An introduction to methodological issues when including non-randomised studies in systematic reviews on the effects of interventions

Barnaby C. Reeves,a*† Julian P. T. Higgins,b,c Craig Ramsay,d Beverley Shea,e Peter Tugwellf and George A. Wellsg

Background: Methods need to be further developed to include non-randomised studies (NRS) in systematic reviews of the effects of health care interventions. NRS are often required to answer questions about harms and interventions for which evidence from randomised controlled trials (RCTs) is not available. Methods used to review randomised controlled trials may be inappropriate or insufficient for NRS.

Aim and methods: A workshop was convened to discuss relevant methodological issues. Participants were invited from important stakeholder constituencies, including methods and review groups of the Cochrane and Campbell Collaborations, the Cochrane Editorial Unit and organisations that commission reviews and make health policy decisions. The aim was to discuss methods for reviewing evidence when including NRS and to formulate methodological guidance for review authors.

Workshop format: The workshop was structured around four sessions on topics considered in advance to be most critical: (i) study designs and bias; (ii) confounding and meta-analysis; (iii) selective reporting; and (iv) applicability. These sessions were scheduled between introductory and concluding sessions.

Summary: This is the first of six papers and provides an overview. Subsequent papers describe the discussions and conclusions from the four main sessions (papers 2 to 5) and summarise the proposed guidance into lists of issues for review authors to consider (paper 6). Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: systematic review; non-randomised study; benefit; harm

1. Introduction

The Non-Randomised Studies Methods Group (NRSMG) of the Cochrane Collaboration [www.cochrane.org] convened an invited workshop in June 2010 to discuss methodological issues raised by the inclusion of primary studies that are not based on the principle of randomization in systematic reviews of the effectiveness of health care interventions. Invited participants were chosen to ensure that important constituencies with an interest in the topic were represented. These constituencies included methods groups of the Cochrane and Campbell Collaborations, the Cochrane Editorial Unit, Cochrane and Campbell review groups and organisations responsible for commissioning reviews and making...
health policy decisions. All participants are listed in Appendix A. This is the first of a series of papers to report the conclusions of the workshop and to offer guidance for review authors.

This topic has become a priority for The Cochrane Collaboration. Health policy stakeholders, who often commission reviews, are advising the Collaboration’s leaders that systematic reviews need to review evidence beyond randomised controlled trials (RCTs) if they are to be considered relevant. Thus, review teams addressing questions of high priority to such stakeholders, but for which there is a paucity of RCTs, urgently need guidance about how they can best inform key health policy decisions. Furthermore, review teams addressing conventional questions increasingly recognise that a review only of the benefits of an intervention is biased and potentially harmful to health. The growing availability of large datasets containing information about adverse effects makes it increasingly feasible to quantify the harms of many interventions. As a result of these influences, the Collaboration has been reconsidering its policies towards non-randomised studies (NRS; for a definition, see Box 1) (Reeves et al., 2008).

A common misconception is that The Cochrane Collaboration is inherently opposed to including primary research studies other than RCTs in systematic reviews. This is not the case. As long ago as 1994, the Collaboration’s documentation (Oxman et al., 1994) recognised that NRS can contribute important information and, in principle, should be considered for inclusion in a review. However, the Collaboration also recognised that primary NRS are at increased risk of bias, and in the past (when systematic reviews tended to focus on the benefits of interventions), concern about the risk of bias to NRS led most review groups in the Collaboration to exclude NRS from their systematic reviews. Notable exceptions have included the Effective Practice and Organisation of Care, Public Health and Injuries review groups, which have often included NRS in their reviews. The situation is dynamic, particularly with respect to the harmful effects of interventions, as evidenced by the Collaboration’s decision to prioritise the inclusion of NRS to address adverse effects at its 2012 mid-year meeting in Paris. [http://www.editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/2012-CC-strategic-session_short-version.pdf; theme 6, paragraph 39] The Campbell Collaboration has sought to include NRS from the time it was established in 2000. [http://www.campbellcollaboration.org/about_us/index.php]

Systematic review authors are likely to need to include NRS to ensure that a review (i) addresses questions about serious, rare and long-term harmful or unintended effects of an intervention as well as the beneficial, frequent, intended or short-term effects; (ii) adequately considers the applicability of the findings of a review to a range of settings; and (iii) provides an evidence synthesis for prioritised research questions to inform policy and other decisions in the absence of RCTs (where decisions about implementation are going to be taken imminently, and an ‘empty review’ (Yaffe et al., 2012) does not help). However, for most research questions about beneficial effects, the findings of NRS are likely to be more biased than RCTs. Combining the results of NRS at high risk of bias in a meta-analysis runs the risk of producing a biased effect estimate with unwarranted precision (Egger et al., 1998).

