University of BRISTOL

# Announcement of revised tools to assess risk of bias in randomized trials and in non-randomized studies

Julian Higgins, Jonathan Sterne, Jelena Savovic

University of Bristol, UK

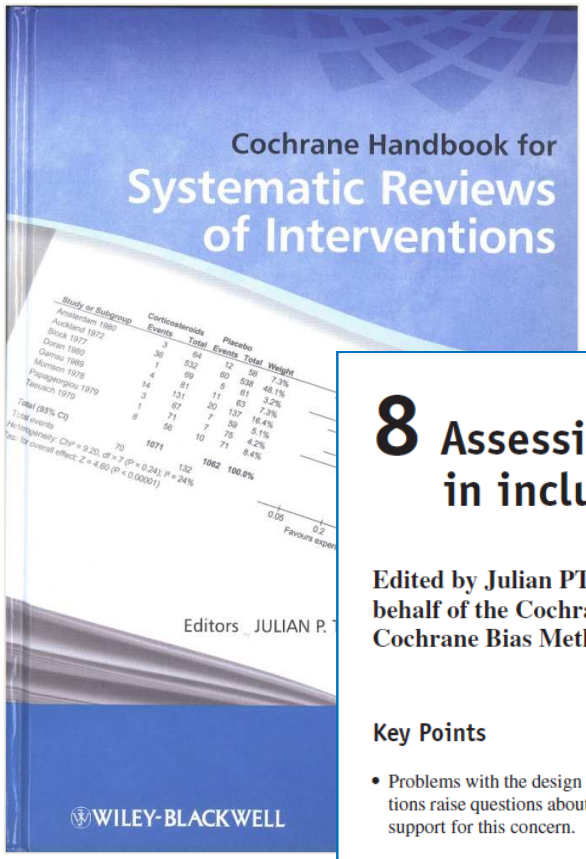
University of BRISTOL

# Purpose of this session

- To announce the availability of two new tools for assessing risk of bias
  - RoB 2.0: a revised tool for risk of bias in randomized trials
  - ROBINS-I: a new tool for Risk Of Bias In Non-randomized Studies - of Interventions

- Outline
  - Why new tools?
  - Innovations common to both tools
  - Remarks on the RoB 2.0 tool (Jelena)
  - Remarks on the ROBINS-I tool (Jonathan)
  - Opportunity for questions

University of BRISTOL




# 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.

*BMJ* 2011; 343: d5928

**RESEARCH METHODS & REPORTING**

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

Julian P T Higgins,[1] Douglas G Altman,[2] Peter C Gøtzsche,[3] Peter Jüni,[4] David Moher,[5 6] Andrew D Oxman,[7] Jelena Savović,[8] Kenneth F Schulz,[9] Laura Weeks,[5] Jonathan A C Sterne,[8] 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

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.[1]

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).[2] 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[3] 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 trials without producing a score).[4-7] Until recently, Cochrane reviews used a variety of these tools, mainly checklists.[8] 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.

**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

### 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 specialty. 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.

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

*Risk of bias*

Foam dressings for venous leg ulcers

| 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."<br><br>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."<br><br>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."<br><br>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."<br><br>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. |

# Current Cochrane tool for risk of bias in randomized trials

- 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)

University of BRISTOL

# Some issues raised with existing tool

- 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**

University of BRISTOL

# Need for a tool for non-randomized studies

- Systematic reviews on the effects of interventions may need to including non-randomized studies (NRSI)
  - long-term or rare outcomes (esp. adverse effects)
  - interventions at population or organization level
  - lack of randomized trials

- Reviews need to critique included studies, but existing tools for NRSI were
  - not adequate
  - or not closely aligned with the approach of the existing Cochrane tool for trials

University of BRISTOL

# Two new tools

- Revised tool for risk of bias in randomized trials
  - Current working title **RoB 2.0**

- New tool for risk of bias in non-randomized studies of interventions
  - Initially called ACROBAT-NRSI
  - Now called **ROBINS-I**

