Cochrane Methods

Cochrane Scientific Committee

AGENDA

18th October 2017

11.00 UK BST

Teleconference

Cochrane Scientific Committee

Wed, Oct 18, 2017 11:00 AM - 1:00 PM BST

Please join my meeting from your computer, tablet or smartphone.
https://global.gotomeeting.com/join/995725261

You can also dial in using your phone.
United States (Toll Free): 1 866 899 4679
United Kingdom: +44 330 221 0097

Access Code: 995-725-261

More phone numbers
Australia (Toll Free): 1 800 191 358
Chile (Toll Free): 800 395 146
Denmark (Toll Free): 8025 3112
France (Toll Free): 0 800 919 896
Germany (Toll Free): 0 800 184 4230
Netherlands (Toll Free): 0 800 020 0178
United Kingdom (Toll Free): 0 800 389 5276

First GoToMeeting? Try a test session: https://care.citrixonline.com/g2m/getready

AGENDA

DT – David Tovey, Editor in Chief, JC – Jackie Chandler, Methods Co-ordinator

Committee members:

Corinna Dressler (CD)
Research Associate at the Division of Evidence-Based Medicine (dEBM) at the Charité – Universitätsmedizin Berlin, Germany

Donna Gilles (DG)
Senior Researcher, Clinical Performance Mental Health Network, Western Sydney, Australia and editor for both the Cochrane Developmental, Psychosocial and Learning Problems Group and Diagnostic Test Accuracy Review Group.

Julian Higgins (JH)
Professor of Evidence Synthesis at the School of Social and Community Medicine, at the University of Bristol, Bristol, UK, and current Senior Scientific Editor of the *Cochrane Handbook of Systematic Reviews for Interventions.*

AsbjØrn Hróbjartsson (AH)
Professor of Evidence-Based Medicine and Clinical Research Methodology at the University of Southern Denmark, and Head of Research for the Center for Evidence-Based Medicine at Odense University Hospital, which hosts the secretariat of the Cochrane Bias Methods Group.

Ana Marusic (AM)
Professor of Anatomy and Chair of the Department of Research in Biomedicine and Health at the University of Split School of Medicine, Split, Croatia and founder of Cochrane Croatia.

Jane Noyes (JN)
Professor of Health and Social Services Research and Child Health, Bangor University, Wales, UK, lead Convenor of the Cochrane Qualitative and Implementation Methods Group, and a UK Cochrane Fellow.

Tomas Pantoja (TP)
Associate Professor, Family Medicine Department, School of Medicine, Pontificia Universidad Católica de Chile and Editor of the Cochrane Effective Practice and Organisation of Care (EPOC) Group.

Philippe Ravaud (PR)
Professor of Epidemiology, Faculty of Medicine, Head of the Clinical Epidemiology Centre, Hôtel-Dieu Hospital, Paris Descartes University, France and Director of Cochrane France.

Johannes Reistma (JR)
Associate Professor at the Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands and a member of both the Cochrane Diagnostic Test Accuracy Working Group and the Screening and Diagnostic Tests Methods Group.

Rebecca Ryan (RR)
Research Fellow at the School of Psychology and Public Health, La Trobe University, Australia and Deputy Co-ordinating Editor of the Cochrane Consumers and Communication Group.

Christopher Schmid (CS)
Professor of Biostatistics, founding member and Co-Director of the Center for Evidence Synthesis in Health, Brown School of Public Health, US, Fellow of the American Statistical Association (ASA) and Founding Co-Editor of *Research Synthesis Methods*.

Nicole Skoetz (NS)
Scientific Co-ordinator, Working Group Standard Operating Procedures of the Comprehensive Cancer Centers, Center of Integrative Oncology Köln Bonn, and Co-ordinating Editor Cochrane Haematological Malignancies Group, Department of Internal Medicine, University Hospital of Cologne.

Nichole Taske (NT)
Associate Director (Methodology), Centre for Guidelines, NICE, U

Chairs Philippe (PR) and Ana (AM) will jointly chair the meeting

AGENDA ITEM	Details and links to documents	Responsibility for item
1) Welcome and apologies received (Agree agenda and items for A.o.B)		AM & JC
2) Approval of previous minutes	Minutes dated 18th May 2017 **PAPER 1 & Statements**	AM
a) Matters arising	Feedback on Expert panel on cumulative meta-analyses	AM & JC
3) CSC Business matters	Clarifying role of CSC to the wider Cochrane Community	DT
4) Submissions	First call and submission list **PAPER 2** For discussion and priority for further work up for future agendas	PR
5) Methods for CSC Review	Follow up comments for ROBINS I	PR
6) Methods for CSC sign off and recommendation	Follow up comments for RoB 2.0	PR
7) Special items		
a) Research priorities and strategy	(i) Developing future agendas (ii) Future agenda items for consideration and prioritisation: 1. Intervention Complexity Assessment tool 2. Guidance for when to include Clinical Study Reports and other regularity data in CR's 3. Methods for prognosis reviews and full roll. 4. Methods for addressing missing participant data 5. Assessing the quality of evidence and presenting the results of Non-randomised Studies in CR's 6. Evaluation and validation of the RCT classifier	AM
8) Any Other Business		AM
9) Meeting schedule	List of meetings: 28th February 2018 @ 8pm UK GMT	JC

Cochrane Scientific Committee

Teleconference 18th May 2017

Notes and abbreviations
EiC – Editor in Chief
CSC – Cochrane Scientific Committee
CRG – Cochrane Review Group
MG – Methods Group

