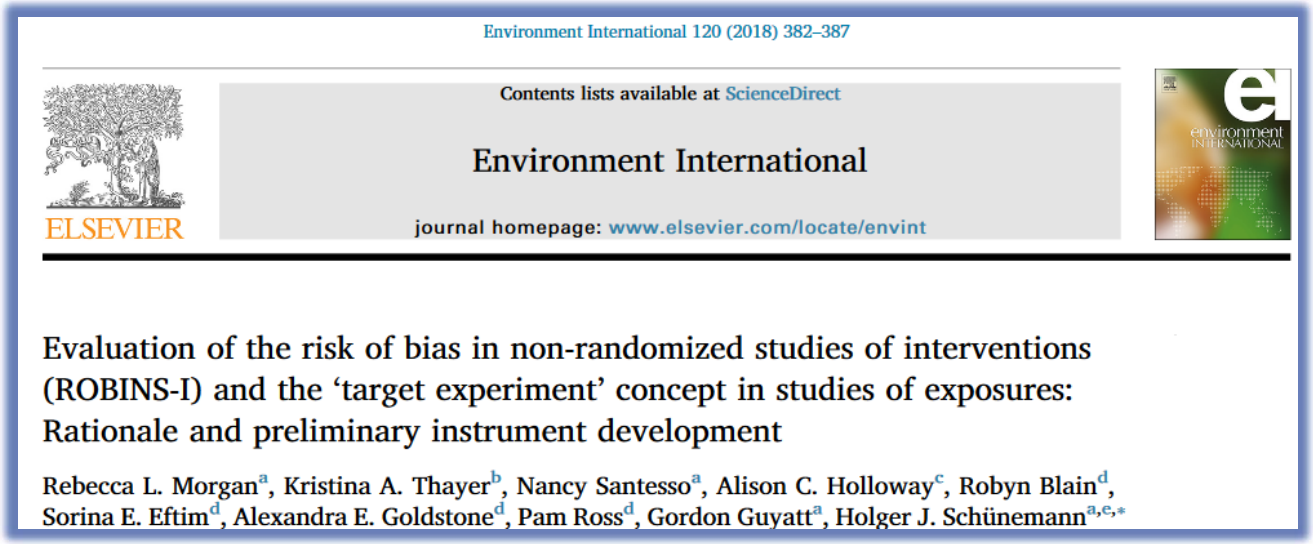
- GRADE with use of ROBINS-I (or similar) tool to evaluate risk of bias in non-randomized studies



Schunemann, H. J., et al. (2018). "GRADE Guidelines: 18. How ROBINS-I and other tools to assess risk of bias in non-randomized studies should be used to rate the certainty of a body of evidence." J Clin Epidemiol.

ROBINS-E for Exposures

- The evaluation of exposures sufficiently different from interventions to warrant separate tool



Environment International 120 (2018) 382–387

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Evaluation of the risk of bias in non-randomized studies of interventions (ROBINS-I) and the 'target experiment' concept in studies of exposures: Rationale and preliminary instrument development

Rebecca L. Morgan[a], Kristina A. Thayer[b], Nancy Santesso[a], Alison C. Holloway[c], Robyn Blain[d], Sorina E. Eftim[d], Alexandra E. Goldstone[d], Pam Ross[d], Gordon Guyatt[a], Holger J. Schünemann[a,e,*]