The decision to include NRS or not in a particular review is both complex and critical to the usefulness of a review. Like any piece of primary research, a review can appropriately inform a clinical or health policy decision if the findings are valid and applicable. It can also misinform a health policy decision with harmful or wasteful consequences if the findings are biased, omit important effects that need to be weighed up against other effects or are of limited applicability. A systematic review has a privileged position in the evidence base, typically sitting between primary research studies and guidelines that frequently cite them. There may be long-term undesirable consequences of reviewing evidence when it is inadequate: An evidence synthesis may make it less likely that less biased research will be carried out in the future, increasing the potential harm arising from a misinformed decision (Stampfer and Colditz, 1991; Siegfried et al., 2005). These considerations place a major responsibility on review authors, who should take expert clinical and methodological advice when choosing the eligibility criteria for studies to address a particular review question and deciding whether or not to include NRS.

<table>
<thead>
<tr>
<th>Box 1. Workshop definition of non-randomized studies (NRS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In this introductory paper, and in the papers that follow, ‘non-randomised studies’ (NRS) are defined as follows: ‘Any quantitative study estimating the effectiveness of an intervention (benefit or harm) that does not use randomization to allocate units to comparison groups. This includes studies where allocation occurs in the course of usual treatment decisions or peoples’ choices, i.e. studies usually called observational. In the scientific literature, there are many types of named non-randomized intervention study, including cohort studies, case-control studies, controlled before-and-after studies, interrupted-time-series studies and controlled trials that use inappropriate randomization strategies (sometimes called quasi-randomized studies).’ (Reeves et al., 2008)</td>
</tr>
</tbody>
</table>

2. Aim and structure of the workshop

The aim of the workshop was to discuss how standard methods for reviewing evidence may need to be adapted when including NRS and to formulate methodological guidance for review authors. Participants were invited so that they brought a diverse range of perspectives. Six Cochrane and Campbell methods groups were represented, covering expertise and interest in NRS, statistics, bias, adverse effects, applicability and recommendations, and equity. The editor-in-chief of the Cochrane Collaboration and editors of Cochrane (EPOC, Public Health, Musculoskeletal) and Campbell (welfare, education, international development) coordinating groups took part, ensuring that discussions considered the practicality of potential review methods. There were also representatives of the Agency for Health Care Quality, the National Institute for Health and Clinical Excellence, the Institute for Clinical Evaluative Studies and the International Initiative for Impact Evaluation. Some participants represented more than one constituency (Appendix A).

The workshop was structured around half-day sessions on four methodological topics considered in advance to be most critical or most likely to influence the decision to include NRS:

- Study design and bias
- Confounding and meta-analysis
- Selective reporting
- Applicability

These sessions were scheduled between two additional shorter sessions designed to set the scene, in which some session leaders and chairs presented specific perspectives and to sum up the main conclusions and achievements of the workshop.

Participants received short notes about each session, key references and priority questions for discussion (not all are addressed in this series) in advance of the workshop. The workshop structure is mirrored in this series of papers. The workshop was funded by grants from the Cochrane Collaboration and the Ottawa Collaborating Centre of the Agency for Healthcare Quality and Research. Many participants contribute to the activities of one or both of these organisations, and some are authors on the papers that follow. However, neither organisation had any influence over the content of the papers.

3. Context for the workshop and cross-cutting issues

The distinction between intended (beneficial) effects and important (serious) unintended (adverse) effects was a consistent theme throughout the workshop and is a major factor influencing whether to include NRS or not to answer a particular research question. Research questions about harms are widely considered to be less at risk of selection bias/confounding because these outcomes are less likely to be indications for administering or withholding an intervention. There is some recent evidence to support this view (Golder et al., 2011). However, confounding can still occur, especially in circumstances in which the adverse event arises in the same biological pathway/organ system as the one targeted by the intended effects of an intervention (Stampfer, 2004). There may also be circumstances, such as surgery, where concern about an adverse event such as a post-operative complication is an important indication for selecting one or other intervention (Concato et al., 1992).

In Box 1, NRS are defined as studies used to evaluate health care interventions. Discussions at the workshop were also confined to NRS used in this way. Studies without random assignment to groups are also used to study aetiological or other important research questions about health care, for example, diagnostic accuracy or prognosis, when RCTs cannot be used (Haynes et al., 2005; Oxford Centre for Evidence-based Medicine, 2001). The distinction between effectiveness and aetiological research questions may seem straightforward but, in some circumstances, can be difficult to make (Box 2). The distinction is important because it is likely that additional issues will need to be considered when performing systematic reviews that include primary NRS of aetiological research questions.