Members of the CSC

Corinna Dressler (CD)	Present
Donna Gilles (DG)	Present
Julian Higgins (JH)	Present
Asbjørn Hróbjartsson (AH)	Present
Ana Marušić (AM)	Present
Jane Noyes (JN)	Present
Tomas Pantoja (TP)	Present
Philippe Ravaud (PR)	Present
Johannes Reistma (JR)	Present
Rebecca Ryan (RR)	Present
Christopher Schmid (CS)	Present
Nicole Skoetz (NS)	Present
Nichole Taske (NT)	Present
David Tovey (DT)	Present

Other attendees	
Louisa Dunn (LD)	Minutes
Hilary Simmonds (HS)	Minutes
Jackie Chandler (JC)	Apology
Jonathan Sterne (JS)	Invited speaker
Christian Gluud (CG)	Invited speaker
Mark Simmonds (MS)	Invited speaker

1. **Welcome and introduction**
   All members introduced themselves. CSC members agreed PR undertake the role of Chair and would do so at subsequent meetings. DT chaired this meeting stepping in on agenda items due to JC's absence. CSC members agreed two Co-chairs were preferable and hoped to identify another chair from the membership before the next meeting.

   ACTION: DT and PR to identify another Chair.

   All members would complete a 3-year cycle, after which terms of office are staggered to allow continuity up to a maximum of a 5 year term.

ACTION: A reminder to all members to ensure all declaration of interest forms are completed.

2. Approval of previous minutes and matters arising

None

3. CSC Business matters

3.1 Current governance arrangements and where the CSC is positioned

DT outlined the rationale in setting up the CSC within the wider framework of Cochrane's organisational structure. An expert body is required to advise the EiC on methods for use in Cochrane Reviews. The CSC will work with the Governing Board, the new Editorial Board, Cochrane Council, CRGs and MGs to ensure that Cochrane implements best available methods. This involves working strategically to review methodological issues on the horizon. Implementation of methods is a further consideration, but not the responsibility of the CSC. Although, members are asked to consider feasibility and implications of methods recommended. CSC could propose piloting of methods before wholesale implementation. In depth discussions on methods that need additional time can use email discussion or task based (subcommittee) groups.

3.2 Draft terms and conditions

These were agreed. These will adapt with experience.

3.3 Processes for submissions and review

Templates were approved. Their clarity, with clear questions, was appreciated. These will develop as we gain experience with using them.

4. Submissions

Options for submissions can come from CSC members along with an open call inviting submissions from the wider Cochrane community. NS has already provided a list of items of interest including agenda item cumulative meta-analyses. CSC members will need to agree how to manage and prioritise agenda items. The Co-chairs (PR and nominated Co-chair), DT and JC will manage call items initially, forming a waiting list if necessary. Efficient implementation will be aided by methods gaining a CSC recommendation. Such decisions will need to follow through with the *Handbook* editors. JH raised the point that the *Handbook* editors might identify methods that will need CSC approval before publication. This also requires a 'judgement' on their part as to what method or update of a method would need CSC approval. Members discussed conflicts of interests on submitted items, noting that they will recuse themselves from the decision making but not necessarily the discussion.

DT clarified that at this point most enhancements to methods would, unless sufficiently uncontested, come to the CSC for approval. He further elaborated the CSC was an important gateway to implementation, however, the threshold on items for consideration may rise with committee experience.

5. Methods for CSC Review

5.1 Review of the development of the risk of bias tool for non-randomised studies for interventions – ROBINS-I: Presented by Jonathan Sterne – slides attached

JN raised the concern regarding implementation of this tool and required expertise. JH and JS agreed it needs content and epidemiological expertise. They were also working with Co-Ed Paul Garner to develop a triage tool to assist implementation. This will assist with decision making as to whether to apply assessment of all domains to all studies, if studies are assessed early as high risk of bias. JR asked about the availability of formal piloting results. JS reported piloting feedback was iteratively incorporated into tool and that a formal evaluation report is not available. There was support for the tool and the web version underway welcomed. Future discussions on the harmonisation of assessment tools was also proposed. RR requested guidance on competency to use the tool.
DECISION: The CSC (using GTM chat to confirm) agreed a recommendation that the ROBINS-I tool should be the preferred tool when including nonrandomised studies in Cochrane Reviews. The tool however, is not mandated and other tools could be used in keeping with *Handbook* guidance, such as the Newcastle Ottawa Scale.
ACTION: JS will work on a form of words to advise on review team competency to undertake use of the tool. Wider implementation awaits completion of the web tool. Further guidance will assist with defining the 'Target Trial'.

5.2 Review of approaches to cumulative meta-analyses for systematic reviews: Presented by Christian Gluud and Mark Simmonds – CG slides attached/MS slides attached.
In summary, there are two distinct approaches to choose from Trials Sequential Analysis and Sequential Meta-Analysis which both address Type I and II errors.
DECISION: The CSC propose an expert panel to peer review these methods and propose recommendations for consideration by the CSC.
ACTION: Members to provide names of potential panellists.

6. Methods for CSC sign off and recommendation

6.1 Review of the updated 'Risk of bias' tool RoB 2.0.
Updates to the revised tool were outlined as in keeping with the ROBINS-I development. Not quite ready for implementation as RevMan needs updating to allow authors access to the tool. Members were positive about the addition of signalling questions.
DECISION: CSC agreed to recommend the mandatory implementation of RoB 2 for new reviews when it is ready for implementation. It will not be applied retrospectively to old reviews and updates.
ACTION: JS & JH to notify Cochrane when final version completed. Co-ordinating Editors need to develop a strategy for its implementation.

7. Special items

No further discussion meeting ran over and closed.

