White Rose Research Online
eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/132719/

Version: Published Version

Article:
Sterne, Jonathan, Hernán, Miguel A, Reeves, Barnaby C et al. (2016) ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. British medical journal. ISSN: 1756-1833

https://doi.org/10.1136/bmj.i4919

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

OPEN ACCESS

ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions

Jonathan AC Sterne,[1] Miguel A Hernán,[2] Barnaby C Reeves,[3] Jelena Savović,[1,4] Nancy D Berkman,[5] Meera Viswanathan,[6] David Henry,[7] Douglas G Altman,[8] Mohammed T Ansari,[9] Isabelle Boutron,[10] James R Carpenter,[11] An-Wen Chan,[12] Rachel Churchill,[13] Jonathan J Deeks,[14] Asbjørn Hróbjartsson,[15] Jamie Kirkham,[16] Peter Jüni,[17] Yoon K Loke,[18] Theresa D Pigott,[19] Craig R Ramsay,[20] Deborah Regidor,[21] Hannah R Rothstein,[22] Lakhbir Sandhu,[23] Pasqualina L Santaguida,[24] Holger J Schünemann,[25] Beverly Shea,[26] Ian Shrier,[27] Peter Tugwell,[28] Lucy Turner,[29] Jeffrey C Valentine,[30] Hugh Waddington,[31] Elizabeth Waters,[32] George A Wells,[33] Penny F Whiting,[34] Julian PT Higgins[35]

For numbered affiliations see end of article.

Correspondence to: J A C Sterne
jonathan.sterne@bristol.ac.uk

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;355:i4919
http://dx.doi.org/10.1136/bmj.i4919

Non-randomised studies of the effects of interventions are critical to many areas of healthcare evaluation, but their results may be biased. It is therefore important to understand and appraise their strengths and weaknesses. We developed ROBINS-I ("Risk Of Bias In Non-randomised Studies - of Interventions"), a new tool for evaluating risk of bias in estimates of the comparative effectiveness (harm or benefit) of interventions from studies that did not use randomisation to allocate units (individuals or clusters of individuals) to comparison groups. The tool will be particularly useful to those undertaking systematic reviews that include non-randomised studies.

SUMMARY POINTS

- Non-randomised studies of the effects of interventions are critical to many areas of healthcare evaluation but are subject to confounding and a range of other potential biases
- We developed, piloted, and refined a new tool, ROBINS-I, to assess "Risk Of Bias In Non-randomised Studies - of Interventions"
- The tool views each study as an attempt to emulate (mimic) a hypothetical pragmatic randomised trial, and covers seven distinct domains through which bias might be introduced
- We use "signalling questions" to help users of ROBINS-I to judge risk of bias within each domain
- The judgements within each domain carry forward to an overall risk of bias judgement across bias domains for the outcome being assessed

Non-randomised studies of the effects of interventions (NRSI) are critical to many areas of healthcare evaluation. Designs of NRSI that can be used to evaluate the effects of interventions include observational studies such as cohort studies and case-control studies in which intervention groups are allocated during the course of usual treatment decisions, and quasi-randomised studies in which the method of allocation falls short of full randomisation. Non-randomised studies can provide evidence additional to that available from randomised trials about long term outcomes, rare events, adverse effects and populations that are typical of real world practice.[1 2] The availability of linked databases and compilations of electronic health records has enabled NRSI to be conducted in large representative population cohorts.[3] For many types of organisational or public health interventions, NRSI are the main source of evidence about the likely impact of the intervention because randomised trials are difficult or impossible to conduct on an area-wide basis. Therefore systematic reviews addressing the effects of health related interventions often include NRSI. It is essential that methods are available to evaluate these studies, so that clinical, policy, and individual decisions are transparent and based on a full understanding of the strengths and weaknesses of the evidence.

