

Value and challenges of using observational studies in systematic reviews of public health interventions

Hilton Boon, Michele; Burns, Jacob; Craig, Peter; Griebler, Ursula; Heise, Thomas L.; Katikireddi, S. Vittal; Rehfuess, Eva; Shepperd, Sasha; Thomson, Hilary; Bero, Lisa

Published in:
American Journal of Public Health

DOI:
[10.2105/AJPH.2021.306658](https://doi.org/10.2105/AJPH.2021.306658)

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in ResearchOnline](#)

Citation for published version (Harvard):
Hilton Boon, M, Burns, J, Craig, P, Griebler, U, Heise, TL, Katikireddi, SV, Rehfuess, E, Shepperd, S, Thomson, H & Bero, L 2022, 'Value and challenges of using observational studies in systematic reviews of public health interventions', *American Journal of Public Health*, vol. 112, no. 4, pp. 548-552.
<https://doi.org/10.2105/AJPH.2021.306658>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please view our takedown policy at <https://edshare.gcu.ac.uk/id/eprint/5179> for details of how to contact us.

Value and Challenges of Using Observational Studies in Systematic Reviews of Public Health Interventions

Michele Hilton Boon, PhD, MPH, Jacob Burns, MSc, Peter Craig, PhD, Ursula Griebler, PhD, MPH, Thomas L. Heise, Diploma in Health Economics, S. Vittal Katikireddi, MBChB, PhD, Eva Rehfuess, PhD, Sasha Shepperd, DPhil, Hilary Thomson, PhD, and Lisa Bero, PhD

ABOUT THE AUTHORS

Michele Hilton Boon, Peter Craig, S. Vittal Katikireddi, and Hilary Thomson are with the Medical Research Council/Scottish Government Chief Scientist Office Social and Public Health Sciences Unit, University of Glasgow, Glasgow, UK. Jacob Burns and Eva Rehfuess are with the Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig Maximilian University of Munich, Munich, Germany. Ursula Griebler is with the Department for Evidence-Based Medicine and Evaluation, Danube University Krems, Krems an der Donau, Austria. Thomas L. Heise is with the Leibniz Institute for Prevention Research and Epidemiology-BIPS, University of Bremen, Bremen, Germany. Sasha Shepperd is with the Nuffield Department of Population Health, University of Oxford, Oxford, UK. Lisa Bero is with the School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO.

Recent developments in systematic review methods provide opportunities to draw more robust conclusions from observational studies of interventions and increase the public health relevance of reviews. Cochrane public health and health systems reviews have expanded in scope and methods, supported by new chapters on non-randomized studies in the updated *Cochrane Handbook* (2021)¹ and the development of new, related guidance by the Cochrane Methods Executive. We illustrate these changes while also summarizing the most recent guidance and research on deciding when to include observational studies, identifying and selecting studies, extracting and synthesizing data, assessing risk of bias, and grading certainty of evidence.

These developments are particularly important for systematic reviews in public health, in which randomized trials to assess health outcomes are often unfeasible, for example, in the case of large and irreversible infrastructure interventions; unethical, for example, where the primary aim of the intervention is to prevent certain harm; unavailable when decisions are urgently required; or unable to detect harms at the population level, for example, observational pharmacovigilance studies that use adverse event data sets to detect harms after a drug is already on the market and in widespread use. We draw on our experience as editors and authors of Cochrane public health and health systems reviews and as methodologists, supplemented by a hand search of the past five years

of key methodology journals. Observational studies of exposures (e.g., environmental exposures) also constitute an important area of current methodological development but are outside the scope of this editorial.

Including observational studies in systematic reviews of interventions produces challenges at every stage of designing and conducting a review, beginning with the terminology used to define and identify these studies. Despite efforts to encourage classification according to study design elements rather than labels,² agreement on terminology is elusive. By “observational studies of interventions,” we intend to encompass the range of classifications that may be encountered when considering quantitative evidence of intervention effects other than from randomized trials. These terms include (but are not limited to) nonrandomized studies of interventions,¹ quasiexperiments,³ natural experiments,⁴ and the many specific study design labels that fall within these categories.

