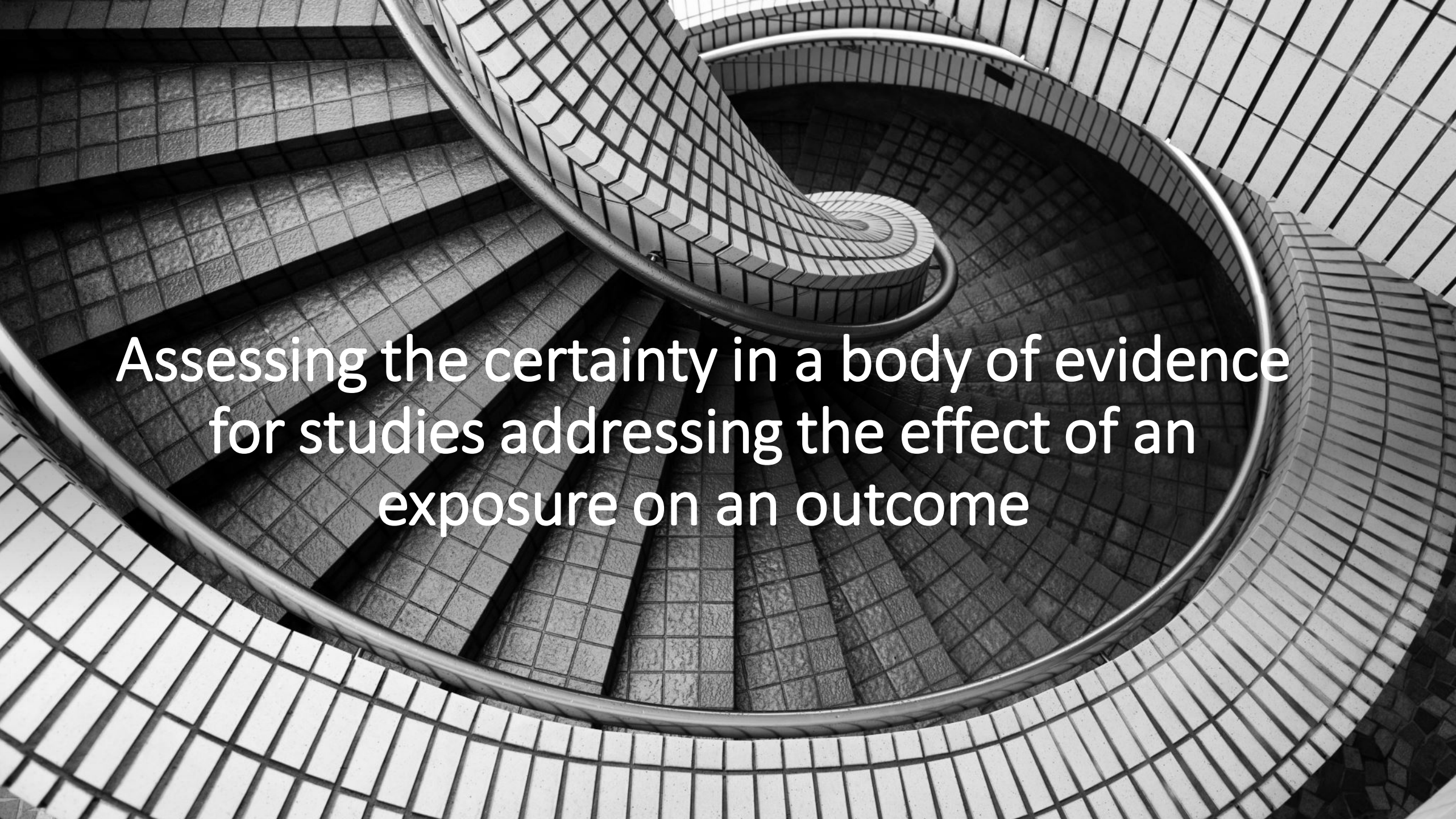

Assessing the certainty in a body of evidence for studies addressing the effect of an exposure on an outcome

Holger Schünemann, MD, MSc, PhD, FRCPC
@schunemann_mac | schuneh@mcmaster.ca
On behalf of the eCOVID RecMap team

HEI @McMaster
CERC @Humanitas

BFR, Berlin 10|11|2023

Disclosures

No direct financial conflicts

GRADE Working Group Co-Chair

Cochrane Canada - Director

Guidelines International Network – chair

INGUIDE – steering committee lead

Research grants from Canadian Institutes of Health Research (FRN VR4-172741, GA3-177732 & REC 183153), American Society of Hematology (ASH), WHO, Public Health Agency of Canada

Colleagues: R. Morgan, E. Senerth

Views expressed my own

Land Acknowledgment

McMaster University sits on the traditional territories of the Mississauga and Haudenosaunee nations and within the lands protected by the Dish With One Spoon
wampum agreement.

Today’s talk

Considerations for identifying the best body of evidence related to exposure studies

GRADE thoughts on assessing risk of bias across a body of evidence

How to use evidence about exposures in decision-making

- GRADE Evidence to Decision (EtD) frameworks

GRADE is a method/system/approach to operationalize:

- the assessment of the certainty in a body of evidence
- the criteria and process for making transparent decisions and recommendations

Environment International 122 (2019) 168–184

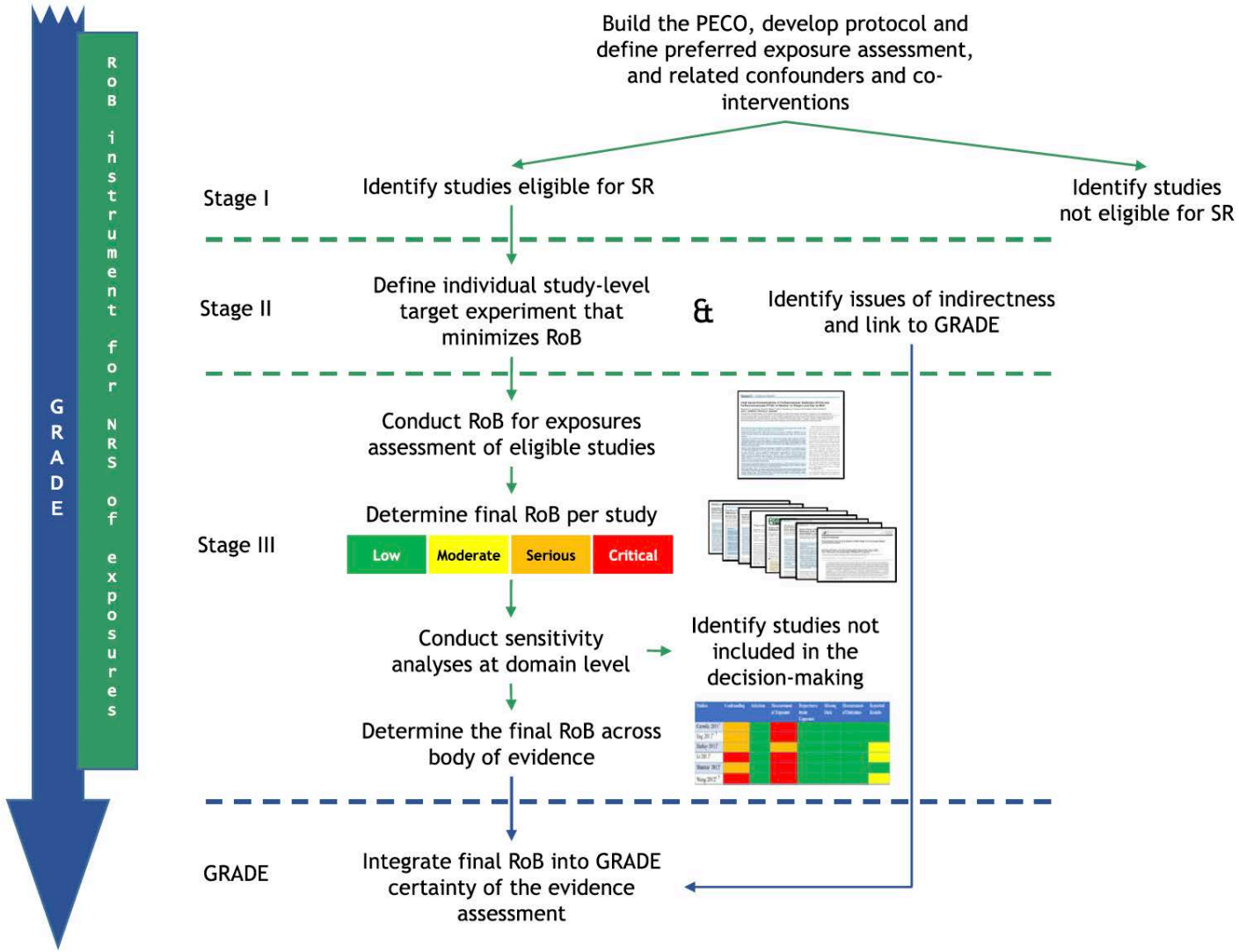
Contents lists available at ScienceDirect

Environment International

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

A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE

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], Mohammed Ansari[e], Elie A Akl[a,f], Tommaso Filippini[g], Anna Hansell[h,i,j], Joerg J. Meerpohl[k], Reem A. Mustafa[a,l], Jos Verbeek[m], Marco Vinceti[g,n], Paul Whaley[o], Holger J. Schünemann[a,p,*], GRADE Working Group



GRADE: Grading of Recommendations Assessment, Development and Evaluation; PECO: population, exposure, comparator, outcome; RoB: risk of bias; SR: systematic review.