Box 2. Distinguishing between aetiology and effectiveness research questions

Including non-randomised studies (NRS) in a Cochrane review allows, in principle, the inclusion of truly observational studies where the use of an intervention has occurred in the course of usual health care or daily life. For interventions that are not restricted to a medical setting, this may mean interventions that a study participant chooses to take, for example, over-the-counter preparations. Including observational studies in a review also allows exposures to be studied that are not obviously interventions, for example, nutritional choices and other behaviours that may affect health.

This introduces a grey area between evidence about effectiveness and aetiology.

It is important to distinguish carefully between different aetiological and effectiveness research questions related to a particular exposure. For example, nutritionists may be interested in the health-related effects of a diet that includes a minimum of five portions of fruit or vegetables per day (five-a-day), an aetiological question. On the other hand,
public health professionals may be interested in the health-related effects of interventions to promote a change in diet to include five-a-day, an effectiveness question. Because of other differences between these two kinds of studies, e.g. duration of the intervention or the timing of assessment of outcome, the former are often perceived as being better or more relevant without acknowledging or realising that they are addressing different research questions. In other instances, the health intervention being evaluated in the NRS will have been undertaken for a purpose other than improving health. For example, a review of circumcision for preventing transmission of HIV included NRS where circumcision had been undertaken for cultural or religious reasons (Siegfried et al., 2005), and it was unclear whether using the intervention for health purposes would have the same effect.

The workshop also considered the importance of the wider context in relation to whether a qualitative or quantitative answer about the relationship between an intervention and adverse effect is required. In some circumstances, inferring causality because of the extreme magnitude of an association may be sufficient, without precise quantification of the relative increase in the risk of the adverse event conferred. In other circumstances, precise quantification may be required, for example, in order to weigh up the balance between harms and benefits or to choose between competing treatments.

Another important consideration discussed was the way in which the nature of included evidence might need to vary depending on whether a review aimed to synthesise evidence about all harms, both suspected and unsuspected, or only suspected harms (Loke et al., 2008; Loke et al., 2011). When reviewing questions about suspected harms, it may be sufficient to consider conventional sources of primary studies, searching on the basis of the population and intervention of interest and the suspected outcomes. When reviewing questions about unsuspected harms, it may be necessary to search more widely, including unconventional sources of primary studies and types of evidence, for example, pharmacological vigilance databases and case reports (where there may not even be a denominator). The discussion of the relevance of varying types of evidence to these contrasting situations led participants to contrast review questions that have an ‘exploratory’ aim with those that are ‘confirmatory’ (Box 3).

Box 3. Distinguishing between confirmatory and exploratory studies
In reality, this concept represents a dimension rather than a dichotomy. The descriptions that follow are intended to describe the poles of the dimension. Without a protocol written in advance of analyses of the data, it can often be difficult to know from publication of the findings of a study what the primary researchers’ original objectives were. Review authors should be aware that primary researchers may change their position during analyses of the data and writing a manuscript for publication, giving rise to selective reporting.

Confirmatory:
A confirmatory study has as its primary objective the estimation of intervention effects for one or more defined outcomes (to a specified degree of precision based on a pre-specified sample size justification) from comparison of two or more defined interventions. All randomised controlled trials should fall into this category. The key feature of such studies is a focused research question that is clearly pre-specified in terms of study population, intervention, comparator and outcome (PICO) (Haynes et al., 2005), clearly distinguishing a primary and other outcomes. Confirmatory studies are usually designed to refute a null hypothesis of no difference between groups, although some are designed to test other hypothesis such as non-inferiority. A precisely defined research question allows the target sample size to be set, which ensures that a study has adequate power to test the hypothesis of interest; thus, justification of the sample size studied should be an important feature of the protocol of a confirmatory study.

Exploratory:
An exploratory study has as its primary objective the investigation of associations between one or more outcomes and two or more interventions. Outcomes and interventions may not be clearly defined at the outset, and the study may not have a pre-specified sample size justification at the outset, that is, be carried out on the basis of ‘available data’. Because exploratory research questions are imprecisely specified, they are unlikely to be able to set specific objectives in PICO terms (although an exploratory study may subsequently be written up to appear as though such objectives were set in advance of doing the study, a risk that is heightened by the lack of a protocol). This type of study is more likely to be non-randomised, to be carried out at an early stage of evaluation of an intervention, or to have as its primary focus the investigation of harmful effects. The danger of selective reporting in such studies is greater because the primary researchers are likely to carry out a wide range of analyses (multiple outcomes, multiple interventions, adjusted in multiple ways) and statistical tests, and choose to report the analyses that can be written up to make a coherent and interesting manuscript. It is important that primary researchers declare transparently when a study is exploratory and a further confirmatory study is needed to attempt to replicate important findings; in a replication, one or more findings from the exploratory study may become the precisely specified objectives in what is then a confirmatory study). Less confidence should be attached to the findings of an exploratory study (substantially because of the risk of bias arising from selective reporting and the risk of reaching false positive conclusions because of multiple statistical testing). Given these features of an exploratory study, review authors should consider carefully whether it is appropriate to include them in a systematic review alongside other, confirmatory studies.
4. Study design and bias (session 1)