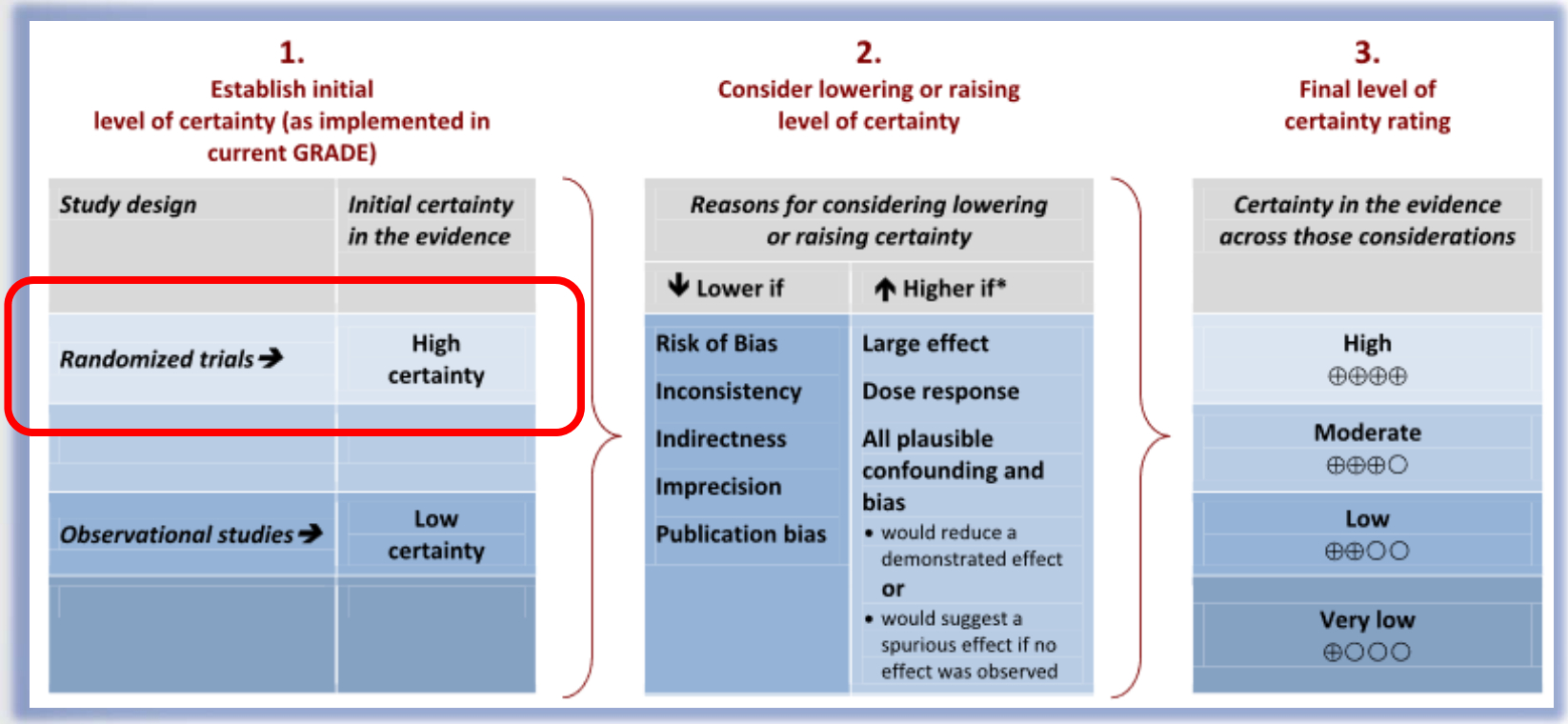
- ROBINS-E Workshop: Developing ROBINS-I for studies of exposures (ROBINS-E). January 30-31, 2017
  - Currently revising based on comments and starting pilot testing
  - Future work: Assessing impact of potential confounding and other factors for an individual study (purview of systematic review) and across a body of evidence (GRADE purview)