8. Any Other Business

None

9. Meeting schedule

Additional meetings to be scheduled with an increase length of meeting.

Cochrane Scientific Committee

Recommendation statement/report

Date: July 2017

Relates to agenda item and meeting reference: 5ii 18th May 2017

Priority: **Medium**

Open access/restricted: **Open**

Review of approaches to cumulative meta-analyses for systematic reviews

Lead developers/investigators: Christian Gluud, Jørn Wetterslev, Julian Higgins, Mark Simmonds and many other colleagues

Abstract:

The problem

The CSC were asked to consider whether methods are required to manage the occurrence of both Type I and Type II errors in cumulative meta-analyses. If so, which of the proposed methods should Cochrane use.

Type I error: Repeatedly updating meta-analyses to incorporate more studies leads to the probability of type I error occurring, that is the false conclusion that an intervention has an effect when it does not (false positive). False positive results can occur due either to systematic errors, or random errors due to repeat testing.

Type II error: False negative results can occur when assuming there is no benefit before the meta-analysis has reached a sufficiently powered information size (sample size).

Summary

Julian Higgins introduced sequential approaches for meta-analyses to Cochrane at the Rome Colloquium in 1999, based on previous work by Anne Whitehead. This led to a publication in 2011 reporting a simulation study comparing six approaches and providing a worked example for **"Sequential methods for random-effects meta-analysis"**. The Higgins and colleagues' approach uses an approximate semi-bayes procedure to update evidence on the among study variance, starting with an informative prior distribution possibly based on findings from previous meta-analyses. Other work led by Jørn Wetterslev, Christian Gluud and colleagues (2005, 2008, 2013) uses **"Trial Sequential Analysis in Systematic Reviews with meta-analysis" (TSA).** This work received the Thomas Chalmers award for a Cochrane Colloquium abstract. TSA is akin to the process for assessing interim analyses in trials to see whether a large enough effect (benefit) is achieved warranting trial discontinuation (stopping rules). They extend the method and test on six randomly selected meta-analyses. An important aspect to their work is the assumption that 'information size', the total number of participants across all included trials in a meta-analysis, is usually underpowered. So, they argue these MA's represent interim analyses rather than an endpoint. They suggest that this information size (when MA is underpowered), heterogeneity across studies, and bias assessment are used to provide an adjustment to the naïve 95% confidence intervals and 5% thresholds for statistical significance in meta-analysis. The Lan-DeMets' sequential monitoring boundaries in TSA provide the adjusted, expanded confidence

intervals and adjusted restrictive thresholds for statistical significance before the diversity-adjusted required information size is reached.

In 2012, *Cochrane Methods* published a discussion between Higgins on one hand and Jørn Wetterslev, Christian Gluud and colleagues on the other as to the issues raised by these methodological developments. See extract from Cochrane Methods (2012) attached.

Additional work under investigation is Shuster and Neu (2013) **"Pocock approach to sequential meta-analysis of clinical trials"** and Hu and colleagues (2007) **"Applying the law of iterated logarithm to control type I error in cumulative met-analysis of binary outcomes"**. These study reports are simulation studies with worked examples. These key approaches, are evaluated in a Cochrane funded (Methods Innovation Fund) research project led by Mark Simmonds, York University, UK. We expect this work to complete in 2018 and the CSC will receive an interim report on this work.

The documentation list provides references to these key studies and other relevant work. Methodologists do not yet agree on the approach, although they agree the principle problem of the increased probability of rejection of the null hypothesis on repeated meta-analysis and the problems with early results before the meta-analysis has reached a sufficiently powered information size. There is a mix of caution (methods not ready) and pragmatism (problem needs addressing now). Methodologists suggest Bayesian meta-analysis shows some promise (Spence et al, 2016), however, several issues need resolving, including access to software and methodological expertise.

The table below highlights some issues from key references.

Questions:

- Is the problem with too little power in most meta-analysis when a required information is not reached with false positive support for the null hypothesis a sufficient problem that undermines the evidence produced by Cochrane reviews?
- Is the problem of false positive meta-analytic conclusions due to random error introduced by underpowered meta-analysis and the probability of repeated analyses rejecting the null hypothesis a sufficient problem that undermines the evidence produced by Cochrane Reviews?
- Is the current state of development for adjustment in cumulative meta-analyses to address, specifically, type II and type I errors sufficient to recommend their implementation in Cochrane Reviews?
- If so, can the CSC recommend one or more techniques?
- If not, what further knowledge or development does the CSC need to reach a satisfactory point to decide?

Critique	By who	Reference
Sequential approaches encourage the use of significance tests and the inappropriate division of results as 'significant' or 'not significant' rather than the direct interpretation of intervention effect estimates and corresponding confidence intervals.	Higgins	*Cochrane Methods* (2012) P32-33
Problem of creating inappropriate 'stopping rules' in MA.	Higgins	*Cochrane Methods* (2012) P32-33