University of BRISTOL

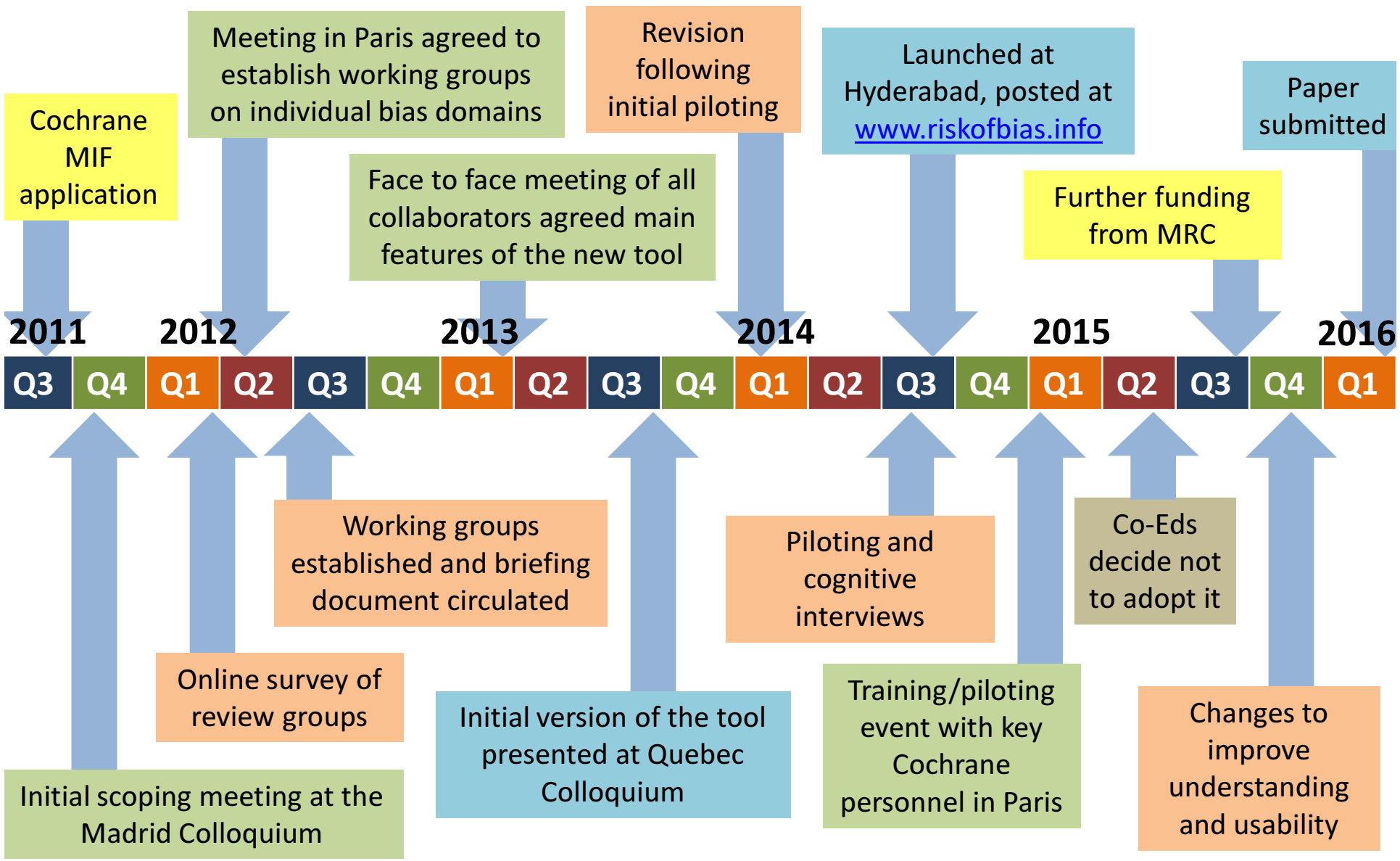
# Funding


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

- Initial development of the tool for non-randomized studies (**ROBINS-I**) and extensions for cross-over and cluster trials was funded by the **Cochrane** Methods Innovation Fund
- Ongoing work on **ROBINS-I** is funded by the UK **Medical Research Council** Methodology Panel (MR/M025209/1)

