



# EPA Public Access

Author manuscript

*Environ Int.* Author manuscript; available in PMC 2022 October 19.

About author manuscripts

Submit a manuscript

Published in final edited form as:

*Environ Int.* 2018 November ; 120: 382–387. doi:10.1016/j.envint.2018.08.018.

## 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<sup>a</sup>, Kristina A. Thayer<sup>b</sup>, Nancy Santesso<sup>a</sup>, Alison C. Holloway<sup>c</sup>, Robyn Blain<sup>d</sup>, Sorina E. Eftim<sup>d</sup>, Alexandra E. Goldstone<sup>d</sup>, Pam Ross<sup>d</sup>, Gordon Guyatt<sup>a</sup>, Holger J. Schünemann<sup>a,e,\*</sup>

<sup>a</sup>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

<sup>b</sup>Integrated Risk Information System (IRIS) Division, National Center for Environmental Assessment (NCEA), Office of Research and Development, US Environmental Protection Agency, Building B (Room 211i), Research Triangle Park, NC 27711, USA

<sup>c</sup>Department of Obstetrics and Gynecology, McMaster University, Health Sciences Centre, Room 3N52A, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada

<sup>d</sup>ICF, 9300 Lee Highway, Fairfax, VA 22031, USA

<sup>e</sup>Department of Medicine, McMaster University, Health Sciences Centre, Room 2C14, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada

### Abstract

Assessing the risk of bias (RoB) of individual studies is a critical part in determining the certainty of a body of evidence from non-randomized studies (NRS) that evaluate potential health effects due to environmental exposures. The recently released RoB in NRS of Interventions (ROBINS-I) instrument has undergone careful development for health interventions. Using the fundamental design of ROBINS-I, which includes evaluating RoB against an ideal target trial, we explored developing a version of the instrument to evaluate RoB in exposure studies. During three sequential rounds of assessment, two or three raters (evaluators) independently applied ROBINS-I

\*Corresponding author at: Department of Health Research Methods, Evidence, and Impact, Health Sciences Centre, Room 2C14, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada. morganrl@mcmaster.ca (R.L. Morgan).

#### Authors' contributions

RLM, KAT, and HJS designed and conceived of the study. RB, SEE, AEG, and PR piloted the risk of bias instrument and provided feedback on its use. RLM collected feedback and modified the instrument. RLM, KAT, NS, ACH, and HJS reviewed suggestions for modifications to the intermediate and final instrument. RLM drafted the manuscript. KAT, NS, ACH, and HJS reviewed the manuscript and provided major revisions. All authors read and approved the final manuscript.

#### Appendix A to Appendix D

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envint.2018.08.018>.

#### Conflict of interest

The authors declare they have no competing financial interests with respect to this manuscript, or its content, or subject matter. The views expressed are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

to studies from two systematic reviews and one case-study protocol that evaluated the relationship between environmental exposures and health outcomes. Feedback from raters, methodologists, and topic-specific experts informed important modifications to tailor the instrument to exposure studies. We identified the following areas of distinction for the modified instrument: terminology, formulation of the ideal target randomized experiment, guidance for cross-sectional studies and exposure assessment (both quality of measurement method and concern for potential exposure misclassification), and evaluation of issues related to study sensitivity. Using the target experiment approach significantly impacts the process for how environmental and occupational health studies are considered in the Grading of Recommendations Assessment, Development and Evaluation (GRADE) evidence-synthesis framework.

---

## 1. Introduction

Assessing study-level risk of bias (RoB), also called internal validity or limitations in the detailed study design, in the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework, is an essential part of the process used to determine the certainty of a body of evidence from a systematic review or other summary of studies. In comparison to instruments to assess RoB in randomized control trials (RCTs), instruments to assess RoB in non-randomized (i.e., observational) studies (NRS) are typically more complex and challenging to implement (Higgins and Green, 2011).

The RoB in NRS of Interventions (ROBINS-I) instrument, formerly named A Cochrane Risk of Bias Tool for Non-randomized Studies of Interventions (ACROBAT-NRSI), was released in 2016 to assess RoB in NRS of health interventions (e.g. antiretroviral therapy for treatment of persons living with HIV) (Sterne et al., 2016; ACROBAT-NRSI, n.d.). This instrument allows users to identify and assess RoB in NRS that evaluate the effects of one or more interventions at the individual outcome level. Based on concepts presented by William Cochran in 1965, ROBINS-I facilitates a structured comparison of NRS to an unbiased experiment represented by a target randomized trial (Cochran and Chambers, 1965). For example, confounding and selection bias in a cohort study would be evaluated against an often hypothetical, ideal RCT that appropriately considered randomization and allocation concealment.