Measurement of accumulated information: • The sum of the study weights in the meta-analysis. (Higgins) • Numbers of participants (Wetterslev et al.) is less sensible because the sample size needs to convert into statistical information for the analyses, and the conversion requires the additional prespecification not only of quantities such as the control group risk for dichotomous data but also of the anticipated amount of heterogeneity when a random effects meta-analysis is planned.	Higgins	*Cochrane Methods* (2012) P32-33
Sequential methods should be applied prospectively with a full analysis plan in the protocol.	Higgins	*Cochrane Methods* (2012) P32-33
Assumptions underlying the sequential design are clearly conveyed and justified, including the parameters determining the design such as the clinically important effect size, assumptions about heterogeneity, and both the type I and type II error rates.	Higgins	*Cochrane Methods* (2012) P32-33
Major disagreement lies in whether the use of the traditional significance level of 0.05 and unadjusted 95% confidence interval is valid in MAs where the available information has not yet reached a required information size. MA results should be interpreted in the light of a realistic required information size and therefore adjustments made to ensure appropriate inference.	Wetterslev & colleagues	*Cochrane Methods* (2012) P33-35.
Response to critique for transferring TSA methods to sequential analysis in MAs – MAs impact on decisions to continue to update or not based on the level of significance. Also, the traditional unadjusted confidence interval will represent a too narrow confidence interval which by chance does not include the null effect, and so the observed effect of the intervention may be misleading and premature.	Wetterslev & colleagues	Trial Sequential Analysis in systematic reviews with meta-analysis *BMC Medical Research Methodology* (2017) 17:39. See paper for further discussion on calculating the required information size.
To overcome the type I error inflation problem Hu et al propose a way to estimate and penalize the Z statistic using the law of iterated logarithm. The penalty to the Z statistic accounts for multiple tests in a cumulative meta-analysis of binary outcomes and, in addition, accounts for estimation of heterogeneity in treatment effects across studies and the unpredictable nature of information from clinical trials. It does not require the pre-specification of the maximum information.	Hu and colleagues	Applying the law of iterated logarithm to control type I error in cumulative meta-analysis of binary outcomes *Clinical Trials* (2007) 4:329-340.
In reference to methods developed by Wetterslev et al, Van der Tweel, and Bollen, and Higgins, Shuster & Neu state: None of these methods allow for the effect sizes to be dynamic. Random effects are drawn from the same conceptual urn from trial to	Shuster & Neu	A Pocock approach to sequential meta-analysis of clinical trials.

trial. These competitors to our methods reweight the relative contributions of the included trials after each trial is added. This violates the critical independent increment property. A potential shortcoming of all methods (including ours) lies in the lack of knowledge of the true information fraction (the ratio of the variance of the estimate at the final look presuming no stopping to that after the current look). 'Look' refers to the moment of meta-analysis in time – updating.		*Research synthesis Methods* (2013) 4 269-279. See paper for further explanation and methods proposed. Please see also further information in *Current controversies in data monitoring for clinical trials* (Pocock, 2006),

SUPPORTING DOCUMENTATION

[Extract] Wetterslev & colleagues and Higgins JP. Trial sequential analysis: Methods and software for cumulative meta-analyses. In Chandler J, Clarke M, Higgins JP, editors. Cochrane Methods, *Cochrane DB Syst Rev* 2012 Suppl 1:29-35.

Presentation providing an interim report on the evaluation of these methods by Mark Simmonds to the Methods symposium at the Seoul Colloquium 2016 on Living Systematic Reviews.

Relevant publications

Higgins, J P, A. Whitehead A, Simmonds M. (2011). Sequential methods for random-effects meta-analysis. *Stat Med* **30**(9): 903-921.

Hu M, Cappelleri JC, Lan KK (2007). Applying the law of iterated logarithm to control type I error in cumulative meta-analysis of binary outcomes. *Clin Trials* **4**(4): 329-340.

Imberger G, Gluud C, Boylan J, Wetterslev J.(2015). Systematic Reviews of Anesthesiologic Interventions Reported as Statistically Significant: Problems with Power, Precision, and Type 1 Error Protection. *Anesth Analg* **121**(6): 1611-1622.

Imberger G, Gluud , Boylan J, Wetterslev J. (2015). Systematic Reviews of Anesthesiologic Interventions Reported as Statistically Significant: Problems with Power, Precision, and Type 1 Error Protection. *Anesth Analg* **121**(6): 1611-1622.

Imberger G, Thorlund K, Gluud C, Wetterslev J. (2016). False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. *BMJ Open* **6**(8): e011890.

Jackson D, Turner R. (2017). Power analysis for random-effects meta-analysis. *Res Synth Methods.*

Mascha, E J. (2015). Alpha, Beta, Meta: Guidelines for Assessing Power and Type I Error in Meta-

Analyses. *Anesth Analg* **121**(6): 1430-1433.

Pereira, TV, Horwitz RI, Ioannidis JP.(2012). Empirical evaluation of very large treatment effects of medical interventions. *JAMA* **308**(16): 1676-1684.

Pocock SJ. (2006). Current controversies in data monitoring for clinical trials. *Clin Trials* **3**(6): 513-521.

Shuster, J. J. and J. Neu (2013). A Pocock approach to sequential meta-analysis of clinical trials. *Res Synth Methods* **4**(3): 269-279.

Spence GT, Steinsaltz D, Fanshawe TR. (2016). A Bayesian approach to sequential meta-analysis. *Stat Med* **35**(29): 5356-5375.

Thorlund, K, Devereaux PJ, Wetterslev J, Guyatt G, Ioannidis JP, Thabane L, Gluud LL, Als-Nielsen B,
Gluud C. (2009). Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? *Int J Epidemiol* **38**(1): 276-286.

Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, Guyatt G, Devereaux PJ, Thabane L.(2011). The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis--a simulation study. *PLoS One* **6**(10): e25491.

Thorlund K, Engstrøm J, Wetterslev J, Brok J, Imberger G, Gluud C. (2011) User Manual for Trials Sequential Analysis (TSA), Copenhagen Trial Unit, Centre for Clinical Intervention Research.

Turner R, Bird M, Higgins JP. (2013). The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. *PLoS One* **8**(3): e59202.