EPA United States Environmental Protection Agency

Animal Studies

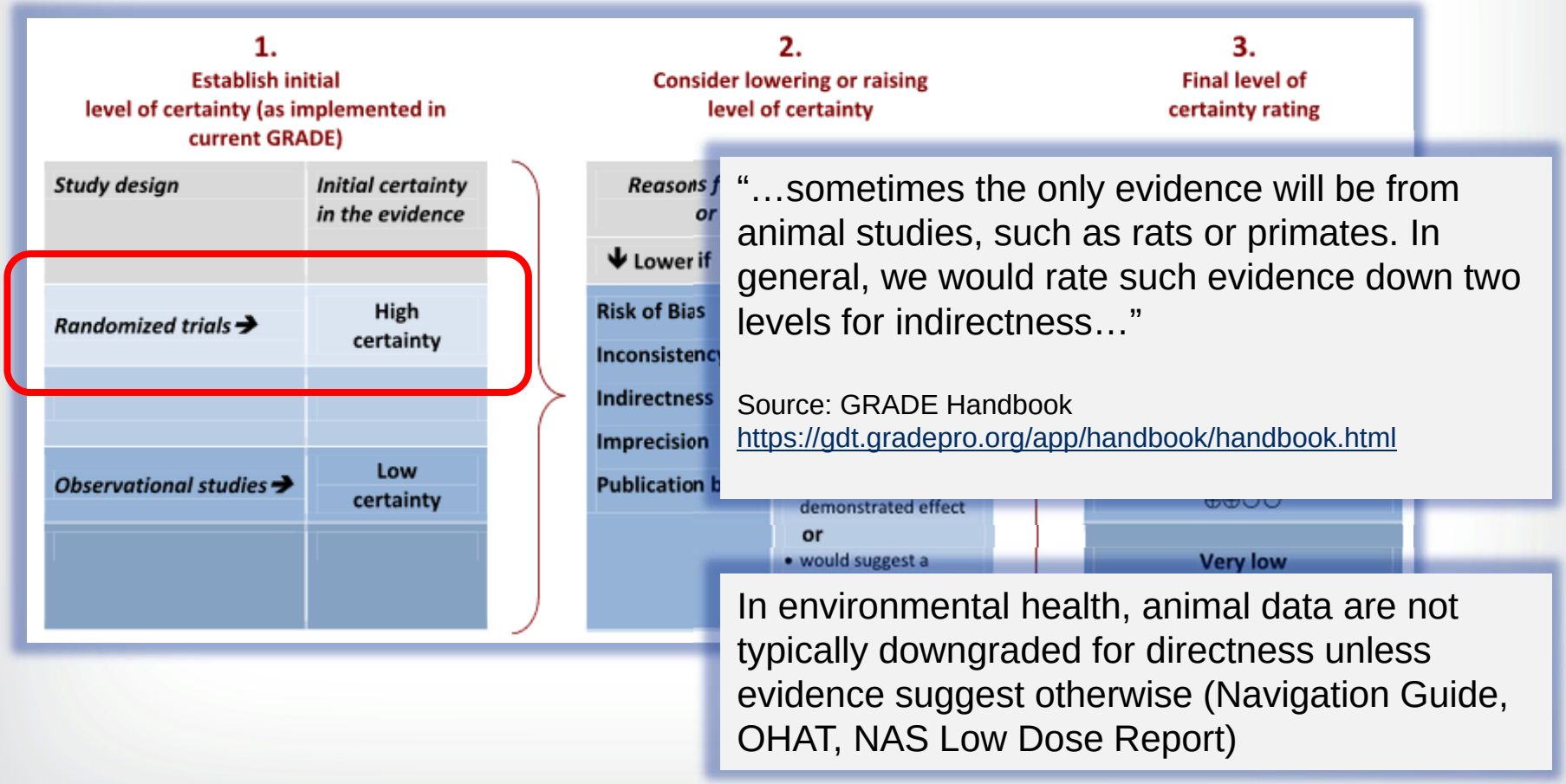
Animal Studies in GRADE

- Animal studies are essentially treated like randomized trials in humans



Animal Studies in GRADE

- Animal studies are essentially treated like randomized trials in humans

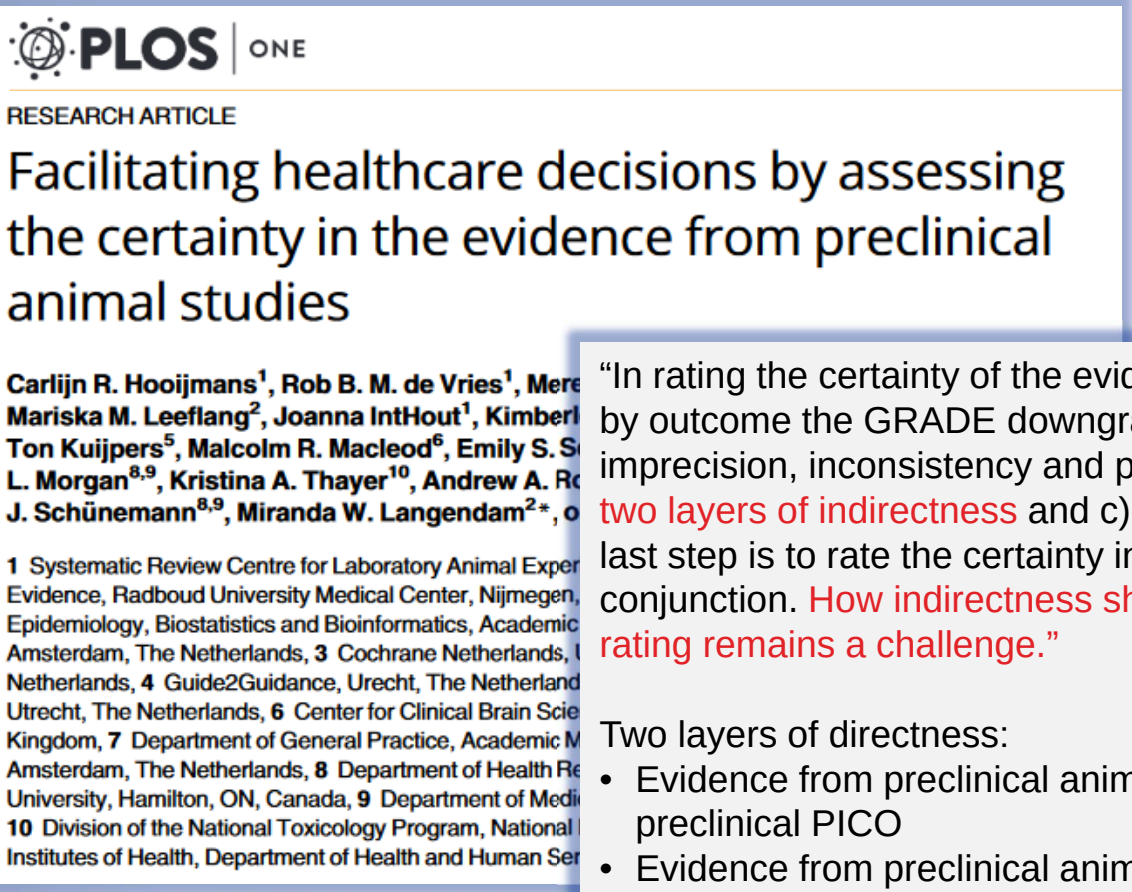


"…sometimes the only evidence will be from animal studies, such as rats or primates. In general, we would rate such evidence down two levels for indirectness…"

Source: GRADE Handbook
https://gdt.gradepro.org/app/handbook/handbook.html

In environmental health, animal data are not typically downgraded for directness unless evidence suggest otherwise (Navigation Guide, OHAT, NAS Low Dose Report)

Animal Studies and Directness



PLOS ONE

RESEARCH ARTICLE

Facilitating healthcare decisions by assessing the certainty in the evidence from preclinical animal studies

Carlijn R. Hooijmans[1], Rob B. M. de Vries[1], Mere
Mariska M. Leeflang[2], Joanna IntHout[1], Kimberl
Ton Kuijpers[5], Malcolm R. Macleod[6], Emily S. S
L. Morgan[8,9], Kristina A. Thayer[10], Andrew A. R
J. Schünemann[8,9], Miranda W. Langendam[2]*, o