University of BRISTOL

# ROBINS-I: development chronology



University of BRISTOL

# ROBINS-I: contributors

- Core group:
  - **Jonathan Sterne, Barney Reeves, Jelena Savović, Lucy Turner, Julian Higgins**
- Collaborators:
  - **David Moher, Yoon Loke, Elizabeth Waters, Craig Ramsay, Peter Tugwell, George Wells, Vivian Welch**
- Additional working group members:
  - **Doug Altman, Mohammed Ansari, Nancy Berkman, Isabelle Boutron, Belinda Burford, James Carpenter, An-Wen Chan, David Henry, Miguel Hernán, Asbjørn Hróbjartsson, Peter Jüni, Jamie Kirkham, Terri Piggott, Deborah Regidor, Hannah Rothstein, Lakho Sandhu, Lina Santaguida, Bev Shea, Ian Shrier, Jeff Valentine, Meera Viswanathan**

- And: **Jan Vandenbroucke, Jon Deeks, Toby Lasserson, Rachel Churchill, Alexandra McAleenan, Roy Elbers, Matthew Page, Rebecca Armstrong, Sasha Shepperd, Hugh Waddington, Su Golder** ...

# RoB 2.0: development chronology

- Revision of the RoB tool started in May 2015
- 1st 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'
- 2nd Development meting held in Bristol on 21-22 April 2016
- Development of further guidance and piloting
- Released for Seoul Colloquium

University of BRISTOL

# RoB 2.0: contributors

- 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**

# Key innovations common to both tools

- **Result-focussed** assessments
- Fixed (inclusive) bias domains, **not modifiable**
- "**Signalling questions**" to facilitate risk of bias judgements
- New **response options** for risk of bias, without 'Unclear' option
- Formal **overall** risk of bias judgement, as worst rating of any individual domain

- Some rethinking of the assessment:
  - Important distinction between **effects of interest**
  - Selective reporting focuses on **reported result**

University of BRISTOL

# The effect of interest

- The current tool has very little to say about situations in 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

University of BRISTOL

# The effect of interest

- The current tool has very little to say about situations in 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**

University of BRISTOL

# The effect of interest

- When interested in effect of **assignment** to intervention
  - Deviations from intended intervention **are not important**
    - e.g. some 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

- When interested in effect **starting and adhering to** intervention
  - Deviations such as poor adherence, poor implementation and co-interventions may **lead to risk of bias**

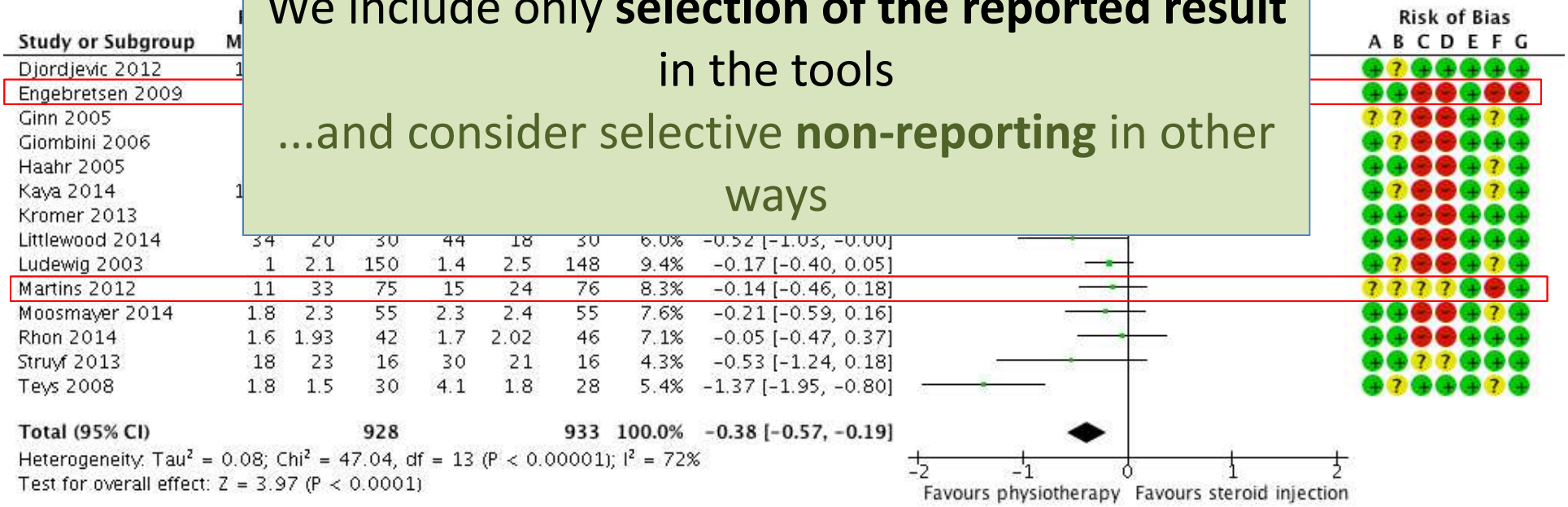
- We therefore have different tools for these two effects of interest