Wetterslev J, Jakobsen JC, Gluud C. (2017). Trial Sequential Analysis in systematic reviews with meta-analysis. *BMC Med Res Methodol* **17**(1): 39.

Wetterslev J, Thorlund K, Brok J, Gluud C. (2008). Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol* **61**(1): 64-75.

CSC RECOMMENDATION

☐ ***Highly recommended***

☐ ***Recommended with provisions***

☐ ***Optional/advisory (one among several options)***

☐ ***Not recommended***

☒ ***Further evaluation required***

The CSC agreed that further technical examination of the key approaches was required to ascertain whether there is a preferred method, or whether the methods provide value to

managing random error and are needed at all, or only in certain scenarios. An expert panel will be asked to consider the work completed by colleagues to date and will report to a future CSC.

CSC STATEMENT

Summary statement

Both presenters concurred over the problem of Type I and Type II random errors in repeated meta-analyses. These errors tend towards overestimation and the problem of leading to early false conclusions in meta-analysis. An appropriate information or sample size protects against this and was illustrated by Christian Guuld's simulation work. He presented a specific approach, Trial Sequential Analysis, to address these errors. Mark Simmonds, as part of the Cochrane Methods Innovation Fund project, presented the findings of simulation studies evaluating four methods (above). In summary, any possible benefit is derived from either Trial Sequential Analysis, or Sequential Meta-analysis with or without using the approximate Bayesian Heterogeneity. These approaches are applied haphazardly in Cochrane and DT would like advice on the best approach.

Key points to note are:

- Both TSA and SMA control for Type I error well in Mark Simmonds's simulations, except for a few trials with high heterogeneity. Therefore, approximate Bayesian Heterogeneity is not required in most circumstances.
- Many analyses showing positive significant results at the 0.5 or lower level are based on too little evidence. And most Cochrane Reviews are underpowered.
- There is no difference in TSA or SMA in controlling for Type II error.
- Other approaches that control Type I error only, are considered overly conservative and are not recommended by Mark Simmonds's work. In addition, controlling for Type I error impacts and lessens power for Type II.
- JH (SMA) summarized his key points:
    - There are two candidates for these methods and it is problematic to suggest we select one over the other at this point.
    - He suggested we should abandon significant testing rather than create methods to correct errors that occur in their use.
    - In addition, repeated confidence intervals are an area of statistical debate, however, these methods could be converted to address repeated confidence intervals
    - Possibly not ban use of these methods, but not encourage them either and explain that they are a comparison of two hypotheses in the traditional Neyman-Pearson paradigm in statistics.
- CS stated there was other work not included here with the use of the random effects model between study variance changing overtime and how that can be accounted for, so these methods should not be used with less than five studies because you will not have a good estimate of all parameters unless you use prior information.
- Following on from whether P Values should be used at all it was noted (JS) that both confidence intervals and P Values are derived from the effect size and the standard error. The problem is that the P Value decides an arbitrary cut point (0.5) into whether a result is positive or negative. The American Association guidelines state this is not scientific and does not have utility. Confidence intervals are not used to accept or reject the null hypothesis.

The discussion indicated that these methods were driven by both technical and theoretical issues that warranted greater examination with other experts in the field.

Credibility & validity

Further work required to assess utility of the approaches.

Limitations/caveats

This issue resides within current theoretical debates amongst statisticians.

Areas of concern/uncertainty

Unclear at this point.

Impact on Cochrane

Training and guidance and utility of method to make a difference.

Cochrane resources needed

None currently.

Implementation

CSC members are not responsible for managing implementation of these recommendations which will require an implementation plan to ensure co-ordination for a smooth introduction. This will include launch, timescales and roll out strategy. Therefore, this statement does not signify immediate implementation.

Cochrane Scientific Committee

Recommendation statement/report

Date: July 2017

Relates to agenda item and meeting reference: 6i 18th May 2017

Priority: Medium

Open access/restricted: Open

Review of the updated 'Risk of bias' tool RoB 2.0

Lead developers/investigators: Jonathan Sterne and Julian Higgins

Summary of development: Developers initiated revisions to the current tool based on work developing the ROBINS I tool. Development involved expert working groups (for different domains of bias and different trial designs) and consensus, with piloting of draft versions with Cochrane collaborators and revisions made. Higgins and colleagues (2016) describe 10 key changes to the original tool (2008, 2011). Please see extract from *Cochrane Methods* for summary and qualification of these changes. There is also a table that shows changes to the domain terminology between the current and new tools. Some of the key changes are (i) the assessment is at the level of a specific result (i.e. a specific comparison at a specific time point and using a specific statistical analysis); (ii) the assessment is specific to whether interest focusses on the effect of assignment to intervention or the effect of starting and adhering to intervention; (iii) the domain of selective outcome reporting has been re-focussed. As with the ROBINS I tool, signalling questions are introduced. The new tool also provides a procedure to reach an overall risk of bias. Finally, there are different templates for different trial designs.

There remain some outstanding issues. These are:

- How many results should be assessed for each study?
- How best can the assessment be integrated into the data extraction process, given that some relevant information is study-level, some is outcome-level and some is result-specific?

Developers have introduced the tool to Cochrane members at both the Seoul and Geneva meetings. They have yet to publish this development and thus undergo peer review.

Caveats: There is increased complexity and changes that impact on updating of reviews particularly with many included studies. Balancing the implementation demands might compromise methodological integrity when applying the RoB 2.0. Consideration therefore is given to allow both tools operate but not in the same review, including updates.

Impact: We expect the transition between tools may pose both practical and technical issues.