Many tools to assess the methodological quality of observational studies in the context of a systematic review have been proposed.[4 5] The Newcastle-Ottawa[6] and Downs-Black[7] tools have been two of the most popular: both were on a shortlist of methodologically sound tools,[5] but each includes items relating to external as well as internal validity and a lack of comprehensive manuals means that instructions may be interpreted differently by different users.[5]

In the past decade, major developments have been made in tools to assess study validity. A shift in focus from methodological quality to risk of bias has been accompanied by a move from checklists and numeric scores towards domain-based assessments in which different types of bias are considered in turn. Examples are the Cochrane Risk of Bias tool for randomised trials,[8] the QUADAS 2 tool for diagnostic test accuracy studies,[9] and the ROBIS tool for systematic reviews.[10] However, there is no satisfactory domain-based assessment tool for NRSI.[4]

In this paper we describe the development of ROBINS-I ("Risk Of Bias In Non-randomised Studies - of Interventions"), which is concerned with evaluating risk of bias in estimates of the

bmj.i4919.indd 1 11-10-2016 17:40:06

effectiveness or safety (benefit or harm) of an intervention from studies that did not use randomisation to allocate interventions.

Development of a new tool

We developed the tool over three years, largely by expert consensus, and following the seven principles we previously described for assessing risk of bias in clinical trials.[8] A core group coordinated development of the tool, including recruitment of collaborators, preparation and revision of documents, and administrative support. An initial scoping meeting in October 2011 was followed by a survey of Cochrane Review Groups in March 2012 to gather information about the methods they were using to assess risk of bias in NRSI. A meeting in April 2012 identified the relevant bias domains and established working groups focusing on each of these. We agreed at this stage to use the approach previously adopted in the QUADAS-2 tool, in which answers to "signalling questions" help reviewers judge the risk of bias within each domain.[9] We distributed briefing documents to working groups in June 2012, specifying considerations for how signalling questions should be formulated and how answers to these would lead to a risk of bias judgement. We also identified methodological issues that would underpin the new tool: these are described below.

After collation and harmonisation by the core group of the working groups' contributions, all collaborators considered draft signalling questions and agreed on the main features of the new tool during a two-day face-to-face meeting in March 2013. A preliminary version of the tool was piloted within the working groups between September 2013 and March 2014, using NRSI in several review topic areas. Substantial revisions, based on results of the piloting, were agreed by leads of working groups in June 2014. Further piloting took place, along with a series of telephone interviews with people using the tool for the first time that explored whether they were interpreting the tool and the guidance as intended. We posted version 1.0.0, along with detailed guidance, at www.riskofbias.info in September 2014. We explained the tool during a three-day workshop involving members of Cochrane Review Groups in December 2014, and applied it in small groups to six papers reporting NRSI. Further modifications to the tool, particularly regarding wording, were based on feedback from this event and from subsequent training events conducted during 2015.

Methodological issues in assessing risk of bias in non-randomised studies

The target trial

Evaluations of risk of bias in the results of NRSI are facilitated by considering each NRSI as an attempt to emulate (mimic) a "target" trial. This is the hypothetical pragmatic randomised trial, conducted on the same participant group and without features putting it at risk of bias, whose results would answer the question addressed by the NRSI.[11 12] Such a "target" trial need not be feasible or ethical: for example, it could compare individuals who were and were not assigned to start smoking. Description of the target trial for the NRSI being assessed includes details of the population, experimental intervention, comparator, and outcomes of interest. Correspondingly, we define bias as a systematic difference between the results of the NRSI and the results expected from the target trial. Such bias is distinct from issues of generalisability (applicability or transportability) to types of individuals who were not included in the study.

The effect of interest

In the target trial, the effect of interest will typically be that of either:

1. Assignment to intervention at baseline (start of follow up), regardless of the extent to which the intervention was received during the follow-up (sometimes referred to as the "intention-to-treat" effect)
2. Starting and adhering to the intervention as indicated in the trial protocol (sometimes referred to as the "per-protocol" effect).

For example, in a trial of cancer screening, our interest might be in the effect of either sending an invitation to attend screening or of responding to the invitation and undergoing screening.

Analogues of these effects can be defined for NRSI. For example, the intention-to-treat effect in a study comparing aspirin with no aspirin can be approximated by the effect of being prescribed aspirin or (if using dispensing rather than prescription data) the effect of starting aspirin (this corresponds to the intention-to-treat effect in a trial in which participants assigned to an intervention always start that intervention). Alternatively, we might be interested in the effect of starting and adhering to aspirin.

The type of effect of interest influences assessments of risk of bias related to deviations from intervention. When the effect of interest is that of assignment to (or starting) intervention, risk of bias assessments generally need not be concerned with post-baseline deviations from interventions.[13] By contrast, unbiased estimation of the effect of starting and adhering to intervention requires consideration of both adherence and differences in additional interventions ("co-interventions") between intervention groups.

