Published in final edited form as:
Environ Int. 2024 April ; 186: 108602. doi:10.1016/j.envint.2024.108602.

A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E)

Julian P.T. Higgins[a,b,c,*], **Rebecca L. Morgan**[d], **Andrew A. Rooney**[e], **Kyla W. Taylor**[e], **Kristina A. Thayer**[f], **Raquel A. Silva**[g], **Courtney Lemeris**[g], **Elie A. Akl**[h], **Thomas F. Bateson**[i], **Nancy D. Berkman**[j], **Barbara S Glenn**[i], **Asbjørn Hróbjartsson**[k], **Judy S. LaKind**[l], **Alexandra McAleenan**[a], **Joerg J. Meerpohl**[m,n], **Rebecca M. Nachman**[i], **Julie E Obbagy**[o], **Annette O'Connor**[p], **Elizabeth G. Radke**[i], **Jelena Savović**[a,b,c], **Holger J. Schünemann**[q], **Beverley Shea**[r], **Kate Tilling**[a,c,s,u], **Jos Verbeek**[t], **Meera Viswanathan**[j], **Jonathan A.C. Sterne**[a,u,v]

[a]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[b]NIHR Bristol Evidence Synthesis Group, University of Bristol, Bristol, UK

[c]NIHR Applied Research Collaboration West (ARC West) at University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK

[*]Corresponding author at: Population Health Sciences, Bristol Medical School, University of Bristol, 39 Whatley Road, Bristol BS8 2PS, UK. julian.higgins@bristol.ac.uk (J.P.T. Higgins).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Judy LaKind reports receipt of support for research on systematic reviews and assessments of review processes from NCASI, the American Petroleum Institute and the National Institute of Environmental Health Sciences and served on the GRADE panel for assessing the certainty of modeled evidence. Annette O'Connor reports funding from the US National Pork Board and the United Soybean Board that includes: consulting or advisory. Jelena Savović reports a relationship with JEMMDx Limited that includes: consulting or advisory. Jelena Savović reports a relationship with Core Models Limited that includes: teaching. Kate Tilling reports a relationship with CHDI Foundation that includes: consulting or advisory. Kate Tilling reports a relationship with UK MHRA that includes: paid expert testimony. Jos Verbeek reports a relationship with World Health Organization that includes: consulting or advisory. The remaining authors report no financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Julian P.T. Higgins: Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Rebecca L. Morgan:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **Andrew A. Rooney:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Kyla W. Taylor:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **Kristina A. Thayer:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **Raquel A. Silva:** Resources, Project administration. **Courtney Lemeris:** Visualization, Resources. **Elie A. Akl:** Writing – review & editing, Investigation. **Thomas F. Bateson:** Writing – review & editing, Investigation. **Nancy D. Berkman:** Writing – review & editing, Investigation. **Barbara S. Glenn:** Writing – review & editing, Investigation. **Asbjørn Hróbjartsson:** Writing – review & editing, Investigation. **Judy S. LaKind:** Writing – review & editing, Investigation. **Alexandra McAleenan:** Writing – review & editing, Investigation. **Joerg J. Meerpohl:** Writing – review & editing, Investigation. **Rebecca M. Nachman:** Writing – review & editing, Investigation. **Julie E. Obbagy:** Writing – review & editing, Investigation. **Annette O'Connor:** Writing – review & editing, Investigation. **Elizabeth G. Radke:** Writing – review & editing, Investigation. **Jelena Savović:** Writing – review & editing, Investigation. **Holger J. Schünemann:** Writing – review & editing, Investigation, Conceptualization. **Beverley Shea:** Writing – review & editing, Investigation. **Kate Tilling:** Writing – review & editing, Investigation. **Jos Verbeek:** Writing – review & editing, Investigation. **Meera Viswanathan:** Writing – review & editing, Investigation. **Jonathan A.C. Sterne:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2024.108602.

[d]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada

[e]National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

[f]Center for Public Health and Environmental Assessment, Chemical & Pollutant Assessment Division, US Environmental Protection Agency, Research Triangle Park, NC, USA

[g]ICF, Reston, VA, USA

[h]Faculty of Medicine, American University of Beirut, Riad El-Solh, Lebanon

[i]Center for Public Health and Environmental Assessment, Chemical and Pollutant Assessment Division, US Environmental Protection Agency, Washington, DC, USA

[j]RTI International, Research Triangle Park, NC, USA

[k]Centre for Evidence-Based Medicine Odense (CEBMO) and Cochrane Denmark, University of Southern Denmark, Odense, Denmark

[l]LaKind Associates, LLC, Catonsville, MD, USA

[m]Institute for Evidence in Medicine, Medical Center & Faculty of Medicine, University of Freiburg, Freiburg, Germany

[n]Cochrane Germany, Cochrane Germany Foundation, Freiburg, Germany

[o]Nutrition Evidence Systematic Review Branch, Center for Nutrition Policy and Promotion, Food and Nutrition Service, US Department of Agriculture, Alexandria, VA, USA

[p]Department of Large Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA

[q]Cochrane Canada and McMaster GRADE Centres, McMaster University, Hamilton, ON, Canada

[r]Ottawa Hospital Research Institute, Ottawa, ON, Canada

[s]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

[t]Cochrane Work, Department of Public and Occupational Health, Academic Medical Centers Amsterdam, University of Amsterdam, Amsterdam, the Netherlands

[u]NIHR Bristol Biomedical Research Centre, Bristol, UK

[v]Health Data Research UK South-West, Bristol, UK

Abstract

Background: Observational epidemiologic studies provide critical data for the evaluation of the potential effects of environmental, occupational and behavioural exposures on human health. Systematic reviews of these studies play a key role in informing policy and practice. Systematic reviews should incorporate assessments of the risk of bias in results of the included studies.

