

# Introducing a revised tool for assessing risk of bias in randomized trials (RoB 2.0)

Jelena Savović

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

With special thanks to Julian Higgins, Matt Page, Jonathan Sterne, Roy Elbers, Barney Reeves, Asbjørn Hróbjartsson, Isabelle Boutron and all RoB 2.0 collaborators

## RESEARCH METHODS & REPORTING

### The Cochrane Collaboration's tool for assessing risk of bias in randomised trials

Julian P T Higgins,<sup>1</sup> Douglas G Altman,<sup>2</sup> Peter C Gøtzsche,<sup>3</sup> Peter Jüni,<sup>4</sup> David Moher,<sup>5,6</sup> Andrew D Oxman,<sup>7</sup> Jelena Savović,<sup>8</sup> Kenneth F Schulz,<sup>9</sup> Laura Weeks,<sup>9</sup> Jonathan A C Sterne,<sup>8</sup> Cochrane Bias Methods Group  
Cochrane Statistical Methods Group

Flaws in the design, conduct, analysis, and reporting of randomised trials can cause the effect of an intervention to be underestimated or overestimated. The Cochrane Collaboration's tool for assessing risk of bias aims to make the process clearer and more accurate

als without producing a score).<sup>6,7</sup> Until recently, Cochrane reviews used a variety of these tools, mainly checklists.<sup>8</sup> In 2005 the Cochrane Collaboration's methods groups embarked on a new strategy for assessing the quality of randomised trials. In this paper we describe the collaboration's new risk of bias assessment tool, and the process by which it was developed and evaluated.

#### Development of risk assessment tool

In May 2005, 16 statisticians, epidemiologists, and review authors attended a three day meeting to develop the new tool. Before the meeting, JPTH and DGA compiled an extensive list of potential sources of bias in clinical trials. The items on the list were divided into seven areas: generation of the allocation sequence; concealment of the allocation sequence; blinding; attrition and exclusions; other generic sources of bias; biases specific to the trial design (such as crossover or cluster randomised trials); and biases that might be specific to a clinical speciality. For each of the seven areas, a nominated meeting participant prepared a review of the empirical evidence, a discussion of specific issues and uncertainties, and a proposed set of criteria for assessing protection from bias as adequate, inadequate, or unclear, supported by examples.

Randomised trials, and systematic reviews of such trials, provide the most reliable evidence about the effects of healthcare interventions. Provided that there are enough participants, randomisation should ensure that participants in the intervention and comparison groups are similar with respect to both known and unknown prognostic factors. Differences in outcomes of interest between the different groups can then in principle be ascribed to the causal effect of the intervention.<sup>1</sup>

Causal inferences from randomised trials can, however, be undermined by flaws in design, conduct, analyses, and reporting, leading to underestimation or overestimation of the true intervention effect (bias).<sup>2</sup> However, it is usually impossible to know the extent to which biases have affected the results of a particular trial.

Systematic reviews aim to collate and synthesise all studies that meet prespecified eligibility criteria<sup>3</sup> using methods that attempt to minimise bias. To obtain reliable conclusions, review authors must carefully consider the potential limitations of the included studies. The notion of study "quality" is not well defined but relates to the extent to which its design, conduct, analysis, and presentation were appropriate to answer its research question. Many tools for assessing the quality of randomised trials are available, including scales (which score the trials) and checklists (which assess tri-

During the meeting decisions were made by informal consensus regarding items that were truly potential biases rather than sources of heterogeneity or imprecision. Potential biases were then divided into domains, and strategies for their assessment were agreed, again by informal consensus, leading to the creation of a new tool for assessing potential for bias. Meeting participants also discussed how to summarise assessments across domains, how to illustrate assessments, and how to incorporate assessments into analyses and conclusions. Minutes of the meeting were transcribed from an audio recording in conjunction with written notes.