In addition to conceiving a target randomized trial for each eligible study, raters evaluate the potential for bias by responding to a series of 30 signaling questions addressing specific elements of bias. Signaling questions in the ROBINS-I instrument prompt raters to assess RoB in the domains of: 1) bias due to confounding, 2) bias in selection of participants into the study, 3) bias in classification of interventions, 4) bias due to departures from intended interventions, 5) bias due to missing data, 6) bias in measurement of outcomes, and 7) bias in selection of reported results. Additional elements of the ROBINS-I instrument include an optional component to judge the direction of the bias for each domain.

ROBINS-I was designed to inform the rating of the certainty of a body of evidence regarding health interventions. We recently described how ROBINS-I applies to the GRADE framework, allowing reviewers to systematically and transparently assess the certainty of a body of evidence to inform decision making, rating these bodies of evidence as ‘High’,

‘Moderate’, ‘Low’, or ‘Very low’ certainty (Guyatt et al., 2008). A number of domains are considered to reach these ratings. The following domains in GRADE may decrease the certainty in a body of evidence: RoB, inconsistency, indirectness, imprecision, and publication bias. In addition, three domains increase one’s certainty in a body of evidence: large magnitude of effect, dose-response gradient or effects, and plausible opposing residual confounding.

In GRADE, RCTs start at ‘High’ initial certainty due to the balance of prognostic factors that RCTs can provide, but if randomization fails to achieve this protection they are rated down. Based on the presence (or lack thereof) of risk of bias or other domains decreasing one’s certainty, RCTs can potentially reach a final level of ‘High’, ‘Moderate’, ‘Low’, or ‘Very low’. Without the use of a RoB instruments that allow for an explicit comparison against randomized studies, such as ROBINS-I, NRS start at ‘Low’ certainty, with the potential to reach a final level of ‘Low’ or ‘Very low’ if there are serious concerns in one of the domains listed above. RoB instruments that allow for the comparison of NRS to RCTs eliminate the GRADE requirement for starting an assessment of a body of evidence as ‘High’ or ‘Low’ certainty based on study design (Schünemann et al., 2018).

Since the ROBINS-I instrument was developed for health intervention studies, which typically involve intentional ‘exposures’, the usefulness of ROBINS-I to evaluate RoB in non-randomized studies of exposures to environmental factors, which are typically unintentional exposures, is unclear. Conceptually, ROBINS-I may be applicable to environmental exposure studies because the RoB domains overlap with instruments and concepts used in the field, such as the Newcastle-Ottawa Scale (NOS) and study evaluation tools developed by the U.S. Environmental Protection Agency (EPA) Integrated Risk Information System (IRIS), European Food Safety Authority (EFSA), the National Research Council, the National Toxicology Program’s Office of Health Assessment and Translation (OHAT) and Office of the Report on Carcinogen (ORoC), and University of California in San Francisco’s Navigation Guide (LaKind et al., 2014; Koustas et al., 2014; NRC (National Research Council), 2014; NTP (National Toxicology Program), 2015a; NTP (National Toxicology Program), 2015b; Wells et al., 2000; Authority EFS, 2015; Council NR, 2014; National Academies of Sciences E, Medicine, 2017; Council NR, 2018).

Although the content of the domains are broadly similar across these instruments, some focus on different study-design and labelling features such as cohort versus case-control studies; some include items related to the sensitivity of the study; and not all use signaling questions (Methods Commentary: Risk of Bias in Cohort Studies, n.d.; Rooney et al., 2016). Most importantly, with some exception of the EPA IRIS and ORoC tools, which require the identification of a practical exemplar study, the ideal study design concept is not explicitly outlined. As the GRADE working group has intensified work in the field of exposure assessment and the impact of the use of ROBINS-I on the certainty of a body of evidence, development of a RoB instrument for studies dealing with exposures will facilitate harmonization of rating NRS of interventions and environmental or occupational exposures in the context of GRADE. Thus, we evaluated the concept of the target randomized experiment for assessing RoB in exposure studies.