Objective: To develop a new tool, Risk Of Bias In Non-randomized Studies - of Exposures (ROBINS-E) to assess risk of bias in estimates from cohort studies of the causal effect of an exposure on an outcome.

Methods and results: ROBINS-E was developed by a large group of researchers from diverse research and public health disciplines through a series of working groups, in-person meetings and pilot testing phases. The tool aims to assess the risk of bias in a specific result (exposure effect estimate) from an individual observational study that examines the effect of an exposure on an outcome. A series of preliminary considerations informs the core ROBINS-E assessment, including details of the result being assessed and the causal effect being estimated. The assessment addresses bias within seven domains, through a series of 'signalling questions'. Domain-level judgements about risk of bias are derived from the answers to these questions, then combined to produce an overall risk of bias judgement for the result, together with judgements about the direction of bias.

Conclusion: ROBINS-E provides a standardized framework for examining potential biases in results from cohort studies. Future work will produce variants of the tool for other epidemiologic study designs (e.g. case-control studies). We believe that ROBINS-E represents an important development in the integration of exposure assessment, evidence synthesis and causal inference.

Keywords

Risk of bias; Confounding; Selection bias; Misclassification/measurement bias; Exposure; Epidemiology; Environmental

1. Introduction

Observational epidemiologic studies play a vital role in evaluation of the effects of environmental, occupational and behavioural exposures on human health. Randomized controlled trials (RCTs) are often not feasible, ethical, large enough or of sufficient duration to evaluate chronic health effects.

Systematic reviews should incorporate assessments of the risk of bias in results of the included studies, that is, the extent to which results may deviate systematically from the truth. Such assessments are useful to label results included in meta-analyses as at lower or higher risk of bias and to inform assessments of the certainty of the body of evidence provided by a meta-analysis or narrative synthesis. They may also be useful to select results for inclusion in meta-analyses. Recently-developed tools for this purpose have moved away from numeric scores based on checklists (which often combined elements of risk of bias with other considerations such as of applicability) to tools based on assessments within specific domains of bias. Several such tools for observational epidemiologic studies provide significant methodological advances over previously used tools (National Toxicology Program, 2015, 2015; Woodruff and Sutton, 2014; U.S. Environmental Protection Agency, 2022). Comparisons between these tools have been published previously (National Toxicology Program, 2015; Steenland et al., 2020; Eick et al., 2020; Radke et al., 2021; Eick et al., 2022) and have observed that different tools, and different ways of addressing risk-of-bias assessments, can lead to differences in conclusions of systematic reviews (Eick et al., 2020; Radke et al., 2021).

In this paper we consider risk of bias in estimates of the causal effect of an exposure on an outcome, rather than the association between them. The Risk Of Bias In Non-randomized

Studies - of Interventions (ROBINS-I) tool was developed to evaluate estimates of the effects of deliberate interventions aimed to alter health outcomes (Sterne et al., 2016). We adapted and built on the ROBINS-I approach to propose a tool to evaluate risk of bias in study-specific estimates of the effects of other types of exposures on outcomes. The Risk of Bias in Non-Randomized Studies – of Exposure (ROBINS-E) tool aims to assess the risk of bias in a result of an observational epidemiologic study. The first version of ROBINS-E, described here, targets cohort (follow-up) studies.

2. Development of ROBINS-E

Early work by a subset of the authors considered the extent to which the ROBINS-I tool was suitable for observational studies of exposures and suggested modifications to the tool (Morgan et al., 2018; Morgan et al., 2019). The team was widened to include the ROBINS-I core developers, and 43 researchers with an interest in developing a new ROBINS-E tool were identified. We followed the principles adopted for the revised risk-of-bias tool for randomized trials and for the ROBINS-I tool (Sterne et al., 2016; Sterne et al., 2019), aiming to produce a tool that was based on domains of bias, with signalling questions leading to judgements about risk of bias in a specific result, for each domain and overall. Working groups were set up to cover types of studies to be addressed by ROBINS-E; specification of the causal effect of interest; confounding and selection bias; deviations from the exposure of interest; measurement of exposure; measurement of outcomes; attrition and selective reporting; integration of ROBINS-E with the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework (Guyatt et al., 2008); and issues in statistical analysis (see Online Supplement). The groups considered empirical evidence of bias where this was available, for example from meta-epidemiological studies (Sterne et al., 2002), although all such studies identified related to randomized trials. We therefore based our choice of bias domains on a combination of (i) empirical evidence from randomized trials, with the additional assumption that issues common to both study designs (missing data, outcome assessment, selection of the reported result) are likely to be similar; (ii) theoretical considerations, including widespread understandings in epidemiology; and (iii) knowledge of instances of demonstrated biases in individual observational studies, which demonstrate the possibility of the bias occurring.