Fig. 1. Approach for conducting an assessment using the RoB instrument for NRS of exposures and the integration into GRADE when conducting systematic reviews of exposure.
GRADE: Grading of Recommendations Assessment, Development and Evaluation; PECO: population, exposure, comparator, outcome; RoB: risk of bias; SR: systematic review.

Formulating questions

- No guiding framework for operationalizing the PECO approach and the types of PECO questions researchers and decision-makers existed
- In environmental, public and occupational health research, specific challenges exist with identifying the exposure and comparator within the PECO
- Five paradigmatic approaches and examples for identifying the exposure and comparator in systematic review and decision-making questions.

ELSEVIER

Environment International

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

Preface

Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes

Rebecca L. Morgan[a], Paul Whaley[b], Kristina A. Thayer[c], Holger J. Schünemann[a,d,*]

[a] Department of Health Research Methods, Evidence, and Impact (Formerly the Department of Clinical Epidemiology & Biostatistics) & Michael G. DeGroote Cochrane

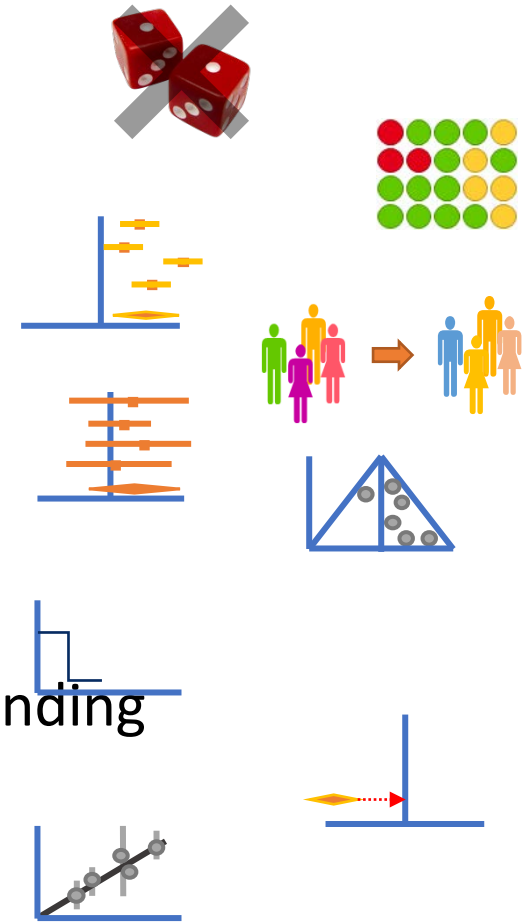
Table 1
Five paradigmatic approaches and examples for identifying the exposure and comparator in systematic review and decision-making questions (from Morgan RL, Whaley P, Thayer KA, Schünemann HJ: Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. Environment International 2018. (Morgan et al., 2018b))

Potential systematic-review or research context	Approach	PECO example
1. Calculate the health effect from an exposure; describing the dose-effect relationship between an exposure and an outcome for risk characterisation.	Explore the shape and distribution of the relationship between the exposure and the outcome in the systematic review.	Among newborns, what is the incremental effect of 10 dB increase during gestation on postnatal hearing impairment?
2. Evaluate the effect of an exposure cut-off on health outcomes, when the cut-off can be informed iteratively by the results of the systematic review.	Use cut-offs defined based on distribution in the studies identified in the systematic review.	Among newborns, what is the effect of the highest dB exposure compared to the lowest dB exposure (e.g. identified tertiles, quartiles, or quintiles) during pregnancy on postnatal hearing impairment?
3. Evaluate the association between an exposure cut-off and a comparison cut-off, when the cut-offs can be identified or are known from other populations.	Use mean cut-offs from external or other populations (may come from other research).	Among commercial pilots, what is the effect of noise corresponding to occupational exposure compared to noise exposure experienced in other occupations on hearing impairment?
4. Identify an exposure cut-off that ameliorates the effects on health outcomes.	Use existing exposure cut-offs associated with known health outcomes of interest.	Among industrial workers, what is the effect of exposure to <80 dB compared to ≥ 80 dB on hearing impairment?
5. Evaluate the potential effect of a cut-off* that can be achieved through an intervention to ameliorate the effects of exposure on health outcomes.	Select the comparator based on what exposure cut-offs can be achieved through an intervention.	Among the general population, what is the effect of an intervention that reduces noise levels by 20 dB compared to no intervention on hearing impairment?

Determinants of certainty in a body of evidence: GRADE

- A body of evidence starts as: high | ⨁⨁⨁⨁

- 5 factors that can lower certainty
  1. Risk of bias
  2. Inconsistency (or heterogeneity)
  3. Indirectness (PICO and applicability)
  4. Imprecision
  5. Publication bias
- 3 factors may increase certainty
  1. large magnitude of effect
  2. opposing plausible residual bias or confounding
  3. dose-response gradient



Any risk of bias tool can be used:

- Should be validated
- Cover the items of interest
- ROBINS-E good candidate

Environment International 120 (2018) 382–387

Contents lists available at ScienceDirect

Environment International

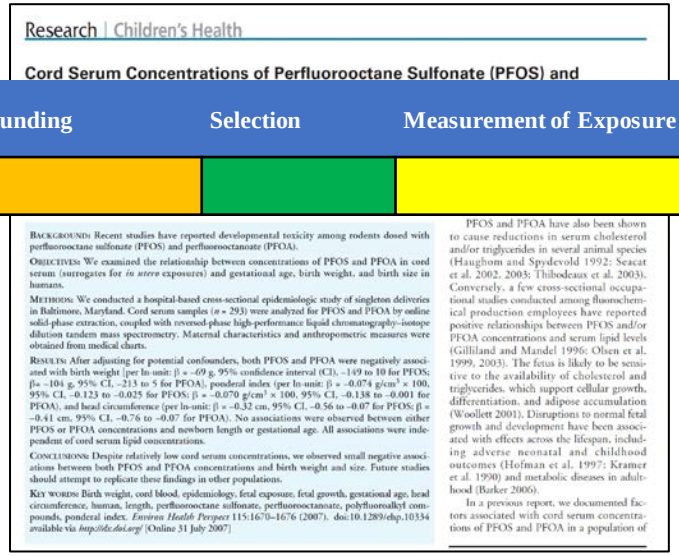
journal homepage: www.elsevier.com/locate/envint

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,*]

[a] Department of Health Research Methods, Evidence, and Impact (formerly the Department of Clinical Epidemiology & Biostatistics), McMaster University, Health Sciences Centre, Room 2C14, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada
[b] Integrated Risk Information System (IRIS) Division, National Center for Environmental Assessment (NCEA), Office of Research and Development, US Environmental Protection Agency, Building 8 (Room 2110), Research Triangle Park, NC 27711, USA
[c] Department of Obstetrics and Gynecology, McMaster University, Health Sciences Centre, Room 3N52A, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada
[d] ICF, 9300 Lee Highway, Fairfax, VA 22031, USA
[e] Department of Medicine, McMaster University, Health Sciences Centre, Room 2C14, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada

Evaluate RoB per outcome using the RoB instrument for NRS of exposures



Research | Children's Health

Cord Serum Concentrations of Perfluorooctane Sulfonate (PFOS) and

BACKGROUND: Recent studies have reported developmental toxicity among rodents dosed with perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA).

OBJECTIVES: We examined the relationship between concentrations of PFOS and PFOA in cord serum (surrogates for *in utero* exposures) and gestational age, birth weight, and birth size in humans.

METHODS: We conducted a hospital-based cross-sectional epidemiologic study of singleton deliveries in Baltimore, Maryland. Cord serum samples ($n = 293$) were analyzed for PFOS and PFOA by online solid-phase extraction, coupled with reversed-phase high-performance liquid chromatography–isotope dilution tandem mass spectrometry. Maternal characteristics and anthropometric measures were obtained from medical charts.

RESULTS: After adjusting for potential confounders, both PFOS and PFOA were negatively associated with birth weight [per ln-unit: β = –69 g, 95% confidence interval (CI), –149 to 10 for PFOS; β= –104 g, 95% CI, –213 to 5 for PFOA], ponderal index (per ln-unit: β = –0.074 g/cm³ × 100, 95% CI, –0.123 to –0.025 for PFOS; β = –0.070 g/cm³ × 100, 95% CI, –0.138 to –0.001 for PFOA), and head circumference (per ln-unit: β = –0.32 cm, 95% CI, –0.56 to –0.07 for PFOS; β = –0.41 cm, 95% CI, –0.76 to –0.07 for PFOA). No associations were observed between either PFOS or PFOA concentrations and newborn length or gestational age. All associations were independent of cord serum lipid concentrations.

CONCLUSIONS: Despite relatively low cord serum concentrations, we observed small negative associations between both PFOS and PFOA concentrations and birth weight and size. Future studies should attempt to replicate these findings in other populations.