Domains of bias

We achieved consensus on seven domains through which bias might be introduced into a NRSI (see table 1 and appendix in supplementary data). The first two domains, covering confounding and selection of participants into the study, address issues before the start of the interventions that are to be compared ("baseline"). The third domain addresses classification of the interventions themselves. The other four domains address issues after the start of interventions: biases due to deviations from intended interventions, missing data, measurement of outcomes, and selection of the reported result.

doi: 10.1136/bmj.i4919 | *BMJ* 2016;355:i4919 | the bmj

bmj.i4919.indd 2 11-10-2016 17:40:06

Table 1 | Bias domains included in ROBINS-I

Domain	Explanation
Pre-intervention	**Risk of bias assessment is mainly distinct from assessments of randomised trials**
Bias due to confounding	Baseline confounding occurs when one or more prognostic variables (factors that predict the outcome of interest) also predicts the intervention received at baseline ROBINS-I can also address time-varying confounding, which occurs when individuals switch between the interventions being compared and when post-baseline prognostic factors affect the intervention received after baseline
Bias in selection of participants into the study	When exclusion of some eligible participants, or the initial follow-up time of some participants, or some outcome events is related to both intervention and outcome, there will be an association between interventions and outcome even if the effects of the interventions are identical This form of selection bias is distinct from confounding—A specific example is bias due to the inclusion of prevalent users, rather than new users, of an intervention
At intervention	**Risk of bias assessment is mainly distinct from assessments of randomised trials**
Bias in classification of interventions	Bias introduced by either differential or non-differential misclassification of intervention status Non-differential misclassification is unrelated to the outcome and will usually bias the estimated effect of intervention towards the null Differential misclassification occurs when misclassification of intervention status is related to the outcome or the risk of the outcome, and is likely to lead to bias
Post-intervention	**Risk of bias assessment has substantial overlap with assessments of randomised trials**
Bias due to deviations from intended interventions	Bias that arises when there are systematic differences between experimental intervention and comparator groups in the care provided, which represent a deviation from the intended intervention(s) Assessment of bias in this domain will depend on the type of effect of interest (either the effect of assignment to intervention or the effect of starting and adhering to intervention).
Bias due to missing data	Bias that arises when later follow-up is missing for individuals initially included and followed (such as differential loss to follow-up that is affected by prognostic factors); bias due to exclusion of individuals with missing information about intervention status or other variables such as confounders
Bias in measurement of outcomes	Bias introduced by either differential or non-differential errors in measurement of outcome data. Such bias can arise when outcome assessors are aware of intervention status, if different methods are used to assess outcomes in different intervention groups, or if measurement errors are related to intervention status or effects
Bias in selection of the reported result	Selective reporting of results in a way that depends on the findings and prevents the estimate from being included in a meta-analysis (or other synthesis)

For the first three domains, risk of bias assessments for NRSI are mainly distinct from assessments of randomised trials because randomisation, if properly implemented, protects against biases that arise before the start of intervention. However, randomisation does not protect against biases that arise after the start of intervention. Therefore, there is substantial overlap for the last four domains between bias assessments in NRSI and randomised trials.

Variation in terminology proved a challenge to development of ROBINS-I. The same terms are sometimes used to refer to different types of bias in randomised trials and NRSI literature,[13] and different types of bias are often described by a host of different terms: those used within ROBINS-I are shown in the first column of table 1.

The risk of bias tool, ROBINS-I

The full ROBINS-I tool is shown in tables A, B, and C in the supplementary data.

Planning the risk of bias assessment

It is very important that experts in both subject matter and epidemiological methods are included in any team evaluating a NRSI. The risk of bias assessment should begin with consideration of what problems might arise, in the context of the research question, in making a causal assessment of the effect of the intervention(s) of interest on the basis of NRSI. This will be based on experts' knowledge of the literature: the team should also address whether conflicts of interest might affect experts' judgements.

The research question is conceptualised by defining the population, experimental intervention, comparator, and outcomes of interest (supplementary table A, stage I). The comparator could be "no intervention," "usual care," or an alternative intervention. It is important to consider in advance the confounding factors and co-interventions that have the potential to lead to bias. Relevant confounding domains are the prognostic factors that predict whether an individual receives one or the other intervention of interest. Relevant co-interventions are those that individuals might receive with or after starting the intervention of interest and that are both related to the intervention received and prognostic for the outcome of interest. Both confounding domains and co-interventions are likely to be identified through the expert knowledge of members of the review group and through initial (scoping) reviews of the literature. Discussions with health professionals who make intervention decisions for the target patient or population groups may also help in identification of prognostic factors that influence treatment decisions.