Once a decision has been made to include NRS to address one or more objectives of a systematic review, the next question is ‘which NRS to include’? NRS are widely considered to vary greatly in their risk of bias, and the risk of bias to an NRS is likely to be influenced to some extent by its study design. The question of whether to include NRS, and if so, which kinds of NRS to include, will depend on the precise objective(s) of the review and on the health setting or context. The strategy for deciding how different types of NRS might be included in the review also needs to be decided: should particular types of studies be included only if there are no ‘better’ studies or should an absolute threshold for eligibility be implemented rigorously, even if no studies meet the standard? The answer to this question may depend on the priority to policy makers of the research question. Furthermore, eligibility criteria may be different for different outcomes of interest, such as intended versus unintended effects, benefit versus harm, short versus long term. In particular, if estimation of harmful effects is judged to be less at risk of confounding, it may be appropriate to include study designs that, for beneficial effects, would be considered to be at too high risk of bias.

To define clear eligibility criteria for inclusion of primary studies in the protocol for a review (Higgins and Green, 2008), there is a need to classify NRS. However, there is no established way to do this. Design ‘hierarchies’ exist but are intended for, or have evolved from, hierarchies for aetiological or epidemiological designs to investigate exposure to risk or protective factors rather than health care interventions (Oxford Centre for Evidence-based Medicine; Eccles et al., 1996; National Health and Medical Research Council, 1999). It is not clear that these study design hierarchies are appropriate for studies of the effects of health care interventions because of the greater diversity of the latter and differences between the nature of exposure to aetiological factors and choice of, or allocation to, an intervention.

Non-randomised studies vary in their design features (Reeves et al., 2008; Higgins et al., 2012). These features make the studies more or less susceptible to bias. In NRS, compared with in RCTs, attrition is often worse (and poorly reported); intervention and outcome assessments are less likely to be conducted according to standardised protocols, and outcomes are rarely assessed blind to the group allocation. Too often, these limitations of NRS are seen as part of doing an NRS with the consequence that their implications for risk of bias to a study (Higgins and Altman, 2008), and the way in which limitations vary across studies, are not properly considered. For example, some users of evidence may consider NRS that investigate long-term outcomes to have ‘better quality’ than randomised trials of short-term outcomes, simply on the basis of their relevance without appraising their risk of bias. Schünemann et al. (2013) describe how such a judgement can be made more systematically.

Questions distributed to participants in advance for consideration during the session:

1. The Cochrane Collaboration has typically sought to summarise reliable evidence rather than the best available evidence (i.e. with an absolute threshold rather than a conditional rule for inclusion). Should it continue to do this?
2. In what circumstances should NRS be included in Cochrane reviews? How acceptable is it for this decision to depend on (i) the anticipated needs of certain stakeholders; (ii) knowledge of how much randomised evidence exists; and (iii) the suspected magnitude of effect size?
3. Should a common tool be used to assess the risk of bias in RCTs and non-randomized studies? What sort of tool would allow differentiation in risks of bias: (i) between RCTs and non-randomized studies, and (ii) between different types of NRS?
4. How should review authors draw distinctions between different types of non-randomised evidence, particularly in relation to determining eligibility for the review? Are design labels (e.g. cohort study) or design features (e.g. prospective identification of participants) preferable?
5. When should review authors be encouraged to use different study designs to assess unintended (e.g. adverse) effects compared with intended effects?

5. Confounding and meta-analysis (session 2)

The risk of confounding by selection and other biases is a key difference between randomised and NRS. The NRSMG recommends that authors of reviews assess the risk of bias from residual confounding in primary studies (Reeves et al., 2008); tools can help reviewers to work through multiple considerations (Palmer et al., 2011; MacLennan et al., 2011; Wells et al., 2013).

The term NRS embraces a wide variety of between-group and within-group designs (Reeves et al., 2008; Higgins et al., 2012). However, it is unknown for which types of study the potential for residual confounding by selection, temporal or other biases is generally greater or generally less. Selection biases in NRS evaluating the effectiveness of interventions are often major and can be unpredictable in both magnitude and direction (Deeks et al., 2003). Moreover, regression adjustment using measured risk factors does not usually fully correct for the biases (Deeks et al., 2003). Thus, any ‘allowance’ for bias has to be uncertain and should widen the confidence
interval in a summary meta-analysis. How this can be carried out in practice is unclear because the uncertainty is epistemic rather than statistical (Valentine and Thompson, 2012).