1 Systematic Review Centre for Laboratory Animal Exper
Evidence, Radboud University Medical Center, Nijmegen,
Epidemiology, Biostatistics and Bioinformatics, Academic
Amsterdam, The Netherlands, 3 Cochrane Netherlands,
Netherlands, 4 Guide2Guidance, Urecht, The Netherland
Utrecht, The Netherlands, 6 Center for Clinical Brain Scie
Kingdom, 7 Department of General Practice, Academic M
Amsterdam, The Netherlands, 8 Department of Health Re
University, Hamilton, ON, Canada, 9 Department of Medi
10 Division of the National Toxicology Program, National
Institutes of Health, Department of Health and Human Ser

"In rating the certainty of the evidence we propose to assess– by outcome the GRADE downgrading factors a) risk of bias, imprecision, inconsistency and publication bias, followed by b) two layers of indirectness and c) considering upgrading. The last step is to rate the certainty in the effect taking all factors in conjunction. How indirectness should be weighted in the total rating remains a challenge."

Two layers of directness:

- Evidence from preclinical animal studies compared to preclinical PICO
- Evidence from preclinical animal studies compared to clinical PICO (also called translatability)

Hooijmans, C. R., et al. (2018) Facilitating healthcare decisions by assessing the certainty in the evidence from preclinical animal studies. *PloS one*, 13(1):e0187271.

EPA United States Environmental Protection Agency

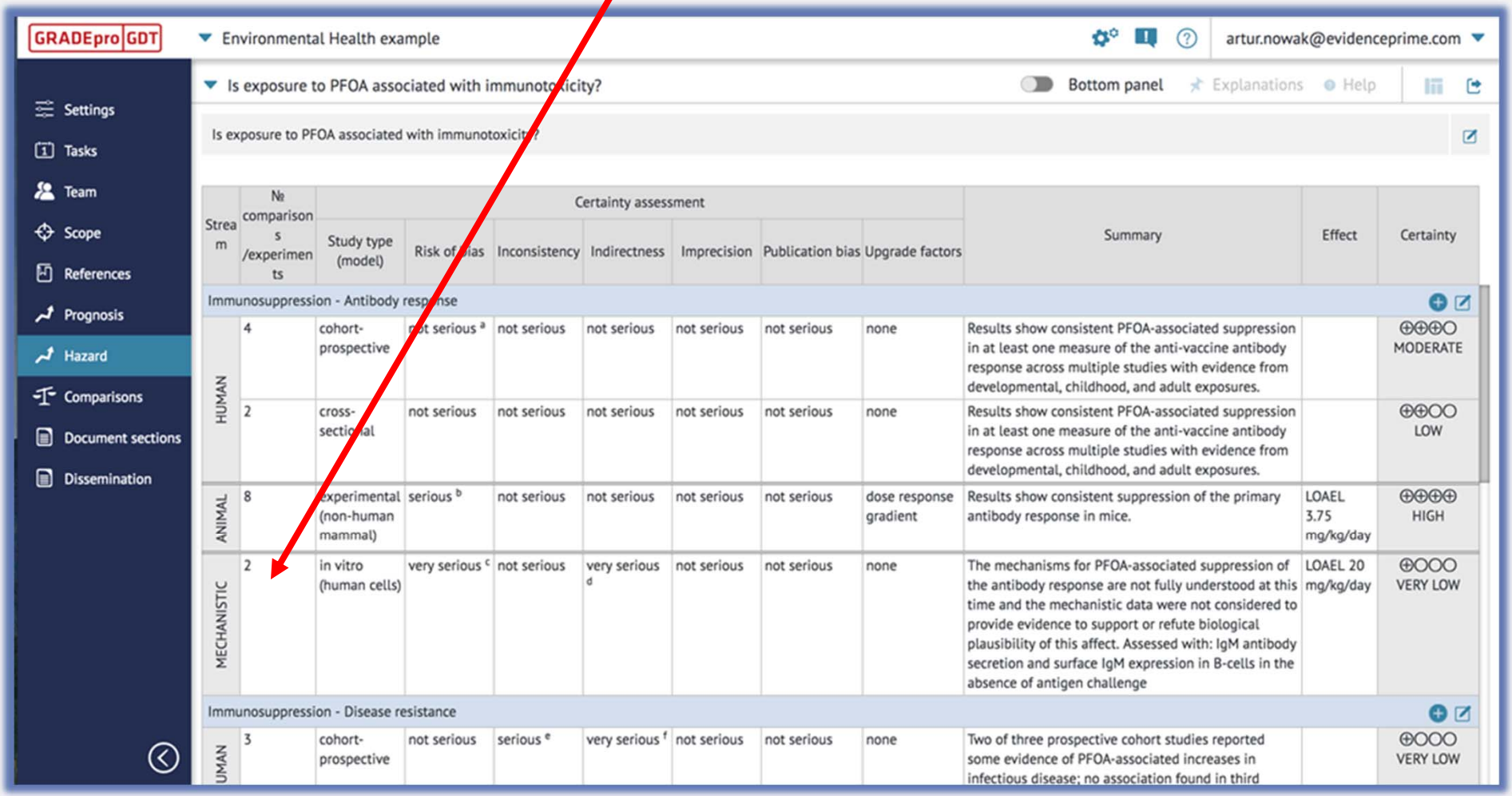
Mechanistic Evidence

Current Discussion Areas

- Assessing risk of bias for mechanistic studies, both at the individual study level (which informs across body of evidence judgements)
  - Risk of bias ≠ reporting quality
  - Pragmatic considerations: the need to evaluate every mechanistic study versus studies assessing key topics, especially for large evidence base topics?
  - How evaluate risk of bias for -omic or high throughput screening data?
- Evaluation of directness
  - Mechanistic evidence as a stand alone evidence stream
  - "Gap filling" in the evaluation of directness of other types of evidence
  - How to communicate increased confidence based on a collection of separate indirect lines of evidence