Resources needed: Software development is required and is important to facilitate easier transition. This includes the ecosystem of authoring tools e.g. Covidence and RevMan. Developers have developed algorithms to map responses to signalling questions to judgements about risk of bias. Training and methods support for implementation are needed, along with consideration of implementation issues.

SUPPORTING DOCUMENTATION

Higgins JPT, Sterne JAC, Savović J, Page MJ, Hróbjartsson A, Boutron I, Reeves B, Eldridge S. A revised tool for assessing risk of bias in randomized trials In: Chandler J, McKenzie J, Boutron I, Welch V (editors). Cochrane Methods. Cochrane Database of Systematic Reviews 2016, Issue 10 (Suppl 1). dx.doi.org/10.1002/14651858.CD201601.

The following table lists the tools and guidance for the different versions, please visit www.riskofbias.info.

Individually randomized, parallel group trials	1. Guidance for using the RoB 2.0 tool for individually randomized trials 2. The tool 3. Blank templates with two variants: a. RoB 2.0 when interest is in the effect of assignment to intervention b. RoB 2.0 when the interest is in the effect of starting and adhering to intervention
Cluster randomized, parallel group trials	1. Guidance for using the RoB 2.0 tool for cluster-randomized trials. 2. The tool (cluster-randomized trials) 3. Blank template with one variant a. RoB 2.0 for cluster randomized r trials when the interest is in the effect of starting and adhering to intervention.
Individually randomized, cross-over trials	1. Guidance for using the RoB 2.0 tool for cross-over trials 2. The tool (cross-over trials). 3. Blank templates with two variants:

	a. RoB 2.0 for cross-over trials when interest is in the effect of assignment to intervention b. RoB 2.0 for cross-over trials when the interest is in the effect of starting and adhering to intervention.

CSC RECOMMENDATION

- [x] ***Highly recommended***
  The is mandatory for new reviews when officially launched. For updates, it is not reasonable to re do previously included studies and a strategy is required to handle these situations.
- [ ] ***Recommended with provisions***
- [ ] ***Optional/advisory (one among several options)***
- [ ] ***Not recommended***

CSC STATEMENT

Summary statement

Members agreed the tool should be implemented. Although, one member raised the definitional difficulty in shifting from 'unclear' to 'serious concerns'. Further explanation was that, unclear covered two distinct points: (i) you cannot ascertain what happened to assess bias, or (ii) you know what happened but it is inadequate (unclear) to assess risk of bias. The new signalling questions will highlight where there is no information and the overall assessment allows a judgement to be made to inform the reader (e.g. serious concerns). The signalling questions are mapped to the risk of bias judgements. Another member of the CSC had applied the tool to fifty different kinds of studies successfully and welcomed the new version of the tool. Recent meetings presenting the tool to the Co-ordinating editors had not raised any issues of concern.

Credibility & validity

This tool has high credibility in its RoB 1.0 version and this version involves developments (signalling questions used in other validated tools (QUADAS 2))

Limitations/caveats

Implementation awaits some final adjustments to the tool and integration into RevMan requires further consideration. Also, implementation of the tool may reveal other issues.

Areas of concern/uncertainty

None specified

Impact on Cochrane

Minor as one tool replaces another for new reviews. There are issues for CRGs and the editorial unit to ensure its implementation when fully released.

Cochrane resources needed

Training, distribution of guidance and software development are key factors for implementation once the developers have produced a final version.

Implementation

CSC members are not responsible for managing implementation of these recommendations which will require an implementation plan to ensure co-ordination for a smooth introduction. This will include launch, timescales and roll out strategy. Therefore, this statement does not signify immediate implementation.

Cochrane Scientific Committee

Recommendation statement/report

Date: July 2017

Relates to agenda item and meeting reference: 5i 18th May 2017

Priority: Medium

Open access/restricted: Open

Review of the development of the risk of bias tool for nonrandomised studies for interventions – ROBINS-I

Lead developers/investigators: Jonathan Sterne and Julian Higgins, Barney Reeves, Jelena Savović and Lucy Turner

Abstract:

Aim & objective

The ROBINS-I tool evaluates the risk of bias (RoB) in the results of nonrandomized studies of interventions (NRSI) that compare the health effects of two or more interventions.

This tool evaluates NRSI that are quantitative studies estimating the effectiveness (harm or benefit) of an intervention, which did not use randomization to allocate units (individuals or clusters of individuals) to comparison groups. These are typically observational studies and include cohort studies, case-control studies, controlled before-and-after studies, interrupted-time-series studies and controlled trials in which intervention groups are allocated using a method that falls short of full randomization (sometimes called “quasi-randomized” studies).

Methods for development

Expert consensus using working groups covering the domains of bias followed the seven principles for assessing risk bias (Higgins et al, 2011). The procedure included a survey of Cochrane Review Groups about current tools used and follow up interviews on a piloted version of the tool to ascertain interpretation and use of guidance. Dissemination activity led to further modifications and the current version.

Results/Development

The tool continues the domain approach used in the current Cochrane ‘Risk of bias’ tool adding three assessment domains specifically related to NRSI: bias due to confounding, bias in selection of participants into the study pre-intervention and bias in classification at intervention. Signalling questions to aid assessor judgements are a key feature, adopted from the QUADAS-2 tool (Whiting et al, 2011). Evaluation commences with considering the target trial. This hypothetical trial provides the assessor with a ‘model’ comparator of a pragmatic randomized trial without the features putting it at risk of bias.

Final product: The currently-published ROBINS-I tool (Word and Access versions) is designed for cohort-like designs, such as cohort studies, quasi-randomized trials and other concurrently controlled studies. Although applicable for case-control studies, cross-sectional studies, interrupted time series and controlled before-after studies further developments to signalling questions are underway. A substantial guidance document is available to support application.