KEY WORDS: Birth weight, cord blood, epidemiology, fetal exposure, fetal growth, gestational age, head circumference, human, length, perfluorooctane sulfonate, perfluorooctanoate, polyfluoroalkyl compounds, ponderal index. *Environ Health Perspect* 115:1670–1676 (2007). doi:10.1289/ehp.10334 available via *http://dx.doi.org/* [Online 31 July 2007]

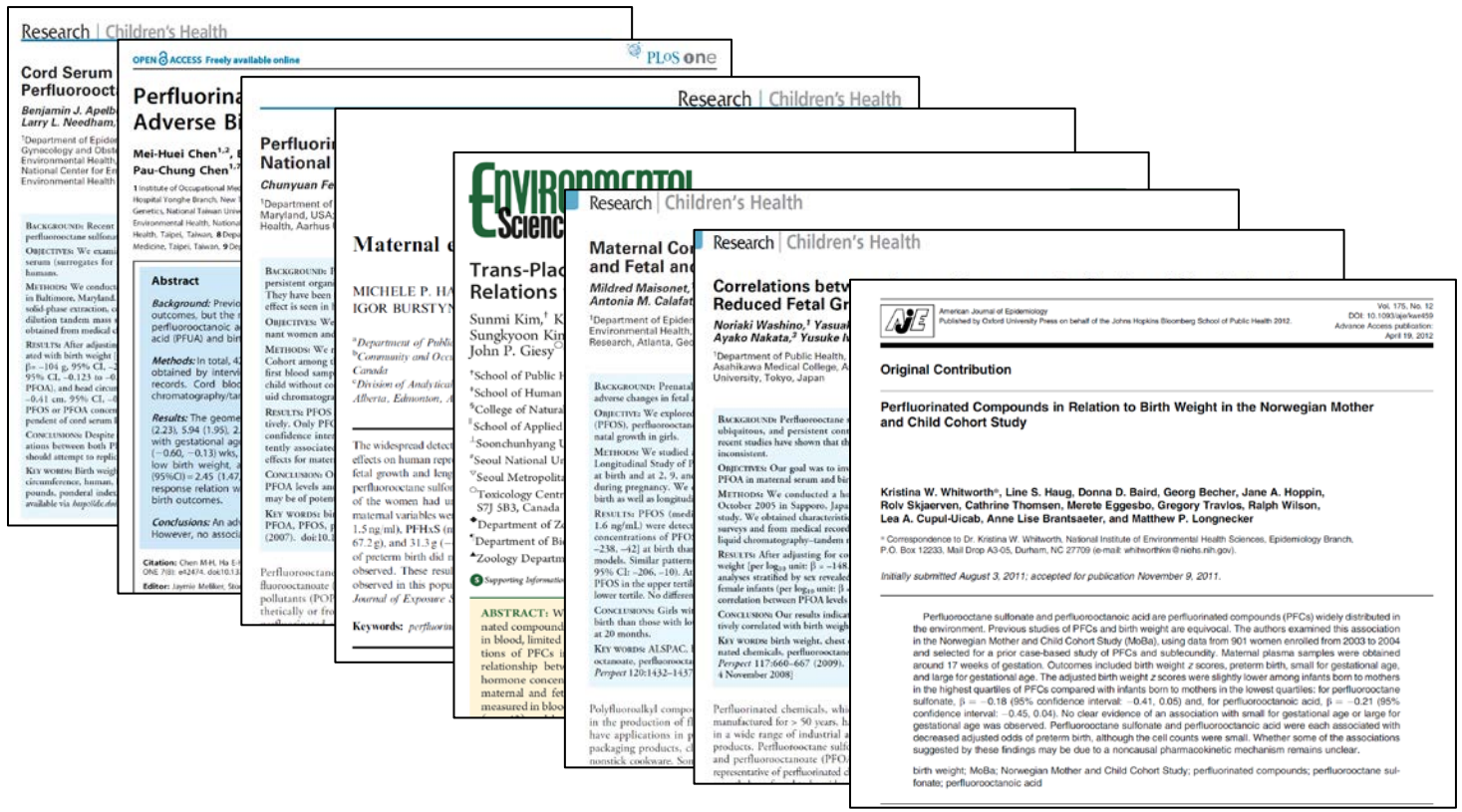
PFOS and PFOA have also been shown to cause reductions in serum cholesterol and/or triglycerides in several animal species (Haughom and Spydevold 1992; Seacat et al. 2002, 2003; Thibodeaux et al. 2003). Conversely, a few cross-sectional occupational studies conducted among fluorochemical production employees have reported positive relationships between PFOS and/or PFOA concentrations and serum lipid levels (Gilliland and Mandel 1996; Olsen et al. 1999, 2003). The fetus is likely to be sensitive to the availability of cholesterol and triglycerides, which support cellular growth, differentiation, and adipose accumulation (Woollett 2001). Disruptions to normal fetal growth and development have been associated with effects across the lifespan, including adverse neonatal and childhood outcomes (Hofman et al. 1997; Kramer et al. 1990) and metabolic diseases in adulthood (Barker 2006).

In a previous report, we documented factors associated with cord serum concentrations of PFOS and PFOA in a population of

Study	Confounding	Selection	Measurement of Exposure	Departures from Exposure	Missing Data	Measurement of Outcomes	Reported Results
Apelberg et al. 2007							

- Items
  - Confounding
  - Selection
  - Measurement of Exposure
  - Departures from Exposure
  - Missing Data
  - Measurement of Outcomes
  - Reported Results

Low	Moderate	Serious	Critical


Research | Children's Health
Cord Serum
Perfluorooct
Benjamin J. Apelb
Larry L. Needham,
OPEN ACCESS Freely available online
PLoS one
Perfluorina
Adverse Bi
Mei-Huei Chen
Pau-Chung Chen
Abstract
Research | Children's Health
Perfluorit
National
Chunyuan Fe
Maternal
MICHELE P. HA
IGOR BURSTYN
Research | Children's Health
Trans-Plac
Relations
Sunmi Kim,
Sungkyoon Kim
John P. Giesy
Research | Children's Health
Maternal Cor
and Fetal and
Mildred Maisonet,
Antonia M. Calafat
Research | Children's Health
Correlations betv
Reduced Fetal Gr
Noriaki Washino,
Ayako Nakata,
American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2012.
Vol. 175, No. 12
DOI: 10.1093/aje/kwr459
Advance Access publication:
April 19, 2012
Original Contribution
Perfluorinated Compounds in Relation to Birth Weight in the Norwegian Mother and Child Cohort Study
Kristina W. Whitworth, Line S. Haug, Donna D. Baird, Georg Becher, Jane A. Hoppin, Rolv Skjaerven, Cathrine Thomsen, Merete Eggesbo, Gregory Travlos, Ralph Wilson, Lea A. Cupul-Uicab, Anne Lise Brantsaeter, and Matthew P. Longnecker
Initially submitted August 3, 2011; accepted for publication November 9, 2011.
Perfluorooctane sulfonate and perfluorooctanoic acid are perfluorinated compounds (PFCs) widely distributed in the environment. Previous studies of PFCs and birth weight are equivocal. The authors examined this association in the Norwegian Mother and Child Cohort Study (MoBa), using data from 901 women enrolled from 2003 to 2004 and selected for a prior case-based study of PFCs and subfecundity. Maternal plasma samples were obtained around 17 weeks of gestation. Outcomes included birth weight z scores, preterm birth, small for gestational age, and large for gestational age. The adjusted birth weight z scores were slightly lower among infants born to mothers in the highest quartiles of PFCs compared with infants born to mothers in the lowest quartiles: for perfluorooctane sulfonate, β = −0.18 (95% confidence interval: −0.41, 0.05) and, for perfluorooctanoic acid, β = −0.21 (95% confidence interval: −0.45, 0.04). No clear evidence of an association with small for gestational age or large for gestational age was observed. Perfluorooctane sulfonate and perfluorooctanoic acid were each associated with decreased adjusted odds of preterm birth, although the cell counts were small. Whether some of the associations suggested by these findings may be due to a noncausal pharmacokinetic mechanism remains unclear.
birth weight; MoBa; Norwegian Mother and Child Cohort Study; perfluorinated compounds; perfluorooctane sulfonate; perfluorooctanoic acid

RoB Matrix: Exposure to BPA on prevalent overweight and obesity

Studies	Confounding	Selection	Measurement of Exposure	Departures from Exposure	Missing Data	Measurement of Outcomes	Reported Results
Carwile 2011*							
Eng 2013*, †							
Harley 2013*							
Li 2013*							
Shankar 2012†							
Wang 2012*, †							

* Prevalent overweight
† Prevalent obesity

Low | Moderate | Serious | Critical

Ranciere, F., Lyons, J. G., Loh, V. H., Botton, J., Galloway, T., Wang, T., . . . Magliano, D. J. (2015). Bisphenol A and the risk of cardiometabolic disorders: a systematic review with meta-analysis of the epidemiological evidence. *Environ Health, 14*(1), 46. doi:10.1186/s12940-015-0036-5