# Selective outcome non-reporting bias

- Current tool takes a broad approach to selective reporting
- Any evidence of it in the trial reports?

We include only **selection of the reported result** in the tools

...and consider selective **non-reporting** in other ways




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

University of BRISTOL

# More about RoB 2.0 for randomized trials

University of BRISTOL

| RoB 1.0 | RoB 2.0 |
|---|---|
| Random sequence generation (*selection bias*) | Bias arising from the randomization process |
| Allocation concealment (*selection bias*) | |
| Blinding of participants and personnel (*perforr* | Bias due to deviations from intended |
| Incomplete (*attri* | |
| Blinding of out (*detec* | |
| Selective reporting (*reporting bias*) | Bias in selection of the reported result |
| Other bias | N/A |
| N/A | Overall bias |

Funding and vested interests to be addressed, but not within this part of the wider framework

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] |
| | **Risk of bias judgement** | 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. 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? | NA / Y / PY / PN / N / NI | [Description] |
| | 2.4. If Y/PY to 2.3: 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 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? | NA / Y / PY / PN / N / NI | [Description] |
| | **Risk of bias judgement** | 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 If N/PN/NI to 3.1: 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 If N/PN/NI to 3.1: Is there evidence that results were robust to the presence of missing outcome data? | NA / Y / PY / PN / N / NI | [Description] |
| | **Risk of bias judgement** | 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 If Y/PY/NI to 4.1: Was the assessment of the outcome likely to be influenced by knowledge of intervention received? | NA / Y / PY / PN / N / NI | [Description] |
| | **Risk of bias judgement** | 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] |
| | **Risk of bias judgement** | Low / High / Some concerns | [Support] |
| | Optional: What is the predicted direction of bias due to selection of the reported result? | | [Rationale] |
| **Overall bias** | **Risk of bias judgement** | Low / High / Some concerns | [Support] |
| | Optional: What is the overall predicted direction of bias for this outcome? | | [Rationale] |