Assessing a specific study

The assessment of each NRSI included in the review involves following the six steps below (supplementary table A, stage II). Steps 3 to 6 should be repeated for each key outcome of interest:

1. Specify the research question through consideration of a target trial
2. Specify the outcome and result being assessed
3. For the specified result, examine how the confounders and co-interventions were addressed
4. Answer signalling questions for the seven bias domains
5. Formulate risk of bias judgements for each of the seven bias domains, informed by answers to the signalling questions
6. Formulate an overall judgement on risk of bias for the outcome and result being assessed.

bmj.i4919.indd 3 11-10-2016 17:40:06

Examination of confounders and co-interventions involves determining whether the important confounders and co-interventions were measured or administered in the study at hand, and whether additional confounders and co-interventions were identified. Supplementary table A provides a structured approach to assessing the potential for bias due to confounding and co-interventions and includes the full tool with the signalling questions to be addressed within each bias domain.

The signalling questions are broadly factual in nature and aim to facilitate judgements about the risk of bias. The response options are: "Yes"; "Probably yes"; "Probably no"; "No"; and "No information". Some questions are answered only if the response to a previous question is "Yes" or "Probably yes" (or "No" or "Probably no"). Responses of "Yes" are intended to have similar implications to responses of "Probably yes" (and similarly for "No" and "Probably no"), but allow for a distinction between something that is known and something that is likely to be the case. Free text should be used to provide support for each answer, using direct quotations from the text of the study where possible.

Responses to signalling questions provide the basis for domain-level judgements about risk of bias, which then provide the basis for an overall risk of bias judgement for a particular outcome. The use of the word "judgement" to describe this process is important and reflects the need for review authors to consider both the severity of the bias in a particular domain and the relative consequences of bias in different domains.

The categories for risk of bias judgements are "Low risk", "Moderate risk", "Serious risk" and "Critical risk" of bias. Importantly, "Low risk" corresponds to the risk of bias in a high quality randomised trial. Only exceptionally will an NRSI be assessed as at low risk of bias due to confounding. Criteria for reaching risk of bias judgements for the seven domains are provided in supplementary tables B and C. If none of the answers to the signalling questions for a domain suggests a potential problem then risk of bias for the domain can be judged to be low. Otherwise, potential for bias exists. Review authors must then make a judgement on the extent to which the results of the study are at risk of bias. "Risk of bias" is to be interpreted as "risk of material bias". That is, concerns should be expressed only about issues that are likely to affect the ability to draw valid conclusions from the study: a serious risk of a very small degree of bias should not be considered "Serious risk" of bias. The "No information" category should be used only when insufficient data are reported to permit a judgement.

The judgements within each domain carry forward to an overall risk of bias judgement for the outcome being assessed (across bias domains, that is), as summarised in table 2 (also saved as supplementary table D). The key to applying the tool is to make domain-level judgements about risk of bias that have the same meaning across domains with respect to concern about the impact of bias on the trustworthiness of the result. If domain-level judgements are made consistently, then judging the overall risk of bias for a particular outcome is relatively straightforward. For instance, a "Serious risk" of bias in one domain means the effect estimate from the study is at serious risk of bias or worse, even if the risk of bias is judged to be lower in the other domains.

It would be highly desirable to know the magnitude and direction of any potential biases identified, but this is considerably more challenging than judging the risk of bias. The tool includes an optional component to predict the direction of the bias for each domain, and overall. For some domains, the bias is most easily thought of as being towards or away from the null. For example, suspicion of selective non-reporting of statistically non-significant results would suggest bias against the null. However, for other domains (in particular confounding, selection bias and forms of measurement bias such as differential misclassification), the bias needs to be thought of as an increase or decrease in the effect estimate and not in relation to the null. For example, confounding bias that decreases the effect estimate would be towards the null if the true risk ratio were greater than 1, and away from the null if the risk ratio were less than 1.