It is important to note that many of these terms have overlapping meanings and are applied in diverse and inconsistent ways, both in primary research and in systematic reviews. However, for the full value of observational studies of interventions to be realized, it is essential that systematic reviewers look beyond traditional study designs, such as cohort and case-control, and consider the relevance of quasiexperimental designs that can adjust for unobserved confounding, or selection on unobservables.⁵ At the same time, systematic reviewers must recognize that observational studies are not all of equal evidentiary value, requiring careful assessment of risk of bias, and that their inclusion may increase the resource requirements of a review.

CHANGES IN UNDERSTANDING

Early methodological guidance recognized that observational studies can fill gaps in the literature, provide long-term follow-up that can identify harms of treatments, and answer questions that cannot (for reasons of ethics or feasibility) be investigated in randomized trials.⁶ However, recognition of this value was tempered by caveats on the vulnerability of observational studies to bias and confounding, increased heterogeneity in meta-analyses, and the assertion that observational studies can estimate associations only between treatment and outcome, rather than unbiased causal effects.⁶ A further concern has been understanding the extent to which observational studies may overestimate the effects of interventions compared with randomized trials. Interestingly, systematic reviews have generally found a lack of statistically significant differences in pooled results when systematically comparing randomized and observational studies,^{7,8} identifying differences in the specific research question, heterogeneity, or risk of bias⁹ as explanations for any dissimilar results.

Although randomized trials remain the gold standard for estimating the effects of interventions, the “causal turn” in epidemiology has facilitated an explicit acknowledgment that observational studies may also aim to estimate causal effects.¹⁰ Causal inference can be strengthened by designs that postulate a plausible counterfactual under certain assumptions, such as the preintervention trend in an interrupted time-series study, Mendelian randomization, instrumental variables, or the untreated control group in a regression discontinuity design, which

allow the causal effect of an intervention to be estimated.⁴ Recent systematic reviews have demonstrated these designs to have been more widely implemented in health research than previously believed; however, these reviews have also noted issues in the quality of the conduct and reporting of these studies.^{11–13}

Recent Cochrane reviews illustrate how including observational studies is essential for providing a comprehensive picture of the range of interventions and evidence available for some public health questions, such as the effects of large-scale primary prevention interventions, particularly when these have been implemented and evaluated across different contexts. For example, a Cochrane review (in progress) of sugar-sweetened beverage taxation has identified no eligible randomized trials, but a large body of at least 39 nonrandomized studies will contribute to a comprehensive evaluation of the effects of these taxes on consumption and sales.¹⁴ A review of interventions to reduce ambient air pollution similarly identified no randomized trials, but 42 nonrandomized studies provide evidence on the effectiveness of 38 different interventions, albeit with low and very low certainty.¹⁵ In a review of environmental interventions to reduce sugar-sweetened beverage consumption, the majority of well-known interventions (including traffic light labels, nutritional rating scores, and price increases) have been evaluated only in observational studies and therefore would not have been represented in the review if only randomized trials had been included.¹⁶

CHALLENGES AND ADVANCES

Although these reviews serve as examples of the value of including

observational studies in evidence synthesis, these studies presented challenges at every stage of the review, from designing the protocol and identifying studies to extracting, evaluating, and synthesizing the results. The online appendix (available as a supplement to the online version of this article at <http://www.ajph.org>) provides a table summarizing the challenges encountered, recent methodological developments that have contributed to meeting these challenges, and priorities for future research.

WHEN TO INCLUDE OBSERVATIONAL STUDIES

The newly revised *Cochrane Handbook* suggests that the inclusion of observational studies is justified when randomized trials answer a review question indirectly or incompletely or when a randomized trial is impossible or unlikely to be conducted.¹ The approach is based on a taxonomy of observational studies that replaces design labels (e.g., controlled before and after) with a breakdown of design elements (e.g., assignment mechanisms and control for confounding) that can enable the study to make causal estimates and minimize risk of bias.² This taxonomy is a helpful shift away from inconsistently applied design labels and toward a recognition of the role of study design elements in supporting causal inference. The *Handbook* recommends that reviewers decide which study design elements would be desirable for the review question, scope the literature to see what studies are available, and set the eligibility criteria in the protocol accordingly. In practice, this strategy requires specialist knowledge of these study designs, and examples of best practice from systematic reviews

that have implemented this strategy are lacking.