The NRSMG has argued that a review protocol should pre-specify important confounders (Reeves et al., 2008), ideally ranked in order of the risk of them introducing confounding. Compiling a single list for a review may be difficult, not least because the list of confounders or the rank order may vary across the outcomes of interest. Some confounders underlying selection bias are likely to be measured imprecisely, unmeasured, or even unmeasurable. Moreover, primary researchers may adjust for different confounders and present their results in different ways. What data about confounders and adjustment methods should be extracted from the primary studies by reviewers? Some individual studies may present adjustments for alternative sets of confounders, which may be useful in judging the possible biases in other studies but may also make it difficult for review authors to decide which analysis should be considered ‘primary’ and extracted for synthesis.

Questions distributed to participants in advance for consideration during the session:

1. Because selection biases in NRS are often major, difficult to control and difficult to assess, should review authors avoid drawing definitive conclusions from meta-analyses that include NRS? Does this apply both to beneficial effects and harms?
2. Are cut-offs based on perceived risk of bias sensible in meta-analyses of NRS? If yes, how should cut-offs be defined? Are there better strategies for analysis, such as sensitivity analyses, in which case, how should results be presented and interpreted?
3. Should the varying degree of adjustment for confounders in different studies simply be regarded as a source of heterogeneity? How else could the varying adjustments in different studies be dealt with?
4. What is the role of judgement as opposed to empirical evidence in judging the effect of biases? How should the extra uncertainty inherent in using potentially biased NRS in a meta-analysis be represented?

6. Selective reporting (session 3)

Researchers carrying out primary studies may report their findings selectively, that is, may choose to report findings based on the results. This process can operate (i) at study level, causing publication bias, or (ii) within studies, with either selective reporting of individual outcomes or selective reporting of analyses from multiple analyses carried out.

Publication bias and selective outcome reporting have been researched in detail and are well-established phenomena (Dwan et al., 2008; Dickersin and Chalmers, 2010). The paradigm for doing such research is to identify an inception cohort of studies, for example, from a funding agency or ethics review board, and compare protocols submitted at the outset (and amendments made during the course of a trial) with subsequent publications.

Publication bias refers to the situation in which studies generating statistically significant results are more likely to be published than studies that do not. A systematic review of data from five inception cohorts (registered clinical trials) found that studies with ‘positive’ findings were about twice as likely to be published as studies with ‘negative’ or ‘null’ results (Hopewell et al., 2009). Publication bias can be more subtle, with positive studies tending to be published more quickly, in journals with higher ‘impact’ and in English (Dickersin and Chalmers, 2010).

Bias from selective outcome reporting in individual studies can be of three basic types: (i) selective (i.e. incomplete) reporting of some of the study outcomes; (ii) selective reporting of a specific outcome; and (iii) incomplete reporting of a specific outcome (Kirkham et al., 2010). Researchers have reviewed cohorts of studies, defined through submissions to funding or regulatory bodies, and compared published reports of findings with descriptions of outcomes in the protocols (Dwan et al., 2008; Dickersin and Chalmers, 2010). In a systematic review of five cohorts, four of which included only RCTs; Dwan and colleagues noted that statistically significant outcomes had a higher odds of being fully reported compared with non-significant outcomes (range of OR, 2.2 to 4.7) and that changes in pre-specified outcomes (not documented in amendments) also frequently occur (Dwan et al., 2008). The impact of bias from selective outcome reporting on the conclusions of Cochrane systematic reviews has also recently been described (Kirkham et al., 2010).]

Selective reporting of analyses has not been researched, although through personal communication with the first author, several analysts have confirmed that they suspect that this is a potentially serious issue, both for RCTs and NRS. Selective reporting of analyses is likely to be easier to identify in RCTs than NRS because summary details may be described when trials are registered, and pre-specified analysis plans tend to be more detailed and the analyses themselves less complex (Norris et al., 2012).

Assessing selective reporting in systematic reviews that include NRS is likely to cause challenges for authors that are more difficult to overcome than for RCTs. Authors may be less likely to obtain responses from researchers for NRS than for RCTs when clarification is required because the infrastructure and audit trails for NRS are less established (for example, with respect to standard operating procedures, study documentation, data management, archiving, and so on). It is less likely that protocols are available (or existed) for NRS than for RCTs. Even when protocols are available, the timing of the protocol with respect to the initial analyses is impossible to establish except for studies that involve prospective data collection or a registered and dated release of the data to the analyst.
Questions distributed to participants in advance for consideration during the session:

1. NRS are unlikely to be registered or to have protocols published in advance of reporting findings. When including NRS in a systematic review, how far should review authors be asked to go when judging the risk of outcome reporting bias (a standard bias domain, see Section 8.13.2 in Higgins et al., 2008)? Are there simple proxy measures of the risk of selective reporting that review authors could attempt to collect, for example, evidence of a protocol or submission for ethics review?