Discussion

We developed a tool for assessing risk of bias in the results of non-randomised studies of interventions that addresses weaknesses in previously available approaches.[4] Our approach builds on recent

Table 2 | Interpretation of domain-level and overall risk of bias judgements in ROBINS-I*

Judgement	Within each domain	Across domains	Criterion
Low risk of bias	The study is comparable to a well performed randomised trial with regard to this domain	The study is comparable to a well performed randomised trial	The study is judged to be at **low risk of bias for all domains**
Moderate risk of bias	The study is sound for a non-randomised study with regard to this domain but cannot be considered comparable to a well performed randomised trial	The study provides sound evidence for a non-randomised study but cannot be considered comparable to a well performed randomised trial	The study is judged to be at **low or moderate risk of bias for all domains**
Serious risk of bias	The study has some important problems in this domain	The study has some important problems	The study is judged to be at **serious risk of bias** in at least one domain, but not at critical risk of bias in any domain
Critical risk of bias	The study is too problematic in this domain to provide any useful evidence on the effects of intervention	The study is too problematic to provide any useful evidence and should not be included in any synthesis	The study is judged to be at **critical risk of bias in at least one domain**
No information	No information on which to base a judgement about risk of bias for this domain	No information on which to base a judgement about risk of bias	There is no clear indication that the study is at serious or critical risk of bias *and* there is a lack of information in one or more key domains of bias (*a judgement is required for this*)

*Also saved as supplementary table D.

doi: 10.1136/bmj.i4919 | *BMJ* 2016;355:i4919 | the bmj

bmj.i4919.indd 4 11-10-2016 17:40:06

developments in risk of bias assessment of randomised trials and diagnostic test accuracy studies.[8 9] Key features of ROBINS-I include specification of the target trial and effect of interest, use of signalling questions to inform judgements of risk of bias, and assessments within seven bias domains.

The ROBINS-I tool was developed through consensus among a group that included both methodological experts and systematic review authors and editors, and was substantially revised based on extensive piloting and user feedback. It includes a structured approach to assessment of risk of bias due to confounding that starts at the review protocol stage. Use of ROBINS-I requires that review groups include members with substantial methodological expertise and familiarity with modern epidemiological thinking. We tried to make ROBINS-I as accessible and easy to use as possible, given the requirement for comprehensive risk of bias assessments that are applicable to a wide range of study designs and analyses. An illustrative assessment using ROBINS-I can be found at www.riskofbias.info; detailed guidance and further training materials will also be available.

ROBINS-I separates relatively factual answers to signalling questions from more subjective judgements about risk of bias. We hope that the explicit links between answers to signalling questions and risk of bias judgements will improve reliability of the domain-specific and overall risk of bias assessments.[14] Nonetheless, we expect that the technical difficulty in making risk of bias judgements will limit reliability. Despite this, ROBINS-I provides a comprehensive and structured approach to assessing non-randomised studies of interventions. It should therefore facilitate debates and improve mutual understanding about the ways in which bias can influence effects estimated in NRSI, and clarify reasons for disagreements about specific risk of bias judgements. Note that the tool focuses specifically on bias and does not address problems related to imprecision of results, for example when statistical analyses fail to account for clustering or matching of participants.

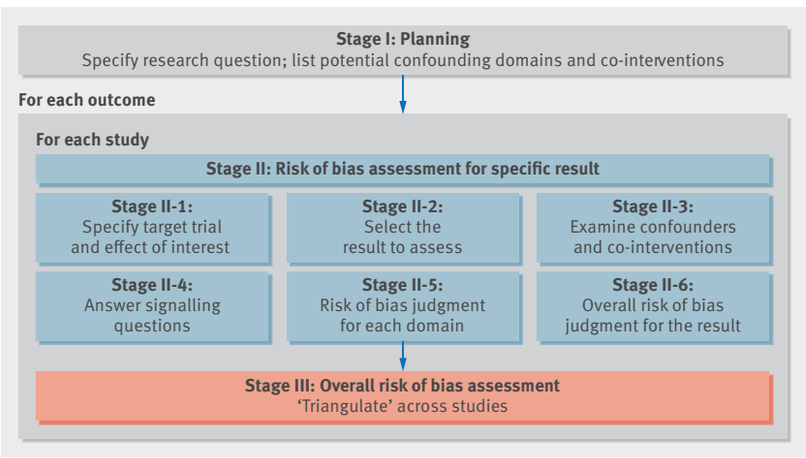

Fig 1 | Summary of the process of assessing risk of bias in a systematic review of non-randomised studies of interventions (NRSI)

We developed the ROBINS-I tool primarily for use in the context of a systematic review. Broader potential uses include the assessment of funding applications and peer review of journal submissions. Furthermore, ROBINS-I may be used to guide researchers about issues to consider when designing a primary study to evaluate the effect(s) of an intervention.