United States Environmental Protection Agency

Example of mechanistic evidence presented as a stand alone evidence stream in GRADEPro software



Is exposure to PFOA associated with immunotoxicity?

Stream	№ comparisons /experiments	Certainty assessment: Study type (model)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Upgrade factors	Summary	Effect	Certainty
Immunosuppression - Antibody response											
HUMAN	4	cohort-prospective	not serious [a]	not serious	not serious	not serious	not serious	none	Results show consistent PFOA-associated suppression in at least one measure of the anti-vaccine antibody response across multiple studies with evidence from developmental, childhood, and adult exposures.		⊕⊕⊕◯ MODERATE
HUMAN	2	cross-sectional	not serious	not serious	not serious	not serious	not serious	none	Results show consistent PFOA-associated suppression in at least one measure of the anti-vaccine antibody response across multiple studies with evidence from developmental, childhood, and adult exposures.		⊕⊕◯◯ LOW
ANIMAL	8	experimental (non-human mammal)	serious [b]	not serious	not serious	not serious	not serious	dose response gradient	Results show consistent suppression of the primary antibody response in mice.	LOAEL 3.75 mg/kg/day	⊕⊕⊕⊕ HIGH
MECHANISTIC	2	in vitro (human cells)	very serious [c]	not serious	very serious [d]	not serious	not serious	none	The mechanisms for PFOA-associated suppression of the antibody response are not fully understood at this time and the mechanistic data were not considered to provide evidence to support or refute biological plausibility of this affect. Assessed with: IgM antibody secretion and surface IgM expression in B-cells in the absence of antigen challenge	LOAEL 20 mg/kg/day	⊕◯◯◯ VERY LOW
Immunosuppression - Disease resistance											
UMAN	3	cohort-prospective	not serious	serious [e]	very serious [f]	not serious	not serious	none	Two of three prospective cohort studies reported some evidence of PFOA-associated increases in infectious disease; no association found in third		⊕◯◯◯ VERY LOW

Courtesy of Brandy Beverly, NTP OHAT (also co-chair of GRADE Environmental Health Project Group, brandy.beverly@nih.gov)

Mechanistic Evidence and GRADE

Issue	Key Events/Activities	Status
Mechanistic Evidence	NAS workshop, December 10-11, 2018 Ottawa, December 17-18, 2018 NAS workshop, June 3-4, 2019 EBTC, June 12, 2019 GRADE, June 13-14, 2019	Examples needed to develop GRADE guidance
Modelled Evidence	Workshop: **“**GRADE for modelled evidence. May 15-16, 2017. McMaster University. Hamilton”	GRADE factors apply; additional examples and discussion needed to develop guidance

For more information contact co-chairs of GRADE Environmental Health Project Group: Rebecca Morgan (morganrl@mcmaster.ca) or Brandy Beverly (brandy.beverly@nih.gov)

EPA United States Environmental Protection Agency

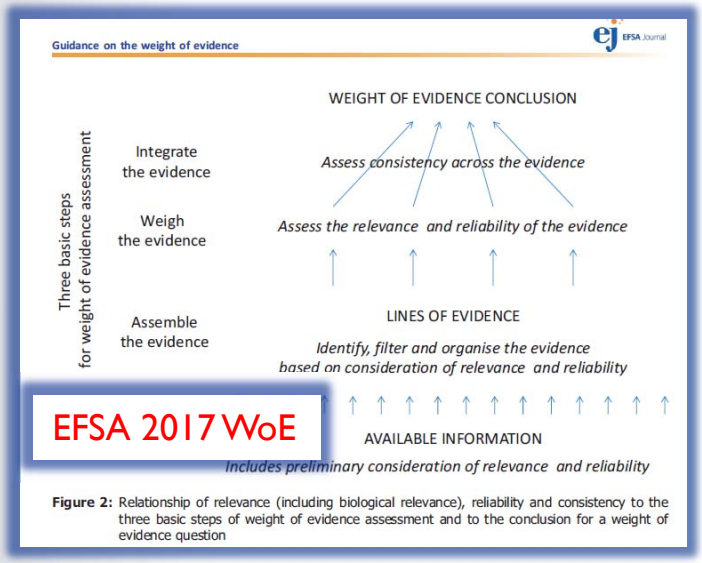
Evidence Integration

Evidence Integration and GRADE

- GRADE does not currently have guidance on integrating across different evidence streams (aka human, animal, in vitro)
- In environmental health, many approaches reach judgements within an evidence stream and then across evidence streams