In this article, we present a RoB instrument for NRS of exposures, developed from a series of pilot tests and external feedback of ROBINS-I, to evaluate studies of environmental exposure studies with this new instrument. We highlight the methodological challenges and considerations that we encountered in its development.

## 2. Development of a RoB instrument to evaluate studies of exposure

### 2.1. Methods

We evaluated the ROBINS-I instrument in environmental studies of exposure by applying it to two existing systematic reviews and one case study protocol during three sequential rounds of assessment (Johnson et al., 2014; Thayer et al., 2013; Zhao et al., 2015) (Appendices A). Detailed methods may be found in Appendix B; however, we provide a brief summary in the following text. Feedback from a group of raters (evaluators) conducting the pilot testing, as well as topic-specific experts and methodologists in the field of environmental health research, identified facilitators and barriers to implementation of the instrument that then informed modifications to the ROBINS-I instrument. A steering group of key investigators (RM, KT, AH, NS, HS) made decisions regarding whether to modify the instrument based on user experience during pilot testing. In addition, the steering committee consulted with other topic-specific experts in environmental health, as well as developers from the ACROBAT-NRSI/ROBINS-I instruments (hence-forth referred to as ROBINS-I), using a modified Delphi (group decision-making) technique conducted in each of the three rounds (Hsu and Sandford, 2007).

**2.1.1. Raters**—To inform modifications to this RoB instrument for NRS of exposures, we selected raters (RB, SE, AG, and PR) with master's or doctoral degrees, training in epidemiological methods, and at least four years (range 4–13 years) of experience as evaluators of epidemiological studies. To model a real-world RoB assessment, raters were not required to have content-specific knowledge of the exposure or outcomes; however, topic-specific experts provided written guidance to aid raters and were available throughout the process as questions arose. During pilot testing three raters independently evaluated each study and discussed judgments to reach a consensus rating.

**2.1.2. Topic-specific experts**—We selected PhD-level topic-specific experts (JL, KS, and KT) who had published articles on the exposure and health outcomes of interest (i.e., BPA and obesity, PFOA and birth weight, and PBDE and thyroid function), to provide the required background information for completion of the evaluation using the ROBINS-I instrument, including confounders and possible co-exposures that were specific to the chosen topic areas.

In addition, we solicited feedback through three iterations of discussion. First, topic-specific experts from the GRADE Environmental Health project group commented on the initial ROBINS-I items and discussions were held with the developers of other study evaluation tools used in environmental health to identify similarities and differences, in particular RoB instruments used by the EPA IRIS, Navigation Guide, OHAT, and ORoC (NRC (National Research Council), 2014; NTP (National Toxicology Program), 2015b; Rooney et al., 2016; Johnson et al., 2014).

Second, three case topic-specific experts (JL, DM, KS) in the field of environmental health assessed one study from the review of BPA and obesity, and one study from the review of PBDE and thyroid function using the modified ROBINS instrument and provided feedback on the signaling questions (Carwile and Michels, 2011; Chevrier et al., 2010). In addition, case topic-specific experts addressed responses to the signaling questions from raters to weigh in on the accuracy and comprehensiveness of their responses.

Lastly, members (including the chairs) of the ROBINS-I instrument development work group provided input on items added to tailor modifications made to evaluate studies of environmental exposure. As ROBINS-I was modified and updated during the conduct of our study, we incorporated these updated versions in our research.

## 2.2. Findings

Feedback from raters solicited during the three rounds of testing identified several barriers to use. These included limited applicability of the instrument to studies of cross-sectional design and shortcomings of shortcomings syntax and double-barreling (i.e. asking two questions) of some signaling questions. In addition, the concept of interventions was felt to be not applicable during the evaluation of exposure studies.

Raters reported concerns regarding the inability of signaling questions to distinguish between the misclassification of different types of exposures. Evaluations of the topic-specific experts' responses revealed these as areas for improvement in the domain for 'Bias in the Classification of Exposures'. For example, when responding to signaling questions within this domain, raters were unsure whether or not to capture the details of the applicability (or directness) of the exposure assessment methods used to identify the exposure as stated in the population, exposure, comparators, outcome (PECO), the quality of the exposure assessment method with respect to measurement accuracy, and the potential for misclassification bias.