Figure 1 summarises the process of assessing risk of bias using the tool in the context of a systematic review of NRSI. To draw conclusions about the extent to which observed intervention effects might be causal, the studies included in the review should be compared and contrasted so that their strengths and weaknesses can be considered jointly. Studies with different designs may present different types of bias, and "triangulation" of findings across these studies may provide assurance either that the biases are minimal or that they are real. Syntheses of findings across studies through meta-analysis must consider the risks of bias in the studies available. We recommend against including studies assessed as at "Critical risk" of bias in any meta-analysis, and advocate caution for studies assessed as at "Serious risk" of bias. Subgroup analyses (in which intervention effects are estimated separately according to risk of bias), meta-regression analyses, and sensitivity analyses (excluding studies at higher risk of bias) might be considered, either within specific bias domains or overall. Risk of bias assessments might alternatively be used as the basis for deriving adjustments for bias through prior distributions in Bayesian meta-analyses.[15 16]

The GRADE system for assessing confidence in estimates of the effects of interventions currently assigns a starting rating of "Low certainty, confidence or quality" to non-randomised studies, a downgrading by default of two levels.[17] ROBINS-I provides a thorough assessment of risk of bias in relation to a hypothetical randomised trial, and "Low risk" of bias corresponds to the risk of bias in a high quality randomised trial. This opens up the possibility of using the risk of bias assessment, rather than the lack of randomisation per se, to determine the degree of downgrading of a study result, and means that results of NRSI and randomised trials could be synthesised if they are assessed to be at similar risks of bias. In general, however, we advocate analysing these study designs separately and focusing on evidence from NRSI when evidence from trials is not available.

Planned developments of ROBINS-I include further consideration of the extent to which it works for specific types of NRSI, such as self-controlled designs, controlled before-and-after studies, interrupted time series studies, and studies based on regression discontinuity and instrumental variable analyses. We also plan to develop interactive software to facilitate use of ROBINS-I. Furthermore, the discussions that led up to the tool will inform a reconsideration of the tool for randomised trials, particularly in the four post-intervention domains.[8]

The role of NRSI in informing treatment decisions remains controversial. Because randomised trials are expensive, time consuming, and may not reflect real

bmj.i4919.indd 5 11-10-2016 17:40:06

world experience with healthcare interventions, research funders are enthusiastic about the possible use of observational studies to provide evidence about the comparative effectiveness of different interventions,[18] and encourage use of large, routinely collected datasets assembled through data linkage.[18] However, fear that evidence from NRSI may be biased, based on misleading results of some NRSI,[19 20] has led to caution in their use in making judgements about efficacy. There is greater confidence in the capacity of NRSI to quantify uncommon adverse effects of interventions.[21] We believe that evidence from NRSI should complement that from randomised trials, such as in providing evidence about effects on rare and adverse outcomes and long term effects to be balanced against the outcomes more readily addressed in randomised trials.[22]

AUTHOR AFFILIATIONS

[1]School of Social and Community Medicine, University of Bristol, Bristol BS8 2PS, UK

[2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; and Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; and Harvard-Massachusetts Institute of Technology Division of Health Sciences and Technology, Boston, Massachusetts, USA

[3]School of Clinical Sciences, University of Bristol, Bristol, BS2 8HW, UK

[4]National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West (NIHR CLAHRC West) at University Hospitals Bristol NHS Foundation Trust, Bristol BS1 2NT, UK

[5]Program on Health Care Quality and Outcomes, Division of Health Services and Social Policy Research, RTI International, Research Triangle Park, NC 27709, USA

[6]RTI-UNC Evidence-based Practice Center, RTI International, Research Triangle Park, NC 27709, USA

[7]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

[8]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK

[9]School of Epidemiology, Public Health and Preventive Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, K1H 8M5, Canada

[10]METHODS Team, Centre of Epidemiology and Statistics Sorbonne Paris Cité Research, INSERM UMR 1153, University Paris Descartes, Paris, France

[11]Department of Medical Statistics, London School of Hygiene and Tropical Medicine and MRC Clinical Trials Unit at UCL, London, UK