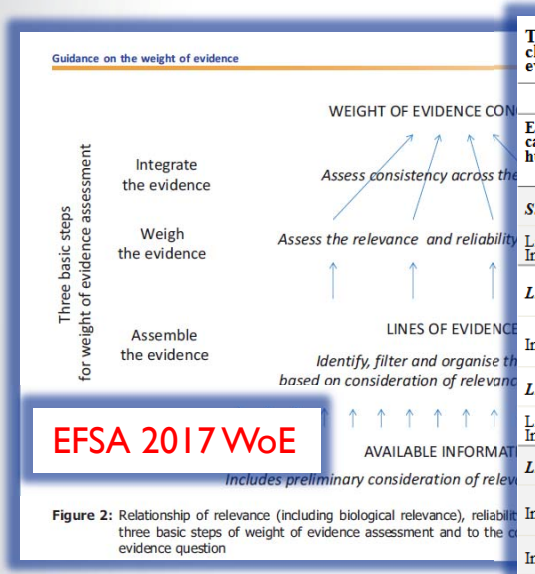
Evidence Integration and GRADE

- GRADE does not currently have guidance on integrating across different evidence streams (aka human, animal, in vitro)
- In environmental health, many approaches reach judgements within an evidence stream and then across evidence streams



Evidence Integration and GRADE

- GRADE does not currently have guidance on integrating across different evidence streams (aka human, animal, in vitro)
- In environmental health, many approaches reach judgements within an evidence stream and then across evidence streams



EFSA 2017 WoE

Table 4. Integration of streams of evidence in reaching overall classifications (the evidence in *bold italic* represents the basis of the overall evaluation)

Stream of evidence			Classification based on strength of evidence
Evidence of cancer in humans[a]	**Evidence of cancer in experimental animals**	**Mechanistic evidence**	
Sufficient	Not necessary	Not necessary	**Carcinogenic to humans (Group 1)**
Limited or Inadequate	***Sufficient***	***Strong (b)(1) (exposed humans)***	
Limited	***Sufficient***	Strong (b)(2–3), Limited, or Inadequate	**Probably carcinogenic to humans (Group 2A)**
Inadequate	***Sufficient***	***Strong (b)(2) (human cells or tissues)***	
Limited	Less than Sufficient	***Strong (b)(1–3)***	
Limited or Inadequate	Not necessary	***Strong (a) (mechanistic class)***	
Limited	Less than Sufficient	Limited or Inadequate	**humans (Group 2B)**
Inadequate	***Sufficient***	Strong (b)(3), Limited, or Inadequate	
Inadequate	Less than Sufficient	***Strong b(1–3)***	
Limited	***Sufficient***	***Strong (c) (does not operate in humans)***[b]	
Inadequate	***Sufficient***	***Strong (c) (does not operate in humans)***[b]	**Not classifiable as to its carcinogenicity to humans (Group 3)**
All other situations not listed above			

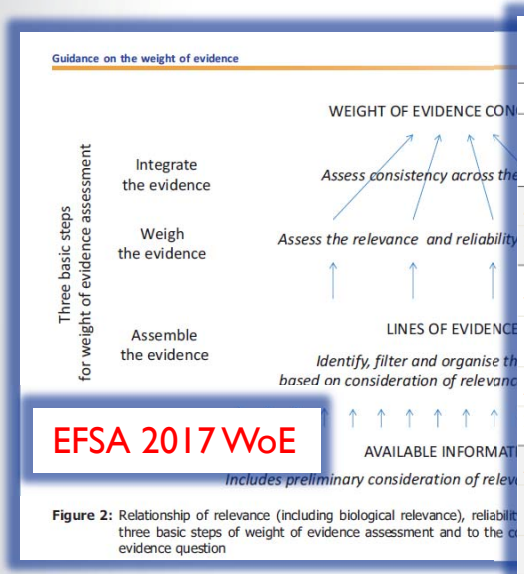
[a] Human cancer(s) with highest evaluation

[b] The *strong evidence that the mechanism of carcinogenicity in experimental animals does not operate in humans* must specifically be for the tumour sites supporting the classification of *sufficient evidence in experimental animals*.

IARC 2019

Evidence Integration and GRADE

- GRADE does not currently have guidance on integrating across different evidence streams (aka human, animal, in vitro)
- In environmental health, many approaches reach judgements within an evidence stream and then across evidence streams