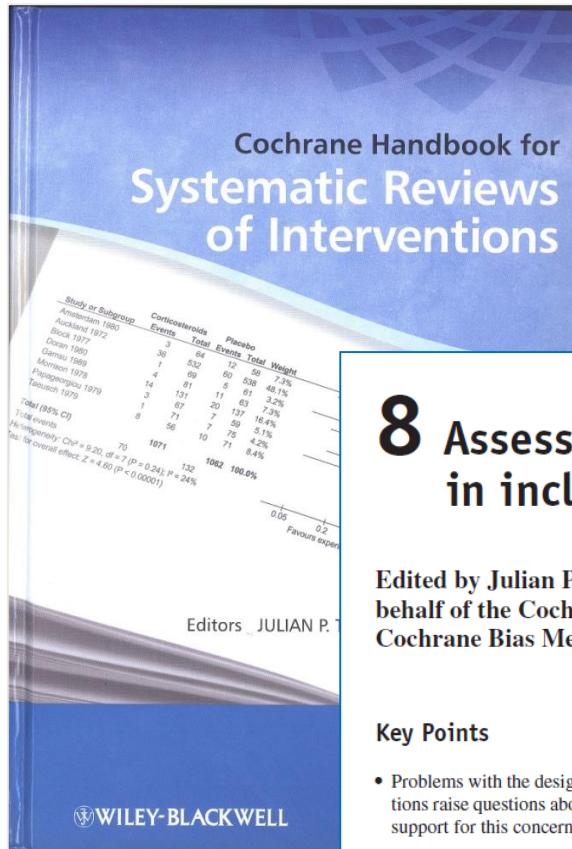
After the meeting, pairs of authors developed detailed criteria for each included item in the tool and guidance for assessing the potential for bias. Documents were shared and feedback requested from the whole working group (including six who could not attend the meeting). Several email iterations took place, which also incorporated feedback from presentations of the proposed guidance at various meetings and workshops within the Cochrane Collaboration and from

## 8 Assessing risk of bias in included studies

Edited by Julian PT Higgins and Douglas G Altman on behalf of the Cochrane Statistical Methods Group and the Cochrane Bias Methods Group

### Key Points

- Problems with the design and execution of individual studies of healthcare interventions raise questions about the validity of their findings; empirical evidence provides support for this concern.
- An assessment of the validity of studies included in a Cochrane review should emphasize the risk of bias in their results, i.e. the risk that they will overestimate or underestimate the true intervention effect.
- Numerous tools are available for assessing methodological quality of clinical trials. We recommend against the use of scales yielding a summary score.
- The Cochrane Collaboration recommends a specific tool for assessing risk of bias in each included study. This comprises a description and a judgement for each entry in a 'Risk of bias' table, where each entry addresses a specific feature of the study. The judgement for each entry involves answering a question, with answers 'Yes' indicating low risk of bias, 'No' indicating high risk of bias, and 'Unclear' indicating either lack of information or uncertainty over the potential for bias.



#### SUMMARY POINTS

Systematic reviews should carefully consider the potential limitations of the studies included

The Cochrane Collaboration has developed a new tool for assessing risk of bias in randomised trials

The tool separates a judgment about risk of bias from a description of the support for that judgment, for a series of items covering different domains of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "Subjects were randomised in blocks of six to one of the two treatment groups using sequentially numbered, sealed opaque envelopes."  Comment: sequence generation not reported.
Allocation concealment (selection bias)	Low risk	Quote: "Subjects were randomised in blocks of six to one of the two treatment groups using sequentially numbered, sealed, opaque envelopes."  Comment: allocation process adequate.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Quote: "Because the study was not blinded, secondary absorbent dressing and peri ulcer treatments used were at the discretion of the investigator."  Comment: stated as not being blinded.
Blinding of outcome assessment (detection bias) All outcomes	High risk	Quote: "Because the study was not blinded, secondary absorbent dressing and peri ulcer treatments used were at the discretion of the investigator."  Comment: stated as not being blinded.
Incomplete outcome data (attrition bias) All outcomes	High risk	Comment: numbers withdrawing and reasons reported by group (Group 1: 14/60 (23%); Group 2: 5/58 (9%)) but a higher proportion of participants withdrew from Group 2 and analysis not undertaken as ITT.
Selective reporting (reporting bias)	Unclear risk	Comment: although all trial outcomes described in the published report are in the supplied RCT protocol, it was unclear from the published report what the primary outcomes were (maceration in the protocol). A secondary outcome of 'ability to adapt' in the protocol (translated from Danish) is not identifiable in the published report.