2. Is analysis reporting bias a serious concern? If yes, how might review authors detect it and assess the threat that it poses? Is this an issue that should extend across both RCTs and NRS?

3. NRS are more likely to be included when reviewing the harms of an intervention than the benefits. What is the likelihood that selective reporting (from publication bias, outcome reporting bias and analysis reporting bias) operates differentially for harms compared with benefits? Is selective reporting likely to operate differentially with respect to other aspects of outcomes, for example, long term versus short term?

4. Are the main reasons for publication bias (decision by researcher not to submit, rejection by journal, others) likely to vary for NRS compared with RCTs?

7. Applicability (session 4)

A key output of the members of the Cochrane Collaboration’s Applicability and Recommendations Methods Group is the summary of findings (SoF) table, which review authors are now strongly encouraged to include as part of Cochrane systematic reviews (Schünemann et al., 2008). SoF tables present the key question, including the intervention, the comparator, the key outcomes together with the results of the review and an evaluation of the quality of the body of evidence, preferably using the GRADE system (Guyatt et al., 2008).

To help users of systematic review make decisions, outcomes for both benefits and harms should be included in the SoF table. This requirement means that evidence from NRS may frequently be required to inform decision making and that review authors may have to rely on NRS for some outcomes. The assessment of the quality of evidence for an outcome in an SoF is intended to be broad, including an evaluation of how directly the question posed by the review is addressed by the evidence reviewed. This directness question is frequently informed by NRS evidence, which is widely considered to be better able to inform whether the intervention can be applied in the real world.

These considerations lead to the need to decide how data from NRS should be presented in SoF tables and what detailed criteria and evidence should be used to evaluate whether the population, intervention, comparator and outcome components of a study (NRS or RCT) represent a sufficiently direct test of the review question.

Particular challenges may arise when evidence from both RCTs and NRS is available for a particular outcome. Review authors are prompted to look for reasons to explain the different findings when RCT and NRS evidence is contradictory. However, the main explanations for such contradictions (i.e. differential bias by study design or substantive differences in the research question) can also cause RCTs and NRS to agree, with the different explanations shifting the effect estimate in opposite directions (MacLehose et al., 2000). Therefore, review authors should examine differences in the contexts of RCT and NRS evidence with equal care to avoid overlooking potentially important determinants of the effectiveness of an intervention in the real world (Schünemann et al., 2013).

Questions distributed to participants in advance for consideration during the session:

1. How should data from observational studies be presented in SoF tables (e.g. how should columns for case control studies look)?

2. The assumptions in studies addressing adverse events are based on the principle that the exposure-outcome association is directly linked to the assumption that removal of the exposure leads to avoidance of the outcome. What are the situations and criteria that decrease our confidence that an exposure-outcome association can be linked to a ‘removal of an exposure–avoidance of outcome’ relationship?

3. What are the criteria in the categories ‘population, intervention, comparison and outcomes’ that review authors should consider when assessing the applicability/ directness of the evidence? How can a metric for assessing the degree of lack of confidence in the criteria be developed that will enable review authors to make these judgements?

4. What efforts should review authors make to review information about the implementability of interventions?

5. What should users of reviews do if the quality of RCTs and NRS is similar but the results are contradictory (e.g. further downgrade the quality)? Preferring one answer over another represents a judgement about the risk of bias and likely applicability to a particular situation.

8. Summary and guide to the subsequent papers in the series

The four papers describing each of the main sessions are structured in the same way. First, the key methodological issues are introduced. Then, the practical consequences of these issues are discussed. Finally, the authors
summarise guidance about the issues, where there was majority consensus at the workshop and set out research priorities where participants disagreed or felt there was an urgent need for evidence.

In the final paper (Wells et al., 2013), the guidance from the preceding four papers has been formulated into lists of issues for review authors to consider. Together with tools developed by others, e.g. the Agency for Healthcare Research and Quality, these lists are providing the starting point for a research project funded by the Methods Innovation fund of the Cochrane Collaboration to extend the existing risk of bias tool (Higgins et al., 2008) to include NRS and crossover and cluster RCTs.

