Environment International 199 (2025) 109463

Contents lists available at ScienceDirect

Environment International

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

Correspondence

Letter: Robins-E risk of bias tool

Risk-of-bias tools are increasingly used as part of systematic reviews, to help make a uniform evaluation of study quality across a variety of studies (NASEM 2021). Several well-known tools, including OHAT and the Navigation Guide, are used to evaluate observational epidemiologic studies (OHAT 2019, Woodruff and Sutton, 2014).

A recently published tool called ROBINS-E (Higgins et al. 2024) was developed to evaluate observational studies on environmental and occupational exposures. We have concerns regarding how this tool can be appropriately used and how it relates to other approaches to evidence synthesis. We are a group of environmental and occupational epidemiologists/exposure experts, nearly all of whom took part in the early discussion and piloting of this new risk-of-bias tool and previously published our general views on risk-of-bias tools (Steenland et al. 2020).

Although some of our initial concerns were addressed in the subsequent development of ROBINS-E, ultimately we declined to participate as co-authors of the ROBINS-E article out of continued concerns, discussed below. We note that it would have been informative and transparent if the ROBINS-E authors had started with a description of the debate about risk-of-bias tools, addressing the various critiques that have been published in the last few years (Boogaard et al. 2023, Eick et al. 2020, Bero et al. 2018, Steenland et al. 2020, Savitz et al. 2019), and explaining how this tool overcomes the previously identified limitations. Our concerns are listed below.

1. The methods described in ROBINS-E evaluate individual studies, and possibly exclude studies before the evidence synthesis stage. Regarding the potential for unwarranted exclusion of studies, we note that seven 'domains' (areas of potential bias) are assessed as low, moderate (some concerns), high, or very high risk of bias – and a very high risk of bias in one domain leads to an overall evaluation of the study having a very high risk of bias. However, this may lead to unnecessary exclusions at the evidence synthesis stage, as there may be information from other studies that may mitigate a bias concern, or a quantitative bias assessment might determine that the alleged bias is of little concern (see below).
2. The ROBINS-E article, although it mentions triangulation, does not identify this technique as a key element of bias assessment (Lawlor et al. 2016). Triangulation involves comparing studies potentially affected by biases operating in different directions (or the same potential bias with different strengths); such contrasts are particularly informative if such studies yield similar outcomes (eg. Lenters et al. 2011, Bhatia et al. 1998). Similarly, such studies may provide empirical information on the presumed biases. Indeed, specific familiarity with the body of literature of a particular exposure-outcome association is a prerequisite when approaching the evaluation of bias in individual studies.
3. We believe the evaluation of the likely direction and magnitude of bias in ROBINS-E is insufficiently addressed. For example, a bias in one direction may cancel out a bias in the other direction in the same study. Furthermore, quantitative bias assessment can be asses the likely direction and its magnitude of bias. A recent example is the E-value (Vanderweele and Ding, 2017), although this method has been debated (eg. Greenland 2020, IARC 2024 (Chapter 3)). The E-value represents the minimum association strength, on the risk ratio scale, that an unmeasured confounder would need with both the treatment and outcome to fully explain a specific treatment-outcome relationship, given the measured covariates. With quantitative information available from the study or literature regarding the likely direction and magnitude of bias, a more thorough quantitative bias assessment can estimate the true effect measure. This approach is rooted in the long-standing tradition of quantitative bias assessment (Cornfield et al. 1959; Axelson, 1978), and has been further developed in recent years (Fox et al. 2005, Lash et al. 2014, IARC 2024 (Chapter 3)). For example, Steenland and Greenland (2004) adjusted the original silica/lung cancer rate ratio of 1.60 (95 % CI: 1.31, 1.93), estimated without data on smoking, by using partial evidence of smoking in the cohort and known smoking/lung cancer rate ratios, to derive a likely smoking-adjusted rate ratio of 1.43 (95 % Monte Carlo limits: 1.15, 1.78).
4. For quality assessment of individual observational studies with ROBINS-E, the lowest bias category is "low risk of bias except for concerns about uncontrolled confounding". However, ROBINS-E is also proposed for use in evidence synthesis, particularly within the GRADE framework, where evidence from observational studies starts at "low certainty of evidence" (Guyatt et al., 2008), as they are not randomized controlled trials (RCTs). Essentially, several individual studies, each rated as low risk of bias by ROBINS-E, would, taken together in evidence synthesis by GRADE, start at "low certainty of evidence". While the authors do note that RCTs are generally not possible for environmental toxins, their theoretical approach using a hypothetical target trial fails to recognize that RCTs should not routinely serve as a model for observational environmental studies (Pearce and Vandenbroucke, 2023, Vandenbroucke et al. 2016). We note that concerns about 'uncontrolled confounding', presuming there is a specific confounder of interest (without which concern about uncontrolled confounding is completely speculative), can be addressed by triangulation and quantitative bias assessment (see points 2 and 3 above).
5. We are concerned that non-experts may misapply risk-of-bias tools. Risk-of-bias tools, such as ROBINS-E, evaluate potential bias across a set of specific domains; this can aid across systematic

https://doi.org/10.1016/j.envint.2025.109463
Received 14 June 2024; Received in revised form 8 October 2024; Accepted 11 April 2025
Available online 30 April 2025

reviews by making study evaluations uniform. These domains should be noted *a priori*, before evaluating studies and based on factors that are generally recognized as potentially influencing study outcomes. From this list, key domains can be identified that are most problematic, and then a review of these key domains in each individual study should be done by experts well-versed in the subject. This should be done without any summary classification of each study as having low, moderate, high, or very high risk of bias, because such a summary of overall level of bias can lead to automatic exclusion of a specific study in evidence syntheses. A better approach to evaluating study biases in evidence synthesis is the careful consideration of biases and their likely importance, such as done in systematic literature reviews by IARC (IARC Preamble 2019). Reviews following the IARC model include Filho et al. 2023, DeBono et al. 2023, and a series of four articles assessing potential biases in ionizing radiation studies (Daniels et al. 2020 (measurement error), Schubauer-Berigan et al. 2020 (confounding and selection bias), Linet et al. 2020 (potential outcome misclassification), and Gilbert et al. 2020 (impact of different analytic methods). In these examples, the authors determined the most likely important biases relevant to the specific exposures/outcomes of the systematic review, similar to that recommended by Savitz et al. 2019 and IARC 2024 (Chapter 6). For example, in the DeBono et al. review of firefighting and cancer, the authors developed their own bias assessment approach, in which the key bias domains were misclassification of exposure; misclassification of outcome; healthy worker hire and survivor bias; confounding by lifestyle factors (e.g., tobacco or alcohol consumption, sun exposure) or occupational exposures outside of firefighting; medical surveillance bias; and selection bias. The strength of these appraisals is that they were led by subject matter experts who had an excellent understanding of the specific domains of greatest concern for each exposure-outcome (cancer) pair. The protocols for the *meta*-analyses by DeBono et al. and Filho et al. were registered with the international database of prospectively registered systematic reviews (PROSPERO). IARC has recently published an extended discussion on assessing different potential biases in the context of cancer hazard identification (IARC 2024)

6. Omission of systematic consideration of study sensitivity or informativeness (e.g. relevant exposure contrasts, relevant lag time) (IARC Preamble 2019), and an appropriate statistical analysis are significant omissions to the primary ROBINS-E evaluation. The tool has an optional section where these issues can be considered, but their review should be mandatory. While these issues may not lead to biases in a study, they may render a study uninformative in assessing the exposure-outcome association in question.
7. The selection bias penalization for environmental and occupational studies that did not start follow-up at the time of first exposure to the studied population is inappropriate. This decision stems from an overall preference for the RCT paradigm. Many, if not most, environmental and occupational cohort studies begin follow-up of cohorts after the start of exposure for some or most of the cohort participants. While this can theoretically lead to bias due to missing person-time at risk (left truncation), in practice, this bias is likely to be small (Applebaum et al. 2011) or perhaps not present at all (Barry et al. 2015). The advantage of a 'survival cohort' with follow-up during the most relevant years (usually 10 or more years after first exposure for occupational carcinogens) may outweigh the advantage of 'incident' cohorts, and careful data analysis can also lead to valid and appropriate results (Vandenbroucke and Pearce, 2015). Thus it is wrong to routinely apply this penalization without careful consideration of the specific issue under study.
8. Bias in reporting results is a domain considered in ROBINS-E: "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". In our view, this text fails to distinguish between cherry-picking results and correctly assessing different exposure metrics to report the best-fitting ones. Nor does it reflect the literature in assessing multiple comparisons across many different exposures (e.g., use of Benjamin-Hochberg false detection rates vs. Bonferroni comparisons). A study that assessed different exposure metrics might score as a 'high' risk of bias, when in fact this approach may be more appropriate than reporting results from a pre-determined analysis method, which would receive an automatic 'low' risk of bias in the current algorithm.
9. The development of ROBINS-E has taken many years, but is still only available for cohort studies. Suppose researchers settle on this tool for their risk of bias assessment. In that case, we are concerned that case-control studies and other potentially informative study designs will be excluded from evidence synthesis until a tool is developed for other study designs.
10. We also note that in pilot testing, the ROBINS-E tool often led to more extreme conclusions of bias than the summary judgments of the experts applying the tool. The hypothetical "target experiment" aspect of ROBINS-Ewas particularly problematic. The tool also did not work well at discriminating between studies with a potential for major bias away from the null and those with more minor potential biases, as we noted above. Finally, assessment using this tool took much longer than via other risk-of-bias methods. Efficiency and ease of use are important factors in adopting new methods for systematic review and both were missing during piloting.
11. Finally, we note that ROBINS-E does not have a domain for conflict of interest assuming "that if financial conflicts of interest lead to bias then this will operate through one of the domains already in the tool." While this may be true, it can be subtle and may not be detected in risk-of-bias tools. We note that many authoritative agencies have a screening system that precludes those with a conflict of interest from participating in the evaluation of the literature, including the US National Academy of Sciences, Engineering, and Medicine (NASEM 2022), recognizing the potential bias due to real or perceived conflicts of interest. We also note that other risk-of-bias tools include a domain on potential conflicts of interest (e.g. Woodruff and Sutton, 2014)

In summary, current risk of bias tools are useful in providing a shared list of domains to consider in evaluating individual studies. However it is our view that the bias potential for individual studies can be assessed without the use of currently existing specific risk of bias tools, including the new ROBINS-E, if evaluators consider all domains of interest but then focus on the ones which are potentially problematic. Examples are the systematic reviews cited above from IARC, where a list of key domains is determined for the specific exposure-outcome of interest. These key domains are then evaluated for each study. If currently existing tools are used, they should be used with caution, not mechanically, but by identifying potential sources of bias and ways to think about that bias. These tools could be considered as one way to inform bias judgements in evidence synthesis, within a triangulation framework, without relying on them as the main method of evaluating overall bias in single observational studies.

CRediT authorship contribution statement

K. Steenland: Writing – review & editing, Writing – original draft,

Conceptualization. **K. Straif:** Writing – review & editing, Writing – original draft, Conceptualization. **M.K. Schubauer-Berigan:** Writing – review & editing, Writing – original draft, Conceptualization. **P.A. Demers:** Writing – review & editing, Writing – original draft, Conceptualization. **F. Forastiere:** Writing – review & editing, Writing – original draft, Conceptualization. **T. Stenzel:** Writing – review & editing, Writing – original draft, Conceptualization. **R. Vermeulen:** Writing – review & editing, Writing – original draft, Conceptualization. **W. Arroyave:** Writing – review & editing, Writing – original draft, Conceptualization. **S.H. Zahm:** Writing – review & editing, Writing – original draft, Conceptualization. **N. Pearce:** Writing – review & editing, Writing – original draft, Conceptualization.

Funding

This work was not funded.

Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer / World Health Organization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

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

References

Applebaum, K.M., Malloy, E.J., Eisen, E.A., 2011. Left truncation, susceptibility, and bias in occupational cohort studies. Epidemiology 22, 599–606.

Axelson, O., 1978. Aspects on confounding in occupational health epidemiology. Scand. J. Work Environ. Health 4 (1), 98–102. https://doi.org/10.5271/sjweh.2720. PMID: 644270.

Barry, V., Klein, M., Winquist, A., Darrow, L.A., Steenland, K., 2015. Disease fatality and bias in survival cohorts. Environ. Res. 140, 275–281.

Bero, L., Chartres, N., Diong, J., Fabbri, A., Ghersi, D., Lam, J., Lau, A., McDonald, S., Mintzes, B., Sutton, P., Turton, J.L., Woodruff, T.J., 2018. The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures. Syst. Rev. 7 (1), 242. https://doi.org/10.1186/s13643-018-0915-2. PMID: 30577874; PMCID: PMC6302384.

Bhatia, R., Lopipero, P., Smith, A.H., 1998. Diesel exhaust exposure and lung cancer. Epidemiology 9 (1), 84–91.

Boogaard, H., Atkinson, R.W., Brook, J.R., Chang, H.H., Hoek, G., Hoffmann, B., Sagiv, S. K., Samoli, E., Smargiassi, A., Szpiro, A.A., Vienneau, D., Weuve, J., Lurmann, F.W., Forastiere, F., 2023. Evidence synthesis of observational studies in environmental health: lessons learned from a systematic review on traffic-related air pollution. Environ. Health Perspect. 131 (11). https://doi.org/10.1289/EHP11532. Epub 2023 Nov 22. PMID: 37991444; PMCID: PMC10664749.

Cornfield, J., Haenszel, W., Hammond, E.C., Lilienfeld, A.M., Shimkin, M.B., Wynder, E. L., 1959. Smoking and lung cancer: recent evidence and a discussion of some questions. Int. J. Epidemiol. 38 (5), 1175–1191.

Daniels, R.D., Kendall, G.M., Thierry-Chef, I., Linet, M.S., Cullings, H.M., 2020. Strengths and weaknesses of dosimetry used in studies of low-dose radiation exposure and cancer. J. Natl. Cancer Inst. Monogr. 2020 (56), 114–132. https://doi.org/10.1093/jncimonographs/lgaa001.

DeBono, N.L., Daniels, R.D., Beane Freeman, L.E., Graber, J.M., Hansen, J., Teras, L.R., Driscoll, T., Kjaerheim, K., Demers, P.A., Glass, D.C., Kriebel, D., Kirkham, T.L., Wedekind, R., Filho, A.M., Stayner, L., Schubauer-Berigan, M.K., 2023. Firefighting and cancer: a meta-analysis of cohort studies in the context of cancer hazard identification. Saf. Health Work 14 (2), 141–152.

Eick, S.M., Goin, D.E., Chartres, N., Lam, J., Woodruff, T.J., 2020. Assessing risk of bias in human environmental epidemiology studies using three tools: different conclusions from different tools. Syst. Rev. 9 (1), 249. https://doi.org/10.1186/s13643-020-01490-8. PMID: 33121530; PMCID: PMC7596989.

Filho, A.M., Turner, M.C., Warnakulasuriya, S., Richardson, D.B., Hosseini, B., Kamangar, F., Pourshams, A., Sewram, V., Cronin-Fenton, D., Etemadi, A., Glass, D. C., Rahimi-Movaghar, A., Sheikh, M., Malekzadeh, R., Schubauer-Berigan, M.K., 2023. The carcinogenicity of opium consumption: a systematic review and meta-analysis. Eur. J. Epidemiol. 38 (4), 373–389.

Fox, M.P., Lash, T.L., Greenland, S., 2005. A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int. J. Epidemiol. 34 (6), 1370–1376.

Gilbert, E.S., Little, M.P., Preston, D.L., Stram, D.O., 2020. Issues in interpreting epidemiologic studies of populations exposed to low-dose, high-energy photon radiation. J. Natl. Cancer Inst. Monogr. 2020 (56), 176–187. https://doi.org/10.1093/jncimonographs/lgaa004.

Greenland, S., 2020. Commentary: an argument against E-values for assessing the plausibility that an association could be explained away by residual confounding. Int. J. Epidemiol. 49, 1501–11153.

Guyatt, G.H., Oxman, A.D., Vist, G.E., 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.

Higgins, J.P.T., Morgan, R.L., Rooney, A.A., Taylor, K.W., Thayer, K.A., Silva, R.A., Lemeris, C., Akl, E.A., Bateson, T.F., Berkman, N.D., Glenn, B.S., Hróbjartsson, A., LaKind, J.S., McAleenan, A., Meerpohl, J.J., Nachman, R.M., Obbagy, J.E., O'Connor, A., Radke, E.G., Savović, J., Schünemann, H.J., Shea, B., Tilling, K., Verbeek, J., Viswanathan, M., Sterne, J.A.C., 2024. A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E). Environ. Int. 186, 108602.

IARC, 2019. Preamble to the IARC Monographs (amended January 2019); available at https://monographs.iarc.who.int/iarc-monographs-preamble-preamble-to-the-iarc-monographs/.

IARC, 2024. Statistical Methods in Cancer Research Volume V: Bias Assessment in Case–Control and Cohort Studies for Hazard Identification. Berrington de González A, Richardson DB, Schubauer-Berigan MK, Eds. Lyon, France: IARC Scientific Publication No. 171. Available at https://publications.iarc.who.int/634.

Lash, T.L., Fox, M.P., MacLehose, R.F., Maldonado, G., McCandless, L.C., Greenland, S., 2014. Good practices for quantitative bias analysis. Int. J. Epidemiol. 43 (6), 1969–1985.

Lawlor, D.A., Tilling, K., Davey, S.G., 2016. Triangulation in aetiological epidemiology. Int. J. Epidemiol. 45, 1866–1886.

Lenters, V., Vermeulen, R., Dogger, S., Stayner, L., Portengen, L., Burdorf, A., Heederik, D., 2011. A meta-analysis of asbestos and lung cancer: is better quality exposure assessment associated with steeper slopes of the exposure-response relationships? Environ. Health Perspect. 119 (11), 1547–1555.

Linet, M.S., Schubauer-Berigan, M.K., Berrington de González, A., 2020. Outcome assessment in epidemiological studies of low-dose radiation exposure and cancer risks: sources, level of ascertainment, and misclassification. J. Natl. Cancer Inst. Monogr. (56), 154–175. https://doi.org/10.1093/jncimonographs/lgaa007.

NASEM (National Academies of Sciences, Engineering, and Medicine 2021). The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2022. Review of U.S. EPA's ORD Staff Handbook for Developing IRIS Assessments: 2020 Version. Washington, DC: The National Academies Press. https://doi.org/10.17226/26289.

OHAT (Office of Health Assessment and Translation). 2019. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. https://ntp. niehs.nih.gov/ntp/ohat/pubs/handbook-march2019_508.pdf, accessed November 13, 2020.

Pearce, N., Vandenbroucke, J.P., 2023. Are Target trial emulations the gold standard for observational studies? Epidemiology 34 (5), 614–618. https://doi.org/10.1097/EDE.0000000000001636.

Savitz, D.A., Wellenius, G.A., Trikalinos, T.A., 2019. The problem with mechanistic risk of bias assessments in evidence synthesis of observational studies and a practical alternative: assessing the impact of specific sources of potential bias. Am. J. Epidemiol. 188 (9), 1581–1585. https://doi.org/10.1093/aje/kwz131.

Schubauer-Berigan, M.K., Berrington de Gonzalez, A., Cardis, E., Laurier, D., Lubin, J.H., Hauptmann, M., Richardson, D.B., 2020. Evaluation of confounding and selection bias in epidemiologic studies of populations exposed to low-dose, high-energy photon radiation. J. Natl. Cancer Inst. Monogr. 2020 (56), 133–153. https://doi.org/10.1093/jncimonographs/lgaa008.

Steenland, K., Greenland, S., 2004. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. Am. J. Epidemiol. 160 (4), 384–392. https://doi.org/10.1093/aje/kwh211.

Steenland, K., Schubauer-Berigan, M.K., Vermeulen, R., Lunn, R.M., Straif, K., Zahm, S., Stewart, P., Arroyave, W.D., Mehta, S.S., Pearce, N., 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. https://doi.org/10.1289/EHP6980.

Vandenbroucke, J., Pearce, N., 2015. Point: incident exposures, prevalent exposures, and causal inference: does limiting studies to persons who are followed from first exposure onward damage epidemiology? Am. J. Epidemiol. 182 (10), 826–833.

Vandenbroucke, J., Broadbent, A., Pearce, N., 2016. Causality and causal inference in epidemiology - the need for a pluralistic approach. Int. J. Epidemiol. 45, 1776–1786.

VanderWeele, T.J., Ding, P., 2017. Sensitivity analysis in observational research: introducing the E-value. Ann. Intern. Med. 167 (4), 268–274. https://doi.org/10.7326/M16-2607.

Woodruff, T.J., 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.

K. Steenland[a,*], K. Straif[b], M.K. Schubauer-Berigan[c], P.A. Demers[d], F. Forastiere[e,f], T. Stenzel[g], R. Vermeulen[h], W. Arroyave[i], S.H. Zahm[j], N. Pearce[k]

[a] *Dept. Environmental Health, Rollins School of Public Health, Emory U., Atlanta, USA*

[b] *Global Observatory on Planetary Health, Boston College, MA, and ISGlobal, Barcelona, Spain*

[c] *International Agency for Research on Cancer, Lyon, France*

[d] *Occupational Cancer Research Center, Ontario Health, Toronto Canada*

[e] *National Research Council, IFT, Palermo, Italy*

[f] *Environmental Research Group, School of Public Health, Imperial College London, London, United Kingdom*

[g] *Stewart Exposure Assessments, LLC, Arlington, VA, USA*

[h] *Institute for Risk Assessment Science, University of Utrecht, Utrecht, Netherlands*

[i] *Integrated Laboratory Systems, Morrisville, NC, USA*

[j] *Shelia Zahm Consulting, Hermon, ME, USA*

[k] *London School of Hygiene and Tropical Medicine, London, United Kingdom*

[*] Corresponding author.
E-mail address: nsteenl@emory.edu (K. Steenland).
Handling Editor: N Roth