- 
- Cochrane RoB tool is very widely used (Jørgensen 2016)
    - 100 out of 100 Cochrane reviews from 2014 (100%)
    - 31 out of 81 non-Cochrane review (38%)
  - >2700 citations from non-Cochrane sources
  - The scientific debate on risk of bias has continued
  - Evaluation studies of the tool
    - User experience: survey and focus groups (Savovic 2014)
    - Inter-agreement studies (e.g. Hartling 2009 & 2013)
    - Actual use in reviews and published comments (Jørgensen 2016)

- Used **simplistically**
- Used **inconsistently** (domains added or removed)
- Modest **agreement** rates
- RoB judgements are **difficult** for some domains
- Challenges with **unblinded trials**
- Not well suited to **cross-over trials or cluster-randomized trials**
- Not well set up to assess **overall risk of bias**

- The revised tool for randomized trials (**RoB 2.0**) is supported by the UK **Medical Research Council** Network of Hubs for Trials Methodology Research (MR/L004933/1- N61)

- Revision of the RoB tool started in May 2015
- 1<sup>st</sup> Development meeting held in Bristol in August 2015
- First ‘working draft’ of the tool completed January 2016
- Piloting phase Feb – March 2016
- Revised ‘working draft’
- 2<sup>nd</sup> Development meeting to be held in Bristol on 21-22 April 2016
- Development of further guidance and piloting
- Launch at the **Seoul Colloquium**

- Core group:
  - **Jelena Savović, Julian Higgins, Matthew Page, Asbjørn Hróbjartsson, Isabelle Boutron, Barney Reeves, Roy Elbers, Jonathan Sterne**
- Working Group members:
  - **Doug Altman, Natalie Blencowe, Mike Campbell, Christopher Cates, Rachel Churchill, Mark Corbett, Nicky Cullum, Francois Curtin, Amy Drahota, Sandra Eldridge, Jonathan Emberson, Bruno Giraudeau, Jeremy Grimshaw, Sharea Ijaz, Sally Hopewell, Asbjørn Hróbjartsson, Peter Jüni, Jamie Kirkham, Toby Lasserson, Tianjing Li, Stephen Senn, Sasha Shepperd, Ian Shrier, Nandi Siegfried, Lesley Stewart, Penny Whiting**
- And: **Henning Keinke Andersen, Mike Clarke, Jon Deeks, Geraldine MacDonald, Richard Morris, Mona Nasser, Nishith Patel, Jani Ruotsalainen, Holger Schünemann, Jayne Tierney**



RoB 1.0	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	N/A
N/A	Overall bias

# Proposed domains of assessment

---

- All domains to be mandatory
- No additional domains to be available (i.e. no ‘Other bias’ domain)
  - The domains in the tool should cover all potential issues.
- Funding and vested interests to be addressed but not to contribute to overall risk of bias assessments
  - working group led by Asbjørn Hróbjartsson and Isabelle Boutron

# Signalling questions and judgements

---

- Signalling questions are introduced to make the tool easier (and more transparent)
  - ‘Yes’, ‘Probably yes’, ‘Probably no’, ‘No’, ‘No information’
- Risk of bias judgements follow from answers to signalling questions (can be over-ridden)
  - **‘Low risk of bias’, ‘Some concerns’, ‘High risk of bias’**
- A change in the interpretation of the judgements, so that a ‘High risk of bias’ judgement in one domain puts the whole study at high risk of bias
- Overall risk of bias judgement can then be completed automatically (can be over-ridden)