EFSA 2017 WoE

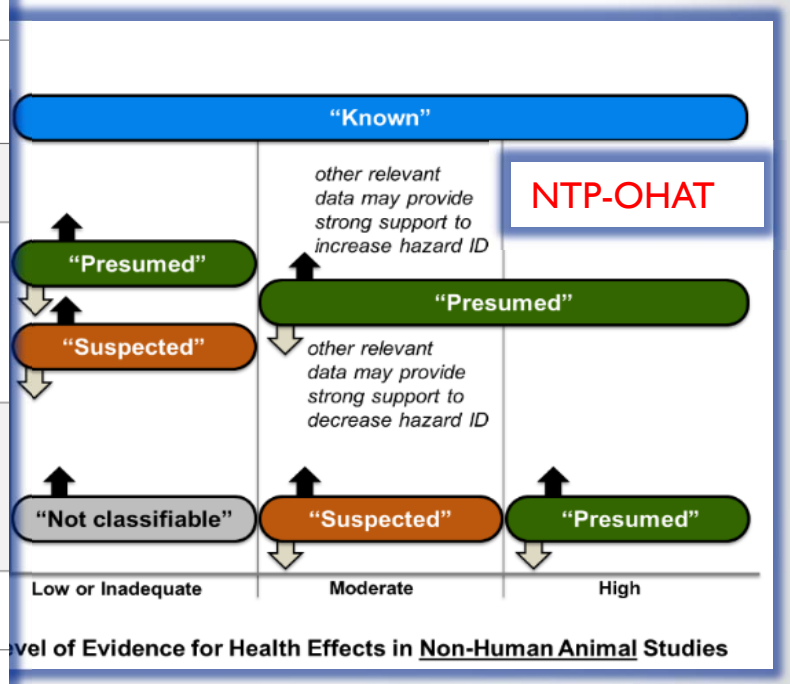
Table 4. Integration of streams of evidence in reaching overall classifications (the evidence in *bold italic* represents the basis of the overall evaluation)

Stream of evidence			Classification based on strength of evidence
Evidence of cancer in humans[a]	Evidence of cancer in experimental animals	Mechanistic evidence	
Sufficient	Not necessary	Not necessary	**Carcinogenic to humans (Group 1)**
Limited or Inadequate	***Sufficient***	***Strong (b)(1) (exposed humans)***	
Limited	***Sufficient***	Strong (b)(2–3), Limited, or Inadequate	**Probably carcinogenic to humans (Group 2A)**
Inadequate	***Sufficient***	***Strong (b)(2) (human cells or tissues)***	
Limited	Less than Sufficient	***Strong (b)(1–3)***	
Limited or Inadequate	Not necessary	***Strong (a) (mechanistic class)***	
Limited	Less than Sufficient	Limited or Inadequate	**humans (Group 2B)**
Inadequate	***Sufficient***	Strong (b)(3), Limited, or Inadequate	
Inadequate	Less than Sufficient	***Strong b(1–3)***	
Limited	***Sufficient***	***Strong (c) (does not operate in humans)***[b]	
Inadequate	***Sufficient***	***Strong (c) (does not operate in humans)***[b]	**Not classifiable as to its carcinogenicity to humans (Group 3)**
All other situations not listed above			

[a] Human cancer(s) with highest evaluation

[b] The *strong evidence that the mechanism of carcinogenicity in experimental animals does not operate in humans* must specifically be for the tumour sites supporting the classification of *sufficient evidence in experimental animals*.

IARC 2019



NTP-OHAT

US EPA IRIS Evidence Profile Table

evidence integration within a line of evidence (the step most analogous to GRADE CiE framework)

evidence integration across lines of evidence

Studies and interpretation	Factors that increase strength	Factors that decrease strength	Summary of findings	Human and animal evidence judgments	Inference across lines of evidence	Overall evidence integration conclusion
[Health effect or outcome grouping]						
Evidence from human studies [route]					• Human relevance of findings in animals* • Coherence across lines of evidence (i.e., for both health* effect-specific and mechanistic data) o Information on susceptibility o MOA analysis inferences* o Other inferences	Describe conclusion(s) and primary basis for the integration of all available evidence (across human, animal, and mechanistic): ⊕⊕⊕ Evidence demonstrates ⊕⊕⊙ Evidence indicates ⊕⊙⊙ Evidence suggests ⊙⊙⊙ Evidence inadequate – – – Strong evidence supports no effect
• References • Study confidence (based on evaluation of risk of bias and sensitivity) • Study design description	• Consistency or replication • Dose-response gradient • Coherence of observed effects (apical studies)* • Effect size (magnitude, severity) • Mechanistic evidence providing plausibility* • Medium or high confidence studies	• Unexplained inconsistency • Imprecision • Low confidence studies or other concerns about methods or design across studies • Other (e.g., single/few studies) • Evidence demonstrating implausibility (e.g., mechanistic)	Results information (general endpoints affected/unaffected) across studies Human mechanistic evidence informing biological plausibility*: discuss how data influenced the human evidence judgment (e.g., evidence of precursors in exposed humans) Could be multiple rows (e.g., grouped by study confidence or population) if this informs heterogeneity of results	Describe strength of the evidence from human studies and primary basis for judgment: ⊕⊕⊕ Robust ⊕⊕⊙ Moderate ⊕⊙⊙ Slight ⊙⊙⊙ Indeterminate – – – Compelling evidence of no effect		