Proposals from the working groups were discussed at a two-day face-to-face meeting of 31 participants in Bristol, UK, in January 2017. The working groups subsequently refined their proposals to produce signalling questions and guidance for answering them. Between April and June 2019, the working groups met in a series of webinars to discuss a preliminary draft of ROBINS-E, and in October 2019 a 2-day piloting event was held in Bristol at which 28 in-person participants, plus 3 remote participants, used the draft tool to assess risk of bias in the results of various observational studies of exposures. Revisions to the tool were made based on feedback from the piloting event, through new working groups addressing confounding, selection bias, exposure measurement and co-exposures/co-interventions. A revised draft was piloted during June-July 2021 and further refinements made before a full ‘launch’ version of ROBINS-E was posted on https://www.riskofbias.info in June 2022. Updates to the tool will be made available on the web site.

3. Overview of ROBINS-E

ROBINS-E aims to assess the risk of bias in a specific result (exposure effect estimate) from an individual observational cohort study that examines the effect of an exposure on an outcome. The latest version of ROBINS-E can be found at https://www.riskofbias.info/welcome/robins-e-tool. Fig. 1 provides an overview of the tool. A series of preliminary considerations informs the main ROBINS-E assessment and checks whether a full assessment is necessary (see Section 4). The preliminary considerations include specification of the result being assessed and the causal effect being estimated.

The ROBINS-E risk-of-bias assessment addresses seven domains of bias (Fig. 1, Section 5). Each domain is addressed using a series of *signalling questions* that aim to gather relevant information about the study and analysis being assessed. Most signalling questions have response options 'Yes', 'Probably yes', 'Probably no', 'No' and 'No information', with 'Yes' and 'Probably yes' having the same implications for risk of bias and similarly for 'No' and 'Probably no'. Some questions have additional response options (a 'weak' and a 'strong' version of 'Yes' or 'No') to help discriminate between higher and lower risk of bias. As an example, the first question in the confounding domain is "Did the authors control for all the important confounding factors for which this was necessary", with response options 'Yes', Probably yes', 'Weak No (No but uncontrolled confounding was probably not substantial)', and 'Strong No (No and uncontrolled confounding was probably substantial)'.

Some signalling questions are not applicable if responses to earlier questions indicate this. For example, question 4.2 ("Is it likely that the analysis corrected for the effect of post-exposure interventions that were influenced by prior exposure?") is not applicable if the answer to question 4.1 ("Were there post-exposure interventions that were influenced by prior exposure during the follow-up period?") is "No", Probably no" or "No information".

After the relevant signalling questions have been completed, three summary 'domain-level' judgements are made for each of the seven assessed domains of bias. First, the assessor judges the *risk of bias* in the result, which refers to the risk of material bias that has the potential to impact on the estimated effect of exposure on outcome. A suggested judgement, based on the answers to the signalling questions, is generated using an algorithm. This suggestion can be overridden, for example if the assessor believes that the totality of problems identified suggests more or less serious risk of bias within that domain than identified by the algorithm. Possible judgements are 'Low risk of bias', 'Some concerns', 'High risk of bias' and 'Very high risk of bias'. Second, the assessor predicts the *direction of bias*, balancing the issues addressed within the domain. Response option for this depend on the bias domain. Third, a judgement is made about the *threats to conclusions* from the bias, specifically whether the risk of bias (arising from this domain) is sufficiently high, in the context of its likely direction and the magnitude of the estimated exposure effect, to threaten conclusions about whether the exposure has an important effect on the outcome. Response options are 'Yes', 'No' and 'Cannot tell'. As described in Section 6, once these three summary judgements are made for each bias domain, they are summarized across domains to create **overall** judgements about risk-of-bias, predicted direction of bias, and threats to conclusions for the result.

4. Preliminary considerations

The starting point for a ROBINS-E assessment is to specify the study result that is being assessed for risk of bias (Fig. 1, part A). A screening section then facilitates identification of results that are at very high risk of bias, allowing the user to avoid a detailed risk of bias assessment (Fig. 1, part B). This might be the case, for example, if the screening questions identify that there is substantial confounding and the study has made no attempt to control for it. These screening questions were originally included within the relevant bias domains, and answers to these had the potential to lead to a judgement of 'Very high risk of bias' for the domain. The motivation for their inclusion in Step B was that full risk-of-bias assessments may not required if the result will in any case be judged as at very high risk of bias.

If proceeding with a full risk of bias assessment, the user of the tool then assembles information about the participants, exposure measure (s), outcome and analysis methods on which that result is based (Fig. 1, part C). To evaluate risk of bias it is important to define the causal effect of exposure on outcome that would be estimated by the result in the absence of bias. The specification involves defining comparisons of what would happen to the same individual (or group of individuals) when exposed to different patterns of exposure, i.e. comparing 'counterfactuals' (Fig. 1, part D). This includes the population in which this effect applies, the exposure whose effect on the outcome is estimated by the study result, the exposure window (or period) of interest, and the comparisons between different patterns of exposure that are of interest. Parts C and D can be completed in either order, but each should be completed before part E.

The final part E of the preliminary considerations is specifying the known important confounding factors that are likely to influence the association between the exposure and the outcome (Fig. 1, part E). These confounding factors should be identified by a combination of reviewing literature and consulting with experts including members of the review team (Pufulete et al., 2022).

5. Bias domains

The seven bias domains included in ROBINS-E are explained below and the considerations addressed by the signalling questions within the domains are summarized in Table 1.

Bias due to confounding occurs when one or more factors that are prognostic for the outcome also predict exposure group or exposure level. ROBINS-E includes a detailed assessment of how the study investigators addressed the important confounding factors that were pre-specified by the ROBINS-E user. Full control of confounding requires that: confounding factors are controlled for when this is necessary; these factors are measured validly and reliably by the variables available in the study; and that effects of exposure (which may either be mediators of the effect of exposure on outcome or 'colliders' (common effects of exposure and outcome) are not controlled for.

It is impossible to exclude the possibility of uncontrolled confounding in a non-randomized study, so the best available risk-of-bias judgement for this domain is labelled "Low risk of

bias (except for concerns about uncontrolled confounding)". By contrast, it is possible to envisage situations in which risk of bias for other domains is low: for example, if exposure is measured without error, there are no missing data, or results are reported in accordance with a pre-specified analysis plan.

There are two variants of the confounding domain. The first variant, which we expect to apply in most studies, applies when only baseline confounding (by prognostic factors present before the start of the exposure window of interest ['baseline']) is an issue. We expect this variant to apply to the majority of exposure studies.

The second variant additionally addresses time-varying confounding, which can occur when the exposure changes over time and follow-up time is split according to exposure level (Hernán and Robins, 2020; Mansournia et al., 2017). In these circumstances, there can be confounding both by baseline factors and by time-varying prognostic factors that vary after the cohort baseline and that affect subsequent exposure during the exposure window, known as 'time-varying' confounders. For example, if participants' time is split into periods exposed and not exposed to a potentially harmful chemical, then there may be confounding by time-varying prognostic factors that predict changes to exposure during follow-up.

Bias arising from measurement of the exposure occurs if exposure mismeasurement or misclassification leads to under- or over-estimation of its effect on the outcome. The domain first addresses whether there is mismeasurement of the exposure (either because the choice of measurement method was not appropriate or there were errors in the measurements made). It then addresses whether the mismeasurement was differential (related to subsequent outcomes) or non-differential (unrelated to subsequent outcomes). There are three variants of this domain depending on whether: a single measurement of exposure was made; multiple exposure measurements were combined to provide a single measure of exposure for each participant; or the analysis was based on splitting participants' follow up time according to exposure status or magnitude.

Bias in selection of participants into the study (or into the analysis) is the first of two domains that address selection bias, which arises when we condition, in the study design or in the analysis, on common effects of exposure and outcome (Hernán and Robins, 2020; Hernán et al., 2004). This is known as 'collider bias'. Conditioning on common effects distorts the association between exposure and outcome, compared with the association in the population in the absence of such conditioning.

The domain covers: exclusion of follow up after the start of the exposure window; other bias arising from selection of participants into the study (or analysis) being related to an effect of either the exposure or a cause of the exposure and an effect of either the outcome or a cause of the outcome; and corrections that might have been made for selection biases in the analysis. The other ROBINS-E domain addressing selection bias is "Bias due to missing data, described below.

Bias due to post-exposure interventions may arise if the presence or level of exposure leads to administration of interventions that change the subsequent risk of the outcome. For example, in a study of individuals exposed to asbestos, the most highly exposed individuals

may be more likely to receive a CT scan, which reduces their risk of lung cancer mortality. It is possible to correct for such bias, for example by censoring at the time of the intervention and using inverse-probability-of-censoring weights (Hernán and Robins, 2020). However, such analyses require that the factors leading to the post-exposure intervention can be modelled. We expect that in most studies the risk of bias for this domain will be assessed to be low.

Bias due to missing data may arise from missing exposure data, outcome data or data on confounding variables. The domain first examines whether complete data on exposure status, the outcome and confounders were available for all or nearly all participants. If this is not the case, the domain examines the potential that the missing data led to bias in results based on either of the most common approaches to analysis: complete case analysis (restricted to participants with complete data on exposure, outcome and confounding variables) and multiple imputation of missing data. In a complete case analysis, the important considerations are whether exclusion of participants from the analysis is related to the true value of the outcome for those participants and, if so, whether the predictors of such exclusion are included in the analysis model. For analyses based on multiple imputation, the important considerations are whether: data are likely to be missing at random; all predictors of missingness were included in the imputation models; and all the variables in the model used for the main analysis were included in the imputation models.

Bias arising from measurement of the outcome is a problem when error in measurement of the outcome is related to the presence or level of exposure, in which case the measurement error is differential rather than non-differential. This domain addressed whether: measurement methods were the same for all levels of exposure; outcome assessors were aware of exposure levels; and knowledge of exposure was likely to influence assessments of the outcome.

Bias in selection of the reported result arises when study authors select results from a multiplicity of analyses, for example from different ways of measuring the exposure, different ways of measuring the outcome, different subsets of the full study sample or different analyses. There is a risk of bias when such selection is based on the magnitude, direction or P value of the result. Ideally a protocol or statistical analysis plan would specify exactly which analyses are to be taken and reported. However, such pre-specified plans are uncommon for observational studies and often the nature of the data need to be taken into account when selecting a suitable statistical model and method.

6. Overall risk of bias

After completing all seven bias domains, overall judgements are made for risk of bias, predicted direction of bias and threat to conclusions. The default ROBINS-E overall risk-of-bias judgement is that for the domain with the greatest risk of bias. For example, if the greatest risk-of-bias judgement across domains is of high risk of bias, then the result is judged as at high risk of bias overall. However, the user may override this, for example if they judge that so many domains were rated as at ‘High’ risk of bias that the overall judgement should be ‘Very high’ risk of bias. Predicting the direction of bias overall may

be difficult. Risk-of-bias judgements for the individual domains might be used to inform judgements about the influence of that domain on the likely direction of bias overall. The default judgement for the threat to conclusions is derived in a similar way to the overall risk-of-bias judgement; for example, the overall judgement is 'Yes' if there is considered to be a threat to conclusions in any of the seven bias domains.

7. Study sensitivity and appropriateness of its design

It is important to describe the general ability of a study to provide useful information about the effect of exposure on outcome. This is often referred to as the study sensitivity, a term used to embrace aspects of risk of bias as well as additional aspects of study design, including whether an appropriate range of exposure levels was examined and whether there was sufficient follow up for outcomes to be affected. These additional considerations, which fall outside the scope of the main ROBINS-E assessment, are addressed in an optional section of ROBINS-E.

8. Discussion

ROBINS-E is a new general-purpose tool for assessing risk of bias in results of non-randomized studies of exposure. Because risk of bias is assessed in relation to a defined causal effect of an exposure on an outcome, a low risk of bias corresponds to an absence of bias in estimating that causal effect. For this reason, ROBINS-E assessments may support causal inferences, if sufficient evidence at 'low' risk of bias or with 'some concerns' about risk of bias can be located. ROBINS-E provides a standardized framework for examining the range of biases common to observational studies of exposures. The instrument was based on the ROBINS-I tool for non-randomized studies of interventions (Sterne et al., 2016) and RoB 2 tool for randomized trials (Sterne et al., 2019), and was informed by previous instruments for assessing studies of exposure. Important elements in ROBINS-E include a definition of the causal effect of interest, signalling questions to guide domain-level risk-of-bias judgements across seven domains, algorithms to derive suggested risk of bias judgements from answers to the signalling questions, and the separation of risk of bias from other important considerations in the evaluation of study sensitivity.

ROBINS-E is designed primarily for use in the context of a systematic review. We hope that it will enable a thorough examination of the strength of evidence about the presence and/or magnitude of the effect of an exposure on an outcome. Fig. 2 outlines a typical process for using ROBINS-E in this context. It is important that risk-of-bias assessments are integrated appropriately into any synthesis across studies. At its simplest, overall risk-of-bias judgements might be used to stratify results in the synthesis, or to perform sensitivity analyses including results at lower risk of bias or excluding results at higher risk of bias (Boutron et al., 2019). Exploratory analyses can investigate whether the risk of bias explains any heterogeneity or inconsistency of results across studies through formal comparison of subgroups or meta-regression. The domain-level granularity from the ROBINS-E assessment allows users to conduct more detailed investigations in relation to different types of bias. Studies taking different methodological approaches, or estimating mathematically connected parameters, might be considered together in a triangulation exercise, in which both the risks

of bias (assessed by ROBINS-E) and differences between causal questions being addressed are considered jointly (Lawlor et al., 2016). In addition, the domain-level predicted direction of bias and assessment of threats to conclusions in ROBINS-E facilitates consideration of the impact of the individual biases on conclusions regarding the direction of the association.

The extent of the potential for bias across a body of evidence for an exposure-outcome effect informs assessment of the certainty of evidence for that effect. ROBINS-E can be used as part of a framework to rate the quality of the evidence for a group of studies. An example of such a framework is the GRADE approach (Guyatt et al., 2011) where users may start at high certainty in the “study limitations” (risk of bias) domain of GRADE and use their ROBINS-E evaluation to determine how many levels to rate down this certainty (Schunemann et al., 2018). An optional section of ROBINS-E allows users to record observations about study sensitivity, aspects of which may additionally feed into the “indirectness” domain within GRADE, as it relates to generalizability of the result of the study to the population, exposure and outcome of interest.

ROBINS-E does not include direct assessment of financial conflicts of interest, which may be important in interpreting the findings and credibility of studies of exposures and have been recommended for inclusion in evaluation of environmental studies (National Academies of Sciences Engineering and Medicine, 2022). The rationale, previously discussed in the context of randomized trials (Sterne, 2013), is that if financial conflicts of interest lead to bias then this will operate through one of the domains already in the tool. Furthermore, there are impacts of such conflicts that are not problems of internal validity (targeted by ROBINS-E) but instead relate to what question is being asked in the study or analysis. This question can be manipulated to give an illusion of presence or lack or effect (for example by studying trivial levels of exposure or irrelevant outcomes), but this need not have implications for internal validity. We believe that financial conflicts of interest should routinely be assessed and displayed within systematic reviews: a tool for assessing conflicts of interest in randomized trials is under development (TACIT: Tool for Addressing Conflicts of Interest in Trials, 2024).