The workshop sessions primarily addressed issues relating to the conduct of a review. In doing so, the need for careful planning of a review that includes NRS was a further issue that participants often volunteered. Including NRS in a review requires review authors to make many, often complex, decisions that would not arise in a review of RCTs only. Therefore, the protocol is likely to be even more important for a review that includes NRS. It is important that the review authors specify in their protocol how these decisions will be made and, where this is not possible, describe transparently in the review how the decisions were made subsequently. Otherwise, there is a danger that a review will attract many of the criticisms levelled against primary NRS.

Appendix A. Ottawa Non-Randomised Studies Workshop Group

<table>
<thead>
<tr>
<th>Participant</th>
<th>Primary affiliation</th>
<th>Other affiliations</th>
<th>Role in meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laurie Anderson</td>
<td>Washington State Institute for Public Policy</td>
<td>Cochrane Public Health Group; Campbell Social Welfare Coordinating Group</td>
<td>Rapporteur, session 3; chair, session 4</td>
</tr>
<tr>
<td>Maria Benkhalti</td>
<td>Centre for Global Health, University of Ottawa</td>
<td>Campbell and Cochrane Equity Methods Group</td>
<td>Facilitator</td>
</tr>
<tr>
<td>Josip Car</td>
<td>Global eHealth Unit, Imperial College London</td>
<td>Cochrane Consumers and Communication Group</td>
<td>Chair, session 2</td>
</tr>
<tr>
<td>Stephanie Chang</td>
<td>Agency for Healthcare Quality and Research</td>
<td></td>
<td>Chair, session 1</td>
</tr>
<tr>
<td>Fabrizio Faggiano</td>
<td>Department of Clinical and Experimental Medicine, Università del Piemonte Orientale Amedeo Avogadro</td>
<td>Cochrane Drugs and Alcohol Group</td>
<td>Rapporteur, session 2; rapporteur, session 3</td>
</tr>
<tr>
<td>John Fletcher</td>
<td>Canadian Medical Association Journal, Ottawa</td>
<td></td>
<td>Rapporteur, session 1</td>
</tr>
<tr>
<td>Sarah Garner</td>
<td>National Institute for Health and Clinical Excellence</td>
<td>Cochrane Skin Group</td>
<td>Rapporteur, session 2; chair, session 3</td>
</tr>
<tr>
<td>Elizabeth Ghogomu</td>
<td>Cochrane Musculoskeletal Group, University of Ottawa</td>
<td></td>
<td>Facilitator</td>
</tr>
<tr>
<td>Jeremy Grimshaw</td>
<td>Centre for Best Practices, Institute of Population Health, University of Ottawa</td>
<td>Cochrane Effective Care and Organisation of Care Group</td>
<td>Chair, session 1; rapporteur, session 4</td>
</tr>
<tr>
<td>Mark Helfand</td>
<td>Oregon Evidence-based Practice Center, Oregon Health and Science University</td>
<td>Cochrane Incontinence Group</td>
<td>Lead, session 4</td>
</tr>
<tr>
<td>Julian Higgins</td>
<td>MRC Biostatistics Unit, Cambridge, UK</td>
<td>Cochrane Methods Board; Cochrane Statistical Methods Group</td>
<td>Lead, session 1; rapporteur, session 2</td>
</tr>
<tr>
<td>David Henry</td>
<td>Institute of Clinical and Evaluative Sciences</td>
<td>Cochrane Injuries Group</td>
<td>Rapporteur, session 1</td>
</tr>
<tr>
<td>Yoon Loke</td>
<td>Medical School, University of East Anglia</td>
<td>Cochrane Adverse Effects Methods Group</td>
<td>Rapporteur, session 3</td>
</tr>
<tr>
<td>Lara Maxwell</td>
<td>Cochrane Musculoskeletal Group, University of Ottawa</td>
<td>Cochrane Musculoskeletal Group</td>
<td>Facilitator</td>
</tr>
</tbody>
</table>