SEARCHES AND STUDY SELECTION

Reviews that include observational studies typically must deal with a large volume of retrieved records. The lack of standardized terminology creates a challenge for information retrieval and for study selection. For example, in a review of taxation of sugar and sugar-added foods to prevent obesity, 24 454 records were retrieved, of which only one interrupted time-series study was eligible for inclusion.¹⁷ Machine learning is a potential solution, although there are considerable problems in applying machine learning to fully automate the selection of nonrandomized studies owing to varying terminology.¹⁸ Machine learning tools that prioritize studies for screening by identifying patterns in human reviewers' decisions (semiautomation) can be useful in reducing screening burden.¹⁹ Search filters allow database-specific strategies to reduce volume but are not yet available to cover the full range of study types (with varying labels) and databases required for public health reviews.²⁰ Furthermore, these strategies depend on the completeness, quality, and uniformity of records and retrievable full texts so that study identification and selection remain labor-intensive steps that cannot be fully automated at present.

DATA EXTRACTION

Observational studies, in particular quasiexperimental and natural experimental designs, typically offer multiple analyses and effect estimates for the same outcome in a single study, again requiring specialist methodological

knowledge on the part of the review team to undertake data extraction. Methods for addressing effect size multiplicity have been described,²¹ but the impact of choice of method and of selection of effect sizes on the results of meta-analysis of observational studies is unknown.²² In some cases the same data set may have been used in more than one secondary analysis, creating a risk of double counting of results in a review, even from independently conducted studies. Additionally, poor data quality may pose a threat to validity that is difficult to detect and assess.⁹ Standardized tools for data extraction are lacking.

RISK OF BIAS

Assessing risk of bias is an essential task that poses a significant challenge for systematic reviews of observational studies, as hundreds of tools exist; no tool applies equally to all study designs, making consistent assessments difficult; and consensus is lacking on which is preferred.²³ Furthermore, there is evidence that the choice of tool can affect the conclusions of reviews of observational studies.²⁴ The ROBINS-I (Risk Of Bias In Non-randomized Studies-of Interventions) tool has been advanced as a solution to this dilemma.²⁵ ROBINS-I uses a series of signaling questions to assess risk of bias in seven domains: confounding, selection of participants, classification of intervention status, deviations from intended interventions, missing data, outcome measurement, and selective reporting.

This rigorously developed tool enables a systematic assessment of risk of bias, with signaling questions currently developed for cohort and case-control studies and versions covering additional study designs in development.

However, use of this tool requires a strong understanding of epidemiological principles and a significant time investment. Early reports indicate that users have difficulty in applying the tool consistently, although this is partly because observational studies of interventions are sometimes poorly reported.²⁶ Selective reporting bias and publication bias are especially difficult to assess, as protocol registration and prespecified analysis plans remain uncommon for observational studies. Cochrane is currently undertaking research into preferred and acceptable risk of bias tools when ROBINS-I is not appropriate. The interactive Tableau risk of bias tool finder (https://ntp.niehs.nih.gov/go/ohat_tools) can help users compare and select from 62 risk of bias tools for observational studies of exposures.

SYNTHESIS

The *Cochrane Handbook* notes that the inclusion of observational studies of interventions, with various design elements and conducted in a range of populations and settings, typically leads to high statistical heterogeneity that may be methodological, contextual, or unclear in origin. In principle, meta-analysis can be conducted using effect estimates from observational studies. The *Handbook* recommends that a random-effects model be the default approach and that separate analyses be conducted for studies with very different design features¹; however, detailed guidance on how and when to do so is lacking. In practice, pooling these studies is often deemed inappropriate because of very large heterogeneity across interventions and outcomes, statistical heterogeneity encountered as a default for population-level interventions, or

outcomes data assessed and reported in a manner that precludes meta-analysis. The little guidance that exists suggests that meta-analysis of observational studies should focus not only on a pooled effect estimate but also on assessing the influence of moderators and potential sources of bias, employing subgroup analysis and meta-regression,^{27,28} leave-one-out meta-analysis would be another option to identify exaggerated effect sizes that stem from a particular study.

Reviewers face a considerable challenge in structuring and reporting a nonstatistical synthesis, which may be narrative, tabular, or graphical. New reporting guidelines on SWIM (Synthesis Without Meta-analysis) help to address this challenge by detailing how reporting can be improved in aspects of methods and results that often lack transparency in such reviews, including how studies have been grouped and how heterogeneity has been investigated.²⁹

CERTAINTY OF EVIDENCE

The introduction of the target trial concept and ROBINS-I have contributed to a significant advance in GRADE (Grading of Recommendations Assessment, Development and Evaluation), a methodology widely used to assess the certainty of a body of evidence in systematic reviews and guidelines. GRADE originally reflected the traditional hierarchy of evidence by having all bodies of observational evidence start with a low rating and bodies of randomized trial evidence start as high certainty. Although, crucially, observational evidence could be upgraded in certain circumstances, upgrading rarely occurred, and concerns were raised that GRADE underrated the certainty of evidence in areas lacking in randomized trials, such as population health.³⁰ New GRADE

guidance indicates that when ROBINS-I is used, observational studies also start with a high certainty rating, allowing better comparison and integration of randomized and nonrandomized evidence; however, examples are lacking.³¹ To address this gap, the GRADE Public Health Group is undertaking research into the conditions under which evidence from designs such as interrupted time series can produce a body of high- or moderate-certainty evidence.³⁰

FURTHER DEVELOPMENTS NEEDED

Systematic reviews vary in methodological and reporting quality.³² Including observational studies introduces additional challenges and resource requirements but can also increase the public health relevance of a review if study quality is rigorously assessed and guidelines for producing a high-quality systematic review are followed. Given both the value and the challenges of including observational studies in systematic reviews of interventions, we look forward to further development of methods and tools to ensure that such studies are identified, assessed, and incorporated into public health and health systems reviews in the best possible manner. Machine learning algorithms and search filters, data extraction tools, and ROBINS-I extensions will help to address these challenges. A greater focus on study design elements that reduce bias and confounding and on investigation of whether underlying design assumptions have been met, rather than design labels that are inconsistently applied in the literature, may contribute to producing tools that are easier to use and apply. Meta-epidemiological research

on less familiar study designs, such as natural experiment and quasiexperimental designs, is needed to support the development of tools and reporting standards with empirical evidence. The online appendix summarizes challenges, developments, and priorities for further research.

RECOMMENDATIONS FOR REVIEWERS

Systematic reviews in public health and health systems should be designed at the protocol stage to consider the potential relevance of different observational study types, notably natural experimental and quasiexperimental studies, to the research question and specify inclusion and exclusion criteria, search strategies, risk of bias assessment, and synthesis plans accordingly. Risk of bias tools should be comprehensive in addressing selection bias, confounding, information bias, and selective reporting. Where possible, they should specifically apply to the study designs included in the review, and the rationale for the choice of tool should be reported. Data extraction should identify the data set used in secondary analyses, as reviewers will need to guard against double counting if multiple independent studies have analyzed the same data. Review teams need to be appropriately resourced, given the large amounts of search results, methodological expertise, and time required for reviews of observational studies. Finally, along with implications for the education and training of researchers to appropriately conduct observational studies of the effects of interventions, systematic review authors should be appropriately trained to identify, analyze, and assess these studies.

AJPH

CORRESPONDENCE

Correspondence should be sent to Dr. Michele Hilton Boon, WISE Centre for Economic Justice, Hamish Wood Building, Glasgow Caledonian University, Cowcaddens Rd, Glasgow G4 0BA, UK (e-mail: mbo5@gcu.ac.uk). Reprints can be ordered at <http://www.ajph.org> by clicking the "Reprints" link.

PUBLICATION INFORMATION

Full Citation: Hilton Boon M, Burns J, Craig P, et al. Value and challenges of using observational studies in systematic reviews of public health interventions. *Am J Public Health*. 2022;112(4):548–552. Acceptance Date: December 6, 2021.