[12]Women's College Research Institute, Department of Medicine, University of Toronto, Canada

[13]Centre for Reviews and Dissemination, University of York, York, YO10 5DD, UK

[14]Institute of Applied Health Research, University of Birmingham, Birmingham B15 2TT, UK

[15]Center for Evidence-Based Medicine, University of Southern Denmark & Odense University Hospital, 5000 Odense C, Denmark

[16]Department of Biostatistics, University of Liverpool, Liverpool, L69 3GL, UK

[17]Applied Health Research Centre (AHRC), Li Ka Shing Knowledge Institute of St Michael's Hospital, and Department of Medicine, University of Toronto, Toronto, Ontario, Canada

[18]Norwich Medical School, University of East Anglia, Norwich NR4 7TJ, UK

[19]School of Education, Loyola University Chicago, Chicago, IL 60611, USA

[20]Health Services Research Unit, University of Aberdeen, Aberdeen, AB25 2ZD, UK.

[21]Evidence Services, Kaiser Permanente, Care Management Institute, Oakland, CA 94612, USA

[22]Department of Management, Zicklin School of Business, Baruch College—CUNY, New York, NY 10010, USA

[23]Division of General Surgery, University of Toronto, Toronto, Canada

[24]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, L8S 4K1, Canada

[25]Departments of Clinical Epidemiology and Biostatistics and of Medicine, Cochrane Applicability and Recommendations Methods (GRADEing) Group, MacGRADE center, Ontario, L8N 4K1, Canada

[26]Ottawa Hospital Research Institute, Center for Practice Changing Research and School of Epidemiology, Public Health and Preventive Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, K1H 8M5, Canada

[27]Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Quebec, Canada

[28]Department of Medicine and School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Ontario, Canada

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

[30]University of Louisville, Louisville, KY 40292, USA

[31]International Initiative for Impact Evaluation, London School of Hygiene and Tropical Medicine, and London International Development Centre, London, UK

[32]Jack Brockhoff Child Health & Wellbeing Program, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, VIC 3010, Australia

[33]School of Epidemiology, Public Health and Preventive Medicine and Director, Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Ontario, K1Y 4W7, Canada

[34]School of Social and Community Medicine, University of Bristol, Bristol BS8 2PS, UK; and National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West (NIHR CLAHRC West) at University Hospitals Bristol NHS Foundation Trust, Bristol BS1 2NT, UK

[35]School of Social and Community Medicine, University of Bristol, Bristol BS8 2PS, UK

The ROBINS-I tool is reproduced from riskofbias.info with the permission of the authors. The tool should not be modified for use.

We thank Professor Jan Vandenbroucke for his contributions to discussions within working groups and during face-to-face meetings. Professor David Moher and Dr Vivan Welch contributed to the grant application and to initial discussions. We thank all those individuals who contributed to development of ROBINS-I through discussions during workshops and training events, and through their work on piloting.

Contributions of authors: JACS, BCR, JS, LT, YKL, EW, CRR, PT, GAW, and JPTH conceived the project. JACS, JPTH, BCR, JS, and LT oversaw the project. JACS, DH, JPTH, IB, and BRR led working groups. MAH developed the idea of the target trial and its role in ROBINS-I. NDB and MV undertook cognitive testing of previous drafts of the tool. All authors contributed to development of ROBINS-I and to writing associated guidance. JACS, MAH, and JPTH led on drafting the manuscript. All authors reviewed and commented on drafts of the manuscript.

Funding: Development of ROBINS-I was funded by a Methods Innovation Fund grant from Cochrane and Medical Research Council (MRC) grant MR/M025209/1. Sterne and Higgins are members of the MRC Integrative Epidemiology Unit at the University of Bristol, which is supported by the MRC and the University of Bristol (grant MC_UU_12013/9). This research was partly funded by NIH grant P01 CA134294. Sterne was supported by National Institute for Health Research (NIHR) Senior Investigator award NF-SI-0611-10168. Savović and Whiting were supported by NIHR Collaboration for Leadership in Applied Health Research and Care West (NIHR CLAHRC West). Reeves was supported by the NIHR Bristol Biomedical Research Unit in Cardiovascular Disease. None of the funders had a role in the development of the ROBINS-I tool, although employees of Cochrane contributed to some of the meetings and workshops. The views expressed are those of the authors and not necessarily those of Cochrane, the NHS, NIHR, or Department of Health.