Evaluations of the target trial within ROBINS-I by topic-specific experts and methodologists revealed initial confusion and criticism when applied to studies of environmental exposure. Topic-specific experts reported concerns about the use of the phrase 'target trial'. Concern was expressed that they may not always be plausible, due to ethical and practical challenges in conducting such studies; however, this is exactly the purpose of considering the ideal trial, even if hypothetical (i.e. a counterfactual statement). Randomization to balance prognostic factors, known and unknown, remains a mainstay of achieving the ideal experiment whether or not it is a toxin, another substance, a medication, device or a skill-dependent intervention. To alleviate concerns, we opted to use the term target experiment for this work. The emphasis must be on the fact that this experiment may be hypothetical and might only be achievable through animal experiments.

## 2.3. Modifications to the instrument

Modifications to ROBINS-I included both semantic and conceptual issues (Appendix C), including: replacement of the term 'intervention' with 'exposure'; renaming of 'target trial' to 'target experiment' to emphasize that the unintentional exposures assessed could also be compared to a hypothetical animal experiment; additional protocol fields to collect

information on measurement of exposures and outcomes; and inclusion of additional signaling questions to assess bias in exposure measurement. The semantic and conceptual modifications to the ROBINS-I instrument were significant enough to necessitate a distinct instrument for assessing RoB in NRS of exposure (Appendix D). Although ROBINS-I served as the platform for our initial assessment we subsequently referred to the modified instrument as a RoB instrument for NRS of exposures to acknowledge the extent of modifications. We posted this preliminary version of the RoB instrument for NRS of exposures on the University of Bristol website in 2017 and the McMaster GRADE Centre website in 2018, so that interested organizations could pilot and provide feedback for further development (The ROBINS-E tool, n.d.; The Risk of Bias Instrument for Non-randomized Studies of Exposures, n.d.).

### **3. Evaluating RoB for NRS of exposures**

#### **3.1. Conducting the RoB assessment**

A RoB instrument for NRS of exposures has three steps (Fig. 1): In Step I, the review group describes the question of the review: the PECO, as well as topic-specific information regarding confounders, co-exposures, and assessment of the exposure and health outcome. This includes identifying the specific outcome of interest that raters will evaluate with the instrument, as RoB instruments should be applied at the outcome level. In Step II, for each eligible study, the review group describes the ideal target experiment, including confounders and co-exposures that could have occurred in the study. Sensitivity-related issues may be considered at this stage, such as what constitutes sufficient variation in the range of exposure levels to detect an association if present. Lastly, in Step III, raters evaluate each study using the RoB signaling questions and make domain-level judgments.

At each step in the process, the involvement of topic-specific experts is paramount to identify confounders and co-exposures, nuances of the exposure and outcome measurements, as well as review the eligibility of studies, and the accuracy and completeness of rationales provided in response to the signaling questions.

#### **3.2. Operationalizing the target experiment**

ROBINS-I presents a distinguishing feature from previously published RoB instruments for evaluating NRS: the target trial (Sterne et al., 2016). Each review question is formulated to emulate a hypothetical pragmatic RCT. Pivoting the conceptual model from the ideal target trial (RCT) to the ideal experimental exposure study is philosophically aligned with the intent of ROBINS-I (i.e. bias should be assessed against a randomized experimental study design); however, an additional point of confusion relates to the notion of pragmatic. Given that RCTs are generally considered unethical for the assessment of potential hazardous environmental or occupational exposures in humans, a pragmatic study could be interpreted by users to mean the ideal NRS study rather than ideal randomized experimental study. This is not the intended meaning of our instrument. Considering randomization to reduce known and unknown imbalance in prognostic factors and confounding in the (hypothetical) target experiment is critical. The underlying conceptual underpinning of using a target

(randomized) trial to reduce confounding and other bias is identical whether or not it is a toxin, another substance, a medication, device or a skill-dependent intervention.

We did, however, identify issues related to the exposure and comparator of the target experiment. The ROBINS-I instrument is used to evaluate NRS that compare outcomes in at least two groups (i.e. comparative studies). Primary studies should include an ‘intervention group’ and an alternate control or comparison, whether provided in the study or by the initial research question. The comparator could address different thresholds, levels, durations, ranges, means, or medians of exposure, including an incremental increase in exposure (Morgan et al., accepted for publication). A common challenge faced by users of the RoB instrument for NRS of exposures may be trying to evaluate whether an association is present at any exposure level, as this is required to progress to the comparison of the results from different levels of exposures. Findings from complementary streams of evidence, such as animal studies or in vitro/in vivo, may facilitate the identification of a clear exposure-response relationship.

### 3.3. Misclassification of the exposure

Environmental exposures can often be conceptualized as unintentional interventions, which poses a challenge for accurate ascertainment of the exposure. One fundamental challenge is evaluating confidence in exposure characterization (Rooney et al., 2016). For example, how certain are we that an individual in the lower exposure group is correctly classified to that group? In addition, information about the timing (start, finish, and degree of consistency during that period) of exposure is often unavailable. While RoB instruments evaluating RCTs or intentional interventions assume that researchers have control over or knowledge of the initiation of the intervention, this, apart from randomized experiments in animals and trials in nutrition, is not frequently the case for studies of exposures. Within the RoB instrument for NRS of exposures, the rater collects two different types of information to characterize the exposure. Prior to answering the RoB signaling questions, in Step II, the rater collects information about how applicable the measurement of the exposure is to the ‘E’ of the PECO question. When responding to questions about bias in each study, in Step III, the rater applies the information outlined in Steps I and II to evaluate the validity (i.e. the most robust exposure assessment methods) of the measures used and whether or not they distinguish between the exposed and comparative groups.

Raters who pilot tested ROBINS-I and early drafts of the RoB instrument for NRS of exposures conflated issues of indirectness (applicability, generalizability, transferability, translatability, or external validity) and RoB. Indirectness refers to how directly the evidence measures the PECO question (i.e., is the population, exposure, comparator, or outcome of the eligible study the same as the original PECO), whereas RoB evaluates the potential for error based on the methods used to derive the evidence (Schünemann et al., 2018). Per best practices, raters captured these concepts within their responses to the RoB signaling questions. Steps I and II in the RoB instrument for NRS of exposures contain fields to capture information on the accuracy and precision of the exposure measurement that can inform how directly the ‘E’ and ‘C’ in the PECO are measured. The domain dedicated to ‘Misclassification of the Exposure’ includes signaling questions regarding how the exposure

status is defined (i.e. is there concern for exposure misclassification) and the robustness of the exposure assessment methods (i.e. were the methods used to measure the exposure conducted correctly).

Confusion arose when raters evaluated studies with a cross-sectional design. Cross-sectional design studies are common in the environmental health literature; they allow measurement of environmental exposures, health outcomes, as well as certain confounders, at a single time point, may require less time to complete, and be more feasible to conduct than other study designs (Levy, 2006). However, cross-sectional designs are at particularly high risk of bias. Specific concerns of bias in cross-sectional studies include the inability to establish temporality of the potential association between exposure and outcome, and the lack of access to multiple measurements of an exposure to assess stability of the exposure levels over time. Many exposures require measurement at multiple time points (i.e. non-persistent chemicals) to more accurately characterize the extent of exposure. The RoB instrument for NRS of exposures delineates such concerns by distinguishing the measurements needed to determine exposure accuracy in Step I and the degree of RoB if such measures are applied incorrectly. In addition, current guidance suggests that raters evaluating cross-sectional studies respond to questions of temporality as ‘Low RoB’ (i.e. less severe RoB) not ‘N/A’ to reflect that in this situation there is no potential for RoB. This prioritizes the completion of all signaling questions and inappropriate avoidance of answering the signaling question by selecting ‘N/A’.

#### 4. Discussion

We developed an instrument to evaluate the RoB of environmental exposure studies as an ideal target experiment, building on the methodological foundation established by the ROBINS-I: The RoB instrument for NRS of exposures (Sterne et al., 2016). One conclusion from the work presented here is that revisions to ROBINS-I are needed to create an instrument for assessing exposures. Ideally, additional pilot testing and refinement will lead to the development of a RoB instrument for NRS that could be applied to environmental, nutritional or occupational exposures. Development of such an instrument is important so that recent guidelines describing use of ROBINS-I in the GRADE certainty rating process could also be applied to NRS of exposures (Schünemann et al., 2018). This is significant because use of ROBINS-I addresses a commonly noted barrier to adoption of GRADE within environmental health, namely that NRS studies are rated down to ‘Low’ for lack of randomization without explicit consideration of the types of bias at play. Recent GRADE guidance states that when using ROBINS-I for assessing risk of bias in NRS, the initial GRADE certainty in the evidence from a body of studies using an NRS design starts as ‘High’ (Schünemann et al., 2018); however, GRADE emphasizes that risk of bias due to confounding and selection bias will lead to rating down to ‘Low’ unless clear rationale can be provided for why these domains do not cause risk of bias. Moving away from study label-based (based on study design) RoB instruments may increase the transparency of the specific risk of bias, maintaining awareness of the risk caused by that imbalance in prognostic factors and confounders (Morgan et al., 2016).



#### 4.1. Strengths and weaknesses of the study

Strengths of this work include our undertaking of a multi-step process to determine the extent to which the ROBINS-I instrument is applicable to assessing the health impact of environmental exposures (Appendix B). To assess and improve the presentation and relevance of the instrument, at least one topic-specific expert in the field of interest populated Step I during an iterative process. Topic-specific experts provided suggestions for improved applicability of the instrument in three ways during the evaluation and instrument adaptation process: 1) by providing background information for raters when applying ROBINS-I and the modified instrument; 2) by providing guidance on additional questions to add to ROBINS-I specific to environmental exposure; and 3) by performing additional piloting of the modified instrument. We found that information specific to identifying and measuring the exposures and outcomes was important when responding to the corresponding signaling questions.

Based on the narrative responses to the signaling questions, even with the initial adjustments made to the ROBINS-I instrument to address exposures, raters reported misunderstanding the concepts in the questions and the information in the studies. Modifications to the instrument and instructions from the three rounds of pilot testing and external feedback improved understanding. While the same raters applied the instrument to individual studies within the systematic reviews, we did not formally test or calculate reliability of responses. This required a final version of the modified instrument. We will address the validity and reliability of the instrument in future work.

#### 4.2. Strengths and weaknesses in relation to other instruments

Through suggestions identified during pilot-testing feedback and external consultation, we modified ROBINS-I to improve how it may effectively evaluate RoB in NRS of exposures (i.e. face validity). Indeed, the RoB instrument for NRS of exposures has both similarities and differences when compared with other RoB instruments used to assess studies of environmental exposure (Rooney et al., 2016). Domains that assess aspects of bias due to confounding, selection of participants, measurement of exposure, intended exposure, missing data, measurement of outcomes, and reported results are also reflected in other instruments (NOS, EPA IRIS, EFSA, Navigation Guide, OHAT, and ORoC instruments); however, the modified instrument remains distinct in its assessment of how well individual studies emulate the hypothetical target experiment, which impacts evaluation of each bias domain and overall RoB judgments (Sterne et al., 2016; Rooney et al., 2016). Other tools also utilize the concept of a target or ideal study (e.g., ORoC, EPA IRIS), but it is anchored to how closely the study emulates the best possible NRS design rather than a randomized target experimental version of the study (in which both known and unknown confounders are balanced).

An additional concept not yet addressed within the development of the randomized target experiment or measurement of the exposure relates to the uncertainty about a study's ability to detect a true association or effect, one aspect of study sensitivity (Cooper et al., 2016). This could happen when the levels of exposure do not include sufficient variation among subjects to allow for the detection of an effect. It may also happen if the analytical methods

are not sensitive enough to detect low levels thus giving many participants a non-detect reading. Depending on the question asked, study sensitivity may be distinct from RoB, as the measures within a study could be applied correctly; however, if the evaluated exposure is not reflective of the intended level of exposure, it would lead to under or over estimate of the effect. Additional work remains on how to best conceptualize study sensitivity as a related but distinct aspect of evaluating RoB for NRS of exposures, as well as within the GRADE certainty in the evidence framework.

#### 4.3. Implications for researchers and policymakers

When using evidence to inform decision making, the RoB instrument for NRS of exposures allows systematic-review authors and guideline developers to evaluate NRS with the target experiment as a reference point. Developers preferring to focus on assessing RoB in ways that reduce reliance on study design labels will favor the RoB instrument for NRS of exposures. Also, using the RoB instrument for NRS of exposures facilitates assessment of the overall certainty of a body of evidence because of its integration with GRADE. Planning for the application of the modified instrument should recognize the resources needed to conduct the RoB assessment as intended, such as time demands and topic-specific expertise (Morgan et al., unpublished).