Bias arising from the randomization process	1.1 Was the allocation sequence random?	Y / PY / PN / N / NI	[Description]
	1.2 Was the allocation sequence concealed until participants were recruited and assigned to interventions?	Y / PY / PN / N / NI	[Description]
	1.3 Were there baseline imbalances that suggest a problem with the randomization process?	Y / PY / PN / N / NI	[Description]
	<b>Risk of bias judgement</b>	Low / High / Some concerns	[Support]
	Optional: What is the predicted direction of bias arising from the randomization process?		[Rationale]
Bias due to deviations from intended interventions	2.1. Were participants aware of their assigned intervention during the trial?	Y / PY / PN / N / NI	[Description]
	2.2. Were carers and trial personnel aware of participants' assigned intervention during the trial?	Y / PY / PN / N / NI	[Description]
	2.3. <u>If Y/PY/NI to 2.1 or 2.2:</u> Were there deviations from the intended intervention beyond what would be expected in usual practice?	NA / Y / PY / PN / N / NI	[Description]
	2.4. <u>If Y/PY to 2.3:</u> Were these deviations from intended intervention unbalanced between groups and likely to have affected the outcome?	NA / Y / PY / PN / N / NI	[Description]
	2.5 Were any participants analysed in a group different from the one to which they were assigned?	Y / PY / PN / N / NI	[Description]
	2.6 <u>If Y/PY/NI to 2.5:</u> Was there potential for a substantial impact (on the estimated effect of intervention) of analysing participants in the wrong group?	NA / Y / PY / PN / N / NI	[Description]
	<b>Risk of bias judgement</b>	Low / High / Some concerns	[Support]
	Optional: What is the predicted direction of bias due to deviations from intended interventions?		[Rationale]
Bias due to missing outcome data	3.1 Were outcome data available for all, or nearly all, participants randomized?	Y / PY / PN / N / NI	[Description]
	3.2 <u>If N/PN/NI to 3.1:</u> Are the proportions of missing outcome data and reasons for missing outcome data similar across intervention groups?	NA / Y / PY / PN / N / NI	[Description]
	3.3 <u>If N/PN/NI to 3.1:</u> Is there evidence that results were robust to the presence of missing outcome data?	NA / Y / PY / PN / N / NI	[Description]
	<b>Risk of bias judgement</b>	Low / High / Some concerns	[Support]
	Optional: What is the predicted direction of bias due to missing outcome data?		[Rationale]
Bias in measurement of the outcome	4.1 Were outcome assessors aware of the intervention received by study participants?	Y / PY / PN / N / NI	[Description]
	4.2 <u>If Y/PY/NI to 4.1:</u> Was the assessment of the outcome likely to be influenced by knowledge of intervention received?	NA / Y / PY / PN / N / NI	[Description]
	<b>Risk of bias judgement</b>	Low / High / Some concerns	[Support]
	Optional: What is the predicted direction of bias due to measurement of the outcome?		[Rationale]
Bias in selection of the reported result	Are the reported outcome data likely to have been selected, on the basis of the results, from...		
	5.1 ... multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	Y / PY / PN / N / NI	[Description]
	5.2 ... multiple analyses of the data?	Y / PY / PN / N / NI	[Description]
	<b>Risk of bias judgement</b>	Low / High / Some concerns	[Support]
	Optional: What is the predicted direction of bias due to selection of the reported result?		[Rationale]
Overall bias	<b>Risk of bias judgement</b>	Low / High / Some concerns	[Support]
	Optional: What is the overall predicted direction of bias for this outcome?		[Rationale]

## Overall risk of bias judgement

---

Low risk of bias	The study is judged to be at <b>low risk of bias for all domains</b> for this result.
Some concerns	The study is judged to be at <b>some concerns</b> in at least one domain for this result.
High risk of bias	The study is judged to be at <b>high risk of bias</b> in at least one domain for this result. OR The study is judged to have <b>some concerns</b> for <b>multiple domains</b> in a way that substantially lowers confidence in the result.

## Some excerpts from the tool

## The RoB 2.0 tool (individually randomized, parallel group trials)

### Study design

- Randomized parallel group trial
- Cluster-randomized trial
- Randomized cross-over or other matched design

Specify which outcome is being assessed for risk of bias

Specify the numerical result being assessed. In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed.

Is your aim for this study...?

- to assess the effect of *assignment to intervention*
- to assess the effect of *starting and adhering to intervention*

Which of the following sources have you obtained to help inform your risk of bias judgements (tick as many as apply)?

- Journal article(s) with results of the trial
- Trial protocol
- Statistical analysis plan (SAP)
- Non-commercial trial registry record (e.g. ClinicalTrials.gov record)
- Company-owned trial registry record (e.g. GSK Clinical Study Register record)
- "Grey literature" (e.g. unpublished thesis)
- Conference abstract(s) about the trial
- Regulatory document (e.g. Clinical Study Report, Drug Approval Package)
- Research ethics application
- Grant database summary (e.g. NIH RePORTER, Research Councils UK Gateway to Research)
- Personal communication with trialist
- Personal communication with the sponsor