Competing interests: All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare: grants from Cochrane, MRC, and NIHR during the conduct of the study. Dr Carpenter reports personal fees from Pfizer, grants and non-financial support from GSK and grants from Novartis, outside the submitted work. Dr Reeves is a co-convenor of the Cochrane Non-Randomised Studies Methods Group. The authors report no other relationships or activities that could appear to have influenced the submitted work.

bmj.i4919.indd 6 11-10-2016 17:40:06

Provenance: The authors are epidemiologists, statisticians, systematic reviewers, trialists, and health services researchers, many of whom are involved with Cochrane systematic reviews, methods groups and training events. Development of ROBINS-I was informed by relevant methodological literature, previously published tools for assessing methodological quality of non-randomised studies, systematic reviews of such tools and relevant literature, and by the authors' experience of developing tools to assess risk of bias in randomised trials, diagnostic test accuracy studies, and systematic reviews. All authors contributed to development of ROBINS-I and to writing associated guidance. All authors reviewed and commented on drafts of the manuscript. J Sterne will act as guarantor.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/.

1 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8. doi:10.1136/bmj.312.7040.1215.
2 Feinstein AR. An additional basic science for clinical medicine: II. The limitations of randomized trials. *Ann Intern Med* 1983;99:544-50. doi:10.7326/0003-4819-99-4-544.
3 Strom B. Overview of automated databases in pharmacoepidemiology. In: Strom BL, Hennessy S, eds. *Pharmacoepidemiology*. 5th ed. Wiley, 2012.
4 Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36:666-76. doi:10.1093/ije/dym018.
5 Deeks JJ, Dinnes J, D'Amico R, et al. International Stroke Trial Collaborative Group European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:iii-x, 1-173. doi:10.3310/hta7270.
6 Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2008. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 1/03/2016).
7 Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377-84. doi:10.1136/jech.52.6.377.
8 Higgins JPT, Altman DG, Gøtzsche PC, et al. Cochrane Bias Methods Group Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928. doi:10.1136/bmj.d5928.
9 Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36. doi:10.7326/0003-4819-155-8-201110180-00009.
10 Whiting P, Savović J, Higgins JPT, et al. ROBIS group. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225-34. doi:10.1016/j.jclinepi.2015.06.005.
11 Institute of Medicine. *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*.The National Academies Press, 2012.
12 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758-64. doi:10.1093/aje/kwv254.
13 Mansournia MA, Higgins JPT, Sterne JAC, et al. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology* [forthcoming].
14 Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 2013;66:973-81. doi:10.1016/j.jclinepi.2012.07.005.
15 Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. *J R Stat Soc Ser A Stat Soc* 2009;172:21-47. doi:10.1111/j.1467-985X.2008.00547.x.
16 Wilks DC, Sharp SJ, Ekelund U, et al. Objectively measured physical activity and fat mass in children: a bias-adjusted meta-analysis of prospective studies. *PLoS One* 2011;6:e17205. doi:10.1371/journal.pone.0017205.
17 Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *J Clin Epidemiol* 2011;64:380-2. doi:10.1016/j.jclinepi.2010.09.011.
18 Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med* 2009;151:203-5. doi:10.7326/0003-4819-151-3-200908040-00125.
19 Bjelakovic G, Nikolova D, Gluud LL, Simonetti RG, Gluud C. Mortality in randomized trials of antioxidant supplements for primary and secondary prevention: systematic review and meta-analysis. *JAMA* 2007;297:842-57. doi:10.1001/jama.297.8.842.
20 Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008;19:766-79. doi:10.1097/EDE.0b013e3181875e61.
21 Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS Med* 2011;8:e1001026. doi:10.1371/journal.pmed.1001026.
22 Schünemann HJ, Tugwell P, Reeves BC, et al. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. *Res Synth Methods* 2013;4:49-62. doi:10.1002/jrsm.1078.

© BMJ Publishing Group Ltd 2016

Appendix: The seven domains of bias addressed in the ROBINS-I assessment tool
Supplementary tables. Table A: Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool. Table B: Reaching risk of bias judgements in ROBINS-I: pre-intervention and at-intervention domains. Table C: Reaching risk of bias judgements in ROBINS-I: post-intervention domains. Table D: Interpretation of domain-level and overall risk of bias judgements in ROBINS-I

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe