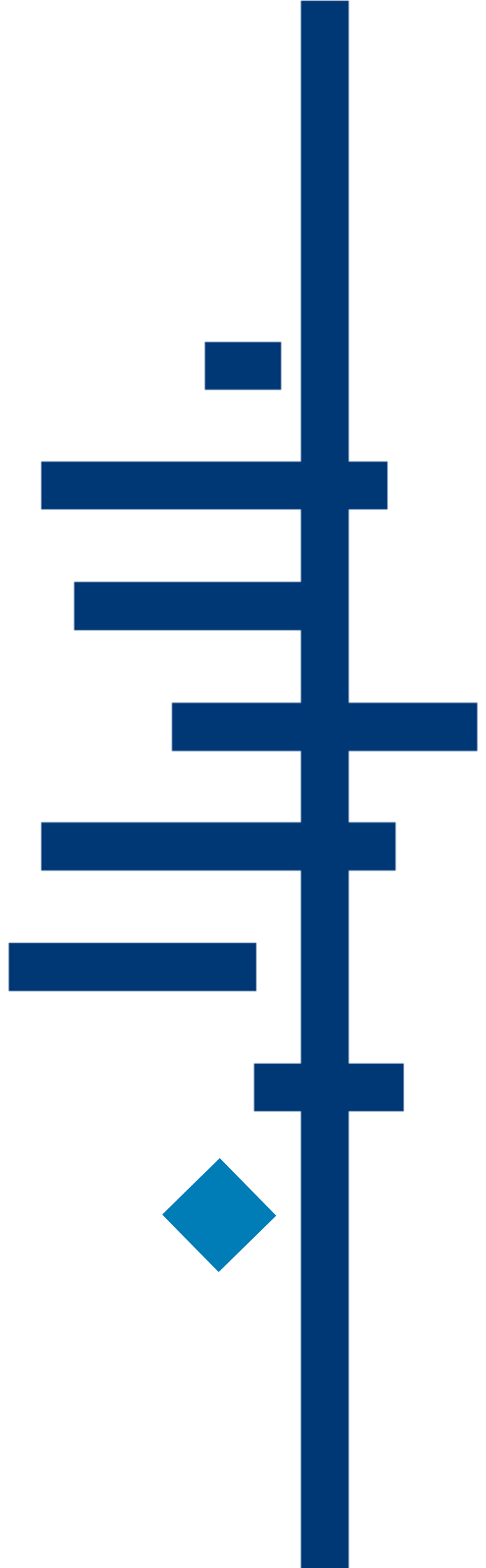


# Cochrane Scientific Committee

AGENDA

18<sup>th</sup> May 2017



# Contents

<b>AGENDA</b>	<b>3</b>
<b>CSC ACCOMPANYING DOCUMENTS</b>	<b>5</b>

# AGENDA

All members are expected to attend although Donna Gilles may not as she is travelling but intends to attend.

## Committee members:

**David Tovey (DT)**, Editor in Chief

**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, UK.

## Cochrane Staff:

**Jackie Chandler (JC)**, Methods Co-ordinator

## First Meeting 18<sup>th</sup> May 2017

Please declare any interests with any item on the agenda at the beginning of the meeting

AGENDA ITEM	Time	Details, links to documents and action required	Responsibility for item
1) Welcome and apologies received	12.00-12.15	I. Introductions II. Elect CSC Chairs III. Agree process to stagger terms of office IV. Declarations of interest for completion	JC
2) Approval of previous minutes	-	Minutes dated - none	-
a) Matters arising	-	List of items - none	-
3) CSC Business matters	12.15-12.30	I. Current governance arrangements & where the CSC is positioned. II. Agree draft Terms & Conditions <a href="#">Document 3i</a> III. Processes for submissions and review <a href="#">Document 3ii</a>	DT  JC  JC
4) Submissions of methods items	12.30-12.40	Nicole Skoetz suggested items of interest – Members may propose items for consideration. CSC need to agree future process and how to prioritise. Future meetings will have results from a call.	NS & CSC members
5) Methods for CSC Review	12.40-13.00	I. Review of the development of the risk of bias tool for non-randomised studies for interventions – ROBINS-I. <a href="#">Document 5i</a> II. Review of approaches to cumulative meta-analyses for systematic reviews <a href="#">Document 5ii</a>	Jonathan Sterne – presentation  Mark Simmonds Christian Gluud - presentations
6) Methods for CSC sign off and recommendation	13.00-13.10	I. Review of the updated ‘Risk of bias’ tool RoB 2.0. <a href="#">Document 6i</a>	JH & Jonathan Sterne CSC members
7) Special items			

---

a) Future meeting agenda	13.00-13.15	Approaches to Network Meta-analysis and GRADE (for methods review) Tool for assessing intervention Complexity (for sign off)	JC
b) Research priorities and strategy		Opening discussion on a formal approach to develop an agenda of research synthesis priorities.	
8) Any Other Business			CSC members
9) Meeting schedule		Scheduling meetings	JC

---

## CSC ACCOMPANYING DOCUMENTS

Attached documents:

Document 3i – Terms and Conditions

Document 3ii – Templates

Document 5i – Review of the development of the risk of bias tool for non-randomised studies for interventions – ROBINS-I

Document 5ii – Review of approaches to cumulative meta-analyses for systematic reviews

Document 6i – Review of the updated ‘Risk of bias’ tool RoB 2.0.

Scientific Committee members Dropbox will contain all documentation including access to relevant publications.



# Cochrane Scientific Committee

## Terms and Conditions for Scientific Practice

Document prepared by: Jackie Chandler

Reviewed by: Carl Moons, Yemisi Takwoingi on behalf of the Methods Board,  
members of the Methods Executive & David Tovey (Editor in Chief)

May 2017

Trusted evidence.  
Informed decisions.  
Better health.



# Contents

1.	Introduction and background	4
2.	CSC purpose and area of responsibility	5
3.	Governance	5
4.	Aims and objectives	5
5.	CSC principles of scientific practice	5
6.	Committee membership and structure	6
6.1.	Committee structure	6
6.2.	CSC membership	7
6.3.	Member criteria	7
6.4.	Selection and role of Co-Chairs	7
6.5.	Terms of office	7
6.6.	Termination of membership	8
6.7.	Officers of the CSC	8
6.8.	Sub Committees and work groups	8
6.9.	Members responsibilities	8
6.10.	Membership support	9
6.11.	Risk and indemnity	9
7.	CSC functions	9
7.1.	Agenda management and submission of items	9
7.1.1.	Responsibility for agenda organization and minutes	9
7.1.2.	Submission of items for discussion	9
7.1.3.	Meetings and symposia	10
7.1.4.	Production of consensus scientific statements or reports	10
7.2.	Processes and work plan	10
7.2.1.	Work plans	11
8.	CSC relations within Cochrane	11
8.1.	Cochrane Board	11

8.2. Methods Groups	11	10.2. Disputes and disagreements	12
8.3. Editor in Chief	12	11. Communications	12
8.4. CDSR Board	12	11.1. The Website	12
8.5. Cochrane Central Executive	12	11.2. Communication	12
9. Methods and Review Development Framework (MRDF)	12	12. Consultants, advisors to the Committee	12
9.1. Implementation of recommendations	12	12.1. Methods Convenors	12
10. Handling disagreements and disputes	12	12.2. External advisors	13
10.1. Diverging opinions	12	13. Dissolving the CSC	13
		13.1. Failure to carry out its obligations	13
		14. Independent review of CSC	13



# Cochrane Scientific Committee

## The Cochrane Scientific Committee (CSC) will provide impartial recommendations on methods for Cochrane Reviews

### 1. Introduction

Evidence synthesis for health care is undergoing rapid change with many methodological advances emerging. These include diagnostic test evaluations, prognosis studies, IPD studies, big data research, data automation techniques, and ‘real world’ evidence studies. Cochrane’s Strategy to 2020 identified the need to continue to identify, critically evaluate and implement methodological advances for the benefit of health care decision makers (health care practitioners, policy makers and consumers).

Following an internal Methods Review in 2015 the Cochrane Governing Board agreed to establish a body to guide strategic decisions on ‘what’ methods Cochrane should employ in its Reviews, and which new methods or types of Cochrane Reviews Cochrane should embark upon, anticipating the 10-year research horizons of numerous influential institutes such as EU (Horizon 2020), NHS (UK) and NIH (US). This will involve new review methods, possibly

addressing different types of review questions, additional methods to those currently in use, new sources of data, technology advances, and changes to current methods. In addition, some methods currently in use could become obsolete or inappropriate.

### 2. Relationship between CSC, Cochrane Methods Groups and other experts within and beyond Cochrane

This new Committee will represent the main decision making body to agree what methods are employed within and outside Cochrane and will advise the Editor in Chief. In making its judgements the Scientific Committee will draw upon expertise within the Cochrane community, including Cochrane’s Methods Groups. The Methods Groups represent networks of expert researchers with expertise and experience in specific areas of methodology who support Cochrane by conducting methods research, developing tools to guide best practice in methods, producing guidance and providing training on methods. Other Cochrane experts, and methods experts not working within Cochrane, will provide additional expertise. The CSC will provide an overview of recommended methods for priority implementation and arbitrate between expert differences of opinion.

### 3. CSC purpose and area of responsibility

The Cochrane Scientific Committee (CSC) provides an independent forum to discuss, debate and agree current and future methodological issues. The Cochrane Board's delegated authority to the CSC will support the Editor in Chief of the Cochrane Library, to determine the most appropriate methods for implementation in Cochrane Reviews. The CSC will work in close association with the Methods Groups. However, whilst the CSC will inevitably wish to consider and advise on implementation, this is not its primary role. The burden of responsibility and justification as to how recommended methods are implemented will rest with review authors and their respective Review Groups, and ultimately the Editor in Chief and the Editorial Board. However, weighing up the implementation impact with the relative value to adopt a new method is useful.

### 4. Governance of the CSC

The CSC is an independent arm's length body reporting to the Editor in Chief who reports to the Cochrane Board.

#### Aims and objectives

**Aim:** The CSC will advise Cochrane on recommended methods for its Reviews, by maintaining vigilance on the ongoing development of systematic reviews and future methodological possibilities collaborating with Cochrane's Methods Groups and other experts in Cochrane.

#### Objectives:

Specifically, the CSC in collaboration with Methods Groups and other experts will:

- Consult, in the first instance, with relevant Methods Groups and other methodological experts internal or external to Cochrane.
- Address current outstanding methodological issues, where there is a difference of opinion or ambiguity regarding whether specific methods should be implemented or not.

- Identify (horizon scan) important methods that Cochrane should prioritize for implementation in the immediate, short to medium and long term.
- Arbitrate between different professional opinions on methods and their application, in support of the Editor in Chief and the Editorial Board in implementing the most appropriate and up to date methods in Cochrane Reviews.
- Provide recommendations on methods advising on whether a method is best practice, recommended, permitted or not permitted from a pool of available methods, clearly stating whether a method is no longer appropriate and should cease.
- Provide recommendations on methods that Cochrane should prioritise for evaluation and development, negotiating with the appropriate Methods Group and other experts.

### 5. CSC principles of scientific practice

The CSC will provide, to the best of its members' ability, impartial advice that benefits Cochrane allowing it to meet its strategic objectives, following principles of scientific practice outlined in these terms and conditions. The CSC's deliberations and recommendations must provide the best available advice. This advice, as far as reasonably possible, should be based on available evidence provided by evaluation, or current expert consensus (Methods Groups and other experts). Importantly, advice and decision-making should be credible ensuring its integrity. The CSC should communicate with clarity and completeness on how it obtained its information, and reached a consensus decision.

#### 5.1. Individual members

Scientific integrity of the CSC, although requiring formal rules and governance structures, also requires individuals to act with intellectual honesty taking personal responsibility for their actions and decisions, and requires individuals engaged in scientific practice to:

- Assume responsibility for their actions
- Display critical thinking and strategic awareness
- Fully disclose any potential bias or conflicts
- Refrain from being affected by outside interference and censorship
- Ensure adequate procedural and information security
- Be prepared to report any breach
- Represent other people's work fairly and accurately and acknowledge the contribution of others
- Incorporate an awareness of equity issues
- Maintain objectivity uninfluenced by one's own prejudices and prior beliefs
- Acknowledge differing views and opinions with respect and sensitivity

### 5.2. Independence and objectivity

CSC members' decisions should not be influenced by any other consideration other than the scientific basis of the advice provided. CSC members should refrain from any political or commercial influence. CSC members engaging in the principle of independence need to ensure that their processes reflect an independent approach.

### 5.3. Transparency

Scientific advice provided, conclusions drawn, limits of their validity and the relevant uncertainties must be clear and understandable. The process and rationale for decisions made must also be clear and understandable. Should members involve any third party or invite in expertise, the process and rationale for doing so must be explicit.

### 5.4. Declaration of interests

Members are required to declare conflicts of interests in line with Cochrane guidance for the three years preceding the start of their term of office. This should include Cochrane methods members have developed. Members are required to do this on an annual basis and are responsible for updating the CSC, should there be a change affecting their previous disclosure in the meantime. We expect members to adhere to the Cochrane Commercial Sponsorship policy (see appendix 1).

### 5.5. Confidentiality

CSC discussions should be conducted in the spirit of transparency. However, CSC members are asked to exercise due diligence and not divulge information deemed to be confidential beyond the CSC and its sub Committees, and if appropriate the relevant Methods Group or expert. In addition, members should maintain, confidentiality of all CSC discussions and opinions for future recommendations before their public announcement and dissemination.

## 6. Committee membership and structure

### 6.1. Committee structure

The CSC is an independent arm's length committee within Cochrane's governance structure, to ensure the independence of its view to the Cochrane membership and its stakeholders. CSC will consist of up to fifteen members. The Editor-in-Chief will sit on the committee as a non-voting member and will take part in the CSC's deliberations and discussions. He/she will advise the CSC on procedural and editorial issues to inform discussions as appropriate. The Methods Co-ordinator will support the activities of the CSC, and is not a member and does not have voting rights.

The CSC needs a quorum of 10 members (excluding the Editor in Chief) for decisions. This should include at least one co-Chair and the Editor in Chief. CSC members can continue to meet when not quorate to continue

with work in progress. Any decisions or developments should be subsequently ratified by a quorate CSC.

## 6.2. CSC membership

CSC members are credible and influential members of the research evidence synthesis community. They will demonstrate active involvement in the practice of conducting reviews and developing review methodology. CSC membership is unlikely to include expertise across all possible methods so the CSC can co-opt appropriate experts at its discretion. These co-opted experts may attend a single meeting or work on sub-Committees for a fixed pre-determined period. The CSC will need to decide based on the 'membership' status of the additional member whether it is appropriate for them to have voting rights or not. The composition of the CSC membership is:

- Six to eight members from within the Cochrane community who either have a strong focus on methods research and development, or editorial skills and healthcare experience with strong methods interests. Evidence of a longstanding leading role in Cochrane is an additional requirement. However, the selected member does not represent any entity in Cochrane.
- Four to six external members for independent balance. These people are senior experienced research leaders within their specialist field, who have a wide knowledge of systematic review methodology, or senior experienced systematic reviewers or editors with a known interest and experience in methodological development. At least two of the external members will also represent stakeholders and end users of reviews e.g. agencies using Cochrane Reviews in guidelines, health research funders and those representing consumer interests.
- The Editor in Chief (or Deputy Editor in Chief)
- An early career researcher who is also within 5 years of completing a PhD, developing a relevant methodological track record.

Selection will consider geographical location, gender and language diversity and any other considerations of equity. The CSC will take responsibility for the selection of members following a process of open nomination for suitable candidates.

## 6.3. Member criteria

A spectrum of expertise is sought. Individuals will not necessarily meet all criteria. The following provides a baseline line that may be subject to amendment by the CSC. Aside from the early career researcher, we expect nominated members to have a high level of experience (5 years+) in their specific field and hold senior positions. The CSC needs the following breadth of experience across its membership:

1. Methodological research
2. Conducting, editing and publishing Systematic Reviews
3. Senior management of evidence producing systems, e.g. guidelines.
4. Similar scientific or research committees.
5. Health research funding, grant management or health policy.
6. Consumer advocacy.
7. More than five years' senior level experience holding a senior institutional position
8. Early career researcher who is also within 5 years of PhD.
9. Expertise in developments using technology for methodological purposes
10. Committee chair experience

## 6.4. Selection and role of Co-Chairs

The CSC will select two Co-Chairs from amongst its membership. Co-chairs are responsible for CSC conduct. Co-Chairs will manage the agenda and briefing of members in consultation with the Editor in Chief and the Methods Co-ordinator. The Co-Chairs are expected to have experience chairing similar bodies, and should have strong communication and conflict resolution skills. The Co-Chairs may be required to represent the CSC on certain occasions.

## 6.5. Terms of office

Terms of office are initially for three years, extended on request for a further two years at the Co-Chair's discretion. No member should serve for more than 5 years. Co-chairs should change every two years. Staggered membership at CSC inception will ensure continuity throughout the CSC life cycle.

### 6.6. Termination of membership

We request members give prior notification, at least one meeting in advance, when resigning from the Committee. The Co-chairs are responsible for membership management. Co-chairs of the CSC are obliged to consider termination if a member has:

- not attended three consecutive meetings, or any meetings in one year.
- acted beyond the scope of Cochrane policy e.g. Spokesperson policy, Commercial Sponsorship Policy.
- acted in a manner that undermines the scientific integrity of the Committee.
- breached any of these terms and conditions.
- acted in a manner considered inappropriate by CSC members (including the Editor in Chief) supported by the Co-chairs of the Cochrane Governing Board.

The CSC Co-Chairs (in consultation with the Editor in Chief) will determine whether the Committee member should leave the CSC immediately or within a given notice period. The Co-Chairs will communicate and record clearly in writing the reasons for their decision.

The CSC Co-chairs can take the option to counsel the member making a written record of their recommendation to allow the member to continue membership.

Should such a breach occur involving a CSC Co-Chair, the Editor-in-Chief will seek the advice of the Governing Board Co-Chairs and support for any subsequent action.

### 6.7. Officers of the CSC

The Methods Co-ordinator will manage CSC support and liaison ensuring the fulfilment of CSC processes and obligations. They will support the Co-chairs and the Editor in Chief with agenda and member management.

This will include:

- managing the agenda and action points from meeting to meeting;
- submission of agenda items from CSC members, Cochrane Methods Groups and other Cochrane members and any external submission if appropriate;
- organizing and supporting scientific reports and statements;
- dissemination of recommendations;
- scheduling a calendar of meetings and any special or extraordinary meetings as required;
- supporting the management of selection and induction of new members.

### 6.8. Sub committees and work groups

The CSC can set up *sub committees* on an ongoing basis, if required, although it retains responsibility for the sub-committee and should ensure proper reporting mechanisms. For short term, focussed work the CSC can establish *work groups* provided with specific, clearly defined tasks and objectives. Again, the CSC retains responsibility for the work group and its output to ensure proper reporting mechanisms. For both sub-committees and work groups the CSC is at liberty to co-opt appropriately skilled people to meet its objectives.

### 6.9. Members responsibilities

All Committee members should maintain the scientific integrity of the CSC maintaining a balanced viewpoint that objectively ascertains the right

approach for Cochrane. Committee statements and reports should be of a high quality reflecting scientific integrity, due diligence and consideration.

CSC members are expected to attend all meetings unless prevented due to unforeseen circumstances or prior engagement and notification. Committee members unable to attend three meetings in a row or no meetings within one year may be asked by the Co-chairs to stand down from the Committee. Cochrane expects Committee members to familiarize themselves with all pertinent Cochrane policies, including the Spokesperson policy (see appendix 2).

For the CSC to function properly members will need to ensure they have capacity within their primary roles to undertake Committee activities as specified here. Expectations are attendance at teleconference meetings, occasional face to face meeting, involvement in sub-Committee activity and contributing to statement or report writing. Also, members are expected to read and familiarize themselves with papers presented in support of agenda items. This may require up to 5-8 days per year. Co-chairs will need to provide some additional support.

#### **6.10. Membership support**

Cochrane will support all virtual communication. We do not anticipate the need for regular face to face meetings, but if these are required, Cochrane will cover travel and accommodation expenditure, within the limits of Cochrane's expense claims policy. Additional out of pocket expenses will be negotiated on a case by case basis.

#### **6.11. Risk and indemnity**

Individual members are professionally liable for their own conduct within their CSC role, and ensuring all Conflicts of Interest are declared. Cochrane accepts all opinions in good faith, it cannot be liable for misappropriation of information.

### **7. CSC functions**

The Committee will convene as often as needed to address the issues tabled for discussion. A CSC discussion forum will allow members to discuss and tease out pertinent issues for further investigation or information from

support staff (Methods Co-ordinator) before meetings. The CSC may on occasion convene in response to urgent issues that arise. It may also require experts to present and address the CSC. CSC agenda, decisions and recommendations are open access unless pre-determined as restricted. Agenda development will involve submissions to the CSC for consideration and scientific opinion.

In order to agree on the adoption of a new method a broad consensus (following input from Methods Groups and other experts) is ideal, however when a consensus cannot be reached, there should be at the minimum a majority (80%) of CSC members supporting a recommendation. If this is not the case, refer to section 10.

### **7.1. Agenda management and submission of items**

#### **7.1.1. Responsibility for agenda organization and minutes**

The Co-chairs and the Editor in Chief with support from the Methods Co-ordinator, will manage the agenda reviewing items submitted, prioritising those for action and organizing items as appropriate. This will require additional input from the Co-chairs, via teleconference, to support CSC function. We will keep additional communication to a minimum using other media other than a conference call. The Methods Co-ordinator will manage minutes, actions and work plan.

#### **7.1.2. Submission of items for discussion**

A submission system through the Methods website will on a regular basis, linked to planned CSC meetings, call for initial expressions of interest. This will only record essential information. The CSC will review these submissions and request further information and supporting evidence to fit in with the CSC schedule of meetings and their agenda priorities. A feedback system will log and a record progress and decisions on submissions and communicate to the originator. Submissions are expected to come from a variety of sources primarily Methods Groups, and then CRGs, Centres, Fields, and other individuals with research interests either members or non-members (interested parties linked to other projects). CSC members



may also submit proposals for consideration. The member will not vote on their specific submitted interest.

### **7.1.3. Meetings and symposia**

Members can anticipate at least two meetings a year. On occasion, it may be necessary to call an urgent extra-ordinary meeting. The CSC may invite an open forum from to time to provide an opportunity for transparency and wider discussion on current debates (e.g. Symposia or workshops – specific funding will need to be agreed).

### **7.1.4. Production of consensus scientific statements or reports**

Minutes of the meeting are required to be succinct and open access (in plain language) and report CSC activity and management. In addition, Cochrane will require a brief open access annual report submitted to the Cochrane Board. The Committee will form its decisions and recommendations on methods in clearly worded ‘CSC Statements’ or for detailed accounts in a ‘CSC Report’. These statements and reports are open access unless otherwise restricted for reasons provided. These open access reports or statements should be widely distributed and made accessible. Following decisions on implementation by the Editorial Board and the Editor in Chief roll out will be support by the Central Executive who will co-ordinate implementation of recommendations (see section 8.5).

## **7.2. Processes and work plan**

The CSC will review, assess and judge the appropriateness of material submitted for discussion, request additional information, discuss and deliberate, or advise further evaluation (See section 9).

We envisage the work of this committee to involve the following steps (see fig.1):

- Review of methods, their assessment and priority
- Co-opting specialist advisers where appropriate
- Reviewing evidence produced in relation to submissions to the Committee
- Seeking additional information
- Advising on additional work, development or evaluation
- Consideration of the likely implementation challenges
- Committee deliberation to form recommendations
- Production of written statements on recommended methods and/or further evaluation required.

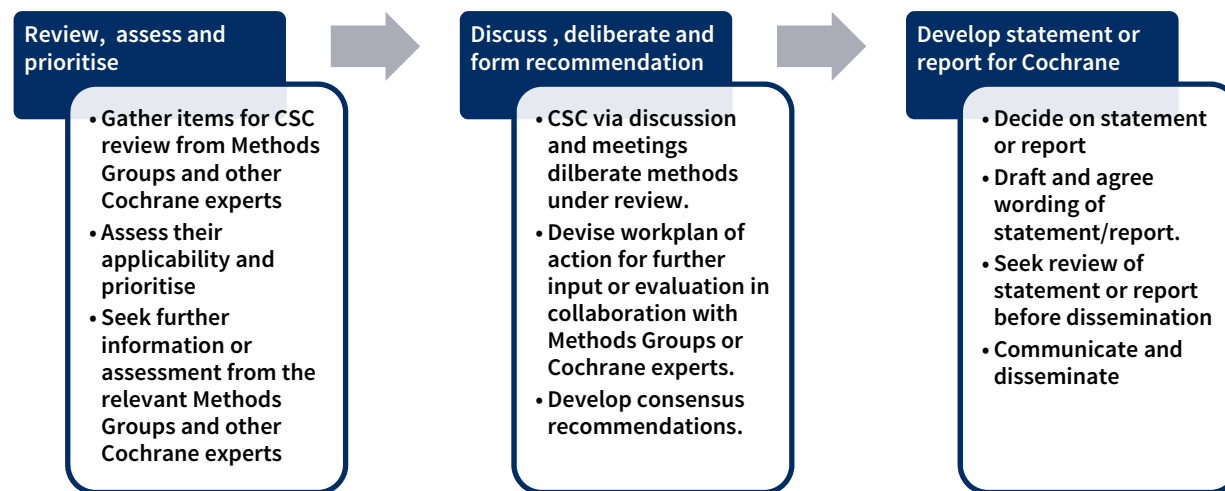


Fig. 1

### 7.2.1. Work plans

A key CSC task is the development of a research agenda mapping and prioritising current developments in the evidence synthesis field. Co-Chairs in co-operation with the Editor in Chief supported by the Methods Co-ordinator will manage the development of the research agenda in cooperation with CSC members.

## 8. CSC relations within Cochrane

### 8.1. Cochrane Governing Board

The CSC is independent. Should the Committee or members of the Committee seriously breach these terms and procedures and bring Cochrane into discredit, the Governing Board and the Editor in Chief have responsibility for managing the situation. Otherwise, all CSC matters and decisions should firstly go through the Editor in Chief, although if this is

not satisfactory to members they can request Governing Board guidance and support.

### 8.2. Methods Groups

Cochrane's Methods Groups are Cochrane's in house experts providing ongoing methodological advice, support and training directly to Cochrane. These researchers also conduct research as part of their individual research agenda's as well as conducting research specifically for Cochrane. These Methods Groups will predominately put forward Methods for CSC consideration. Minor changes to current methods or tools that do not fundamentally change the method but are just enhancements to improve methods already agreed do not necessarily need CSC review. However, Methods Groups may wish to seek the authority of the CSC to facilitate take up by Cochrane.



### 8.3. Editor in Chief

The Editor in Chief is a non-voting, ex-officio member of the CSC. They will keep the CSC advised on current issues regarding CRG methods practice, the quality of reviews, and implications for implementation of methods in Cochrane Reviews.

### 8.4. Cochrane Library Editorial Board

CSC Statements and Reports will go to the Editorial Board (and others) for consideration. The Editorial Board can petition the CSC on their recommendations, however, the Editor in Chief will make the final decision.

### 8.5. Cochrane Central Executive

The primary contact within the Central Executive Team (CET) will be the Methods Co-ordinator. Several departments within Cochrane will support the CSC when required, for example, Communications and External Affairs. The Editor in Chief will provide the necessary communication and reporting structure between the CSC and the CET Senior Management Team.

## 9. Methods implementation

Cochrane agreed processes for experimentation and evaluation of methods before their wide spread implementation. This process of defining the type of methodological change and the processes of development (e.g. pilot, evaluation or exemplar development) and decision-making, include whether adoption is universal or self-selecting by Cochrane Review Groups or Review author teams. The CSC will give due attention to Cochrane's current procedures for testing and evaluating methods before their implementation.

### 9.1. Implementation of recommendations

Implementation of recommendations is not the responsibility of the CSC, although they may advise. The Editorial Board along with support from the Central Executive Team will manage implementation of recommendations. Implementation of methods will most often start with a process of development and evaluation that can involve testing and piloting by Cochrane Review Groups before widespread implementation.

## 10. Handling disagreements and disputes

### 10.1. Diverging opinions

On occasion expert opinions diverge and it may not be possible to resolve these differences of opinion. Further evaluations may be required. The CSC should provide a report or statement clearly outlining the differences and their implications, with any recommendation for action as appropriate.

### 10.2. Disputes and disagreements

Should a dispute occur between CSC members, or between the CSC and members of Cochrane or Central Executive Staff that is unresolved through discussion the following steps should be undertaken.

Both Co-Chairs of the CSC and the Governing Board with the Editor in Chief should decide a plan of action.

This plan should include a meeting with the key Cochrane members (MG and CRG and others based on the nature of issue), CSC members, both sets of Co-chairs, and the Editor in Chief; they should agree a process for resolution. The Board will make the final decision.

## 11. Communications

### 11.1. The Website

A page will be designated on the Cochrane Methods website for managing communications (discussions) between CSC members. The page will also provide open access to any documentation including agendas (work plans), minutes, statements and reports.

### 11.2. Communication

The CSC through the Editor in Chief and the Methods Co-ordinator will liaise with the Communications and External Affairs department on internal and external communications.

## 12. Advisors to the Committee

### 12.1. Methods Convenors

Methods Convenors are an important community of experts attached to Cochrane, who will be very familiar with the Cochrane context for methodological development and implementation. Cochrane expects

their active participation when appropriate methodological discussions arise. The CSC might require additional input from a wider base of advisers. However, these in-house experts should be the first of port of call and there should be ongoing dialogue with active engagement of the relevant Methods Group(s)

### **12.2. External advisors**

The CSC should seek additional expertise, as and when, appropriate. This will ensure a balanced and considered approach to deciding on appropriate methodology. Likewise, including external advisors will ensure Cochrane remains abreast of all developments and future possibilities.

## **13. Dissolving the CSC**

### **13.1. Failure to carry out its obligations**

If there are concerns that the CSC is persistently unable to function and meet its obligations as set out in this document, or that the ambition of this guidance is unrealistic, an independent review conducted at the

## **14. Independent review of CSC**

The CSC will undergo two yearly review of its processes, management and output. A special review team considered ‘independent’ of any interest in the CSC will conduct the review. The Editor in Chief, however, will lead this review supported by the Methods Co-ordinator.

earliest opportunity will recommend either dissolution of the CSC or amendments to these terms.

## COCHRANE SCIENTIFIC COMMITTEE TEMPLATES

The following draft templates provide a process to managing CSC review of methods and decisions.

### 1. Expressions of interest

This will operate through an online system that allows a brief notification of methods requested for CSC review and recommendation. CSC will assess priority, relevance and timetabling in meeting programme.

### 2. Methods Briefing

This is the formal submission for discussion and should provide accompanying supportive evidence giving details of method for review.

### 3. Methods for sign off

Method or tool previously reviewed or deemed uncontroversial and requires formal decision before implementation.

### 4. Scientific Committee statement document

Proposed open access reporting document on CSC decisions.

# Cochrane Scientific Committee

## Expressions of interest submission Form

**Date:** Of CSC meeting targeted

**CSC:** Meeting reference e.g. 1:17, 2:17 (refers to number of meetings in year – will need to think about report citation using this reference at sign off meeting)

**Agenda item:**

Open access/restricted:

[Insert TITLE OF METHOD/DEVELOPMENT]

**Lead developers/investigators:**

**Abstract (100 words):** Aim & objective of methodological development

**Key features:**

**Methods used for evaluation and development**

**Brief details of Method or tool development**

**Conclusions**

**Key publication/guidance document if applicable** Please append

# Cochrane Scientific Committee

## Briefing report – Methods review

**Date:** Of CSC meeting

**CSC:** Meeting reference e.g. 1:17, 2:17 (refers to number of meetings in year – will need to think about report citation using this reference at sign off meeting)

**Agenda item:** Noting attendance & presentations

**Priority:**

**Open access/restricted:**

[Insert TITLE OF METHOD/DEVELOPMENT]

**Lead developers/investigators:**

**Abstract:**

*Aim & objective*

*Methods for development*

*Results/Development*

*Final product: Description, including guidance documentation*

**Impact:**

**Resources needed:**

**Recommendation requested:**

**SUPPORTING DOCUMENTATION**

List

# Cochrane Scientific Committee

## Briefing report 2 – For sign off and recommendation

**Date:** Of CSC meeting

**CSC:** Meeting reference e.g. 1:17, 2:17 (refers to number of meetings in year – will need to think about report citation using this reference at sign off meeting)

**Agenda item:** Noting attendance & presentations, if relevant

**Priority:**

**Open access/restricted:**

[Insert TITLE OF METHOD/DEVELOPMENT]

**Lead developers/investigators:**

**Summary of method or development:**

**Caveats:**

**Impact:**

**Resources needed:**

### SUPPORTING DOCUMENTATION

List

# Cochrane Scientific Committee

## Recommendation statement/report

Date: Of report

Relates to agenda item and meeting reference:

Priority:

Open access/restricted:

[Insert TITLE OF METHOD/DEVELOPMENT]

Lead developers/investigators:

Abstract:

Aim & objective

Methods for development

Results/Development

Final product: Description, including guidance documentation

## SUPPORTING DOCUMENTATION

List

## CSC RECOMMENDATION

- Highly recommended**  
*Because*
- Recommended with provisions**  
*Because*
- Optional/advisory (one among several options)**  
*Because*
- Not recommended**  
*Because*

## CSC STATEMENT

Summary statement

Credibility & validity

Limitations/caveats

Areas of concern/uncertainty

Impact on Cochrane

Cochrane resources needed

# Cochrane Scientific Committee

## Briefing report – Methods review

**Date:** 18<sup>th</sup> May 2017

**CSC:** 1:17

**Agenda item:** 5i Jonathan Sterne will attend giving a brief presentation

**Priority:** Medium

**Open access/restricted:** Open

Review of the development of the risk of bias tool for non-randomised studies for interventions – ROBINS-I (Risk Of Bias In Non-randomized Studies - of Interventions)

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

[Additional members supporting development](#)

### **Abstract:**

#### **Aim & objective**

The ROBINS-I tool evaluates the risk of bias (RoB) in the results of non-randomized 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 randomised 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 [in Dropbox]

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.

**Impact:** High, based on implementation and integration into Cochrane systems.

**Resources needed:** Currently not in RevMan, however, interactive software is in development. Requires training and support for implementation, although it comes with a health warning for completion, in that, epidemiological expertise on the author team is necessary, along with strong content expertise. Training required for CEU – screen team, CRG editors (Co-ed, ME, methodologists) etc.

**Recommendation requested:** The Scientific Committee is asked to consider whether the ROBINS-I tool is applicable and ready for implementation in Cochrane Reviews. Further recommendation is requested on whether updates that previously included NRSI, should follow the same principle given to the ‘Risk of Bias’ tool (See MECIR standard U9) to implement the same tool to all included studies applies.

## SUPPORTING DOCUMENTATION

ROBINS-I tool template

ROBINS-I Guidance

Publication: Sterne et al ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919 including appendix and supplementary tables

# Cochrane Scientific Committee

## Briefing report – Methods review

**Date:** 18<sup>th</sup> May 2017

**CSC:** 1:17

**Agenda item:** 5ii Presentations from Christian Gluud and Mark Simmonds (letters to the Committee attached)

**Priority:** Medium

**Open access/restricted:** Open

### TRIAL SEQUENTIAL ANALYSIS or SEQUENTIAL META-ANALYSIS

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

#### Abstract:

##### The problem

The CSC are 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 meta-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?

**Key critique about methods please see letters from Christian Gluud and Mark Simmonds summing up arguments**

<b>Critique</b>	<b>By who</b>	<b>Reference</b>
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. Problem of creating inappropriate 'stopping rules' in MA.	Higgins  Higgins	<i>Cochrane Methods</i> (2012) P32-33  <i>Cochrane Methods</i> (2012) P32-33
Measurement of accumulated information: <ul style="list-style-type: none"> <li>• The sum of the study weights in the meta-analysis. (Higgins)</li> <li>• Numbers of participants (Wetterslev et al.)</li> </ul> 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	<i>Cochrane Methods</i> (2012) P32-33
Sequential methods should be applied prospectively with a full analysis plan in the protocol.	Higgins	<i>Cochrane Methods</i> (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	<i>Cochrane Methods</i> (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. 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  Wetterslev & colleagues	<i>Cochrane Methods</i> (2012) P33-35.  Trial Sequential Analysis in systematic reviews with meta-analysis <i>BMC Medical Research Methodology</i> (2017) 17:39.
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	Hu and colleagues	Applying the law of iterated logarithm

<p>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.</p>		<p>to control type I error in cumulative meta-analysis of binary outcomes <i>Clinical Trials</i> (2007) 4:329-340.</p>
<p>In reference to methods developed by Wetterslev et al, Van der Tweel, and Bollen, and Higgins, Shuster &amp; 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 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).</p> <p>‘Look’ refers to the moment of meta-analysis in time – updating.</p>	<p>Shuster &amp; Neu</p>	<p>A Pocock approach to sequential meta-analysis of clinical trials. <i>Research synthesis Methods</i> (2013) 4 269-279.</p> <p>See paper for further explanation and methods proposed.</p> <p>Please see also further information in <i>Current controversies in data monitoring for clinical trials</i> (Pocock, 2006),</p>

**Impact:** Adding additional complexity to analysis

**Resources needed:** Implications for training and software integration, implementation and quality control.

**Recommendation requested:** See questions

## SUPPORTING DOCUMENTATION

### Supporting documentation attached

1. Letter addressed to the Committee from Christian Gluud.
2. Letter addressed to the Committee from Mark Simmonds.
3. [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.
4. 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.

### Published papers in Dropbox folder

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.

Christian Gluud's responses to the raised questions regarding the naive Cochrane

Copenhagen May 7, 2017

To the Cochrane Scientific Committee.

Att.: **Jackie Chandler**

Methods Co-ordinator | Cochrane Editorial Unit  
Cochrane Central Executive London

**Re: Does Cochrane reviews reasonably take into account risks of type I errors and type II errors and what can one do about it?**

Dear all,

I have chopped up the above questions into a number of questions to try to give you my personal opinion on them. My short responses are here with relevant supporting literature appended.

**Re: Does Cochrane reviews reasonably take into account risks of type I errors and type II errors?**

No, Cochrane does not! Most Cochrane reviews ignore the problems that can be caused by underpowered meta-analysis by naively and falsely committing type I errors (that is declaring benefits long before a plausible required information size has been reached) or type II errors (that is declaring that something does not benefit long before a plausible required information size has been reached). Numerous earlier studies have shown this. During more recent years especially four publications have rehammered this out. These are:

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

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



Christian Gluud's responses to the raised questions regarding the naive Cochrane

**Turner RM, Bird SM, Higgins JPT. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PLoS One. 2011;8:e59202.**

**Jackson D, Turner R. Power analysis for random-effects meta-analysis. Res Syn Meth. 2017. <https://doi.org/10.1002/jrsm.1240>**

**Re: What should Cochrane do about it?**

Act! Since the mid 1990s (and even before), several people started to see that there was a problem! The majority of Cochrane, however, chose not to respond to the problem. Cochrane loses credibility every day by not having a plan for how to deal with it! Taking action with these methods is better than no action at all!

**Re: How should Cochrane act?**

By introducing methods for systematic reviews that can control risks of type I and type II random errors more effectively than practices in the majority of systematic reviews.

There are two major ways in which Cochrane can act, the frequentist way or the Bayesian way! Or maybe one should consider both.

**Re: How should Cochrane act if it chooses the frequentist way?**

In my mind, the best frequentist way is through the conduct of Trial Sequential Analysis based on prior chosen plausible parameters (at the protocol stage). A large number of empirical studies support this:

**Thorlund K, Devereaux PJ, Wetterslev J, Guyatt G, Ioannidis JP, Thabane L, Gluud LL, Als-Nielsen B, Gluud C. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? Int J Epidemiol. 2009;38:276–86.**

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

Christian Gluud's responses to the raised questions regarding the naive Cochrane

***Imberger G, Gluud C, Boylan J, Wetterslev J. Systematic reviews of anesthesiologic interventions reported as statistically significant: problems with power, precision, and type 1 error protection. Anesth Analg. 2015;121:1611–22.***

***Mascha EJ. Alpha, beta, meta: guidelines for assessing power and type I error in meta-analyses. Anesth Analg. 2015;121:1430–3.***

**Re: How should Cochrane act if it chooses Trial Sequential Analysis?**

Please read the following recent article and our Trial Sequential Analysis Manual.

***Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Medical Research Methodology (2017) 17:39. DOI 10.1186/s12874-017-0315-7.***

***Thorlund K, Engstrøm J, Wetterslev J, Brok J, Imberger G, Gluud C. User manual for Trial Sequential Analysis (TSA). Copenhagen Trial Unit, Centre for Clinical Intervention research, Copenhagen, Denmark. 2011: 1–115 available from [www.ctu.dk/tsa](http://www.ctu.dk/tsa).***

**Re: How should Cochrane act if it chooses the Bayesian way?**

Please ask Julian Higgins!

**Re: How should Cochrane act if it chooses the frequentist way as well as the Bayesian way?**

There are pros and cons of both ways. Both ways need overview to secure that they are implemented correctly. I suggest reassessment within say two years.

CG's COI: Spent much time on Trial Sequential Analysis.

Very best wishes,

Christian Gluud.

# Comments on the briefing document to the Cochrane Scientific Committee

**Mark Simmonds**

**Centre for Reviews and Dissemination, University of York**

## The “Higgins” method

My preferred terminology for the method in Higgins et al 2011 is Sequential Meta-Analysis or SMA for short. It is not, as often assumed, a Bayesian method. It is an application of the sequential monitoring boundaries method of Whitehead to meta-analysis, and so essentially frequentist. As such it is very similar to Trial Sequential Analysis (TSA), but with a different specification of the stopping boundaries. My work suggests that the two methods are almost equivalent in most cases.

The Bayesian component comes in only for estimating heterogeneity, where an “approximate” Bayesian procedure is used to replace the standard (e.g. DerSimonian-Laird) estimate of heterogeneity with a Bayesian one, to avoid mis-estimation of heterogeneity leading to invalid conclusions. This is useful for meta-analyses with very few studies where heterogeneity estimation is unreliable, but is less necessary for meta-analyses with more studies.

My work suggests that, in practice SMA and TSA produce similar results, although there will always be some cases where one method will cross a stopping boundary and the other will not. I think the choice between them is mostly one of personal preference or familiarity.

## Other methods

Shuster’s Pocock approach (2013) and Hu et al’s Law of the Iterated Logarithm approach (2007) are alternatives, but neither controls for Type II error (unlike TSA and SMA). My work suggests that both are very conservative in preserving Type I error rates, at the cost of losing power to detect genuine treatment effects. I can’t see any benefits of these methods over TSA/SMA, and so I wouldn’t recommend them at present.

A true Bayesian method (Spence et al Stat Med 2016) exists, which is a full Bayesian extension of SMA. This is still new and little studied, but appears valid and useful if doing a Bayesian analysis, but, as yet, has no obvious advantages over SMA/TSA.

## What should Cochrane use?

My work suggests that, for typical Cochrane Reviews (that might only be updated 2-3 times, and with moderate or little heterogeneity) the Type I error from using standard meta-analysis rises to around 10-15% rather than the desired 5%. TSA/SMA can avoid this rise in error, but might obviously be unnecessary in many Cochrane Reviews that are unlikely to receive many updates, provided authors are aware of this increased error rate.

I think sample size calculations for meta-analyses as a guide to the robustness of evidence should be much more widely used: at least half of all Cochrane Reviews will be under-powered. TSA/SMA could

be used as a post-hoc check for meta-analyses with conventionally significant results, but low sample size, but this is less satisfactory than building in TSA/SMA use from the protocol onwards.

TSA/SMA are probably most needed in “prospective” reviews, where the review is undertaken while trials are still ongoing and there might be little completed trial evidence at the first review and many updates will be needed. In those cases the risk of error is substantial, and must be controlled.

**Please note:** Some of the results given here are currently unpublished. This represents the reviews of Mark only, and not necessarily the views of other members of this MIF project team.

adequately addressed in the majority of reviews, and the process of measuring, modelling and accounting for between-study heterogeneity is often limited or inadequate. In my experience, these issues can also arise in reviews by other Cochrane Review Groups; improvements are thus essential.

A major problem is the interpretation of random-effects meta-analyses.<sup>2,3</sup> A fixed-effect approach assumes the intervention effect is common (fixed) across studies, and thus the pooled meta-analysis result gives the best estimate of this common effect. However, a random-effects approach allows the intervention effect to vary across studies due to between-study heterogeneity, and so the pooled result provides an estimate of the average intervention effect. None of the CPG reviews using a random-effects model interpreted the pooled result as the average intervention effect, and there was no recognition that the intervention effect could differ from the average in a particular study setting. The newly proposed 95% prediction interval can help address this, as it allows reviewers to calculate a range of potential values for the intervention effect when applied in an individual setting.<sup>2,3</sup>

The decision about when to perform a random-effects rather than a fixed-effect meta-analysis also continues to generate tension. At the moment, many review authors are selecting one of these two methods based on a 'large'  $I^2$  value (e.g. >50%) or on the P value for a  $\chi^2$  test for heterogeneity. I suggest that when deciding between the two approaches, or indeed whether meta-analysis is actually appropriate, review authors should consider both statistical and clinical reasoning. In terms of statistical reasoning, I concur with Rucker and colleagues that the choice of meta-analysis approach should be guided by  $\tau^2$ , the between-

study variance, and not just  $I^2$ .<sup>4</sup> Generally (and assuming studies have low risk of bias and the threat of publication bias is low), if  $\tau^2$  is estimated to be non-zero, I prefer to adopt a random-effects approach, as it is more conservative: accounting for the between-study variance produces wider confidence intervals for the pooled effect, and allows the calculation of prediction intervals to explicitly show how the intervention effects vary across studies. However, I recognise difficulties arise when the number of studies is small, such that  $\tau^2$  is imprecisely estimated. If the decision to choose between fixed- and random-effects is not clear, I see no harm in presenting results from both analyses to show transparently how the choice of model influences estimates and conclusions.

In terms of clinical reasoning, if there is heterogeneity, reviewers must seek advice about whether the distribution of the intervention effects across studies is clinically meaningful. For example, one CPG review stated that not only was a random-effect approach statistically appropriate as heterogeneity existed, but also clinically meaningful as there was much clinical overlap across studies (e.g. in the patient characteristics, dose of drug, etc).<sup>5</sup> In such situations, estimating the average effect of the intervention, the between-study variance, and a 95% prediction interval is clinically helpful.

Given that publication bias is known to be a common threat to the validity of meta-analyses, it was very surprising that few CPG reviews discussed it. This must be addressed, especially in regard to primary analyses, or misleading or overly strong conclusions may be made. I support the recent recommendations for investigating small study effects and funnel plots asymmetry, which may signal publication bias.<sup>6</sup>

In conclusion, I feel there is a pressing need for The Cochrane Collaboration to involve more statisticians within each Cochrane Review Group, not just at each stage of the editorial and refereeing process, but also 'hands-on' within each individual review. This will inevitably require funding, and statisticians with expertise in meta-analysis are clearly not abundant, but improving statistical standards should be a high priority in the coming years.

## References

- Riley RD, Gates SG, Neilson J, Alfirevic Z. Statistical methods can be improved within Cochrane Pregnancy and Childbirth Reviews. *Journal of Clinical Epidemiology* 2011; 64(6): 608-618.
- Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011; 342: d549.
- Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society, Series A* 2009; 172: 137-159.
- Rucker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on  $I^2$  in assessing heterogeneity may mislead. *BMC Medical Research Methodology* 2008; 8: 79.
- King JF, Flenady V, Cole S, Thornton S. Cyclo-oxygenase (COX) inhibitors for treating preterm labour. *Cochrane Database of Systematic Reviews* 2005, Issue 2. Art. No.: CD001992. doi: 10.1002/14651858.CD001992.pub2.
- Sterne JAC, A.J. S., J.P.A. I, Terrin N, Jones DR, Lau J, Carpenter J, Rucker G, Harbord RM, Schmid CH, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomized controlled trials. *BMJ* 2011, 342: d4002.

## Trial sequential analysis: methods and software for cumulative meta-analyses

Jørn Wetterslev, Janus Engstrøm,

Christian Gluud and Kristian Thorlund

Correspondence to: wetterslev@ctu.rh.dk  
Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark.

**Background:** When a meta-analysis includes a small number of trials and a small number of patients, random errors can cause spurious findings.<sup>1-10</sup> The more statistical tests that are employed throughout the accumulation of additional data in cumulative meta-analysis, the higher the likelihood of observing a false positive result. This phenomenon is commonly known as 'multiplicity due to repeated significance testing'.<sup>11-13</sup>

Treatment effects from meta-analyses are typically assessed on the basis of P values and confidence intervals (CIs) for the

effect estimate. Meta-analysts must decide on the threshold at which a P value is sufficiently small or the CI sufficiently narrow to justify a 'positive' conclusion. Any threshold involves a trade-off between the risk of observing a false positive result and the risk of observing a false negative result. For example, if the threshold for statistical significance is lowered for the P value, i.e. increased for the test-statistic, Z, as displayed in Figure 1, the risk of observing an early false 'positive' result (Figure 1(A)) would have decreased while the risk of observing an early false negative result



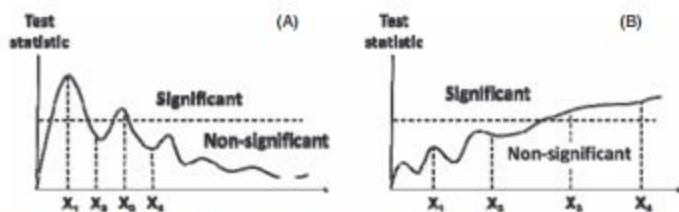
(Figure 1(B)) would have increased. However, as evidence accrues, P values (and test statistics) become increasingly reliable (see Figure 1). Thus, any inferences about the strength of evidence should be measured using the accrued number of patients, the observed number of events, and the impact of multiplicity.<sup>6,13,12,16</sup>

**The heterogeneity-adjusted required information size for a meta-analysis to be conclusive:** In a single randomized trial, we perform a sample size calculation to ensure that a 'sufficient' number

of patients are included. A similar 'yardstick' is needed for a meta-analysis.<sup>1,2,6,14</sup> This yardstick has been referred to as the required information size (RIS) of the meta-analysis.<sup>1,2,14,15</sup> Figure 2 illustrates two typical meta-analytic scenarios where the test statistic has stabilized after the RIS has been reached. The sample size in a single trial is typically calculated based on the expected event proportion in the control group, the anticipated relative risk reduction in the experimental intervention group, and the maximum risks of type I

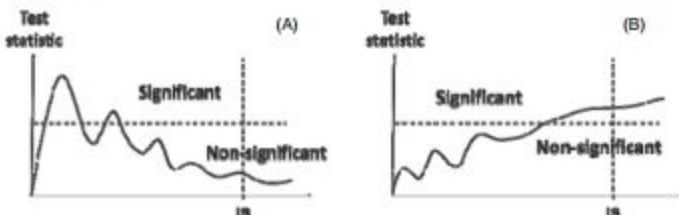
and type II errors. In a meta-analysis, there is often heterogeneity<sup>16</sup> across included trial populations, interventions, duration of intervention, and methods. Because increased variation reduces the precision of results, information size considerations must incorporate all sources of variation in a meta-analysis, including between-trial variation.<sup>13,15</sup> Accordingly, the RIS (the meta-analysis 'sample size'), needs to be adjusted upwards to consider the variance introduced by such heterogeneity.<sup>12,15</sup> One approach for incorporating heterogeneity in information size estimation is to multiply the sample size required for a single clinical trial to be adequately powered by some heterogeneity-adjustment factor.<sup>14</sup> In the fixed-effect model, it is assumed that all included trials have a common effect, and thus, the RIS for a fixed-effect meta-analysis to be conclusive may effectively be calculated as the required sample size for a single adequately powered clinical trial. In the random-effects model the true effect is assumed to vary across included trials and the variance is always greater than that in a fixed-effect model. A heterogeneity-adjustment factor must therefore account for the increase in variation that a meta-analysis incurs from going from the fixed-effect assumption to the random-effects assumption. This is accurately achieved when the heterogeneity-adjustment factor is equal to the ratio of the variance in a random-effects model meta-analysis and the variance in a fixed-effect model meta-analysis.<sup>14,15</sup>

**Adjusting confidence intervals and testing for statistical significance before the required information size has been reached:** A meta-analysis aims to identify the benefit or harm of an intervention as early and as reliably as possible.<sup>4,8,10,15</sup> Therefore, meta-analyses are commonly updated when new trials are published. Authors of Cochrane reviews are meant to update their systematic reviews every second year<sup>16</sup> and when meta-analyses are updated, they are repeatedly subjected to significance testing. In randomized trials, repeated significance testing on accumulating data is known to increase the overall risk of type I error.<sup>12,13</sup> Simulation studies of meta-analyses suggest that if repeated significance testing is done in meta-analyses and P values smaller than 0.05 are considered to be evidence of 'statistical significance', then the actual risk of type I errors will be between 10% and 30%.<sup>6,8</sup>

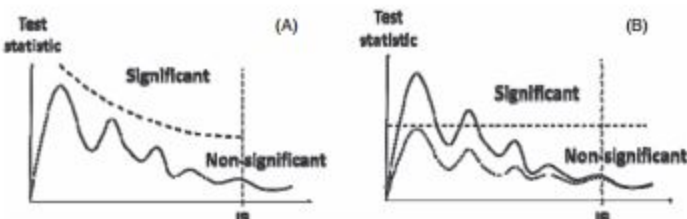


**Figure 1.** Examples of false positive and false negative statistical test results over time.

In figure 1(A), significance testing at times  $X_1$  and  $X_3$  would result in a false declaration of statistical significance, whereas significance testing at  $X_2$  and  $X_4$  would not. Thus, only at times  $X_1$  and  $X_3$  is the impact of random error 'extreme enough' to yield spurious statistically significant results. In figure 1(B), significance testing at  $X_1$  and  $X_2$  could have resulted in a false declaration that the interventions under investigation were not significantly different, whereas significance testing at  $X_3$  and  $X_4$  would not. Thus, only at times  $X_1$  and  $X_2$  is the imprecision of a magnitude that causes spurious absence of statistical significance.



**Figure 2.** Examples of how the required information size ensures reliable significance tests in two cumulative meta-analyses.



**Figure 3.** Examples of significance threshold adjustment (stipulated monitoring boundary) (A) and penalized test statistic (stipulated) (B) to avoid false positive statistical results in cumulative meta-analyses.



Likewise, empirical studies have shown a high false positive proportion of statistically significant meta-analyses when updated.<sup>6,9</sup> To deal with this problem, one can adjust the thresholds for which results are considered statistically significant.<sup>1-13</sup> Such approaches use methodology developed for repeated significance testing in randomized trials with statistical monitoring boundaries.<sup>3,5,8,10,12,13,18</sup> Similar adjustments can be applied to the CIs. Consider the example where the nominal significance level under repeated testing is 5% but the adjusted threshold at some time point corresponds to a single test 1% threshold. Here the adjusted CI would have a 95% (100% to 5%) nominal coverage under repeated testing, but at the same time point it would correspond to a single test 99% CI.

**Testing for futility before the required information size has been reached:** Using conventional thinking from randomized trials, a finding of 'no significant effect' may be considered to be due to lack of power until an appropriate sample size has been reached. In some cases, however, we may be able to conclude earlier that a treatment effect is unlikely to be as large as anticipated, and thus, prevent trial investigators from proceeding with including more unnecessary patients. Of course, the anticipated intervention effect can be revisited so further research may be needed. Trial sequential analysis may provide a technique to identify lack of an anticipated clinical effect as early as possible in a cumulative meta-analysis. 'Futility boundaries' are constructed and used to provide a threshold for 'lack of anticipated effect'.<sup>19</sup>

**Trial Sequential Analysis Software:** The software Trial Sequential Analysis (TSA) facilitates user-friendly calculation of the RIS and corresponding construction of adjusted thresholds for statistical significance of intervention effects, adjusted CIs, and futility thresholds. As discussed, these tools are particularly useful for gauging the strength of the statistical evidence before reaching the meta-analysis information size requirement. The TSA software and manual are freely available and can be downloaded from [www.ctu.dk/tsa](http://www.ctu.dk/tsa). In

addition, the TSA software runs on all operating systems that support Java.

Competing interests: Jørn Wetterslev, Janus Engstrøm, Christian Gluud, and Kristian Thorlund have all joined the working group of developing the TSA software and the TSA manual available at [www.ctu.dk/tsa](http://www.ctu.dk/tsa).

## References

- Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomized controlled trials. *Lancet* 1998; 351: 47-52.
- Pogue J, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Controlled Clinical Trials* 1997; 18: 580-593.
- Thorlund K, Devereaux PJ, Wetterslev J et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? *International Journal of Epidemiology* 2009; 38: 276-286.
- Trikalinos TA, Churchill R, Ferri M et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *Journal of Clinical Epidemiology* 2004; 57: 1124-1130.
- Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *Journal of Clinical Epidemiology* 2008; 61: 64-75.
- Borm GF, Donders AR. Updating meta-analyses leads to larger type I errors than publication bias. *Journal of Clinical Epidemiology* 2009; 62: 825-830.
- Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative meta-analyses. *Proceedings of the National Academy of Sciences USA* 2001; 98: 831-836.
- Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *Journal of Clinical Epidemiology* 2008; 61: 763-769.
- Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials

have modest credibility and inflated effects. *Journal of Clinical Epidemiology* 2011; 64(10): 1060-9.

- Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive - Trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. *International Journal of Epidemiology* 2009; 38: 287-298.
- Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. *Statistics in Medicine* 2011; 30(9): 903-21. doi: 10.1002/sim.4088.
- DeMets D, Lan KK. Interim analysis: the alpha spending function approach. *Statistics in Medicine* 1994; 12: 1341-1352.
- Lan KK, DeMets D. Discrete sequential monitoring boundaries for clinical trials. *Biometrika* 1983; 659-663.
- Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in a random-effects meta-analysis. *BMC Medical Research Methodology* 2009; 9.
- Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, Guyatt G, Devereaux PJ, Thabane L. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis-a simulation study. *PLoS One* 2011; 6(10): e25491.
- Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.0*. John Wiley & Sons, 2009.
- Higgins JP, Thompson S. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002; 21: 1539-1558.
- Reboussin DM, DeMets DL, Kyungmann K, Lan KKG. Programs for Computing Group Sequential Boundaries Using the Lan-DeMets Method, v.2. 2009.
- Jennison C, Turnbull B. *Group sequential methods with applications to clinical trials*. Chapman & Hall/CRC Press, 2000.



# COMMENT on 'Trial sequential analysis: methods and software for cumulative meta-analyses' by Wetterslev and colleagues

Julian PT Higgins

Correspondence to: julian.higgins@mrc-bsu.cam.ac.uk  
MRC Biostatistics Unit, Cambridge, UK and Centre for Reviews and Dissemination, University of York, York, UK.

## Introduction

Wetterslev and colleagues argue for the adoption of sequential approaches to the updating of meta-analyses, primarily in order to reduce the risk of false positive results when the amount of accrued information is not large. Sequential approaches address the problem that if a standard meta-analysis is updated several times with more studies added each time, then the probability that one of the analyses will produce a P value lower than 0.05 is somewhat higher than 5% under the null hypothesis that there is no underlying effect in any study. This may be inflating the rate of false conclusions that an intervention has an effect when it doesn't. So do these sequential approaches provide the answer? Should they be implemented as standard, in some situations, or not at all, in Cochrane reviews?

I have worked intermittently in this methodological area for a number of years. I first suggested the application of sequential approaches to The Cochrane Collaboration at the Rome Colloquium<sup>1</sup> (Higgins 1999) and was honoured to receive the Thomas C Chalmers Award for my talk, but my message was not adopted. Last year I finally co-authored a paper describing the approach I prefer to take.<sup>2</sup> Wetterslev and colleagues have been strong advocates of the application of sequential approaches in Cochrane reviews and have published numerous articles describing a slightly different approach (see the citations in their article). They have implemented their methods in several systematic reviews, published in the *Cochrane Database of Systematic Reviews* and elsewhere. I comment on the technical differences between our approaches later,

but the bigger issue is whether sequential approaches are appropriate at all.

## A controversial area

The application of sequential methods to meta-analysis is a controversial area. Because of increasing pressures from within the Collaboration to consider its widespread adoption – and following a request from Simon Gates of the Pregnancy and Childbirth Group for the Collaboration to formulate a policy – the Cochrane Statistical Methods Group (SMG) dedicated much of its meeting at the Madrid Colloquium in October 2011 to a discussion of this topic. Christian Gluud and Jørn Wetterslev overviewed many of the issues addressed in the article presented here, and Jonathan Sterne and I were invited to respond formally.

Jonathan argued against the use of sequential approaches on the basis that they encourage the use of significance tests, and the consequent inappropriate division of results as 'significant' or 'not significant', rather than the direct interpretation of intervention effect estimates and corresponding confidence intervals (CIs), as recommended in the *Cochrane Handbook for Systematic Reviews of Interventions*. He reviewed some fundamentals of statistical inference, noting the distinction between the Fisher approach and the Neyman-Pearson approach, and pointed out that a division of results into 'significant' and 'non-significant' was not the intention of either. This is discussed in more details in a *BMJ* paper.<sup>3</sup> Jonathan noted that control of rate of type I errors (false positive findings) may come at the cost of increasing the rate of type II errors (falsely concluding that there is no effect when an effect exists). Finally, he questioned the analogy between stopping trials based on interim analyses supervised by a Data Monitoring and Ethics Committee, and 'stopping' meta-analyses. If new, high quality, trials are already under way when the decision to 'stop' a meta-analysis is made, should these be ignored in subsequent meta-analyses? And should no further updating be done?

I responded to Jonathan that the sequential methods usually proposed for meta-analysis follow the Neyman-Pearson approach. In this approach, both a null hypothesis (e.g. the odds ratio is 1) and a specific alternative hypothesis (e.g. the odds ratio is 0.8) are prespecified, along

with values for type I and type II error rates (typically set to 5% and 10% respectively). These lead to a decision framework including a 'stopping rule'. If, at any particular update, there is insufficient evidence to accept or reject the null hypothesis, then more data need to be collected. The null hypothesis can be accepted if the path of the meta-analysis crosses a 'futility' boundary, based on the prespecified power and clinically important effect size, or it is rejected if the path crosses boundaries that reflect benefit or harmful effects. In both situations – according to the prespecified hypothesis testing framework – the question is answered, and no more data are needed. When applying these ideas to a single trial, they can lead directly to the continuation or stopping of the trial.

A key concern with taking sequential approaches to meta-analysis, widely expressed at the meeting, is that the notion of a 'stopping rule' cannot be applied to the meta-analysis situation. The meta-analyst is not in a position to make such decisions about future studies, although could of course decide not to include future studies in the review. My own view on this matter is that an important decision that the meta-analyst is (arguably) in a position to make is about making recommendations whether future studies should be undertaken (or not).

Several quantities need to be prespecified in order to design a sequential approach to meta-analysis, including the type I and type II errors and, most awkwardly, the alternative hypothesis. This of course raises a further key concern with the use of these sequential methods: from where do such values come? The alternative hypothesis should reflect an effect size that is important to detect. However, different effects sizes are likely to be important to different people, so whose perspective should be taken? Inferences can depend markedly on the values given to these quantities.<sup>4</sup> Should several different perspectives be addressed simultaneously? This is probably appropriate for a Cochrane review, although an important implication is that this will lead to a multitude of stopping rules, and hence, in practice, none that can govern the review's implications for research.

## Technical issues

There are many technical issues and difficulties in the application of sequential



methods to meta-analysis. The main technical difference between the methods of Wetterslev and colleagues and the methods I have worked on relates to how accumulating information is measured. I follow the approach first applied to meta-analysis by Anne Whitehead, in which information is measured using standard statistical measures of information (essentially the sum of the study weights in the meta-analysis). Statistical information can be specified unambiguously in advance, since it is closely related to power. Wetterslev and colleagues follow work by Janice Pogue and Salim Yusuf, who used numbers of participants as a measure of accumulating information. I'd argue that this is less sensible, since the sample size needs to be converted 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 (or other quantities for other types of data) but also of the anticipated amount of heterogeneity when a random-effects meta-analysis is planned. The approach requires special attention for different study designs such as cross-over trials and cluster-randomized trials. The issue of heterogeneity is problematic when taking any sequential random-effects approach. The principal value of a sequential approach is in the early stages when there are few studies. But when there are few studies, the amount of heterogeneity is poorly estimated. I have been explicit about this and incorporate a prior distribution for the heterogeneity variance.<sup>2</sup> Wetterslev and colleagues often estimate heterogeneity from large numbers of trials and apply the estimate back in time. I see no value in applying sequential methods retrospectively, because there is then no need for multiple tests to be applied. In short, I do not think we are quite there with the statistical methods at this point in time.

## RESPONSE to comment by Higgins

Jørn Wetterslev, Janus Engstrøm, Christian Glud and Kristian Thorlund  
Correspondence to: wetterslev@ctu.hk Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark.

## What should the Collaboration do?

Where does this leave us? This is an area of methodology about which our statisticians disagree. In such circumstances, it would be inappropriate for us to provide definitive guidance. We should instead communicate the arguments for and against these methods in as balanced a way as possible. I think there is currently a real risk of Cochrane reviews with scant evidence drawing false positive conclusions, and that some of our review teams over-interpret findings that they naively categorise as 'statistically significant'. Sequential approaches offer one means of addressing this problem, but they might be considered to swap one arbitrary specification (use of  $P < 0.05$ ) with a collection of other arbitrary specifications (including power and a clinically important effect size).

There are perhaps more important general lessons to be learned from these discussions. All Cochrane review teams require sufficient expertise to interpret results of a meta-analysis with regard to two important issues. First, the emphasis should be on looking at estimates of the magnitude of effect and the uncertainty surrounding these estimates. For most, this equates to the interpretation of CIs in the light of clinically meaningful (and clinically plausible) magnitudes of effect. Second, we need to be aware of the precariousness of inferences based on small quantities of evidence. CIs that only just exclude null effects are probably insufficient to draw firm conclusions, whether or not they arise from a sequential method. This is particularly the case when the number of studies, or the sizes of the studies, is small. In these situations, the addition of one or two more studies can lead to substantial changes in the effect estimate or CI.

In conclusion, I have no problem in principle with sequential approaches being

applied in Cochrane reviews, providing that (i) they are applied prospectively and not retrospectively, with a full analysis plan provided in the protocol; and (ii) the 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. However, it seems to me problematic to use stopping rules from sequential methods to provide firm recommendations on the need (or not) for further research unless the assumptions feeding into the analysis can be demonstrated to be universally acceptable to the array of users of *The Cochrane Library*. Further exploration of the technicalities of the methodology is warranted, as we do not yet have a method in our toolkit that I would confidently recommend.

## References

1. Higgins J. How should we interpret updated meta-analyses? [contributed paper], Abstract Book, VII Cochrane Colloquium, Rome, October 1999. p. 7.
2. Higgins JPT, Whitehead A, Simmonds MC. Sequential methods for random-effects meta-analyses. *Statistics in Medicine* 2011; 30: 903–921; doi: 10.1002/sim.4088.
3. Sterne JA, Davey Smith G. Sifting the evidence—what's wrong with significance tests? *BMJ* 2001; 322: 226–31. doi: 10.1136/bmj.322.7280.226.
4. Doyle L, Crowther C, Middleton P, Vaysey M, Marret S, Rouse D. Antenatal magnesium sulphate to prevent cerebral palsy in very preterm infants [letter]. *BJOG* 2011; 118: 891–2. doi: 10.1111/j.1471-0528.2011.02971.x.

With thanks to Jonathan Sterne and Doug Altman for helpful comments on an earlier draft of this commentary.

## Introduction

Julian Higgins was one of the first in the world to recommend and develop the adoption of sequential methods in cumulative meta-analysis. Therefore, it should come as no surprise that we agree with the vast majority of points raised in his response.

Our major disagreement lies in whether the use of the traditional significance level

of 0.05 and unadjusted 95% confidence intervals (CIs) is valid in meta-analyses where the available information has not yet reached the required information size. In a traditional meta-analysis there is one hidden assumption: the information available to us represents all we need so therefore we need not adjust the CI or the statistical significance level.



Today no one would ever accept an interpretation of the results from a single trial on a binary outcome without a pre-planned sample size calculation presented in a transparent way according to the control event proportion, the anticipated intervention effect, and the risks of type 1 and 2 errors one is willing to accept. If it is a multicentre trial or a trial with anticipated heterogeneity, researchers also recommend taking the anticipated heterogeneity into account when planning the sample size<sup>1</sup>. And if anyone should look at the data before the sample size is reached, then adjustment of the CI and the statistical significance level are required. This point of view has achieved global consent among investigators and regulators<sup>2</sup>.

It is therefore not reasonable to us why the meta-analyst, although not able to plan the acquired information size of a meta-analysis, should not interpret their results in the light of a realistic required information size (calculated as one calculates the sample size for a well powered trial). Should one seriously advocate that just because we call the accumulated data a meta-analysis then other rules apply rather than what is consented on by both the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) for interim results in a single trial? According to our point of view – and here we seem to agree with Higgins – we should not rely on sparse data and repeated significance testing in a cumulative meta-analysis.

## A controversial area

We never meant to encourage the use of significance tests. We have consistently argued (also at the Madrid meeting), that the main benefit from the use of sequential methods in meta-analysis is the increased validity of CIs when these are 'sequentially adjusted'. However, CIs and P values are close 'siblings' in any statistical analysis. When a CI does not include 1.00 any more than the P value becomes below 0.05. When a CI is far from 1.00 then the P value is far below 0.05. Traditional meta-analyses in Cochrane reviews also provide P values and their relation to 0.05. For a CI to provide a valid basis for statistical inference that an intervention seems to work, it has first of all to be statistically significant. We fully share Higgins' view on the meta-analysts' obligation to make recommendations on whether future trials should be undertaken or not. We have never advocated that firm stopping rules should be

applied for meta-analysis, simply for the reason that no one is able to enforce such an absurd rule and that in a free world any analysis can of course be conducted. Our concern is and has always been the interpretation of a meta-analysis falling short of a reasonable required information size for detecting a realistic and important effect size. In this vein, both sequentially adjusted CIs and adjusted significance levels constitute more sensible inferential measures in the light of the strength of the available evidence.

We very much agree that quantities of control event proportion, anticipated effect size, heterogeneity, and type 1 and 2 error risks have to be prespecified. Different scenarios, prospectively foreseen, can be of huge importance for planning further trials as it allows both trialists and meta-analysts to gauge the level of current evidence and the gap between the current and necessary degree of evidence to reach definite conclusions<sup>3</sup>. Trial sequential meta-analysis and other sequential meta-analyses will not govern but inform future research in an important manner.

## Technical issues

We are aware that different methods are available for sequential meta-analysis, and as presented at the Madrid meeting there are pros and cons to each of them. Applying statistical information may have advantages although it may be less transparent to the average clinician and meta-analyst. Our Trial Sequential Analysis (TSA) software is able to use both required statistical information and required information size ([www.ctu.dk/tsa](http://www.ctu.dk/tsa)). It is our impression that our TSA and Higgins' sequential meta-analysis<sup>4</sup> yield nearly similar results under comparable assumptions but further research is needed to confirm this<sup>5</sup>.

We agree that the prespecification of heterogeneity, different from the one found in the meta-analysis itself, may be important especially when there are few trials (please see our Manual on TSA at [www.ctu.dk/tsa](http://www.ctu.dk/tsa)). An empirical analysis of 16 large Cochrane meta-analyses point to the necessity of including at least 15 trials with more than 500 events in a meta-analysis for reliable estimation of heterogeneity, whether measured as  $I^2$  or  $D^{26}$ . However, totally ignoring the actually observed heterogeneity, as a forecast of the heterogeneity eventually experienced, seems also a bit obscure and we cannot see why it should not be used, e.g. as a sensitivity analysis.

## What should The Cochrane Collaboration do?

We agree with Higgins' general summary of the situation. The number of trials and participants in meta-analyses is generally very small, both within and outside *The Cochrane Library*. This necessitates the use of methods considering adjustments of CIs and significance levels according to this situation in order to make inferences more realistic in the light of a realistic required information size. The sequential methods, with their specification of event control proportion, anticipated intervention effects size, heterogeneity, and type I and type II errors, transparently facilitates a sensible framework for more realistic inferences. This framework is also in full alignment with the recommendations of the GRADE Group<sup>7</sup>. Moreover, we have no intention of swapping one arbitrary specification with a collection of others. However, one huge problem of the traditional meta-analysis is that the specification of the arbitrary  $P < 0.05$  actually does hide that the reliability of such a finding is based on an assumption that the available information is sufficiently large. Most Cochrane meta-analyses include less than seven or eight trials, which are usually small. The sequential methods transparently let the specification of event control proportion, anticipated intervention effects size, heterogeneity, and type I and type II errors out in the open and calculate their implications for inference from the results.

In conclusion, we fully support that TSA should be applied prospectively and that the assumptions should be described transparently. As already stated, we do not advocate the use of firm stopping rules but support the inference from sequential meta-analysis, based on transparent assumptions feeding the analysis.

These requirements to sequential methods should be applied to traditional meta-analysis as well – but this is presently rarely occurring. No traditional meta-analysis highlight the hidden or underlying assumption that the available data analysed represents sufficient information for concluding on a specific intervention effect size. The main problem right now is that inferences from many traditional meta-analyses with unadjusted CIs are obviously unreliable<sup>7,8,9</sup> and the underlying assumption of sufficiently large information sizes is not tested. Therefore, for this and several other reasons presented



above, sequential meta-analyses represent a step forward in transparency and inference from cumulative meta-analysis. In 1993 we made a bold step forward by creating The Cochrane Collaboration. At that time, the first edition of the *Cochrane Handbook for Systematic Reviews of Interventions* was still rudimentary, if at all existing, yet several meta-analyses were conducted due to demand for research synthesis on several important clinical questions. Today, the *Cochrane Database of Systematic Reviews* contains an impressive 5102 systematic reviews. In other words, there is a myriad of clinical answers to act on. However, as the clinical risks and cost of acting on evidence that may not be reliable are high, it becomes essential to assess the certainty surrounding the available clinical answers.

We recognise that the adoption of sequential methods may be daunting to many Cochranites and systematic reviewers in general. The 'risk' that one's apparently statistically significant findings may come into question with use of sequential methods, may not seem appealing. Also the diminishing promise of statistically significant findings may discourage new authors from performing systematic reviews. The evidence, however, sends a clear message: sequential meta-analyses represent a step forward in transparency and reliability of inferences from cumulative

meta-analysis. It is our hope that the Cochranites around the world can find in them the same boldness that led to the creation of The Cochrane Collaboration, and embrace the use of sequential methods as a natural progression in meta-analysis methodology. We think it is not optimal for the Collaboration or patients, if Higgins is not prepared to take the step now. At the time of the first edition of the *Cochrane Handbook for Systematic Reviews of Interventions* it was still rudimentary. If one waits for the perfect *Handbook* on sequential analysis, one runs the risk of letting the perfect become the enemy of the good.

## References

1. Fedorov V, Jones B. The design of multicentre trials. *Statistical Methods in Medical Research* 2005; 14: 205–248.
2. Jennison C, Turnbull B. *Group sequential methods with applications to clinical trials*. Chapman & Hall/CRC Press, 2000.
3. Thorlund K, Anema A, Mills E. Interpreting meta-analysis according to the adequacy of sample size. An example using isoniazid chemoprophylaxis for tuberculosis in purified protein derivative negative HIV-infected individuals. *Clinical Epidemiology* 2010; 2: 57–66.
4. Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-

effects meta-analysis. *Statistics in Medicine*. 2011; 30(9): 903–21. doi: 10.1002/sim.4088.

5. Imberger G, Gluud C, Wetterslev J. Comments on 'Sequential methods for random-effects meta-analysis' by J. P. Higgins, A. Whitehead and M. Simmonds. *Statistics in Medicine* 2010; doi: 10.1002/sim.4088. *Statistics in Medicine*. 2011; 30(24): 2965–6
6. Thorlund K, Imberger G, Johnston BC, Walsh M, Awad T, Thabane L, Gluud G, Devreux PJ, Wetterslev J. Evolution of heterogeneity ( $I^2$ ) estimates and their 95% confidence intervals in large meta-analyses. *PLoS One* 2012; 7(7): e39471. doi: 10.1371/journal.pone.0039471.
7. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6: Rating the quality of evidence – Imprecision. *Journal of Clinical Epidemiology* 2011; 64(12): 1283–1293.
8. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, et al. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis – a simulation study. *PLoS One* 2011; 6(10): e25491.
9. Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *Journal of Clinical Epidemiology*. 2011; 64(10): 1060–9.

## Use of logic models in the context of systematic reviews

Laurie M. Anderson

Correspondence to: landerson@wsipp.wa.gov  
Washington State Institute of Public Policy, Olympia, Washington 98504-0999, USA.

There has been increased discussion recently about conducting reviews of complex interventions (see also pages x). What is or is not a complex intervention may be open to debate, but the difficulty of teasing out the numerous pathways by which an intervention is purported to work is familiar to many systematic reviewers. A well-formulated systematic review question gives clarity to the review process by establishing support for a set of eligibility criteria, identifying key domains for the literature searching, determining the types of information that should be

collected and coded from each study, and by providing a frame for reporting results in light of the evidence assembled by the review.<sup>1</sup> Using a logic model in the context of systematic reviews may help illuminate the review question further by depicting how an intervention is thought to work.

The terms 'logic model', 'analytic framework', 'conceptual framework', 'concept map', and 'influence diagram' have been used interchangeably. In the field of programme evaluation, the term 'logic model' typically refers to schematics that show programme inputs and outputs and are used to identify programme resource requirements and establish programme accountability criteria. Our discussion of the use of logic models in systematic reviews is for a different application – that of depicting models of biological, behavioural, or social processes and systems. The practical contribution of logic models, both conceptual and analytical, is evident at most stages of the systematic review process, as illustrated

in Box 1. These stages range from conceptualising the review focus to drawing conclusions about the cumulative evidence.<sup>2</sup>

How an intervention is thought to work can be hypothesized differently by different stakeholders, often reflecting their disciplinary training and orientation. A logic model can help align multidisciplinary definitions of similar phenomena and bring complementary perspectives to similar problems. Graphically illustrating causal pathways, proximal and distal outcomes, and mediators or moderators of intervention effects, provides a rational starting point and allows a common understanding of the conceptual boundaries of the review among reviewers and their collaborators, which increases the probability that the effort will provide actionable information.

As an example, the logic model in Figure 1 was used in a systematic review

# Statistical methods for reliably updating meta-analyses

Mark Simmonds

Centre for Reviews and Dissemination  
University of York, UK

With:

Julian Elliott, Joanne McKenzie, Georgia Salanti,  
Adriani Nikolakopoulou, Julian Higgins

THE UNIVERSITY *of York*  
Centre for Reviews and Dissemination



# Some issues

- When can we stop updating a review?
- Conclusions can change over time
  - Risk of error if we stop too soon
- Type I error inflated by performing multiple analyses



# Controlling error

- Adapted from sequential clinical trial design
  - Sequential meta-analysis (*Higgins, Simmonds, Whitehead 2010*)
    - Includes Bayesian adjustment of heterogeneity
  - Trial sequential analysis (*Wetterslev, Thorlund, Brok, Gluud 2008*)
- Control Type I error
  - Law of Iterated Logarithm (*Lan, Hu, Cappelleri 2007*)
  - “Shuster-Pocock” method (*Shuster, Neu 2013*)
- Other methods
  - Fully Bayesian analysis
  - Robustness or stability of analysis
  - Consequences of adding new studies
  - Power gains from adding new studies



# Analyses of updated Cochrane reviews

- Searched for Cochrane reviews:
  - Updated in 2014-2015
  - At least one new trial added
  - At least one meta-analysis
    - That is statistically significant
    - At least 3 trials
- Included 76 reviews and 286 meta-analyses
  - 62% had statistically significant results
  - 44% were of sufficient size to have 80% power to detect observed effect.

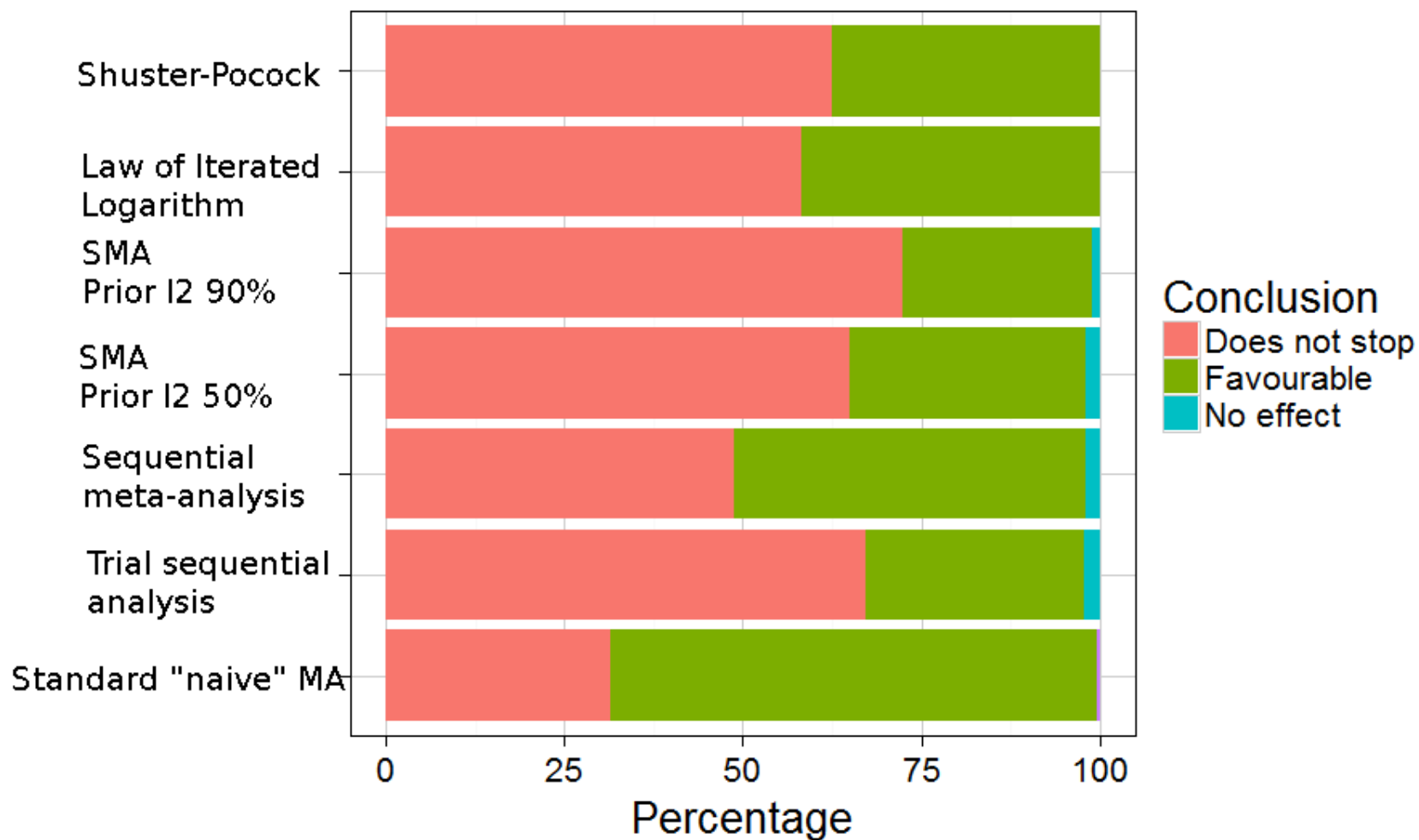
# Assumptions

- Analysis using log odds ratio or SMD
- A new meta-analysis for each added trial
- 5% Type I error, 90% power
- “Desired” effect is same as observed
- Meta-analyses are uncorrelated

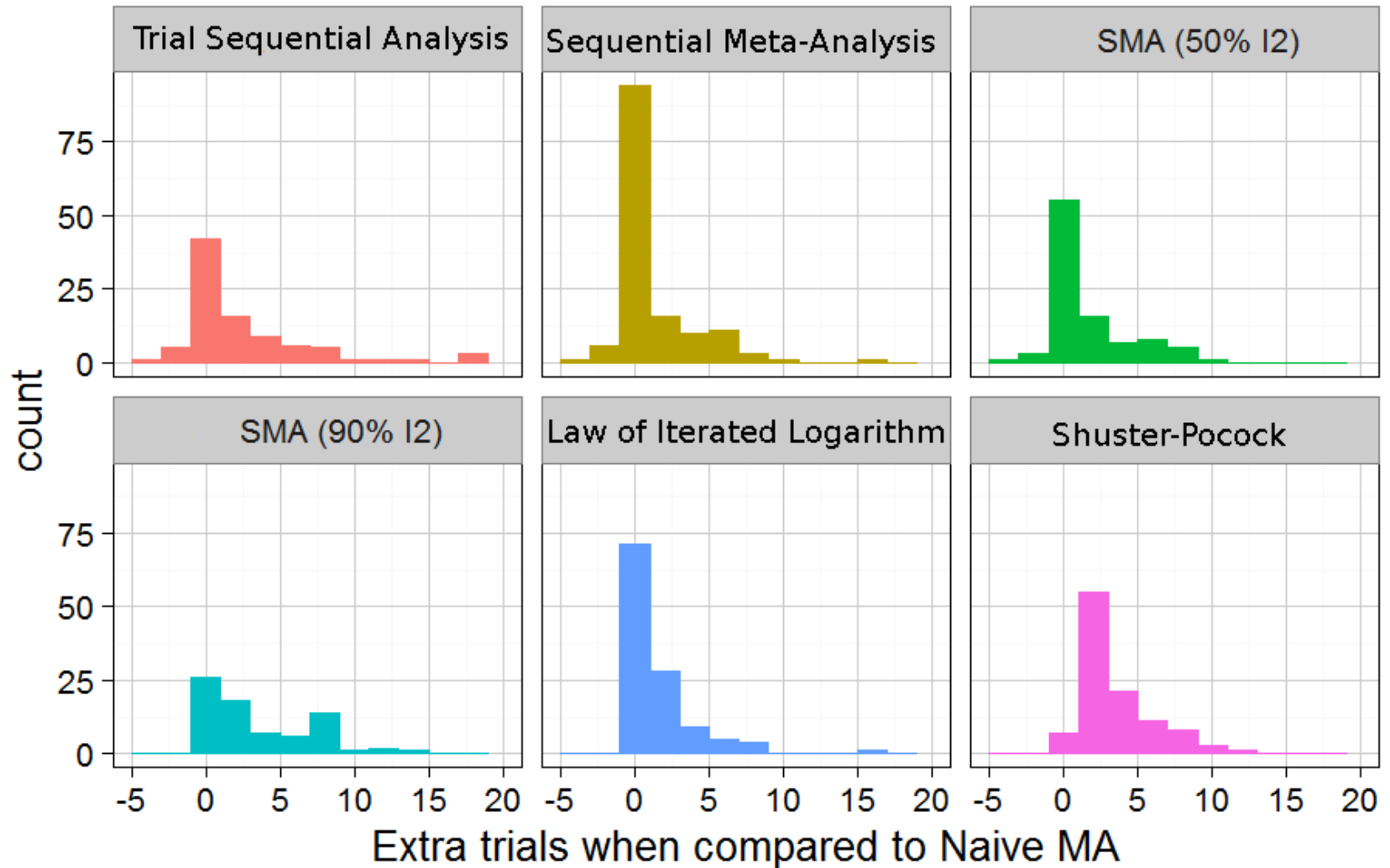




# Conclusions of analyses



# Additional trials to reach a conclusion



# “Inappropriate positives”

Conclusions of updated meta-analysis  
where analysis with all trials is not statistically significant

Method	Does not stop	Evidence of effect	No evidence of effect
Naïve MA	83.8	15.2	-
Trial Sequential Analysis	99.0	0	1.0
Sequential Meta-Analysis	99.0	0	1.0
SMA (50% I2)	99.0	0	1.0
SMA (90% I2)	100	0	0
Law of Iterated Logarithm	98.1	1.9	-
Shuster-Pocock	98.1	1.9	-

# Conventional “Naïve” analysis

- Too many inappropriate positive conclusions
  - Elevated Type I error rate
  - But not vastly elevated for most updated reviews?
- Biased estimates of effect
- Significant results are often based on too little evidence?

# Controlling for error

- All methods appear to control for Type I error
- Increased complexity
- Need to select desired effect size, adjust for heterogeneity etc.
- May take longer before stopping



# Do we need these methods?

- Is the problem with “naïve” analysis serious enough in real Living Systematic Reviews?
- Do the methods needlessly delay a statistically significant result?
- When should they be implemented?
  - As part of protocol?
  - Only with statistically significant results?



# Implications for Living Systematic Reviews

- Reviews with many updates
  - Increased risk of type I error
  - Methods probably needed
- Starting with few trials
  - Need to identify required sample size
  - Methods needed as a caution if results statistically significant?
- Starting with many trials
  - Little new data expected, update for consistency
  - Methods not needed?

# Cochrane Scientific Committee

## Briefing report 2 – For sign off and recommendation

**Date:** 18<sup>th</sup> May 2017

**CSC:** 1:17

**Agenda item:** 6i

**Priority:** Medium

**Open access/restricted:** Open

### REVIEW OF THE UPDATED 'RISK OF BIAS' TOOL ROB 2.0

**Lead developers/investigators:** Julian Higgins, Jonathan Sterne, Jelena Savović, Matt Page, Asbjørn Hróbjartsson, Isabelle Boutron, Barney Reeves, and Sandra Eldridge.

**Summary of method or 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](https://doi.org/10.1002/14651858.CD201601). (Extract in Annex below).

The following table lists the tools and guidance for the different versions and are available in the Dropbox.

<b>Individually randomized, parallel group trials</b>	<ol style="list-style-type: none"> <li>1. Guidance for using the RoB 2.0 tool for individually randomized trials</li> <li>2. The tool</li> <li>3. Blank templates with two variants:               <ol style="list-style-type: none"> <li>a. RoB 2.0 when interest is in the effect of assignment to intervention</li> <li>b. RoB 2.0 when the interest is in the effect of starting and adhering to intervention</li> </ol> </li> </ol>
<b>Cluster randomized, parallel group trials</b>	<ol style="list-style-type: none"> <li>1. Guidance for using the RoB 2.0 tool for cluster-randomized trials.</li> <li>2. The tool (cluster-randomized trials)</li> <li>3. Blank template with one variant               <ol style="list-style-type: none"> <li>a. RoB 2.0 for cluster randomized r trials when the interest is in the effect of starting and adhering to intervention.</li> </ol> </li> </ol>
<b>Individually randomized, cross-over trials</b>	<ol style="list-style-type: none"> <li>1. Guidance for using the RoB 2.0 tool for cross-over trials</li> <li>2. The tool (cross-over trials).</li> <li>3. Blank templates with two variants:</li> </ol>

	<ul style="list-style-type: none"><li>a. RoB 2.0 for cross-over trials when interest is in the effect of assignment to intervention</li><li>b. RoB 2.0 for cross-over trials when the interest is in the effect of starting and adhering to intervention.</li></ul>
--	---

requirements and interests of Cochrane members. Such a project would also require resources.

3. Many Cochrane members are probably using personal current awareness services to keep up to date. These are very varied and can take many forms including RSS feeds, email discussion lists, signing up to receive the contents pages of journals, reading blogs and subscribing to newsletters from key organizations. Saved searches can be set up on large bibliographic databases such as MEDLINE and members receive the results of any new publications meeting the search criteria in an email. These are both passive and active ways to receive new information on methods of interest. A recent innovation via PubMed is a PubMed Systematic Review methods filter (see page 27). This is a tool we will likely wish to evaluate to see how far it meets our individual and general methods needs. There are also a range of academic websites such as ResearchGate, Academia.edu and ORCID as well as professional networking services, such as LinkedIn, that offer ways to keep up to date with methods developments by following authors and their outputs.
4. The *Cochrane Handbooks* (for systematic reviews of interventions and diagnostic test accuracy) are one excellent form of evidence-based guidance and provide summaries of current best evidence-based methods to those of us lacking time to review methods developments. However, we keenly await their update. *Cochrane Handbook* chapters would benefit from transparent reporting of their update

intended to save searchers the time needed to keep up to date on methods development in various searching issues. The SuRe Info approach provides a possible template for other methods topics.

5. Cochrane needs to funnel its scarce resources by promoting awareness of methods developments. A useful approach would be to evaluate current awareness resources for their currency, overlap and added value, and provide Cochrane members with guidance about which are most reliable for their disciplines. This could start with the creation of an inventory or portal with links to the different resources, perhaps via the Cochrane Methods website. If the resources are to be evaluated, it might be helpful to canvas Cochrane members for information about what they need to know about a methods resource in order to decide how to use it. For example, which methods are covered, how the current awareness is carried out, whether it is selective or comprehensive, how are updates supplied and how frequently. There are examples of this type of approach, such as the now retired resource, 'Intute' ([www.webarchive.org.uk/wayback/archive/20140614081921/http://www.jisc.ac.uk/whatwedo/programmes/reppres/irs.aspx](http://www.webarchive.org.uk/wayback/archive/20140614081921/http://www.jisc.ac.uk/whatwedo/programmes/reppres/irs.aspx)), or the EFSA inventory ([www.metaxis.com/EFSAInventory/](http://www.metaxis.com/EFSAInventory/)). Individuals might then amend or develop their own personal approaches to keeping up to date. They can also collaborate by contributing suggestions for new updating resources to the inventory or portal to share with others.

## A revised tool for assessing risk of bias in randomized trials

Julian PT Higgins,<sup>1</sup> Jonathan AC Sterne,<sup>1</sup> Jelena Savović,<sup>1</sup> Matthew J Page,<sup>1</sup> Asbjørn Hróbjartsson,<sup>2</sup> Isabelle Boutron,<sup>3</sup> Barney Reeves,<sup>4</sup> Sandra Eldridge<sup>5</sup>

<sup>1</sup>School of Social and Community Medicine, University of Bristol, Bristol, UK

<sup>2</sup>Research Unit for Evidence-Based Medicine, University of Southern Denmark, Odense, Denmark

<sup>3</sup>University Paris Descartes, Paris, France

<sup>4</sup>School of Clinical Sciences, University of Bristol, Bristol, UK

<sup>5</sup>Centre for Primary Care & Public Health, Queen Mary University of London, London, UK

Correspondence to: Julian.Higgins@bristol.ac.uk

School of Social and Community Medicine, University of Bristol, Bristol, UK

The Cochrane 'Risk of bias' tool for randomized trials is widely used in both Cochrane Reviews and non-Cochrane reviews on the effects of interventions. First released in 2008,<sup>1</sup> and revised slightly in 2011,<sup>2</sup> the tool seeks to determine whether the findings of a randomized trial can be believed. The tool is flexible and is implemented in different ways by different review teams. The default (and recommended) implementation is to examine six items: sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel

(performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias) and selective outcome reporting (reporting bias).<sup>2</sup>

Although it has become the dominant tool for assessing risk of bias in randomized trials, we have identified several ways in which it can be improved.<sup>3,4</sup> Over the last year, we assembled collaborators from across Cochrane and elsewhere to develop a new version of the tool, which we refer to as RoB 2.0. We expect to be announcing RoB 2.0 during the Seoul Colloquium in October 2016. Here we outline 10 key changes in RoB 2.0 compared with the first version of the tool.

### 1. The assessment targets a specific result

It is widely understood that assessments of risk of bias need to be outcome-specific. For example, the importance of masking (or blinding) outcome assessors may depend on how objectively the outcome can be measured. We take this idea one step further in RoB 2.0, and make it clear that the assessment is typically specific to a particular result. For example, if two analyses are presented, one following intention-to-treat principles and the other not, then the risk of bias in the two results may be different. Of course, some items in the tool relate to all outcomes in the trial (e.g. those related to randomization methods), some relate to specific outcomes (e.g. blinded outcome assessment) and some relate to the specific result.

### 2. Nomenclature

The tool will continue to involve domain-based assessments, such that different domains of bias are assessed individually. We have modified our terminology to explain more clearly which



methods and procedures are being addressed in each domain. For example, the first domain relates to 'Bias arising from the randomization process'. Terms such as 'selection bias' and 'performance bias' are problematic and will be dropped, because they are either interpreted differently – or are not known – by many people outside Cochrane.<sup>5</sup>

### 3. A fixed set of bias domains

The revised tool will have a fixed set of domains, which are intended to cover all issues that might lead to a risk of bias. These domains cannot be removed from the tool, and new domains cannot be added. We list the domains in the Table. They are broadly the same as in the existing tool, although the existing items on sequence generation and allocation concealment are now grouped into a single domain about the randomization process. As in the current tool, there is no domain for issues around funding or vested interests. Members of the Cochrane Bias Methods Group are developing a parallel tool to assess this, to be used in conjunction with RoB 2.0.

### 4. Different templates for different types of trial

To retain flexibility to tailor the tool to the type of trial being assessed, we will provide different templates for different trial designs. In the first instance, there will be templates for simple parallel-group trials, cluster-randomized trials, and (two-treatment, two-period) cross-over trials. The template for assessing a cluster-randomized trial includes additional consideration of the possibility of bias in recruiting participants into clusters once the interventions assigned to clusters are known. The template for assessing a cross-over trial includes additional consideration of issues such as carry-over and period effects.

### 5. Modification to response options for 'Risk of bias' judgements

Whereas in the current tool the available response options are 'high risk', 'low risk' and 'unclear risk', in RoB 2.0 the response options are 'high risk', 'low risk' and 'some concerns'. The last option reflects the situation in which the rater has insufficient confidence to assess the risk of bias to be low, but not enough information to classify the risk of bias as high.

### 6. Introduction of 'signalling questions'

To help reach 'Risk of bias' judgements, we are introducing a series of 'signalling questions' within each domain. These questions are reasonably factual in nature, following the approach used in other tools including QUADAS-2<sup>6</sup> and the new ROBINS-1 tool for non-randomized studies of interventions.<sup>7</sup> For example, in the first domain about the randomization process, the opening two questions are 'Was the allocation sequence random?' and 'Was the allocation sequence concealed?'

### 7. Algorithms to reach 'Risk of bias' judgements

We have developed algorithms to map responses to the signalling questions to judgements about risk of bias. When implemented in software, these should facilitate the process of reaching 'Risk of bias' judgements. It will be possible to over-ride the default mappings for situations in which special issues or concerns apply.

### 8. An overall judgement about risk of bias

The current tool does not have a formal procedure for reaching an overall judgement about risk of bias in the result. We have introduced one into RoB 2.0, by implementing a rule that the overall risk of bias for the result is the 'worst' risk of bias recorded across all domains in the tool. Thus, if any domain is assessed to be at high risk of bias, then so is the result overall. If all domains are either assessed to be at low risk of bias or have some concerns, then the overall result for the study has some concerns. It is possible to

over-ride the rule if there are some concerns in multiple domains, which may lead to a judgement of high risk of bias overall.

### 9. Differentiation between the effect of assignment to intervention and the effect of starting and adhering to intervention

Perhaps the most challenging change in the new tool is that we draw an important distinction between two intervention effects that might be of interest to the review author. The first is the effect of participants being assigned to the interventions: this is the effect that is estimated in an intention-to-treat analysis of a trial. It is not essential that individuals are blinded or that participants adhere to the assigned intervention for the result to be unbiased. The second is the effect of participants starting and adhering to the intervention. For this effect, it is essential to examine issues such as adherence, unintended co-interventions and whether the intervention was implemented successfully. We offer two different templates for the 'Risk of bias' assessment, forcing the rater to decide the effect to which 'Risk of bias' assessments relate.

### 10. Reconsideration of selective (non-)reporting

The current tool encourages a study-level judgement about whether there has been selective reporting, in general, of the trial results. The new tool focuses solely on the specific result being assessed for risk of bias. If there is no result (e.g. if it has selectively been omitted from the report) then there is no 'Risk of bias' assessment. Selective non-reporting is therefore not covered by the tool, and should be assessed at the level of the synthesis across studies. Instead, RoB 2.0 examines whether the specific result from the trial is likely to have been selected from multiple possible results on the basis of the findings. This will be either because several alternative outcome measures are available, or because several statistical analyses were performed.

The current tool	The revision: RoB 2.0
Random sequence generation ( <i>selection bias</i> )	Bias arising from the randomization process
Allocation concealment ( <i>selection bias</i> )	
Blinding of participants and personnel ( <i>performance bias</i> )	Bias due to deviations from intended interventions
Incomplete outcome data ( <i>attrition bias</i> )	Bias due to missing outcome data
Blinding of outcome assessment ( <i>detection bias</i> )	Bias in measurement of the outcome
Selective reporting ( <i>reporting bias</i> )	Bias in selection of the reported result
Other bias	(No formal strategy for assessing overall risk of bias)
(No formal strategy for assessing overall risk of bias)	Overall bias

### Closing remarks

We believe the new tool will offer considerable advantages over the existing tool. Extensive piloting has taken place already, including a three-day event held in Bristol in February 2016 and a subsequent round of remote piloting by individuals with varying degrees of relevant experience. The piloting has informed development of the tool as well as of the lengthy written guidance



that accompanies it. Once programmed into software, we expect the tool will be easier to use than the first version. Some issues remain to be resolved, however, such as how many results should be assessed for each study, and how best to integrate the assessment into the data extraction process. We look forward to discussions at the 2016 Seoul Colloquium and beyond about how the tool might be adopted and implemented by Cochrane Review Groups and author teams.

For further details of RoB 2.0, see [www.riskofbias.info](http://www.riskofbias.info)

#### Acknowledgements

Development of RoB 2.0 was supported by the UK Medical Research Council Network of Hubs for Trials Methodology Research (MR/L004933/1-N61).

We thank all of the numerous contributors to our working groups, discussion meetings and piloting initiatives for their help in developing the tool.

#### References

1. Higgins JPT, Altman DG, eds. Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester (UK): John Wiley & Sons, 2008:187-241.

2. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;**343**:d5928.
3. Savović J, Weeks L, Sterne JAC, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. *Systematic Reviews* 2014;**3**:37.
4. Jorgensen L, Paludan-Muller AS, Laursen DR, Savovic J, Boutron I, Stern JA, et al. Evaluation of the Cochrane tool for assessing risk of bias in randomized clinical trials: overview of published comments and analysis of user practice in Cochrane and non-Cochrane reviews. *Systematic Reviews* 2016;**5**(1):80.
5. Mansournia MA, Higgins JPT, Sterne JAC, Hernan MA. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology* (in press).
6. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.
7. Sterne JAC, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ* (in press).

## Cochrane Systematic Reviews in a world with data sharing

Isabelle Boutron

Correspondence to: [isabelle.boutron@aphp.fr](mailto:isabelle.boutron@aphp.fr)  
Cochrane France, Hotel Dieu Hospital, Paris, France

Cochrane systematic reviews aim to identify all available evidence to answer specific questions. Hence, Cochrane requires authors to perform a very extensive search relying on electronic bibliographic databases, trial registries and relevant 'grey' literature sources. Trial registries are particularly useful to identify unpublished studies,<sup>1</sup> complete the information available in a published article when results are posted,<sup>2</sup> and obtain results for unpublished observational studies.<sup>3</sup> Searching clinical trial registries is also crucial to detect the selective reporting of outcomes. For example, the COMPare project ([compare-trials.org](http://compare-trials.org)) led by Ben Goldacre, systematically checked every trial published in the top five medical journals and showed that of the 67 trials evaluated, only nine were perfect; 354 outcomes were not reported and 357 new outcomes were silently added.

However, despite these extensive searches and the availability of results of unpublished studies through trial registries, systematic reviewers have access to only a very limited amount of existing evidence. In fact, only 61% of drug trials, 30% of non-drug trials and 10% of observational studies are registered.<sup>4-6</sup> Furthermore, only 13.4% of trials post their summary results within 12 months after trial completion.<sup>7</sup> We also lack information about the methodological flaws of the trials included in a systematic review. For example, examination of all the trials included in a meta-analysis of a primary outcome in the Cochrane Reviews published in one year showed that 41% had at least one risk of bias domain rated as being at unclear risk of bias. Finally, Jefferson and colleagues demonstrated that the 'devil is in the detail', as only when examining all the informa-

tion available in the clinical study report and other regulatory documents closely can systematic reviewers really understand what happened during a trial and evaluate the possible flaws.<sup>8</sup> This appalling situation<sup>1,9-13</sup> has convinced several stakeholders to engage actively in an era of open science through data sharing. The US Institute of Medicine issued a consensus report that recommends ways to promote responsible data sharing, including when and how to share data to maximize benefits and minimize risks.<sup>14,15</sup> The European Medicine Agency is also engaged in this debate<sup>16</sup> and on 14 April 2015, the World Health Organization published a new statement on the public disclosure of clinical trial results.<sup>17</sup> The International Committee of Medical Journal Editors recently proposed requirements to help meet the obligations for data sharing.<sup>18</sup>

The pharmaceutical companies are part of this movement. Some pharmaceutical companies have committed to share patient data for the studies they sponsor. A website, Clinical Study Data Request (CSDR) ([www.clinicalstudydatarequest.com](http://www.clinicalstudydatarequest.com)), favours centralized communication between sponsors and researchers. Sponsors and funders are invited to list the studies for which they agree to share individual patient data on the website and to define the conditions for sharing these data. Researchers can use this site to request access to anonymized patient-level data and supporting documents from clinical studies to conduct further research. Their research proposals for access are reviewed by an Independent Review Panel. The Yale University Open Data Access (YODA) Project is also committed to open science and data transparency and supports research that attempts to produce concrete benefits to patients, the medical community and society as a whole. For instance, this project has allowed for the release of two systematic reviews on recombinant human bone morphogenetic protein-2 (rhBMP-2), based on patient-level data from all clinical trials conducted by Medtronic.<sup>19-21</sup>

Other initiatives aim to organize the system to help researchers access all information related to a given clinical trial. The threaded publications initiative was recently launched