1.1 Was the allocation sequence random?

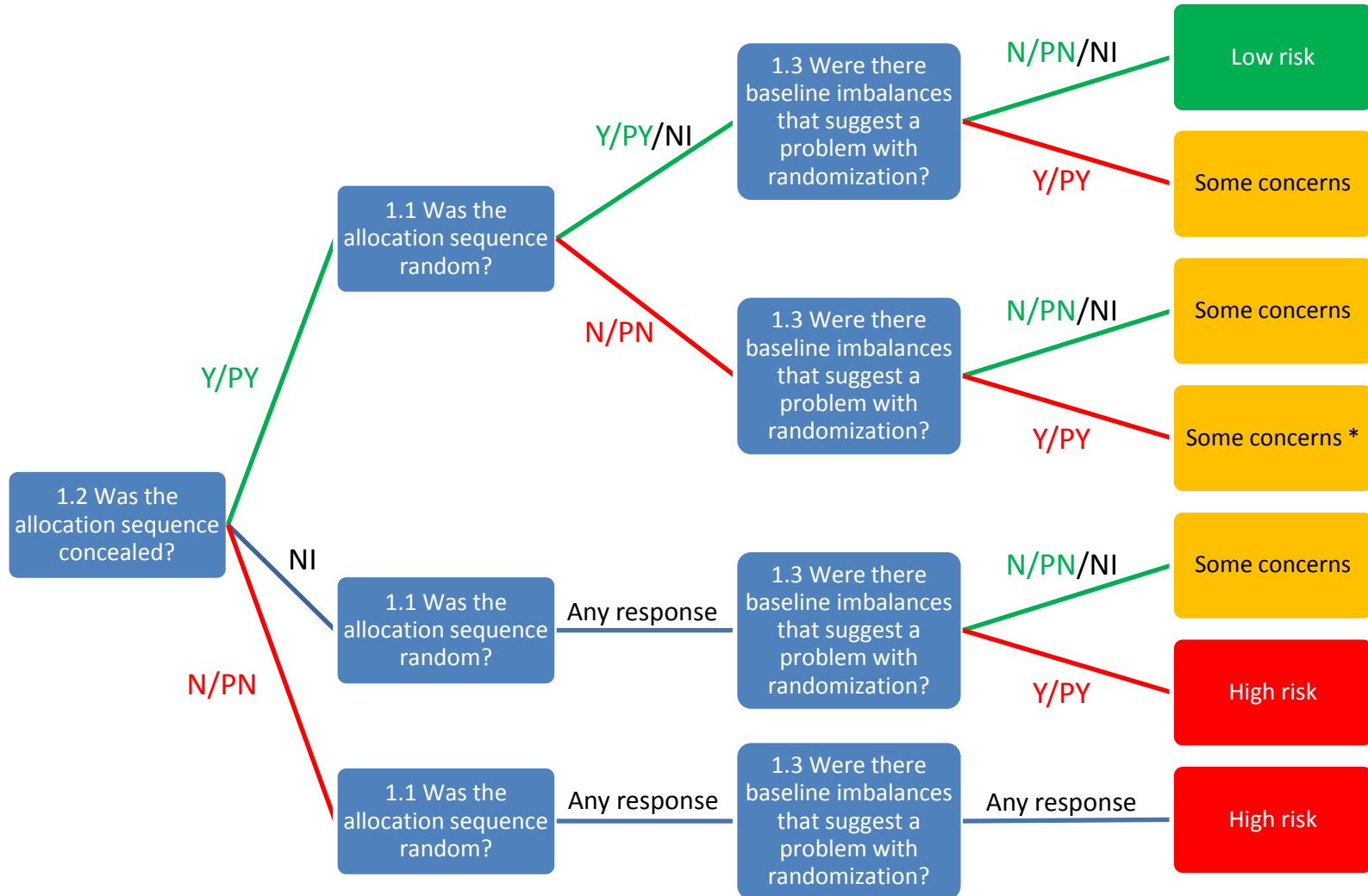
1.2 Was the allocation sequence concealed until participants were recruited and assigned to interventions?

1.3 Were there baseline imbalances that suggest a problem with the randomization process?

**Randomization  
methods**

**Additional  
evidence of  
problems**

# Bias arising from the randomization process



## Effect of assignment to intervention

2.1. Were participants aware of their assigned intervention during the trial?

2.2. Were carers and trial personnel aware of participants' assigned intervention during the trial?

2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention beyond what would be expected in usual practice?