University of BRISTOL

# 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)?**

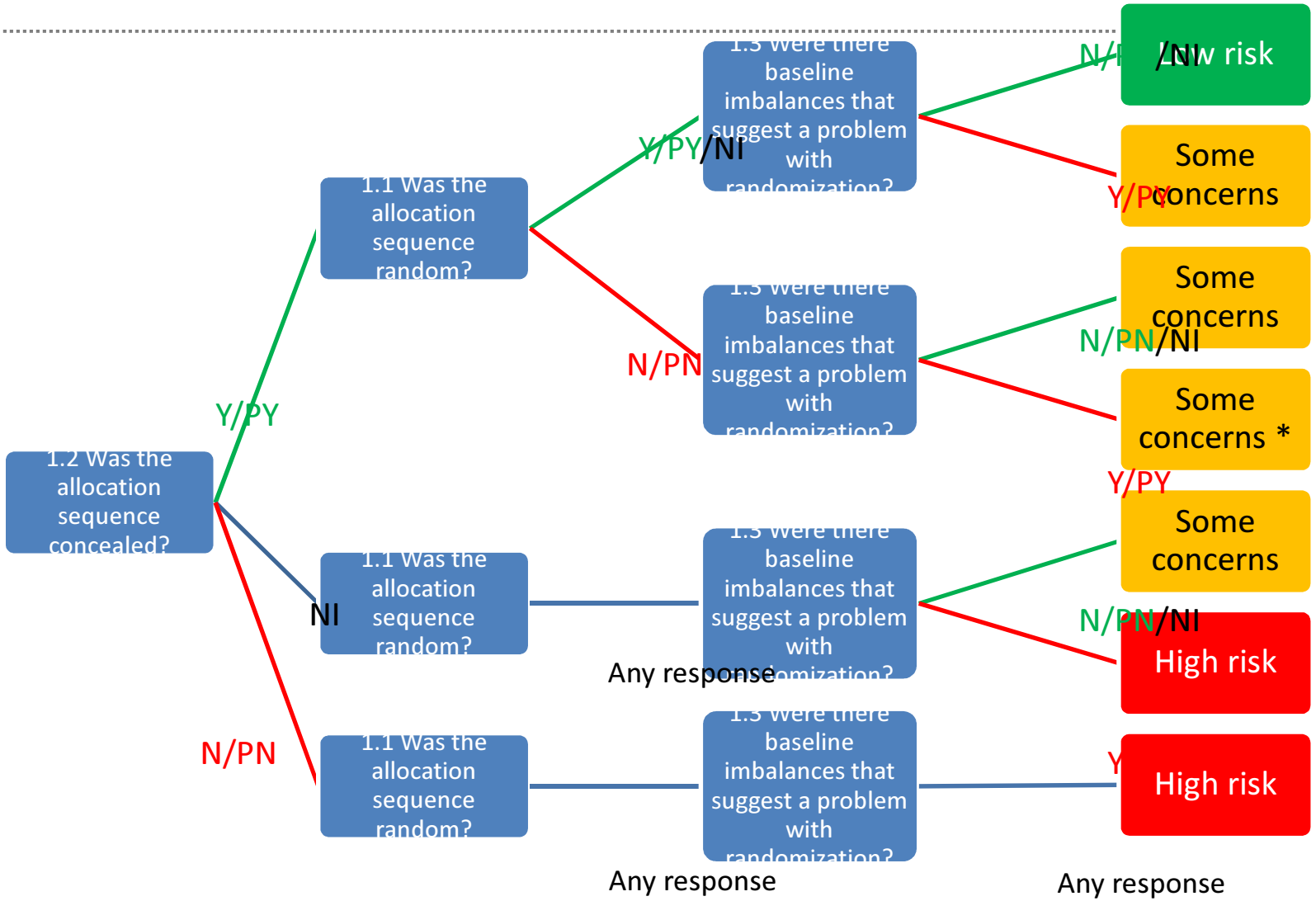
- ☐ Journal article(s) with results of the trial
- ☐ Trial protocol
- ☐ Statistical analysis plan (SAP)
- ☐ Non-commercial trial registry record (e.g. ClinicalTrials.gov record)
- ☐ Company-owned trial registry record (e.g. GSK Clinical Study Register record)
- ☐ "Grey literature" (e.g. unpublished thesis)
- ☐ 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

University of BRISTOL

# Bias arising from the randomization process

| | |
|---|---|
| 1.1 Was the allocation sequence random?<br>1.2 Was the allocation sequence concealed until participants were recruited and assigned to interventions? | **Randomization methods** |
| 1.3 Were there baseline imbalances that suggest a problem with the randomization process? | **Additional evidence of problems** |

# Algorithm


1.2 Was the allocation sequence concealed?
Y/PY
NI
N/PN
1.1 Was the allocation sequence random?
1.1 Was the allocation sequence random?
1.1 Was the allocation sequence random?
Y/PY/NI
N/PN
1.3 Were there baseline imbalances that suggest a problem with randomization?
1.3 Were there baseline imbalances that suggest a problem with randomization?
1.3 Were there baseline imbalances that suggest a problem with randomization?
Any response
Any response
N/PN/NI
Y/PY
Low risk
Some concerns
Some concerns
N/PN/NI
Some concerns *
Y/PY
Some concerns
N/PN/NI
High risk
High risk
Any response

University of BRISTOL

# Bias due to deviations from intended interventions

**Effect of assignment to intervention**

| | |
|---|---|
| 2.1. Were participants aware of their assigned intervention during the trial?<br>2.2. Were carers and trial personnel aware of participants' assigned intervention during the trial? | **Blinding** |
| 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?<br>2.4. If Y/PY to 2.3: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome? | **Deviations reflect usual practice?** |
| 2.5 Were any participants analysed in a group different from the one to which they were assigned?<br>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? | **First ITT principle of ITT** |

University of BRISTOL

# Bias due to deviations from intended interventions

**Effect of starting and adhering to intervention**

| | |
|---|---|
| 2.1. Were participants aware of their assigned intervention during the trial?<br>2.2. Were carers and trial personnel aware of participants' assigned intervention during the trial? | **Blinding** |
| 2.3. If Y/PY/NI to 2.1 or 2.2: Were important co-interventions balanced across intervention groups?<br>2.4. Was the intervention implemented successfully?<br>2.5. Did study participants adhere to the assigned intervention regimen? | **Specific deviations** |
| 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? | **Overcome by analysis?** |

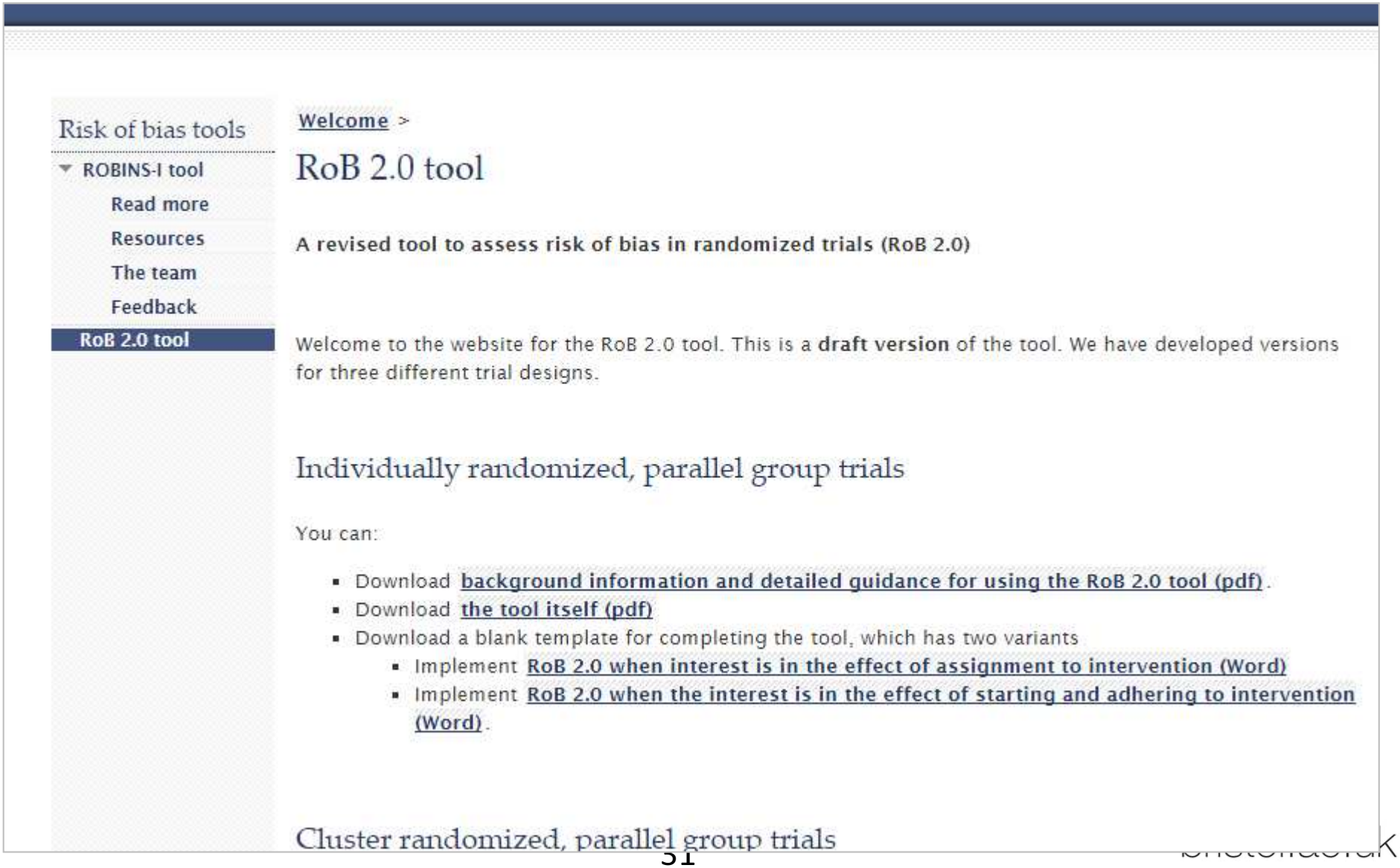
# Bias in selection of the reported result

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? | **Selective outcome reporting** |
| 5.2 ... multiple analyses of the data? | **Selective analysis reporting** |

University of BRISTOL

# riskofbias.info



University of BRISTOL

# More about ROBINS-I

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

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

For numbered affiliations see end of article.

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

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

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

Non-randomised studies of the effects of interventions are critical to many areas of healthcare evaluation, but their results may be biased. It is therefore important to understand and appraise their strengths and weaknesses. We developed ROBINS-I ("Risk Of Bias In Non-randomised Studies - of Interventions"), a new tool for evaluating risk of bias in estimates of the comparative effectiveness (harm or benefit) of interventions from studies that did

such as cohort studies and case-control studies in which intervention groups are allocated during the course of usual treatment decisions, and quasi-randomised studies in which the method of allocation falls short of full randomisation. Non-randomised studies can provide evidence additional to that available from randomised trials about long term outcomes, rare events, adverse effects and populations that are typical of real world practice.[1 2] The availability of linked databases and compilations of electronic health records has enabled NRSI to be conducted in large representative population cohorts.[3] For many types of organisational or public health interventions, NRSI are the main source of evidence about the likely impact of the intervention because randomised trials are difficult or impossible to conduct on an area-wide basis. Therefore systematic reviews addressing the

University of BRISTOL

# Scope

- The tool concerns the risk of bias (RoB) in the results of a NRSI that compares the health effects of two or more interventions
  - quantitative studies
  - estimating effectiveness (harm or benefit) of an intervention
  - did not use randomization to allocate units (individuals or clusters) to comparison groups

# Key differences from assessing randomized trials

- It is very difficult
- The assessment uses the idea of a hypothetical randomized trial as a reference
- There are some things to think about before hand (at protocol stage)
- Careful thinking needed afterwards
  - consistent message despite risk of bias?
  - how to include in syntheses?

# Assessing risk of bias in relation to a target trial

- RoB assessment facilitated by considering NRSI as an attempt to mimic a high quality **hypothetical randomized trial** of interventions of interest
  - "target trial"
  - need not be feasible or ethical

University of BRISTOL

# Overview of the tool

- Preliminary considerations
  - Identify key confounding domains & co-interventions
- Define target (idealized) randomized trial to match the study
    - specify PICO and the effect of interest
- Bias domains of (result-level) assessment
  - Signalling questions
  - Free text descriptions
  - Risk of bias judgements
- Overall (result-level) risk of bias judgement
  - feed into GRADE