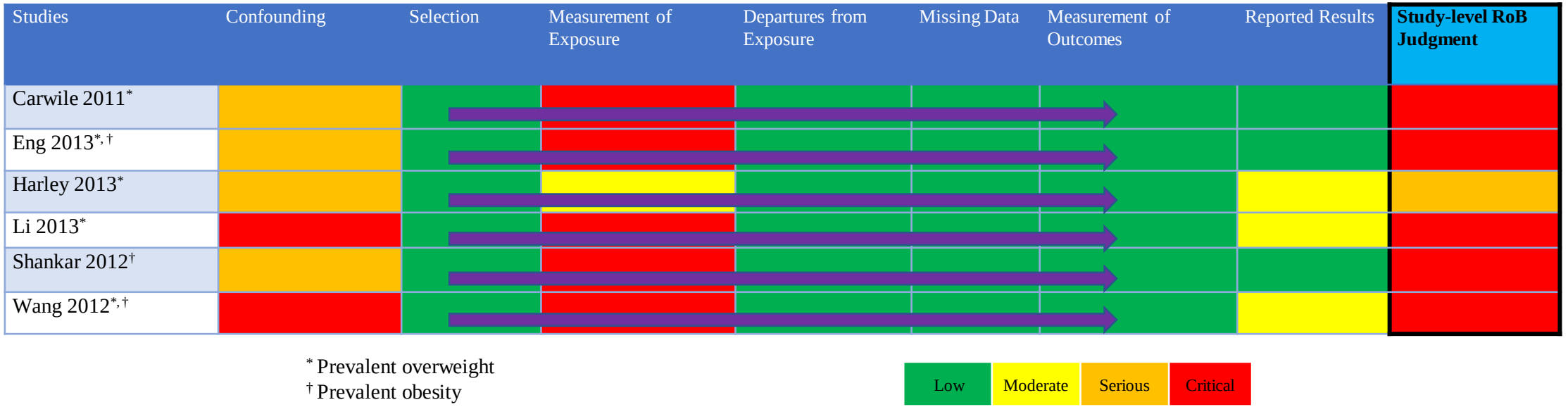
RoB Matrix: Exposure to BPA on prevalent overweight and obesity

Studies	Confounding	Selection	Measurement of Exposure	Departures from Exposure	Missing Data	Measurement of Outcomes	Reported Results
Carwile 2011*							
Eng 2013*, †							
Harley 2013*							
Li 2013*							
Shankar 2012†							
Wang 2012*, †							

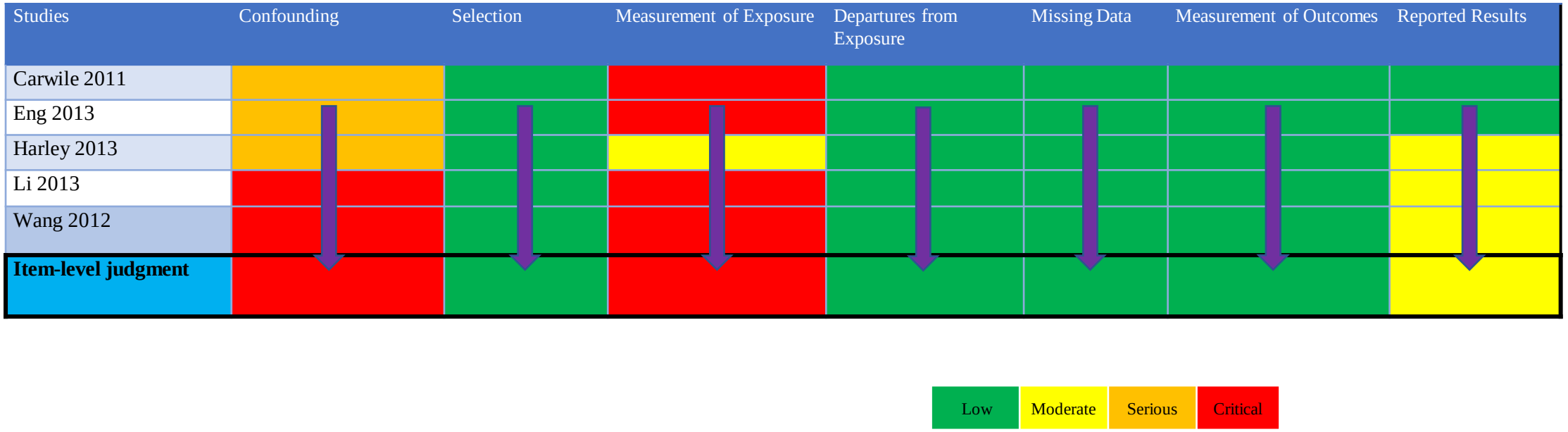
* Prevalent overweight
† Prevalent obesity

Ranciere, F., Lyons, J. G., Loh, V. H., Botton, J., Galloway, T., Wang, T., . . . Magliano, D. J. (2015). Bisphenol A and the risk of cardiometabolic disorders: a systematic review with meta-analysis of the epidemiological evidence. *Environ Health, 14*(1), 46. doi:10.1186/s12940-015-0036-5

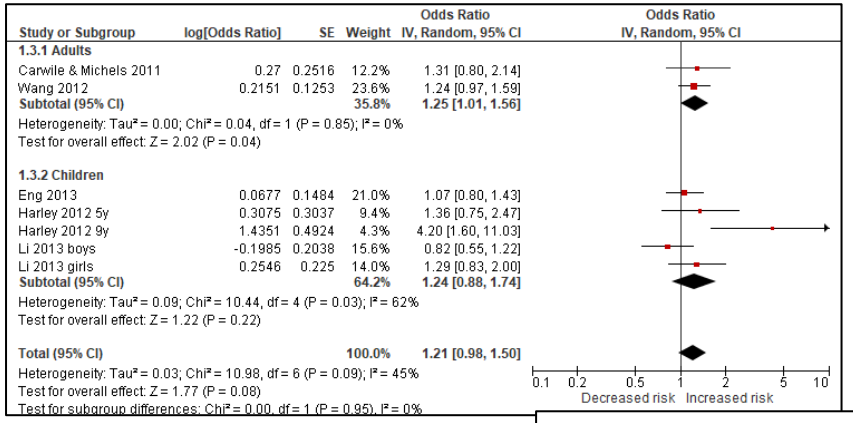
RoB judgment across the body of evidence



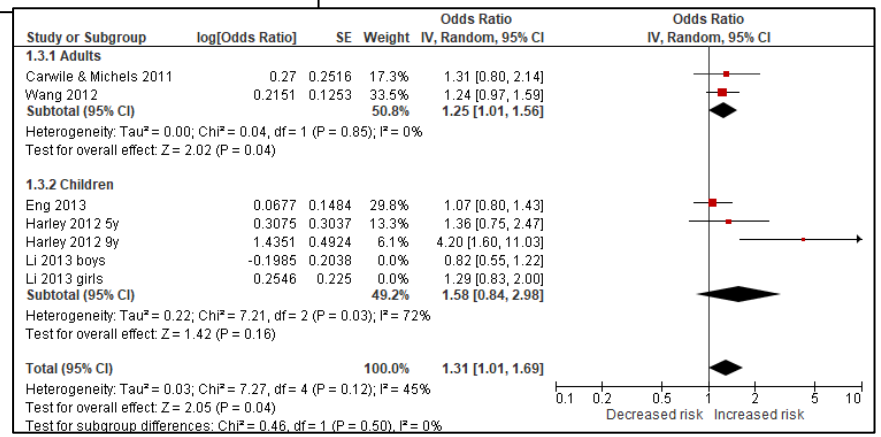
RoB judgment across the body of evidence (prevalent overweight): Part 2



Prevalent overweight

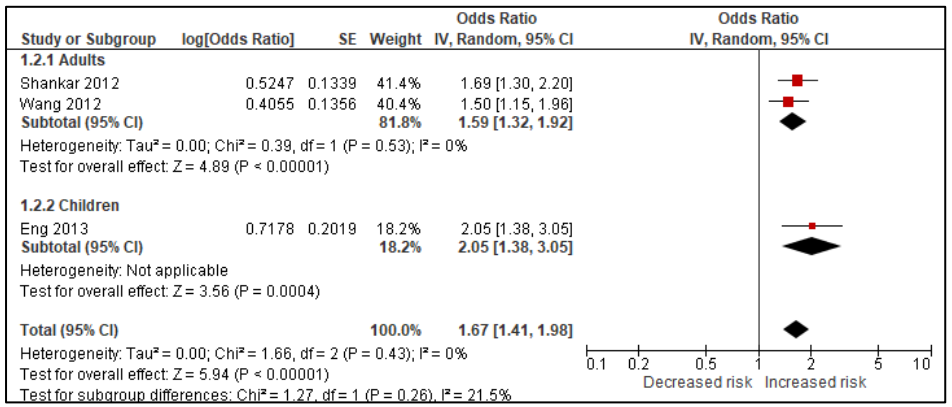


1



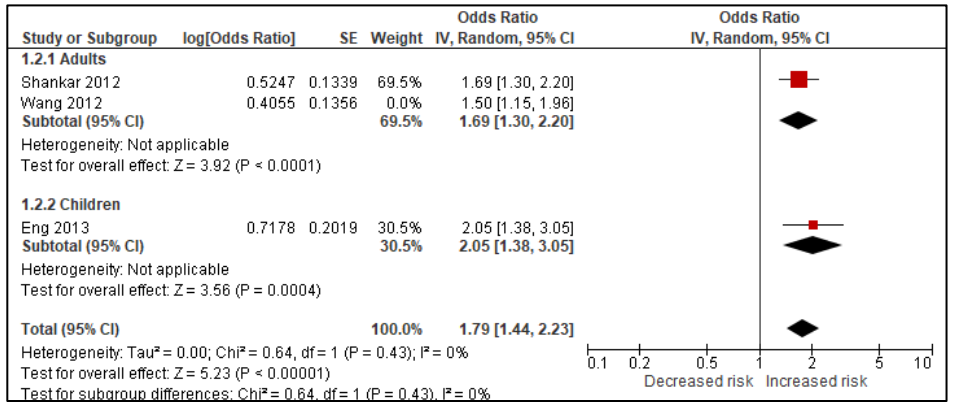
2

By outcome: Prevalent obesity



1

2



Studies	Confounding	Selection	Measurement of Exposure	Departures from Exposure	Missing Data	Measurement of Outcomes	Reported Results	Study level risk of bias
Apelberg et al. 2007*								
Arbuckle et al. 2011								
Chen et al. 2012*								
Fei et al. 2007*								
Fei et al. 2008								
Halldorsson et al. 2011								
Hamm et al. 2010*								
Kim S et al. 2011*								
Kim SK et al. 2011		NI				NI		
Maisonet et al. 2012*								
Monroy et al. 2007								
Nolan et al. 2009*								
Savitz et al. 2012a								
Savitz et al. 2012b								
Stein et al. 2008								
Washino et al. 2008*								
Whitworth et al. 2012*								
Domain-level RoB Judgment								

Low	Moderate	Serious	Critical

- The weighing of the domains for an overall assessment of risk of bias requires considered judgment across domains. The type of bias (domains) should not be equally weighted by default
- Reducing risk of bias through inclusion and exclusion of studies and sensitivity analysis may or may not come at cost of applicability (directness)

- Risk of bias assessment should include an assessment of the direction (and if possible magnitude) of risk of bias
- The GRADE domain of opposing residual plausible confounding is integrated with the risk of bias assessment
- In the context of GRADE confounding bias the default concern is that risk of bias on that domain is serious.
- Look for reasons why this is not the case

Transparency within the Evidence Profile: GRADE assessment

Quality assessment							№ of patients		Effect		Quality
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exposure to BPA (CAS# 80-05-7)	exposure to lower levels of BPA	Relative (95% CI)	Absolute (95% CI)	
Prevalent overweight (assessed with: BMI ≥85th percentile for age/gender in children; BMI 18.5-25/30 kg/m2)											
5	studies	very, very	not serious [b]	not serious	serious	none	1774/5403	1584/5657	OR 1.21	40 more per	⊕◯◯◯
Prevalent obesity											
3	studies										

a. Most studies adjusted for known confounders of body composition (age, ethnicity, gender, height, race), and diet; however, two studies did not account for caloric intake or diet which is relevant for evaluating weight-related outcomes, there is some risk of unmeasured confounding; BPA measurement present potential for bias as the chemical is non-persistent with a short half-life and exposure measurements were not repeated (except in one study), one study measures BPA three months post-BMI measurement, remaining studies measure BPA and BMI at the same time; potential risk of reporting bias because three studies did not report prior publication of a protocol; however, all studies present outcome measures and analyses consistent with a priori plan outlined in the manuscript.

CI: Confide

b. The I^2 value = 45% and exploration of the forest plot suggests some inconsistency introduced by one outlying study contributing 4.3% of the weight to the analysis of children.

c. Imprecision is present because the width of the confidence interval is consistent with both important benefit and harm.

Environment International 157 (2021) 106868

Contents lists available at ScienceDirect

Environment International

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

An approach to quantifying the potential importance of residual confounding in systematic reviews of observational studies: A GRADE concept paper

Jos H. Verbeek [a,*], Paul Whaley [b], Rebecca L. Morgan [c], Kyla W. Taylor [d], Andrew A. Rooney [d], Lukas Schwingshackl [e], Jan L. Hoving [f], S. Vittal Katikireddi [g], Beverley Shea [h], Reem A. Mustafa [i], M. Hassan Murad [j], Holger J. Schünemann [k], GRADE Working Group

A note on residual plausible confounding

Four step approach for using the E-value to judge how likely it is that residual confounding can reduce the observed effect to null or below a threshold of interest

Making certainty assessments transparent

Contents lists available at ScienceDirect

Environment International

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

Preface

Using GRADE to respond to health questions with different levels of urgency

Kristina A. Thayer [a], Holger J. Schünemann [b,*]

[a] Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, P.O. Box 12233, Mail Drop K2-02, Research Triangle Park, NC 27709, USA
[b] Department of Clinical Epidemiology & Biostatistics, Department of Medicine, McMaster University, Health Sciences Centre, Room 2C14, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada

ARTICLE INFO

Article history:
Received 15 March 2016
Received in revised form 21 March 2016
Accepted 21 March 2016
Available online 26 April 2016

ABSTRACT

Increasing interest exists in applying the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to environmental health evidence. While ideally applied to evidence synthesized in systematic reviews and corresponding summary tables, such as evidence profiles, GRADE's correct application requires that "the evidence that was assessed and the methods that were used to identify and appraise that evidence should be clearly described." In this article, we suggest that GRADE could be applied to evidence assembled from narrative reviews, modelled (indirect) evidence, or evidence assembled as part of a rapid response, if the underlying judgments about the certainty in this evidence are based on the relevant GRADE domains and provided transparently. Health questions that require assessing the certainty in a body of evidence to provide trustworthy answers may range from hours, to days or weeks, to a few months to scenarios that allow assessing evidence without short-term time pressures. Time frames of emergent, urgent or rapid evidence assessments will often require relying on existing summaries or rapidly compiling the available evidence and making assessments. Even without available full systematic reviews, expressing the certainty in the evidence can provide useful guidance for users of the evidence and those who evaluate certainty in effects. The ratings also help clarifying disagreement between organizations tackling similar questions about the evidence. Using the structured GRADE domains, narrative or other summaries of the evidence can be presented transparently.

Table 1
Examples of GRADE applied across different time scenarios.

Type of response	Ultra-short emergency response: within one or more hours	Urgent response: one to two weeks	Rapid response: one to three months	Routine response: more than 3 months
Example	West Virginia Elk River spill Population: community exposed to the chemical spill. Intervention/exposure: chemicals in the spill that contaminated water supply. Comparison: no chemicals in the spill. Outcomes: genotoxicity, developmental or reproductive toxicity, liver toxicity and others.	Melamine in composite food products Population: healthy people Intervention/exposure: melamine from composition food products below 0.5 mg/kg body weight per day. Comparison: higher than 0.5 mg/kg body weight of melamine from composition food. Outcomes: renal insufficiency (assessed with renal clearance), urinary tract calculi, urinary tumors (used for this example of the certainty in the evidence).	Avian influenza Population: people with suspected avian influenza infection. Intervention/exposure: oseltamivir. Comparison: no oseltamivir. Outcomes: mortality, duration of hospitalization, incidence of lower respiratory tract complications (used for this example of the certainty assessment below), antiviral drug resistance existing before treatment, and serious adverse events.	PFOA and birth weight Population: women of reproductive age and fetuses (before and/or during pregnancy or development). Intervention/exposure: perfluorooctanoic acid (PFOA; CAS# 335-67-1) or its salts. Comparison: lower levels of PFOA. Outcomes: fetal growth, birth weight, other measures of fetal or newborn size.
Type of evidence	Available evidence: animal toxicology studies in rodents for two chemicals in the spill (a 28-day study and a teratology study) and SAR analyses for other chemicals in the spill with no toxicology data.	Available evidence: animal toxicology studies in rat and mice with exposures to various levels of melamine via feeding, including a control group. The utilized evidence should be supported by a literature search with transparent inclusion and exclusion criteria and a (narrative) summary of that evidence.	Available evidence: five randomized trials in patients with seasonal flu (summarized in systematic reviews), case studies of patients with avian influenza, *in vitro* and *in vivo* animal data.	Available evidence: a systematic review of 18 non-randomized (observational) studies (10 were included in a meta-analysis).
GRADE domains to assess certainty in the evidence: suggested approaches to making judgments or proposed judgments (note these are not necessarily reflecting judgments in the original scenarios).				
Risk of bias	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g. randomization, blinding at outcome assessment, sufficient characterization of test compound, or whether all animals were accounted for). Ideally, RoB assessments would be available for individual studies and summarized across studies. In the Elk River example, the number of animal studies was small and could be assessed at the individual level within a short-time frame. A de novo risk of bias evaluation may not be feasible in cases where evidence is drawn from existing narrative risk assessments that summarize a large body of literature. Nevertheless, it may still be possible to assess risk of bias based on the uncertainties and evidence limitations described in the risk assessment. SAR: could be assessed using OECD model validation or similar guidance that recommends presentation of a defined domain of applicability for a defined endpoint supported by appropriate measures of goodness-of-fit (OECD, 2007).	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g. randomization, pathologists blinded in their assessments or all animals accounted for). In this case it appears that the animal studies did not report that it was randomized and, thus, may be at risk of bias.	Not serious	Serious based on some concern of risk of bias in the included studies (in the original report, the authors used an approach to rating certainty that accounted for risk of bias by lowering the certainty from high to moderate).
Imprecision	Could be assessed for both animal data and SAR (e.g., considering statistical or numerical uncertainty in model parameters).	While no summary estimates are available, an assessment could be guided by the availability of data from only 100 animals in different exposure groups which would result in wide confidence intervals.	Serious	Not serious
Inconsistency	Could be assessed for both animal data and SAR (e.g., assessing similarity of results based on applying different models).	Only one study was included and therefore no inconsistency is present (Guyatt et al., 2011d).	Not serious	Not serious
Publication bias	Could be assessed for both animal studies and SAR. A judgment of undetected might be reasonable if	Could be assessed using guidance for animal studies but a judgment of undetected might be reasonable if	Undetected	Undetected

Table 1
Examples of GRADE applied across different time scenarios.

Type of response	Ultra-short emergency response: within one or more hours	Urgent response: one to two weeks	Rapid response: one to three months	Routine response: more than 3 months
Example	West Virginia Elk River spill Population: community exposed to the chemical spill. Intervention/exposure: chemicals in the spill that contaminated water supply. Comparison: no chemicals in the spill. Outcomes: genotoxicity, developmental or reproductive toxicity, liver toxicity and others.	Melamine in composite food products Population: healthy people Intervention/exposure: melamine from composition food products below 0.5 mg/kg body weight per day. Comparison: higher than 0.5 mg/kg body weight of melamine from composition food. Outcomes: renal insufficiency (assessed with renal clearance), urinary tract calculi, urinary tumors (used for this example of the certainty in the evidence).	Avian influenza Population: people with suspected avian influenza infection. Intervention/exposure: oseltamivir. Comparison: no oseltamivir. Outcomes: mortality, duration of hospitalization, incidence of lower respiratory tract complications (used for certainty antiviral before tr adverse e	PFOA and birth weight Population: women of reproductive age and fetuses (before and/or during pregnancy or development). Intervention/exposure: perfluorooctanoic acid (PFOA; CAS# 335-67-1) or its salts. Comparison: lower levels of PFOA.
Type of evidence	Available evidence: animal toxicology studies in rodents for two chemicals in the spill (a 28-day study and a teratology study) and SAR analyses for other chemicals in the spill with no toxicology data.	Available evidence: animal toxicology studies in rat and mice with exposures to various levels of melamine via feeding, including a control group. The utilized evidence should be supported by a literature search with transparent inclusion and exclusion criteria and a (narrative) summary of that evidence.	Available trials in p (summar reviews), with avia *in vivo* ar	
GRADE domains to assess certainty in the evidence: suggested approaches to making judgments or proposed judg				
original scenarios).				
Risk of bias	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g. randomization, blinding at outcome assessment, sufficient characterization of test compound, or whether all animals	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g. randomization, pathologists blinded in their assessments or all animals accounted for). In this case it	Not serio	

Rest of table summarizes:

- GRADE domains
  - risk of bias, imprecision, indirectness, inconsistency, publication bias, magnitude, etc.
- Certainty in evidence
- Possible summary statements

GRADE domains to assess certainty in the evidence: suggested approaches to making judgments or proposed judgments (note these are not necessarily reflecting judgments in the original scenarios).

Risk of bias	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g. randomization, blinding at outcome assessment, sufficient characterization of test compound, or whether all animals were accounted for). Ideally, RoB assessments would be available for individual studies and summarized across studies. In the Elk River example, the number of animal studies was small and could be assessed at the individual level within a short-time frame. A de novo risk of bias evaluation may not be feasible in cases where evidence is drawn from existing narrative risk assessments that summarize a large body of literature. Nevertheless, it may still be possible to assess risk of bias based on the uncertainties and evidence limitations described in the risk assessment. SAR: could be assessed using OECD model validation or similar guidance that recommends presentation of a defined domain of applicability for a defined endpoint supported by appropriate measures of goodness-of-fit (OECD, 2007).	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g. randomization, pathologists blinded in their assessments or all animals accounted for). In this case it appears that the animal studies did not report that it was randomized and, thus, may be at risk of bias.	Not serious	Serious based on some concern of risk of bias in the included studies (in the original report, the authors used an approach to rating certainty that accounted for risk of bias by lowering the certainty from high to moderate).
Imprecision	Could be assessed for both animal data and SAR (e.g., considering statistical or numerical uncertainty in model parameters).	While no summary estimates are available, an assessment could be guided by the availability of data from only 100 animals in different exposure groups which would result in wide confidence intervals.	Serious	Not serious
Inconsistency	Could be assessed for both animal data and SAR (e.g., assessing similarity of results based on applying different models).	Only one study was included and therefore no inconsistency is present (Guyatt et al., 2011d).	Not serious	Not serious

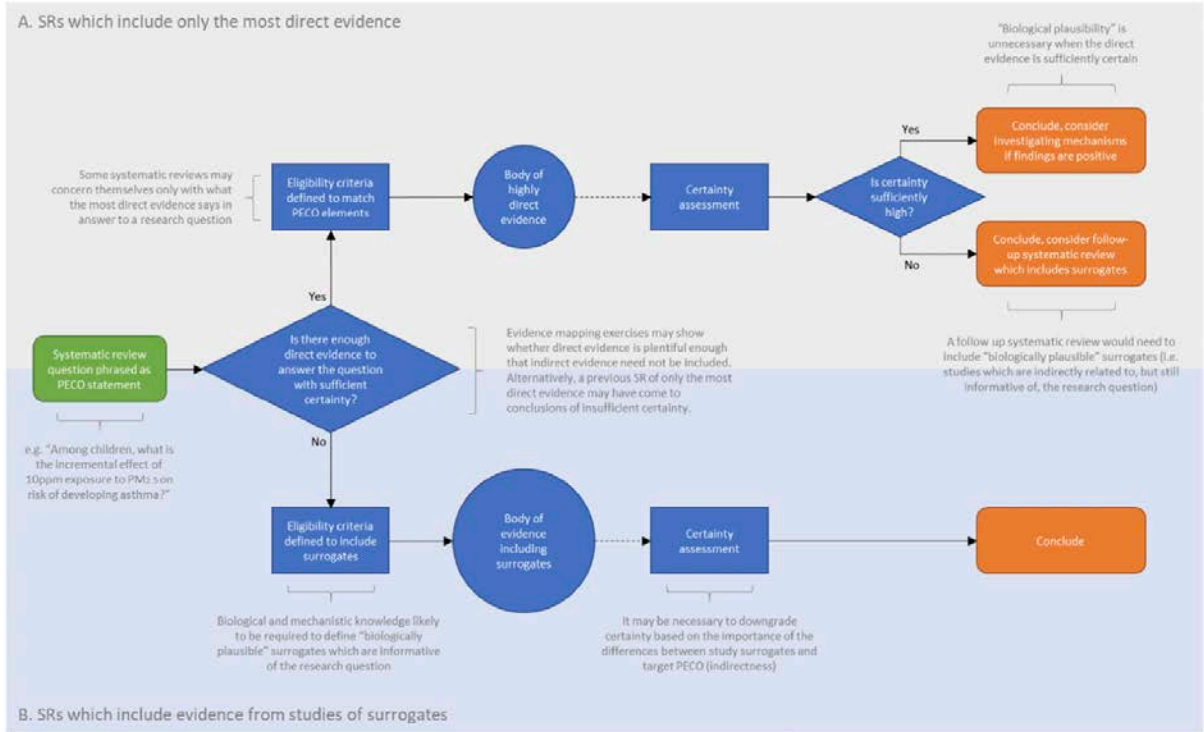


Fig. 2. Schematic representation of how it might be decided to include studies of surrogates in a systematic review.

Table 2
Summary of the 10 examples used in this manuscript to show how discussion of biological pla

	Surrogates of higher biological plausibility, for which indirectness is less important
Population	Animal models for human carcinogenicity of 2-nitropropane
Exposure (dose)	Extrapolating from high doses to low doses of genotoxic substances
Exposure (route)	Oral administration of bisphenol-A via gavage, or availability of a pharmacokinetic model to translate intravenous dose to oral equivalent
Exposure (substance)	Inferring estrogenic potential of other bisphenols and from studies of bisphenol-A
Outcome	Maternal serum thyroxine (T4) for child neurodevelopmental outcomes

Environment International 162 (2022) 107109

Contents lists available at ScienceDirect

Environment International

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

Biological plausibility in environmental health systematic reviews: a GRADE concept paper☆

Paul Whaley [a,b], Thomas Piggott [c], Rebecca L. Morgan [c], Sebastian Hoffmann [b], Katya Tsaioun [b], Lukas Schwingshackl [d], Mohammed T. Ansari [e], Kristina A. Thayer [f], Holger J. Schünemann [c,g,h,*]

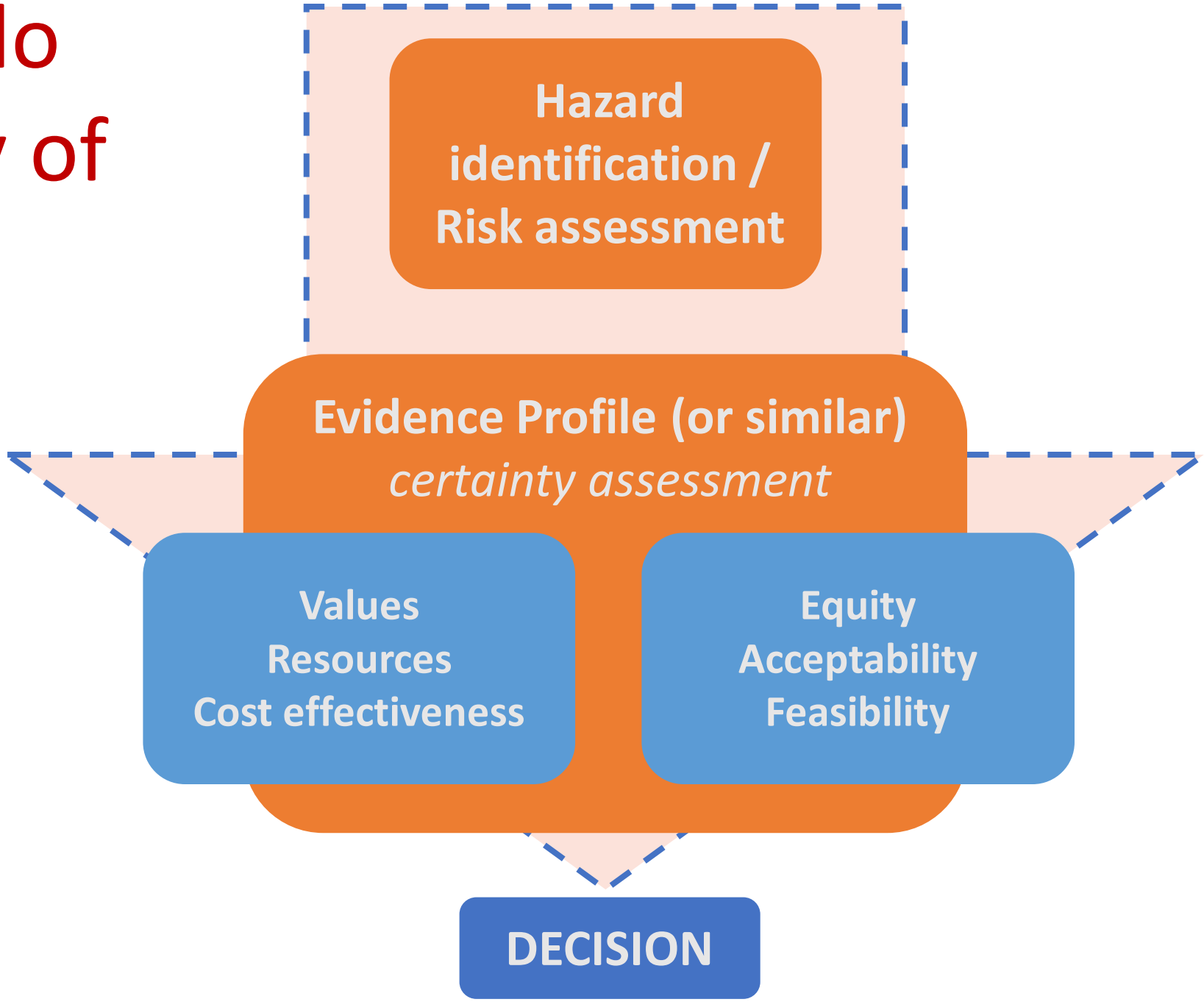
Table 3
Summary of potential influencing factors in judging biological plausibility or external validity of study surrogates, as suggested by the examples in this manuscript.

	Potential influencing factors in judging the biological plausibility or external validity of study surrogates
Population	The extent to which the biological pathway connecting exposure to outcome is operating in both the surrogate population and the target population (Fig. 5A)
Exposure – dose	The similarity of the toxicodynamic and toxicokinetic processes by which the surrogate dose acts in comparison to that of the dose range of interest
Exposure – route	The similarity by which an organism absorbs and metabolises the substance of concern via the surrogate route as opposed to the target route; or the reliability with which exposure from the surrogate route can be transformed to values which match exposure from the route of interest
Exposure –substance	The extent to which the surrogate molecule influences the biological processes by which the target molecule is thought to elicit its biological effects (Fig. 5C)
Outcome	The extent to which a surrogate outcome is predictive of the target outcome of concern (Fig. 5B)

Type of response	Ultra-short emergency response: within one or more hours	Urgent response: one to two weeks	Rapid response: one to three months	Routine response: more than 3 months
Indirectness	Animal studies: could be assessed using GRADE's indirectness assessment (Guyatt et al., 2011c; Schünemann et al., 2013). Animal studies may be rated down for indirectness if concerns exist about extrapolating from animals to humans, e.g., relevance of animal model for the health outcome of interest or route of exposure. SAR: could be assessed based on evidence of direct relation of the model to a defined endpoint. SAR would typically be downgraded for indirectness.	This could be rated down for serious indirectness of extrapolating from animals to humans and uncertainty about the levels of exposure (different levels or routes of exposure evaluated than those one is interested in and modeling of exposure levels based on composition food products from more exact exposures fed to animals). Further concerns would likely be described for the comparator.	Very serious	Not serious
Possible summary statement*	There is low certainty in the evidence suggesting no association between the exposure and toxicity based on SAR analyses.	There is very low certainty in the evidence suggesting no association between levels of melamine exposure from composition food products below 0.5 mg/kg body weight per day and urinary tumors.	There is very low certainty suggesting that oseltamivir reduces hospitalization in patients with avian influenza.	There is moderate certainty in the evidence suggesting that PFOA is associated with harmful effects on fetal growth.

* Note, this hypothetical summary was derived by the authors of this editorial, not those of the original report.

What do we do with the body of evidence?



Environment International

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

Preface

Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes

Rebecca L. Morgan[a], Paul Whaley[b], Kristina A. Thayer[c], Holger J. Schünemann[a,d,*]

Potential systematic-review or research context	Approach
1. Calculate the health effect from an exposure; describing the dose-effect relationship between an exposure and an outcome for risk characterization.	Explore the shape and distribution of the relationship between the exposure and the outcome in the systematic review.
2. Evaluate the effect of an exposure cut-off on health outcomes, when the cut-off can be informed iteratively by the results of the systematic review.	Use cut-offs defined based on distribution in the studies identified in the systematic review.
3. Evaluate the association between an exposure cut-off and a comparison cut-off, when the cut-offs can be identified or are known from other populations.	Use mean cut-offs from external or other populations (may come from other research).
4. Identify an exposure cut-off that ameliorates the effects on health outcomes.	Use existing exposure cut-offs associated with known health outcomes of interest.
5. Evaluate the potential effect of a cut-off that can be achieved through an intervention to ameliorate the effects of exposure on health outcomes.	**Select the comparator based on what exposure cut-offs can be achieved through an intervention.**

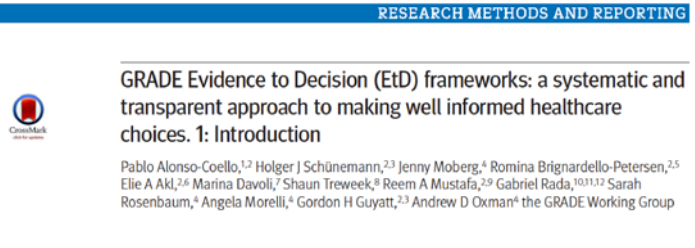
Neumann et al. Implementation Science (2016) 11:93
DOI 10.1186/s13012-016-0462-y

Implementation Science

RESEARCH Open Access

The GRADE evidence-to-decision framework: a report of its testing and application in 15 international guideline panels

Ignacio Neumann[1,2], Romina Brignardello-Petersen[1,3], Wojtek Wiercioch[1], Alonso Carrasco-Labra[1,3], Carlos Cuello[1], Elie Akl[4], Reem A. Mustafa[1,5], Waleed Al-Hazzani[1], Itziar Etxeandia-Ikobaltzeta[1,7], Maria Ximena Rojas[8], Maicon Falavigna[9], Nancy Santesso[1], Jan Brozek[1,6], Alfonso Iorio[1], Pablo Alonso-Coello[1,10] and Holger J. Schünemann[1,6*]



RESEARCH METHODS AND REPORTING

GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction

Pablo Alonso-Coello,[1,2] Holger J Schünemann,[2,3] Jenny Moberg,[4] Romina Brignardello-Petersen,[2,5] Elie A Akl,[2,6] Marina Davoli,[7] Shaun Treweek,[8] Reem A Mustafa,[2,9] Gabriel Rada,[10,11,12] Sarah Rosenbaum,[4] Angela Morelli,[4] Gordon H Guyatt,[2,3] Andrew D Oxman[4] the GRADE Working Group

GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines

Pablo Alonso-Coello,[1,2] Andrew D Oxman,[3] Jenny Moberg,[3] Romina Brignardello-Petersen,[2,4] Elie A Akl,[2,5] Marina Davoli,[6] Shaun Treweek,[7] Reem A Mustafa,[2,8] Per O Vandvik,[3] Joerg Meerpohl,[9] Gordon H Guyatt,[2,10] Holger J Schünemann,[2,10] the GRADE Working Group

GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health

Holger J. Schünemann[a,b,c,*], Reem Mustafa[a,c,d], Jan Brozek[a,b,c], Nancy Santesso[a,c], Pablo Alonso-Coello[a,c,e], Gordon Guyatt[a,b,c], Rob Scholten[f], Miranda Langendam[c,g], Mariska M. Leeflang[g], Elie A. Akl[a,c,h], Jasvinder A. Singh[c,i], Joerg Meerpohl[c,j], Monica Hultcrantz[k], Patrick Bossuyt[g], Andrew D. Oxman[l], GRADE Working Group

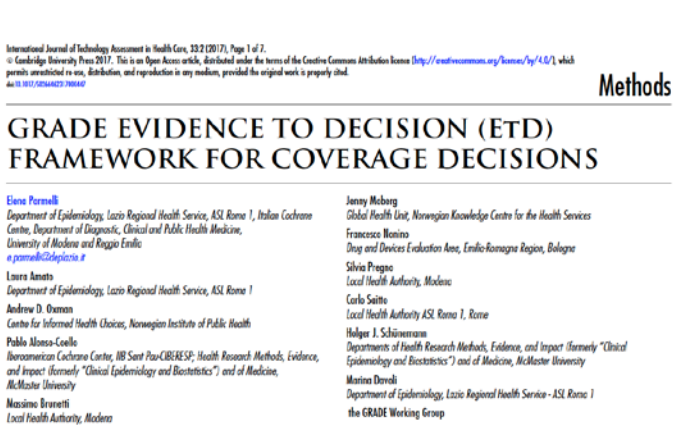
Moberg et al. Health Research Policy and Systems (2018) 16:45
https://doi.org/10.1186/s12961-018-0320-2

Health Research Policy and Systems

REVIEW Open Access

The GRADE Evidence to Decision (EtD) framework for health system and public health decisions

Jenny Moberg[1*], Andrew D. Oxman[1], Sarah Rosenbaum[1], Holger J. Schünemann[2], Gordon Guyatt[3], Signe Flottorp[1], Claire Glenton[1], Simon Lewin[1,4], Angela Morelli[1], Gabriel Rada[5], Pablo Alonso-Coello[6], for the GRADE Working Group



International Journal of Technology Assessment in Health Care, 33:2 (2017), Page 1 of 7.
© Cambridge University Press 2017. This is an Open Access article, distributed under the terms of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Methods

GRADE EVIDENCE TO DECISION (EtD) FRAMEWORK FOR COVERAGE DECISIONS

Elena Pormelli
Department of Epidemiology, Lazio Regional Health Service, ASL Roma 1, Italian Cochrane Centre, Department of Diagnostic, Clinical and Public Health Medicine, University of Modena and Reggio Emilia
e.pormelli@deplazio.it

Laura Amato
Department of Epidemiology, Lazio Regional Health Service, ASL Roma 1

Andrew D. Oxman
Centre for Informed Health Choices, Norwegian Institute of Public Health

Pablo Alonso-Coello
Iberoamerican Cochrane Center, IIB Sant Pau-CIBERESP; Health Research Methods, Evidence, and Impact (formerly "Clinical Epidemiology and Biostatistics") and of Medicine, McMaster University

Massimo Brunetti
Local Health Authority, Modena

Jenny Moberg
Global Health Unit, Norwegian Knowledge Centre for the Health Services

Francesco Nonino
Drug and Devices Evaluation Area, Emilia-Romagna Region, Bologna

Silvia Pregno
Local Health Authority, Modena

Carlo Saitto
Local Health Authority ASL Roma 1, Rome

Holger J. Schünemann
Departments of Health Research Methods, Evidence, and Impact (formerly "Clinical Epidemiology and Biostatistics") and of Medicine, McMaster University

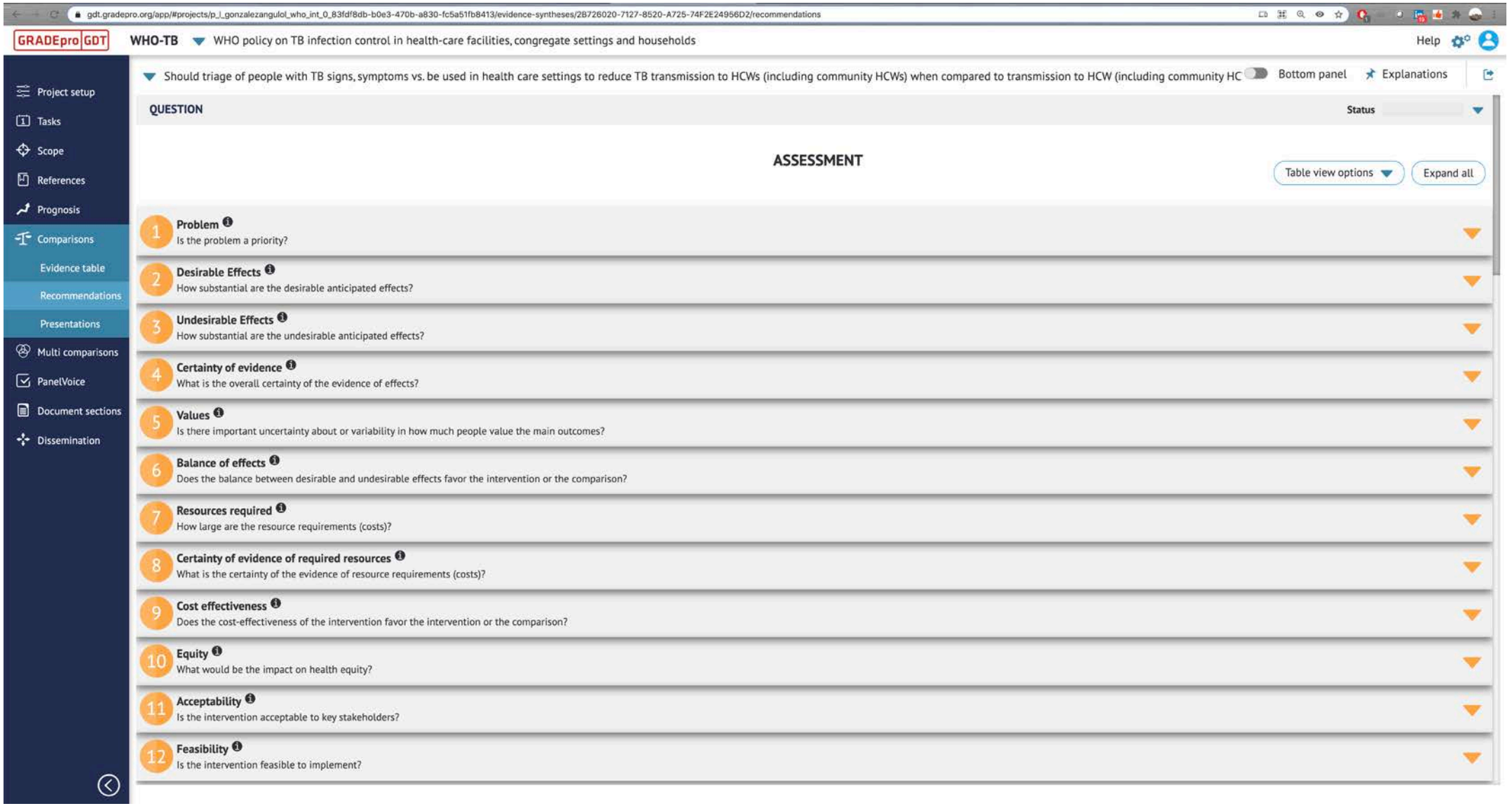
Marina Davoli
Department of Epidemiology, Lazio Regional Health Service - ASL Roma 1

the GRADE Working Group

How should we make decisions

Evaluation of EtD frameworks

Exploring the EtD Frameworks



GRADE findings of which EtDs criteria are relevant

Priority of the problem

Desirable Effects

Undesirable Effects

Values

Balance of Effects

Resources Required

Cost Effectiveness

Equity

Acceptability

Feasibility

Project setup
General information
EtD templates
Tasks
Team
Scope
References
Prognosis
Comparisons
Multi comparisons
PanelVoice
Document sections
Dissemination

Template name

Environmental or occupational health recommendation

Question

Assessment

- Problem
  Is the problem a priority?
- Desirable Effects
  How substantial are the desirable anticipated effects?
- Undesirable Effects
  How substantial are the undesirable anticipated effects?
- Certainty of Evidence
  What is the overall certainty of the evidence of effects?
- Values
  Is there important uncertainty about or variability in how much people value the main outcomes?
- Balance of Effects
  Does the balance between desirable and undesirable effects favor the intervention or the comparison?
- Resources Required
  How large are the resource requirements (costs)?
- Certainty of Evidence of Required Resources
  What is the certainty of the evidence of resource requirements?
- Cost Effectiveness
  Does the cost-effectiveness of the intervention favor the intervention or the comparison?
- Equity
  What would be the impact on equity?
- Acceptability
  Is the intervention acceptable to key stakeholders?
- Feasibility
  Is the intervention feasible to implement?
- Planetary Health
  What is the impact of the interventions on planetary health

Conclusions

Presentations

Summary

- Across body of evidence risk of bias assessment
  - Any instrument can serve
- Across studies and across domains
  - Sensitivity analysis needed
- Integrate with other domains
- But first ask the right (PECO) question