Practical considerations can impact on the successful implementation of ROBINS-E. First, evaluation of risk of bias requires that the review team has expertise in both content (detailed subject-matter knowledge) and methodology (understanding of the different types of bias that may distort the results of observational studies). For example, specification of important confounding factors requires an understanding of what confounding means as well as what confounders (common causes of both exposure and outcome) are likely to be present in the specific context of each study. Second, the accuracy of a risk-of-bias assessment will depend on the quality of information available about the study being assessed. We encourage reference to the maximum possible amount of available information. In addition to published papers describing the study’s methods and results, such information may be derived from unpublished reports or through correspondence with the study investigators. Third, ROBINS-E assessments require judgements based on the available information. It is therefore impossible to avoid subjectivity and between-assessor variability. However, publication of answers to signalling questions and justification for these should make risk of bias judgements transparent. It will be particularly important to justify answers to

signalling questions that lead to a judgement of ‘Very high risk of bias” within Step B of the preliminary considerations, since these imply that a full risk-of-bias assessment will not be conducted.

We believe that the ROBINS-E tool represents an important development in the integration of study assessment of human exposure, evidence synthesis and causal inference. The collaborative development of the tool, between methodologists and stakeholders from fields such as environmental, occupational, chemical, clinical and public health, has highlighted the benefits of sharing ideas and research methods across communities (Hoffmann et al., 2022; Hoffmann et al., 2022). Although ROBINS-E was iteratively developed based on the results of pilot testing with a diverse group, as it becomes more widely used across different exposure topics, it will benefit from more extensive user testing and from formal evaluation of its reliability and validity. The present version addresses only follow-up (cohort) studies, and work is ongoing to develop ROBINS-E variants for other study designs including case-control studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank researchers who pilot tested a draft version of the tool and Whitney Arroyave, Francesco Forastiere, Ellen Kirrane, Tom Luben, Ruth Lunn, Luke McGuinness, Suril Mehta, and Kyle Steenland for contributing to discussions that led to the ROBINS-E tool. We also thank Brandiese Beverly and Ellen Kirrane for helpful comments on the draft manuscript. We thank four anonymous reviewers for their helpful comments.

This work was supported by the Intramural Research Program (ES103379-01) at the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (contract number GS00Q14OADU417 [order number HHSN273201600015U]), by the National Institute for Health and Care Research (NIHR Bristol Biomedical Research Centre; NIHR Applied Research Collaboration West; NIHR203807), by Cancer Research UK (grant number C18281/A29019) and by Health Data Research UK South-West.

Data availability

No data was used for the research described in the article.

References

Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, Hrobjartsson A, 2019. Chapter 7: considering bias and conflicts of interest among the included studies. In: Higgins JPT., Thomas J., Chandler J., Cumpston M., Li T., Page MJ. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions, second ed. John Wiley & Sons, Chichester (UK), pp. 177–204.

Eick SM, Goin DE, Chartres N, Lam J, Woodruff TJ, 2020. Assessing risk of bias in human environmental epidemiology studies using three tools: different conclusions from different tools. Syst. Rev. 9 (1), 249. [PubMed: 33121530]

Eick SM, Goin DE, Lam J, Woodruff TJ, Chartres N, 2022. Authors’ rebuttal to integrated risk information system (IRIS) response to “assessing risk of bias in human environmental epidemiology studies using three tools: different conclusions from different tools”. Syst Rev. 11 (1), 53. [PubMed: 35321722]

Guyatt GH, Oxman AD, Vist GE, Zunz R, Falck-Ytter Y, Alonso-Coello P, et al. , 2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 336 (7650), 924–926. [PubMed: 18436948]

Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. , 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J. Clin. Epidemiol. 64 (4), 383–394. [PubMed: 21195583]

Hernán MA, Hernandez-Diaz S, Robins JM, 2004. A structural approach to selection bias. Epidemiol. 15 (5), 615–625.

Hernán MA, Robins JM, 2020. Causal inference: what if. Chapman & Hall/CRC, Boca Raton.

Hoffmann S, Aiassa E, Angrish M, Beausoleil C, Bois FY, Ciccolallo L, et al. , 2022. Application of evidence-based methods to construct mechanism-driven chemical assessment frameworks. ALTEX 39 (3), 499–518. [PubMed: 35258090]

Hoffmann S, Whaley P, Tsaioun K, 2022. How evidence-based methodologies can help identify and reduce uncertainty in chemical risk assessment. ALTEX 39 (2), 175–182. [PubMed: 35100433]

Lawlor DA, Tilling K, Davey SG, 2016. Triangulation in aetiological epidemiology. Int. J. Epidemiol. 45 (6), 1866–1886. [PubMed: 28108528]

Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G, 2017. Handling time varying confounding in observational research. BMJ 359, j4587. [PubMed: 29038130]

Morgan RL, Thayer KA, Santesso N, Holloway AC, Blain R, Eftim SE, et al. , 2018. 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. Environ. Int. 120, 382–387. [PubMed: 30125855]

Morgan RL, Thayer KA, Santesso N, Holloway AC, Blain R, Eftim SE, et al. , 2019. A risk of bias instrument for non-randomized studies of exposures: a users' guide to its application in the context of GRADE. Environ. Int. 122, 168–184. [PubMed: 30473382]

National Academies of Sciences Engineering and Medicine, 2022. A Review of U.S. EPA's ORD Staff Handbook for Developing IRIS Assessments: 2020 Version. Washington DC: The National Academies Press.

National Toxicology Program, 2015. Handbook for Preparing Report on Carcinogens Monographs 2015. Available from: https://ntp.niehs.nih.gov/go/rochandbook.

National Toxicology Program, 2015. OHAT Risk of Bias Rating Tool for Human and Animal Studies 2015. Available from: https://ntp.niehs.nih.gov/go/riskbias.

Pufulete M, Mahadevan K, Johnson TW, Pithara C, Redwood S, Benedetto U, et al. , 2022. Confounders and co-interventions identified in non-randomized studies of interventions. J. Clin. Epidemiol. 148, 115–123. [PubMed: 35346782]

Radke EG, Glenn BS, Kraft AD, 2021. Integrated risk information system (IRIS) response to "assessing risk of bias in human environmental epidemiology studies using three tools: different conclusions from different tools". Syst Rev. 10 (1), 235. [PubMed: 34419159]

Schunemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. , 2018. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. J. Clin. Epidemiol.

Steenland K, Schubauer-Berigan MK, Vermeulen R, Lunn RM, Straif K, Zahm S, et al. , 2020. Risk of bias assessments and evidence syntheses for observational epidemiologic studies of environmental and occupational exposures: strengths and limitations. Environ. Health Perspect. 128 (9), 95002. [PubMed: 32924579]

Sterne JAC, 2013. Why the Cochrane risk of bias tool should not include funding source as a standard item (Editorial). Cochrane Database of Systematic Reviews. 12, ED000076.

Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. , 2016. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. BMJ. 355, i4919. [PubMed: 27733354]

Sterne JA, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M, 2002. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat. Med. 21 (11), 1513–1524. [PubMed: 12111917]

Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. , 2019. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 366, l4898. [PubMed: 31462531]

TACIT: Tool for Addressing Conflicts of Interest in Trials. https://tacit.one/ (accessed 14 February 2024).

Environmental Protection Agency US, 2022. ORD staff handbook for developing IRIS assessments (EPA/600/R-22/268). U.S. EPA Office of Research and Development, Washington, DC.

Woodruff TJ, Sutton P, 2014. The navigation guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ. Health Perspect. 122 (10), 1007–1014. [PubMed: 24968373]

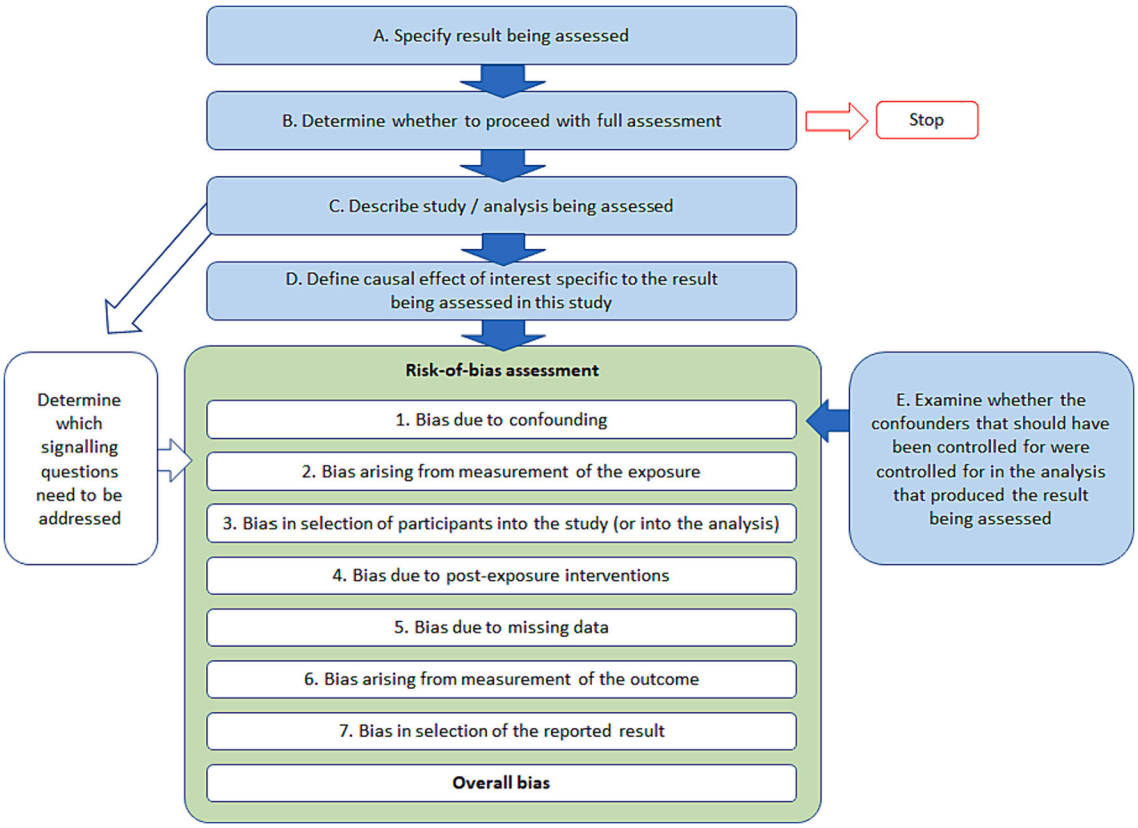

Fig. 1.
Overview of a ROBINS-E in a specific study.

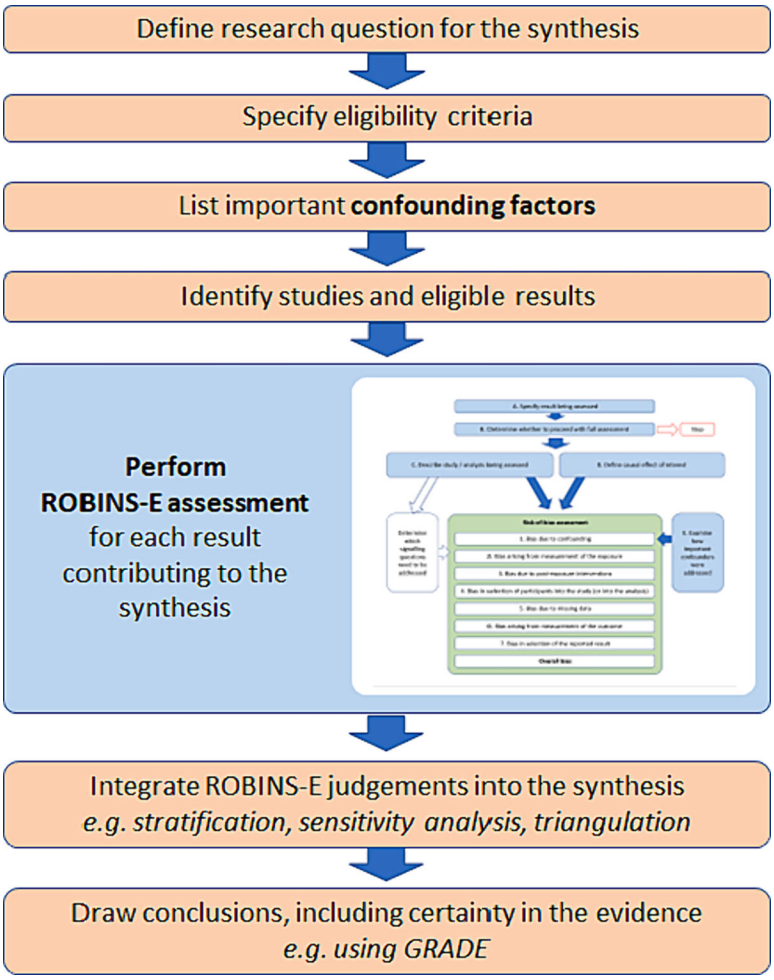

Fig. 2.

ROBINS-E in the context of a systematic review. The 'synthesis' will typically involve a meta-analysis, although it may be a more complex evidence synthesis or a narrative integration of findings across studies.

Table 1

Bias domains included in the ROBINS-E tool, with a summary of the issues addressed for cohort studies.

Bias domain	Issues addressed
Bias due to confounding	Whether: • all important confounding factors were controlled for using appropriate methods;
	• the confounding factors were measured validly and reliably by the variables available; and
	• variables after the start of the exposure window (and that could have been affected by the exposure) were inappropriately controlled for.
Bias arising from measurement of the exposure	Whether: • the measure of exposure used in the study well characterizes the exposure metric of interest;
	• there was likely to be error in, or misclassification of, the exposure measurements in the study;
	• there was differential measurement (or misclassification) error; and
	• non-differential measurement (or misclassification) error would have biased the effect estimate.
Bias in selection of participants into the study	Whether: • start of follow-up and start of the exposure window were the same;
	• selection of participants into the study (or into the analysis) was based on participant characteristics observed after the start of the exposure window;
	• (if applicable) these characteristics were influenced by exposure (or a cause of exposure) and influenced by outcome (or a cause of the outcome); and
	• (if applicable) adjustment techniques were used to correct for the presence of selection biases.
Bias due to post-exposure interventions	Whether: • there were post-exposure interventions influenced by prior exposure; and
	• (if applicable) the analysis corrected for the effect of these post-exposure interventions.
Bias due to missing data	Whether:
	• complete data on exposure status, the outcome, and confounders were available for all or nearly all participants;
	• (for complete case analyses) omission from the analysis is likely to be related to the true value of the outcome and predictors of missingness were included in the analysis model; and
	• (for analyses with imputed data) imputation was performed appropriately.
Bias in measurement of the outcome	Whether: • measurement or ascertainment of the outcome is likely to have differed between exposure groups or levels of exposure;
	• outcome assessors were aware of study participants' exposure history; and
	• (if applicable) assessment of the outcome were likely to have been influenced by knowledge of participants' exposure history.
Bias in selection of the reported result	Whether: • the numerical result being assessed is likely to have been selected, on the basis of the results, from multiple exposure measurements within the outcome domain;
	• the numerical result being assessed is likely to have been selected, on the basis of the results, from multiple outcome measurements within the outcome domain;
	• the numerical result being assessed is likely to have been selected, on the basis of the results, from multiple analyses of the data; and
	• the numerical result being assessed is likely to have been selected, on the basis of the results, from multiple subgroups of a larger cohort.