2.4. If Y/PY to 2.3: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?

2.5. Were any participants analysed in a group different from the one to which they were assigned?

2.6. If Y/PY/NI to 2.5: Was there potential for a substantial impact (on the estimated effect of intervention) of analysing participants in the wrong group?

**Blinding**

**Deviations  
reflect  
usual  
practice?**

**First ITT  
principle of  
ITT**

## Effect of starting and adhering to intervention

2.1. Were participants aware of their assigned intervention during the trial?

2.2. Were carers and trial personnel aware of participants' assigned intervention during the trial?

2.3. If Y/PY/NI to 2.1 or 2.2: Were important co-interventions balanced across intervention groups?

2.4. Was the intervention implemented successfully?

2.5. Did study participants adhere to the assigned intervention regimen?

2.6. If N/PN/NI to 2.3, 2.4 or 2.5: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?

**Blinding**

**Specific  
deviations**

**Overcome by  
analysis?**

Are the reported outcome data likely to have been selected, on the basis of the results, from...

5.1. ... multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain?

**Selective outcome reporting**

5.2 ... multiple analyses of the data?

**Selective analysis reporting**

## Risk of bias tools

## ▼ ROBINS-I tool

[Read more](#)[Resources](#)[The team](#)[Feedback](#)

## RoB 2.0 tool

[Welcome](#) >

## RoB 2.0 tool

A revised tool to assess risk of bias in randomized trials (RoB 2.0)

Welcome to the website for the RoB 2.0 tool. This is a **draft version** of the tool. We have developed versions for three different trial designs.

## Individually randomized, parallel group trials

You can:

- Download [background information and detailed guidance for using the RoB 2.0 tool \(pdf\)](#).
- Download [the tool itself \(pdf\)](#)
- Download a blank template for completing the tool, which has two variants
  - Implement [RoB 2.0 when interest is in the effect of assignment to intervention \(Word\)](#)
  - Implement [RoB 2.0 when the interest is in the effect of starting and adhering to intervention \(Word\)](#).

## Cluster randomized, parallel group trials

- The current tool has nothing to say about review questions for which blinding is not feasible
  - (other than to classify as not blind hence high risk of bias)
- Issues of *performance bias* very different for “ITT effects” and “per-protocol” effects, yet poorly addressed in current RoB tool
- “ITT effect”: **effect of assignment to intervention**
  - e.g. the question of interest to a policy maker about whether to introduce a screening programme
- “Per protocol effect”: **effect of starting and adhering to intervention**
  - e.g. the question of interest to an individual about whether to attend screening

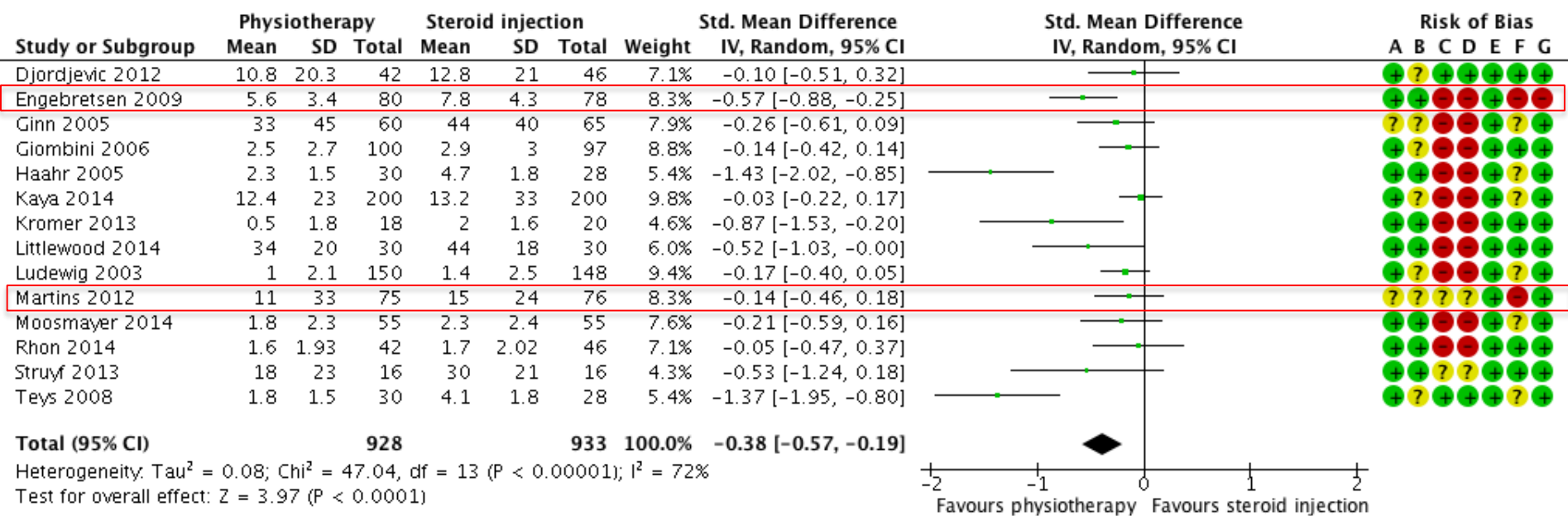
 Not to be confused with ITT or per protocol **analyses**