(Continues)
<table>
<thead>
<tr>
<th>Participant</th>
<th>Primary affiliation</th>
<th>Other affiliations</th>
<th>Role in meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al Mayhew</td>
<td>Cochrane Effective Care and Organisation of Care Group, University of Ottawa</td>
<td></td>
<td>Facilitator</td>
</tr>
<tr>
<td>Susan Norris</td>
<td>Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University</td>
<td>Cochrane Metabolic and Endocrine Disorders Group</td>
<td>Lead, session 3</td>
</tr>
<tr>
<td>Steve Phurrough</td>
<td>Center for Outcomes and Effectiveness, Agency for Healthcare Research and Quality (AHRQ)</td>
<td></td>
<td>Chair, session 3; rapporteur, session 4</td>
</tr>
<tr>
<td>Craig Ramsay</td>
<td>School of Medicine and Dentistry, University of Aberdeen</td>
<td>Cochrane Effective Practice and Organisation of Care Group</td>
<td>Lead, session 1; chair, session 2</td>
</tr>
<tr>
<td>Barney Reeves</td>
<td>School of Clinical Sciences, University of Bristol</td>
<td>Cochrane Non-Randomised Studies Methods Group</td>
<td>Overall workshop lead</td>
</tr>
<tr>
<td>Holger Schünemann</td>
<td>Department of Clinical Epidemiology and Biostatistics, McMaster University</td>
<td>Cochrane Applicability and Recommendations Methods Group; GRADE Working Group</td>
<td>Lead, session 4</td>
</tr>
<tr>
<td>Bev Shea</td>
<td>Community Information Epidemiological Technologies and Network Environment for Aboriginal Health Research, Canada</td>
<td>Cochrane Musculoskeletal Group</td>
<td>Chair, session 3</td>
</tr>
<tr>
<td>Jean Slutsky</td>
<td>Center for Outcomes and Effectiveness, Agency for Healthcare Research and Quality (AHRQ)</td>
<td></td>
<td>Chair, session 4</td>
</tr>
<tr>
<td>Simon Thompson</td>
<td>Department of Public Health, University of Cambridge, UK</td>
<td></td>
<td>Lead, session 2</td>
</tr>
<tr>
<td>David Tovey</td>
<td>Cochrane Editorial Unit, London UK</td>
<td>Editor in Chief of The Cochrane Library</td>
<td>Chair, session 4</td>
</tr>
<tr>
<td>Peter Tugwell</td>
<td>Centre for Global Health, Institute of Population Health, University of Ottawa</td>
<td>Cochrane Musculoskeletal Group; Campbell and Cochrane Equity Methods Group</td>
<td>Chair, session 1</td>
</tr>
<tr>
<td>Erin Ueffing</td>
<td>Institute of Population Health, University of Ottawa</td>
<td>Campbell and Cochrane Equity Methods Group</td>
<td>Facilitator</td>
</tr>
<tr>
<td>Jan Vandenbroucke</td>
<td>Department of Clinical Epidemiology, Leiden University Hospital</td>
<td>Cochrane Metabolic and Endocrine Disorders Group</td>
<td>Chair, session 2</td>
</tr>
<tr>
<td>Jeff Valentine</td>
<td>College of Education and Human Development, University of Louisville</td>
<td>Campbell Training Group</td>
<td>Lead, session 2</td>
</tr>
<tr>
<td>Hugh Waddington</td>
<td>International Initiative for Impact Evaluation</td>
<td>Campbell International Development Coordinating Group</td>
<td>Rapporteur, session 1; rapporteur, session 4</td>
</tr>
<tr>
<td>Vivian Welch</td>
<td>Centre for Global Health, Institute of Population Health, University of Ottawa</td>
<td>Campbell and Cochrane Equity Methods Group; Cochrane Musculoskeletal Group</td>
<td>Facilitator</td>
</tr>
<tr>
<td>George Wells</td>
<td>Cardiovascular Research Methods Centre, University of Ottawa Heart Institute</td>
<td>Cochrane Non-Randomised Studies Methods Group; Cochrane Musculoskeletal Group</td>
<td>Overall workshop lead</td>
</tr>
</tbody>
</table>
Acknowledgements

We are grateful to the following:

- All of the workshop participants (Appendix A), all of whom contributed to the discussions that provided the foundation for this paper.
- Sir Iain Chalmers, for insights about the potential benefits and harms of publishing a systematic review.
- Members of the Cochrane Non-Randomised Studies Methods Group, whose experiences of conducting or reviewing NRS have informed the Group’s guidance.

The Workshop was supported financially by the Agency for Healthcare Quality and Research (through the Ottawa Collaborating Centre of the Agency for Healthcare Quality and Research) and by a grant from the Cochrane Collaboration Discretionary Fund. BCR is supported in part by the UK National Institute for Health Research Bristol Cardiovascular Biomedical Research Unit. JPTH is supported by MRC Grant U105285807. The Health Services Research Unit is funded by the Scottish Government Executive Health Department. The views expressed in this article are those of the authors and not necessarily those of the funding bodies, The Cochrane Collaboration or its registered entities, committees or working groups or The Campbell Collaboration.

Boxes

1. Examples of selective outcome reporting
2. Difficulty of distinguishing between effectiveness and aetiological questions
3. Importance of specifying in a protocol whether review questions are confirmatory or exploratory

Disclosures

Some authors contribute to the Cochrane (BCR, JPTH, CR, BS, PT, GW) and Campbell Collaborations (PT, GW). BCR is funded in part by the National Institute for Health Research Bristol Cardiovascular Biomedical Research Unit.

References