#### 4.4. Unanswered questions and future research

Studies need to evaluate the validity of the modified instrument to assess RoB in a variety of exposure scenarios, including occupational exposures and exposures characterized by techniques other than biomonitoring. While we expect that this instrument could apply to other exposures, such as those studied in an occupational setting, our pilot testing was limited to environmental exposures. Similar to methods used for the development of this instrument, the involvement of topic-specific experts and iterative rounds of pilot testing will be needed. Comprehensive guidance with examples is needed for raters and for decision makers using the output from the RoB instrument for NRS of exposures.

The RoB instrument for NRS of exposures requires an independent control or comparative group to provide an evaluation that emulates a target experiment. We recently described a framework for the formulation of PECO in Step 1 (Morgan et al., accepted for publication). Piloting of ROBINS-I and the modified instrument identified continued confusion on the topic of RoB and other factors related to assessing the certainty of a body of evidence (Methods Commentary: Risk of Bias in Cohort Studies, n.d.). Some consider one's ability to determine an exposure's true effect (e.g., 'study sensitivity'), as distinct from RoB and issues of indirectness but this depends on the PECO question that is asked in the systematic review (Cooper et al., 2016). In this instrument, fields to address study sensitivity could be added to Steps I or II in the instrument and to the separate domain of indirectness.

## 5. Conclusions

We evaluated the application of the ROBINS-I instrument to NRS of exposures by applying it to two existing systematic reviews and one case-study protocol. Based on a three-stage, pilot-testing study that involved numerous raters, topic-specific experts, and collaboration

with the original instrument developers, we modified an existing RoB instrument for evaluation of environmental studies of exposure. Modifications made to the ROBINS-I instrument to tailor it to studies of environmental exposure increased understanding and application of the instrument. The modifications made to the instrument were important enough to recommend an instrument distinct from ROBINS-I: The RoB instrument for NRS of exposure. This RoB instrument for NRS of exposures can serve as a standardized, transparent, and rigorous instrument for evaluating RoB of environmental exposure studies. This instrument lends itself to use in the context of GRADE to assess the overall certainty in a body of evidence, but users should be aware of the special consideration around the initial certainty level.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We would like to acknowledge contributions from Juleen Lam, Danielle Mandrioli, Kavita Singh for their review and application of the modified ROBINS-I instrument; and Ruth Lunn and Glinda Cooper, for their input when comparing signaling questions with other instruments used for RoB assessment in environmental health; and Julian Higgins and Jonathan Sterne, for comments on the final instruments.

## Funding sources

This research was supported by the Intramural Research Program of the National Institute of Environmental Health Sciences and the MacGRADE Centre at the McMaster University.

## References

- ACROBAT-NRSI A Cochrane risk of bias assessment tool for non-randomized studies of interventions. <https://sites.google.com/site/riskofbiastool/>, Accessed date: 24 September 2014.
- Authority EFS, 2015. Tools For Critically Appraising Different Study Designs, Systematic Review and Literature Searches. 12(7) EFSA Supporting Publication.
- Carwile JL, Michels KB, 2011. Urinary bisphenol A and obesity: NHANES 2003–2006. *Environ. Res.* 111 (6), 825–830. [PubMed: 21676388]
- Chevrier J, Harley KG, Bradman A, Gharbi M, Sjödin A, Eskenazi B, 2010. Polybrominated diphenyl ether (PBDE) flame retardants and thyroid hormone during pregnancy. *Environ. Health Perspect.* 118 (10), 1444. [PubMed: 20562054]
- Cochran WG, Chambers SP, 1965. The planning of observational studies of human populations. *J. R. Stat. Soc. Ser. A* 128 (2), 234–266.
- Cooper GS, Lunn RM, Agerstrand M, Glenn BS, Kraft AD, Luke AM, Ratcliffe JM, 2016. Study sensitivity: evaluating the ability to detect effects in systematic reviews of chemical exposures. *Environ. Int.* 92–93, 605–610.
- Council NR, 2014. Review of the Formaldehyde Assessment in the National Toxicology Program 12th Report on Carcinogens. National Academies Press.
- Council NR, 2018. Review of Advances Made to the IRIS Process. National Academies Press.
- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schunemann HJ, Group GW, 2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336 (7650), 924–926. [PubMed: 18436948]
- Higgins J, Green S, 2011. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 (updated March 2011). <http://handbook.cochrane.org>, Accessed date: 3 February 2013.
- Hsu C-C, Sandford BA, 2007. The Delphi technique: making sense of consensus. *Pract. Assess. Res. Eval.* 12 (10), 1–8.

- 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 (10), 1028–1039. [PubMed: 24968388]
- Koustas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S, Robinson KA, Axelrad DA, Woodruff TJ, 2014. The navigation guide - evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. *Environ. Health Perspect.* 122 (10), 1015–1027. [PubMed: 24968374]
- LaKind JS, Sobus JR, Goodman M, Barr DB, Furst P, Albertini RJ, Arbuckle TE, Schoeters G, Tan YM, Teeguarden J, et al. , 2014. A proposal for assessing study quality: biomonitoring, environmental epidemiology, and short-lived chemicals (BEES-C) instrument. *Environ. Int.* 73, 195–207. [PubMed: 25137624]
- Levy BS, 2006. *Occupational and Environmental Health: Recognizing and Preventing Disease and Injury*. Lippincott Williams & Wilkins.
- Methods Commentary: Risk of Bias in Cohort Studies. <https://distillercer.com/resources/methodological-resources/risk-of-bias-commentary/>.
- Morgan RL, Thayer KA, Bero L, Bruce N, Falck-Ytter Y, Ghersi D, Guyatt G, Hooijmans C, Langendam M, Mandrioli D, et al. , 2016. GRADE: assessing the quality of evidence in environmental and occupational health. *Environ. Int.* 92–93, 611–616.
- Morgan RL, Thayer K, Whaley P, Schünemann H, 2018a. Identifying the PECO: a framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. *Environ. Int* Unpublished, Accepted for publication.
- Morgan R, Thayer K, Holloway A, Santesso N, Blain R, Eftim S, Goldstone A, Ross P, Guyatt G, Schünemann H, 2018b. Risk of Bias Instrument for Non-randomized Studies of Exposures: a Users' Guide. Unpublished.
- National Academies of Sciences E, Medicine, 2017. *Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-dose Toxicity From Endocrine Active Chemicals*. National Academies Press.
- NRC (National Research Council), 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. [http://www.nap.edu/catalog.php?record\\_id=18764](http://www.nap.edu/catalog.php?record_id=18764), Accessed date: 1 January 2015.
- NTP (National Toxicology Program), 2015a. Handbook for Conducting a Literature-Based Health Assessment Using Office of Health Assessment and Translation (OHAT) Approach for Systematic Review and Evidence Integration. January 9, 2015 release. Available at. <http://ntp.niehs.nih.gov/go/38673>.
- NTP (National Toxicology Program), 2015b. Handbook for Preparing Report on Carcinogens Monographs—July 2015. Available at. <http://ntp.niehs.nih.gov/go/rochandbook>, Accessed date: 3 January 2017.
- Rooney AA, Cooper GS, Jahnke GD, Lam J, Morgan RL, Boyles AL, Ratcliffe JM, Kraft AD, Schünemann HJ, Schwingl P, 2016. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. *Environ. Int.* 92, 617–629. [PubMed: 26857180]
- Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, Morgan RL, Gartlehner G, Kunz R, Katikireddi SV, 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.*
- Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, et al. , 2016. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 355, i4919. [PubMed: 27733354]
- Thayer K, Rooney A, Boyles A, Holmgren S, Walker V, Kissling G, U.S. Department of Health and Human Services, 2013. Draft protocol for systematic review to evaluate the evidence for an association between bisphenol A (BPA) exposure and obesity. In: National Toxicology Program. *The Risk of Bias Instrument for Non-randomized Studies of Exposures*. <http://heigrade.mcmaster.ca/guideline-development/rob-instrument-for-nrs-of-exposures>.

- The ROBINS-E tool Risk Of Bias In Non-randomized Studies - of Exposures. <https://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/robins-e/>.
- Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P, 2000. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-analyses.
- Zhao X, Wang H, Li J, Shan Z, Teng W, Teng X, 2015. The correlation between polybrominated diphenyl ethers (PBDEs) and thyroid hormones in the general population: a meta-analysis. PLoS One 10 (5), e0126989.

---

Step I: Clarify the review question and identify topic-specific considerations (i.e., potential confounders, co-interventions, and exposure and outcome measurement accuracy information) important to assess bias;

Step II: Describe a target trial or experimental version of the study, including specific confounders and co-interventions from that study that will require consideration; and

Step III: Evaluate the study by completing signaling questions across seven RoB domains.

---

**Fig. 1.**  
Steps for applying the RoB instrument for NRS of exposures.