DOI: <https://doi.org/10.2105/AJPH.2021.306658>

CONTRIBUTORS

M. Hilton Boon drafted the first version of the editorial. All authors revised, completed, and approved the editorial.

ACKNOWLEDGMENTS

This work received no specific funding, but M. Hilton Boon, P. Craig, S. V. Katikireddi, and H. Thomson would like to acknowledge funding from the UK Medical Research Council (grant MC_UU_00022/2) and the Scottish Government Chief Scientist Office (grant SPHSU17). S. V. Katikireddi additionally acknowledges funding from a National Health Service Research Scotland Senior Clinical Fellowship (grant SCAF/15/02).

The authors would like to acknowledge the three anonymous reviewers and the associate editor for their constructive comments and suggestions.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

- Higgins JPT, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions*, version 6.2. February 2021. Available at: www.training.cochrane.org/handbook. Accessed January 25, 2022.
- Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series—paper 5: a checklist for classifying studies evaluating the effects on health interventions—a taxonomy without labels. *J Clin Epidemiol*. 2017;89:30–42. <https://doi.org/10.1016/j.jclinepi.2017.02.016>
- Bärnighausen T, Tugwell P, Røttingen J-A, et al. Quasi-experimental study designs series—paper 4: uses and value. *J Clin Epidemiol*. 2017;89:21–29. <https://doi.org/10.1016/j.jclinepi.2017.03.012>
- Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annu Rev Public Health*. 2017;38:39–56. <https://doi.org/10.1146/annurev-publhealth-031816-044327>
- Waddington H, Aloe AM, Becker BJ, et al. Quasi-experimental study designs series—paper 6: risk of bias assessment. *J Clin Epidemiol*. 2017;89:43–52. <https://doi.org/10.1016/j.jclinepi.2017.02.015>
- Egger M, Schneider M, Smith GD. Spurious precision? Meta-analysis of observational studies. *BMJ*. 1998;316(7125):140–144. <https://doi.org/10.1136/bmj.316.7125.140>
- Anglemeyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev*. 2014;2014(4):MR000034. <https://doi.org/10.1002/14651858.MR000034.pub2>
- Schwingshackl L, Balduzzi S, Beyerbach J, et al. Evaluating agreement between bodies of evidence from randomised controlled trials and cohort studies in nutrition research: meta-epidemiological study. *BMJ*. 2021;374:n1864. <https://doi.org/10.1136/bmj.n1864>
- Mathes T, Rombey T, Kuss O, Pieper D. No inexplorable disagreements between real-world data-based nonrandomized controlled studies and randomized controlled trials were found. *J Clin Epidemiol*. 2021;133:1–13. <https://doi.org/10.1016/j.jclinepi.2020.12.019>
- Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health*. 2018;108(5):616–619. <https://doi.org/10.2105/AJPH.2018.304337>
- Turner SL, Karahalios A, Forbes AB, et al. Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: a review. *J Clin Epidemiol*. 2020;122:1–11. <https://doi.org/10.1016/j.jclinepi.2020.02.006>
- Widding-Havneraas T, Chaulagain A, Lyhmann I, et al. Preference-based instrumental variables in health research rely on important and underreported assumptions: a systematic review. *J Clin Epidemiol*. 2021;139: 269–278. <https://doi.org/10.1016/j.jclinepi.2021.06.006>
- Hilton Boon M, Craig P, Thomson H, Campbell M, Moore L. Regression discontinuity designs in health: a systematic review. *Epidemiology*. 2021; 32(1):87–93. <https://doi.org/10.1097/EDE.0000000000001274>
- Heise TL, Katikireddi SV, Pega F, et al. Taxation of sugar-sweetened beverages for reducing their consumption and preventing obesity or other adverse health outcomes. *Cochrane Database Syst Rev*. 2016;8:CD012319. <https://doi.org/10.1002/14651858.CD012319>
- Burns J, Boogaard H, Polus S, et al. Interventions to reduce ambient particulate matter air pollution and their effect on health. *Cochrane Database Syst Rev*. 2019;5(5):CD010919. <https://doi.org/10.1002/14651858.CD010919.pub2>
- von Philipsborn P, Stratil JM, Burns J, et al. Environmental interventions to reduce the consumption of sugar-sweetened beverages and their effects on health. *Cochrane Database Syst Rev*. 2019;6(6):CD012292. <https://doi.org/10.1002/14651858.CD012292.pub2>
- Pfänder M, Heise TL, Hilton Boon M, et al. Taxation of unprocessed sugar or sugar-added foods for reducing their consumption and preventing obesity or other adverse health outcomes. *Cochrane Database Syst Rev*. 2020;4(4):CD012333. <https://doi.org/10.1002/14651858.CD012333.pub2>
- Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Syst Rev*. 2019;8:163. <https://doi.org/10.1186/s13643-019-1074-9>
- Gates A, Guitard S, Pillay J, et al. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. *Syst Rev*. 2019;8:278. <https://doi.org/10.1186/s13643-019-1222-2>
- Waffenschmidt S, Navarro-Ruan T, Hobson N, et al. Development and validation of study filters for identifying controlled non-randomized studies in PubMed and Ovid MEDLINE. *Res Synth Methods*. 2020;11(5):617–626. <https://doi.org/10.1002/jrsm.1425>
- López-López JA, Page MJ, Lipsey MW, Higgins JPT. Dealing with effect size multiplicity in systematic reviews and meta-analyses. *Res Synth Methods*. 2018; 9(3):336–351. <https://doi.org/10.1002/jrsm.1310>
- Page MJ, Bero L, Kroeger CM, et al. Investigation of Risk Of Bias due to Unreported and Selectively included results in meta-analyses of nutrition research: the ROBUST study protocol. *F1000Research*. 2020;8:1760. <https://doi.org/10.12688/f1000research.20726.2>
- Quigley JM, Thompson JC, Halfpenny NJ, Scott DA. Critical appraisal of nonrandomized studies: a review of recommended and commonly used tools. *J Eval Clin Pract*. 2019;25(1):44–52. <https://doi.org/10.1111/jep.12889>
- Losilla J-M, Oliveras I, Marin-Garcia JA, Vives J. Three risk of bias tools lead to opposite conclusions in observational research synthesis. *J Clin Epidemiol*. 2018;101:61–72. <https://doi.org/10.1016/j.jclinepi.2018.05.021>
- Sterne JAC, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ*. 2016;355:i4919. <https://doi.org/10.1136/bmj.i4919>
- Thomson H, Craig P, Hilton-Boon M, Campbell M, Katikireddi SV. Applying the ROBINS-I tool to natural experiments: an example from public health. *Syst Rev*. 2018;7(1):15. <https://doi.org/10.1186/s13643-017-0659-4>
- Becker BJ, Aloe AM, Duvendack M, et al. Quasi-experimental study designs series—paper 10: synthesizing evidence for effects collected from quasi-experimental studies presents surmountable challenges. *J Clin Epidemiol*. 2017;89:84–91. <https://doi.org/10.1016/j.jclinepi.2017.02.014>
- Metelli S, Chaimani A. Challenges in meta-analyses with observational studies. *Evid Based Ment Health*. 2020;23(2):83–87. <https://doi.org/10.1136/ebmental-2019-300129>
- Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*. 2020;368:l6890. <https://doi.org/10.1136/bmj.l6890>
- Hilton Boon M, Thomson H, Shaw B, et al. Challenges in applying the GRADE approach in public health guidelines and systematic reviews: a concept article from the GRADE Public Health Group. *J Clin Epidemiol*. 2021;135:42–53. <https://doi.org/10.1016/j.jclinepi.2021.01.001>
- Schünemann HJ, Cuello C, Akl EA, et al. 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*. 2019;111:105–114. <https://doi.org/10.1016/j.jclinepi.2018.01.012>
- Pussegoda K, Turner L, Garrity C, et al. Systematic review adherence to methodological or reporting quality. *Syst Rev*. 2017;6(1):131. <https://doi.org/10.1186/s13643-017-0527-2>