Concept maps to GRADE

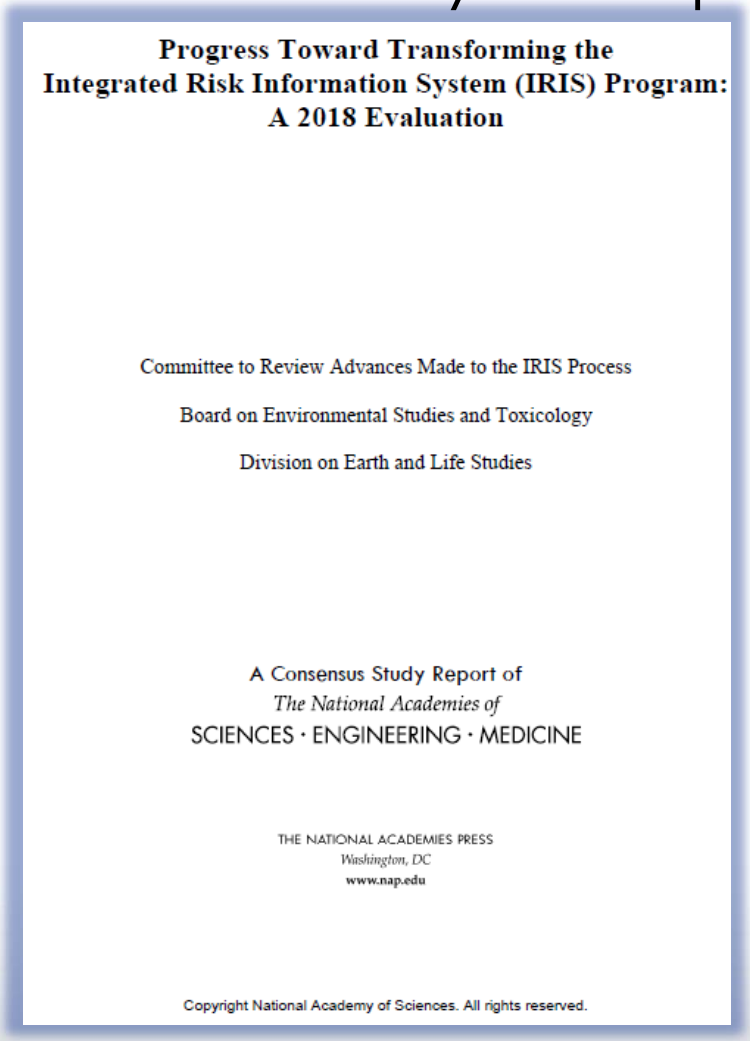
Concept present in GRADE, but presented differently in IRIS

Concept not explicitly outlined in GRADE

***Could be considered as part of directness (Schünemann et al., 2011) in GRADE, but more discussion and examples needed**

IRIS Evidence Synthesis and Integration

- IRIS structured framework reviewed by NAS in April 2018 report



Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation

Committee to Review Advances Made to the IRIS Process

Board on Environmental Studies and Toxicology

Division on Earth and Life Studies

A Consensus Study Report of
The National Academies of
SCIENCES · ENGINEERING · MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

IRIS Evidence Synthesis and Integration

- IRIS structured framework reviewed by NAS in April 2018 report

The major recommendation in Chapter 6 of the 2014 report guided IRIS to choose between making its current guided expert process more transparent and adopting a more structured, GRADE-like,[5] process along the lines...on, using structure...les toward more s...nd D-5, Appendi...ng human eviden...of evidence (see ...nd use mechanist...in- tegration in th...on and to build o...ce synthesis,[6] the...ent with the state ...ce similar challe...ut further transparency would be obtained with completion of a handbook that provides more details about processes, reasoning behind decisions, and approaches for presenting results. In the interim, while EPA is completing its handbook, it is releasing protocols for each assessment that include a description of how evidence within each data stream will be synthesized and how evidence from multiple data streams will be integrated. The draft protocol for the IRIS assessment of chloroform (EPA 2018b) was provided as an example. The committee supports EPA's approach.

"The major recommendation in Chapter 6 of the 2014 report guided IRIS to choose between making its current guided expert process more transparent and adopting a more structured, GRADE-like...The IRIS program has explicitly chosen the first option using structured categories with criteria to guide expert judgment, and EPA has made substantial strides toward more systematic and transparent evidence synthesis...the IRIS program has created a process for evidence synthesis that is scientifically consistent with the state of the art and that effectively leverages approaches of other programs, such as NTP, that face similar challenges...The committee supports EPA's approach." (page 9)

IRIS Evidence Synthesis and Integration

- Released for public comment in several protocols
  - Chloroform (January 2018), chromium (April 2019), arsenic (May 2019)
- Presented in peer-reviewed journals
  - Radke EG et al. (2018) Phthalate exposure and male reproductive effects: a systematic review of epidemiological studies. Environ Int. 2018 Dec;121(Pt 1):764-793.
  - Yost EE et al. (2019) Hazards of diisobutyl phthalate (DIBP) exposure: A systematic review of animal toxicology studies. Environ Int. 2019 Apr;125:579-594.
  - Radke EG et al. (2019, accepted Enviro Int) Phthalate exposure and metabolic effects: a systematic review of the human epidemiological evidence.
  - Radke EG et al. (2019, accepted Enviro Int) Phthalate exposure and female reproductive and developmental outcomes: a systematic review of the human epidemiological evidence
- See posters for additional examples, software and discussion
- Continued engagement in method development and refinement with GRADE, NTP, EBTC, EPA Systematic Review Community of Practice, others

Conclusions

- Engagement between the environmental health community and GRADE has increased in the past several years
  - Consideration of mechanistic evidence on the agenda for the June 13-14, 2019 GRADE meeting (Hamilton, ON)
- The GRADE domain of indirectness appears to be most relevant to mechanistic evidence and evidence integration (?)