- Deviations from intended intervention **are not important** when interest is on the effect of **assignment to** intervention
  - e.g. some people don't respond to invitations to be screened
- ...providing these deviations reflect routine care
  - rather than behaviour that reflects expectations of a difference between intervention and comparator
- But deviations such as poor adherence, poor implementation and co-interventions may lead to bias when interest is in the effect **starting and adhering to** intervention
- We therefore have different tools for these two effects of interest



## Bias in selection of the reported result

- Current tool emphasises assessment of selective non-reporting or partial reporting of outcomes:
  - e.g. trialists measure pain, function and QoL, but only report data for pain
  - e.g. trialists report P values but no means & SDs for pain
- Review authors often rate a study at high risk of bias if one outcome is not reported
  - e.g. “All outcomes were reported except for pain”
  - e.g. “Some outcomes were not reported”



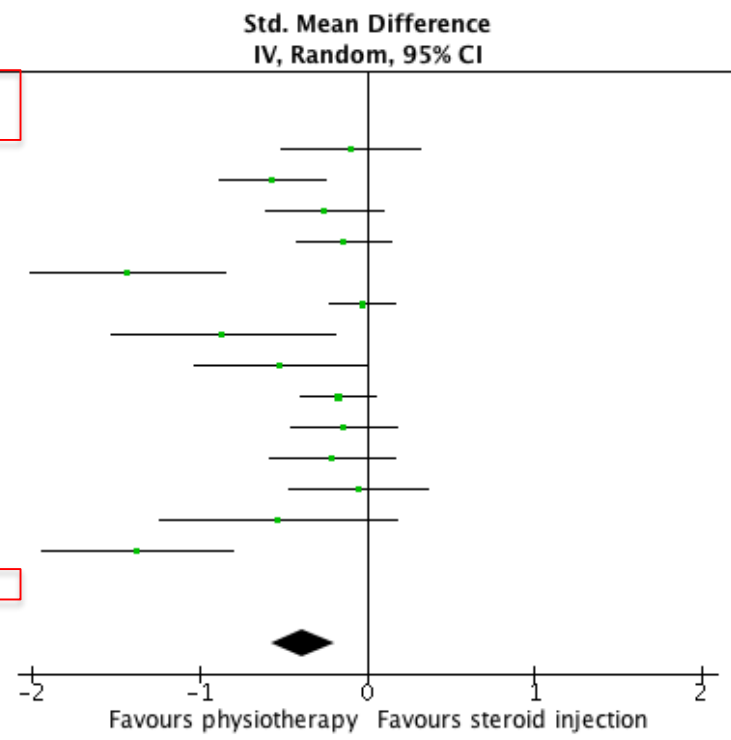
**Risk of bias legend**

(A) Random sequence generation (selection bias)  
 (B) Allocation concealment (selection bias)  
 (C) Blinding of participants and personnel (performance bias)  
 (D) Blinding of outcome assessment (detection bias): Self-reported outcomes  
 (E) Incomplete outcome data (attrition bias)  
 (F) Selective reporting (reporting bias)  
 (G) Other bias

- 2 trials are rated at high risk of bias because **pain** was not reported
- But this is a meta-analysis of **function**, so it does not make sense to display these high risk ratings here

Study or Subgroup	Physiotherapy			Steroid injection			Weight	Std. Mean Difference	
	Mean	SD	Total	Mean	SD	Total		IV, Random, 95% CI	Std. Mean Difference
Barra Lopez 2013	0	0	0	0	0	0		Not estimable	
Blume 2014	0	0	0	0	0	0		Not estimable	
Djordjevic 2012	10.8	20.3	42	12.8	21	46	7.1%	-0.10 [-0.51, 0.32]	
Engelbretsen 2009	5.6	3.4	80	7.8	4.3	78	8.3%	-0.57 [-0.88, -0.25]	
Ginn 2005	33	45	60	44	40	65	7.9%	-0.26 [-0.61, 0.09]	
Giombini 2006	2.5	2.7	100	2.9	3	97	8.8%	-0.14 [-0.42, 0.14]	
Haahr 2005	2.3	1.5	30	4.7	1.8	28	5.4%	-1.43 [-2.02, -0.85]	
Kaya 2014	12.4	23	200	13.2	33	200	9.8%	-0.03 [-0.22, 0.17]	
Kromer 2013	0.5	1.8	18	2	1.6	20	4.6%	-0.87 [-1.53, -0.20]	
Littlewood 2014	34	20	30	44	18	30	6.0%	-0.52 [-1.03, -0.00]	
Ludewig 2003	1	2.1	150	1.4	2.5	148	9.4%	-0.17 [-0.40, 0.05]	
Martins 2012	11	33	75	15	24	76	8.3%	-0.14 [-0.46, 0.18]	
Moosmayer 2014	1.8	2.3	55	2.3	2.4	55	7.6%	-0.21 [-0.59, 0.16]	
Rhon 2014	1.6	1.93	42	1.7	2.02	46	7.1%	-0.05 [-0.47, 0.37]	
Struyf 2013	18	23	16	30	21	16	4.3%	-0.53 [-1.24, 0.18]	
Teys 2008	1.8	1.5	30	4.1	1.8	28	5.4%	-1.37 [-1.95, -0.80]	
Walther 2004	0	0	0	0	0	0		Not estimable	

**Total (95% CI)**    **928**    **933** **100.0%**    **-0.38 [-0.57, -0.19]**  
Heterogeneity:  $\tau^2 = 0.08$ ;  $\chi^2 = 47.04$ ,  $df = 13$  ( $P < 0.00001$ );  $I^2 = 72\%$   
Test for overall effect:  $Z = 3.97$  ( $P < 0.0001$ )



- Selective non-reporting biases the result of the **meta-analysis** which cannot include the trial that omitted the outcome; it does not bias the trial result
- This is similar to publication bias (non-reporting of a study)

Trial result is biased because it has been selected on the basis of the results from multiple:

- Outcome measurements
  - Scales
  - Definitions of/criteria for an event
  - Time points
- Analyses
  - Unadjusted vs adjusted models
  - Different sets of covariates in adjusted models
  - Final values vs change from baseline vs analysis of covariance
  - Continuous scale converted to categorical data with different cut-points

We propose that:

- Bias in selection of the reported result be addressed in the revised risk of bias tool
- Selective non-reporting (and partial reporting) of outcomes be addressed elsewhere, in a new tool to assess the risk of reporting biases in meta-analyses

- Cluster-randomized trials:
  - Key issue is recruitment / identification of participants after interventions have been allocated to clusters
  - Also consideration of missing data at cluster and individual level
- Cross-over trials (AB/BA design)
  - Key issue is carry-over of effect from 1<sup>st</sup> period to 2<sup>nd</sup> period
  - Also period effects, selective reporting of 1<sup>st</sup> period data

- How many results to assess per study?
- How much free text to include to support assessments?
- How should it be presented in the review?
- Implementation
  - RoB 2.0 will need careful consideration to make the process efficient for multiple outcomes
    - Discussions to be initiated with RevMan team at Seoul Colloquium



- We believe RoB 2.0 offers considerable advantages over the existing tool
- Once programmed into software, we expect the tool will be easier to use than the first one
- We are extremely grateful to all those who have contributed to the development of RoB 2.0
- RoB 2.0 is available at [www.riskofbias.info](http://www.riskofbias.info)