References:

Higgins JPT, Altman DG, Goetzche P, et al. (2011)The Cochrane Collaboration's tool for assessing risk of bias in randomised trials *BMJ*; 343:d5928

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM. (2011) QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*; 18;155(8):529-36.

SUPPORTING DOCUMENTATION

Stern JAC, Herńan MA, Reeves BC, Savoić J, Berkman ND, Viswanathan M et al. (2016) ROBINS-I: a tool for assessing risk of bias in non-randomised studies of intervention; BMJ 2016; 355@i4919. http://dx.doi.org/10.1136/bmj.i4919

Access to tool documentation please see www.riskofbias.info

CSC RECOMMENDATION

☐ ***Highly recommended***

☒ ***Recommended with provisions***
The ROBINS-I tool is recommended as the preferred tool for new reviews. It is not mandatory. The importance of competency to use the tool will be highlighted in guidance.

☐ ***Optional/advisory (one among several options)***

☐ ***Not recommended***

CSC STATEMENT

Summary statement

Jonathan Sterne elaborated the key features of the current version of the tool. This starts with specifying a hypothetical randomised trial based on PICO information drawn from the nonrandomised study. Key areas of bias that map onto key epidemiological terms are confounding, selection bias and misclassification bias, however, selective reporting bias does not have an epidemiological analogue and is dealt with separately. Risk judgements are *low, moderate, serious, critical* and *no information.* A nonrandomised study is most likely to make *moderate* at best. Web development is

underway and will permit question skipping. Versions adapted for case control, interrupted time series and before after designs are underway as certain aspects of each differ from the Cohort version. For example, for case control design confounding is the same, whereas selection bias differs because of the way the controls are selected.

CSC discussion focussed on implementation issues and acceptability amongst Co-eds. Opportunities for input into the development has enabled greater acceptance and led to a proposal to develop a triage approach whereby early identification of very low quality studies deemed to be critical, for example, could be removed from further assessment. However, it is not possible to radically simplify the tool and still conduct a proper assessment. Expertise in epidemiology is a key competency to undertake a review of nonrandomised studies. Further guidance to specify competence level needed is forthcoming. Some members were interested in some formal empirical testing of the tool. This is likely to occur over time when used in enough studies to warrant empirical research, permitting further improvements. Developments across similar risk of bias/quality assessment tools (DTA and Prognosis) stimulate feature changes such as the signalling questions, however, a suggestion was made to harmonise across all tools at some point.

The CSC agreed the ROBINS-I tool was the preferred tool to assess nonrandomised studies in Cochrane Reviews. It may be mandated at a later point after further evaluation and development. The development of the web version will assist with implementation. Further guidance will cover required competency of the author team. In addition, a triage tool will identify those studies at serious risk of bias and therefore further evaluation of all domains will not be required. So, there is a clear expectation that review authors where possible should use this tool. In some cases, it may seem appropriate to use another tool, such as the currently recommended Newcastle Ottawa Scale.

Credibility & validity

The tool starting from first principles has undergone iterative development following expert review and pilot testing. Further evaluation will take place during implementation that may lead to future empirical work.

Limitations/caveats

Strong level of competency required in epidemiology to use the tool.

Areas of concern/uncertainty

None noted

Impact on Cochrane

Tool is complicated to complete and needs a level of expertise. It is also time consuming to complete. However, this might be mitigated by the development of the triage tool. Training and

support could be high based on the number of reviews likely to include nonrandomised studies, which may increase due to stakeholder requirements.

Cochrane resources needed

Separate software (not RevMan) is in development. Publication of 'Risk of bias' tables and integration into GRADE will need further consideration.

Implementation

CSC members are not responsible for managing implementation of these recommendations which will require an implementation plan to ensure co-ordination for a smooth introduction. This will include launch, timescales and roll out strategy. Therefore, this statement does not signify immediate implementation.

PAPER 2

Cochrane Scientific Call for agenda items

Thirteen submissions received – reviewed by Ana, Philippe and David

The following items are put forward for consideration by the Committee members for future agenda's. The Chairs and DT removed items not considered ready or appropriate. Full list (Excel spreadsheet) available in Dropbox with additional information on publications and guidance etc.

	Submitted by and title	**Aims and objectives**	**Key features and elaboration**
1.	Nicole Skoetz Inclusion of results from searching study registries in Cochrane reviews: completed but not published studies	To give authors guidance on how to include completed but not published studies identified in study registries	Currently, there is one section in the Cochrane review called "ongoing trials", but what about all the completed trials authors identified in trial registries? Especially those without any published results, where to report them? Still in the ongoing section? This name is misleading, as some might not be ongoing any more. How do review authors search in trial registries? For the "status" ongoing only? Then they will not identify completed, but not published results. Should review authors include completed but not published trials in "included trials" section? Should review authors impute data for these unpublished trials?
	DT, AM, PR commentary	**Advice required on how to proceed** Another related question of interest is, in what situations where there are published reports/journal articles as well as data in trial registries, should authors be expected to examine all sources related to a study and comment on inconsistencies? How best to capture data from multiple sources? Also track changes with trials overtime e.g. outcomes How best to manage and to account for discrepancies that occur between sources and approach systematically? Problems with subsequently imputing data, if inconsistent. This covers a broad topic and improved guidance required.	
2.	Donna Gilles Meta-analyses of prevalence and risk	To broaden the scope of Cochrane reviews to include the best meta-	Meta-analysis of prevalence and risk - specific types of reviews. Cochrane does not support meta-analysis of

		analytic methods of studies of prevalence and risk	prevalence and risk. This is a growing field and high-quality methods need to be developed. In addition, supporting reviews of prevalence and risk could cover many of the areas which users of Cochrane have identified as gaps in our product.
	DT, AM, PR commentary	**Advise on whether a paper outlining challenges and benefits of including this review type should be presented in consultation with the prognosis Methods Group** New review type in terms of resourcing requires serious consideration.	
3.	Donna Gilles Meta-regression	To support meta-regression in order to improve the quality of analytic methods particularly in relation to continuous study variables and potentially confounding variables	Meta-regression - all reviews. Because of the lack of available meta-regression software and support, analyses of many large-scale Cochrane reviews inadequately address continuous factors such as dosage and longitudinal follow-up, as well as potential covariates.
	DT, AM, PR commentary	**It would help to have a collective view on the importance of this method to encourage its application especially for updates.** Meta-regression should be done Currently, RevMan does not support meta-regression. However, should we encourage use of other software, such as R.	
4.	Jayne Tierney Timely and Reliable Evaluation of the Effects of Interventions: A Framework for Adaptive Meta-analysis (FAME)	Aims to develop a prospective approach to Aggregated Data (AD) systematic review that takes all relevant trials into account and allows us to quickly respond and adapt to emerging trial results. The novel Framework for Adaptive Meta-analysis (FAME) allows us to anticipate the earliest opportunity for reliable AD meta-analysis, often years in advance of all trial results being available.	Most systematic reviews of efficacy are retrospective and based aggregate data (AD) from trial reports, meaning they can lag behind therapeutic developments and fail to influence ongoing or new trials. As unpublished and particularly ongoing trials are often overlooked, this can lead to reporting biases, hamper interpretation of meta-analysis results, and means updating is often inefficiently regarded as a separate process. Against this backdrop, unplanned duplication of systematic reviews has flourished. Further information available in Dropbox
	DT, AM, PR commentary	**CSC are asked to review this proposal for future agenda discussion**. Need to agree the scope of the review as it changes.	

5.	Jayne Tierney Determining when meta-analyses of published time-to-event outcomes reliable enough to form robust clinical conclusions. An evidence-based approach	Currently, it is not clear when meta-analyses of published time-to-event outcomes are reliable enough to form robust clinical conclusions. We aim to provide substantial and systematic empirical evidence on the reliability of meta-analyses based on HRs from published AD in comparison to those from IPD, so as to inform when IPD might be required.	Effects of treatments on time-to-event outcomes are usually measured using a hazard ratio (HR). If HRs are not explicitly reported, they can be calculated or estimated indirectly from other published statistics, or from data extracted from Kaplan-Meier (KM) curves. Each require assumptions that may affect the reliability of aggregate data (AD) meta-analyses including HRs. Further, AD meta-analyses of HRs are at risk of reporting biases, including follow-up bias, which the collection of individual participant data (IPD) may overcome. However, the IPD approach is lengthy, not always feasible and still rare. Therefore, when an answer is needed quickly or until IPD becomes more readily available, we will continue to rely on meta-analysis of published HRs. We aimed to provide substantial and systematic empirical evidence on the reliability of HRs derived from published AD and IPD, so as to inform when IPD may be required. Based on an unselected cohort of 18 IPD systematic reviews (238 unique trials), we compared HRs from AD with their IPD equivalents at the trial and meta-analysis level. The IPD represent >80% of eligible trials and ~90% of eligible patients, often with updated follow-up, providing a 'gold standard' with which to compare HRs from AD. Further information available in Dropbox.
	DT, AM, PR commentary	**CSC asked whether leads should submit a paper on providing recommendations as to how to implement and when.** It is now possible to calculate data extracted from Kaplan-Meir curves.	
6.	Rebecca Turner Data-based predictive distributions for between-study heterogeneity	Many meta-analyses contain only a small number of studies, which makes it difficult to estimate the extent of between-study heterogeneity. Bayesian meta-analysis allows	Meta-analyses from the Cochrane Database of Systematic Reviews (Issue 1, 2008) were classified according to the type of outcome, type of intervention comparison and medical specialty. The impact of meta-analysis characteristics on the underlying between-study

		incorporation of external evidence on heterogeneity and offers advantages over conventional random effects meta-analysis (Higgins and Whitehead 1996). To assist with implementation of Bayesian meta-analysis, we have provided empirical evidence on the likely extent of heterogeneity in particular areas of healthcare.	heterogeneity variance was investigated by modelling the study data from all meta-analyses simultaneously. Predictive distributions were obtained for the heterogeneity expected in future meta-analyses. These can be used directly as data-based informative prior distributions for heterogeneity in Bayesian meta-analyses. Between-study heterogeneity was found to be strongly associated with the type of outcome measured in the meta-analysis and somewhat associated with the types of interventions compared. We have published predictive distributions for heterogeneity in meta-analyses of binary outcomes (Turner et al. 2012; Turner et al. 2015) and for heterogeneity in meta-analyses of continuous outcomes (Rhodes et al. 2015). In addition, we have proposed accessible methods for implementing Bayesian meta-analysis with informative priors, avoiding the need for specialist Bayesian software (Turner et al. 2015; Rhodes et al. 2016). Using informative priors for heterogeneity would be beneficial in meta-analyses including few studies. These methods could be applied in standard Cochrane reviews.
	DT, AM, PR commentary	**Seek a view from the CSC as to whether this should be mandatory or discretionary, and therefore consider implications for implementation.**	