| | | |
|---|---|---|
| **Bias due to confounding** | | 1.1 Is there potential for confounding of the effect of intervention in this study?<br>**If N/PN to 1.1:** the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered<br>**If Y/PY to 1.1**: determine whether there is a need to assess time-varying confounding:<br>1.2. Was the analysis based on splitting participants' follow up time according to intervention received?<br>**If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.<br>1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome?<br>**If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6).<br>**If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8) |
| | *For baseline confounding only* | 1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?<br>1.5. If **Y/PY to 1.4**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?<br>1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention? |
| | *For baseline and time-varying confounding* | 1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?<br>1.8. If **Y/PY to 1.7**: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study? |
| | | **Risk of bias judgement** (Optional: What is the predicted direction of bias due to confounding?) |
| **Bias in selection of participants into the study** | | 2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?<br>**If N/PN to 2.1:** go to 2.4<br>2.2. If **Y/PY to 2.1**: Were the post-intervention variables that influenced selection likely to be associated with intervention?<br>2.3 If **Y/PY to 2.2**: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?<br>2.4. Do start of follow-up and start of intervention coincide for most participants?<br>2.5. If **Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?<br>**Risk of bias judgement** (Optional: What is the predicted direction of bias due to selection of participants into the study?) |
| **Bias in classification of interventions** | | 3.1 Were intervention groups clearly defined?<br>3.2 Was the information used to define intervention groups recorded at the start of the intervention?<br>3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?<br>**Risk of bias judgement** (Optional: What is the predicted direction of bias due to classification of interventions?) |
| **Bias due to deviations from intended interventions** | *For effect of assignment to intervention* | 4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?<br>4.2. If **Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome? |
| | *For effect of starting and adhering to intervention* | 4.3. Were important co-interventions balanced across intervention groups?<br>4.4. Was the intervention implemented successfully for most participants?<br>4.5. Did study participants adhere to the assigned intervention regimen?<br>4.6. If **N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention? |
| | | **Risk of bias judgement** (Optional: What is the predicted direction of bias due to deviations from the intended interventions?) |
| **Bias due to missing data** | | 5.1 Were outcome data available for all, or nearly all, participants?<br>5.2 Were participants excluded due to missing data on intervention status?<br>5.3 Were participants excluded due to missing data on other variables needed for the analysis?<br>5.4 If **PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?<br>5.5 If **PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?<br>**Risk of bias judgement** (Optional: What is the predicted direction of bias due to missing data?) |
| **Bias in measurement of outcomes** | | 6.1 Could the outcome measure have been influenced by knowledge of the intervention received?<br>6.2 Were outcome assessors aware of the intervention received by study participants?<br>6.3 Were the methods of outcome assessment comparable across intervention groups?<br>6.4 Were any systematic errors in measurement of the outcome related to intervention received?<br>**Risk of bias judgement** (Optional: What is the predicted direction of bias due to measurement of outcomes?) |
| **Bias in selection of the reported result** | *Is reported estimate selected, on the basis of the results, from...* | 7.1. ... multiple outcome *measurements* within the outcome domain?<br>7.2 ... multiple *analyses* of the intervention-outcome relationship?<br>7.3 ... different *subgroups*?<br>**Risk of bias judgement** (Optional: What is the predicted direction of bias due to selection of the reported result?) |
| **Overall bias** | | **Risk of bias judgement** (Optional: What is the predicted direction of bias for this outcome?) |

University of BRISTOL

# Dimensions of bias

| Bias dimension | Note |
|---|---|
| Bias due to confounding | Selection bias *as it is often used in relation to clinical trials* (and currently in widespread use within The Cochrane Collaboration) |
| Bias in selection of participants into the study | Selection bias *as it is usually used in relation to observational studies*; Inception bias; Lead-time bias; Immortal time bias |
| Bias in classification of interventions | Awareness of treatment when measuring outcome; Objectivity and comparability of outcome measurement |
| Bias due to deviations from intended interventions | Treatment switches; Co-interventions; Fidelity; Performance bias |
| Bias due to missing data | Completeness of outcome data; Imbalance and reasons for missing data; Completeness of intervention (exposure) data; Other missing data; Statistical methods; Attrition bias |
| Bias in measurement of the outcome | Awareness of outcome when measuring intervention; Detection bias |
| Bias in selection of the reported result | Multiple outcomes/time points; Multiple analyses; Reporting a subset of participants |

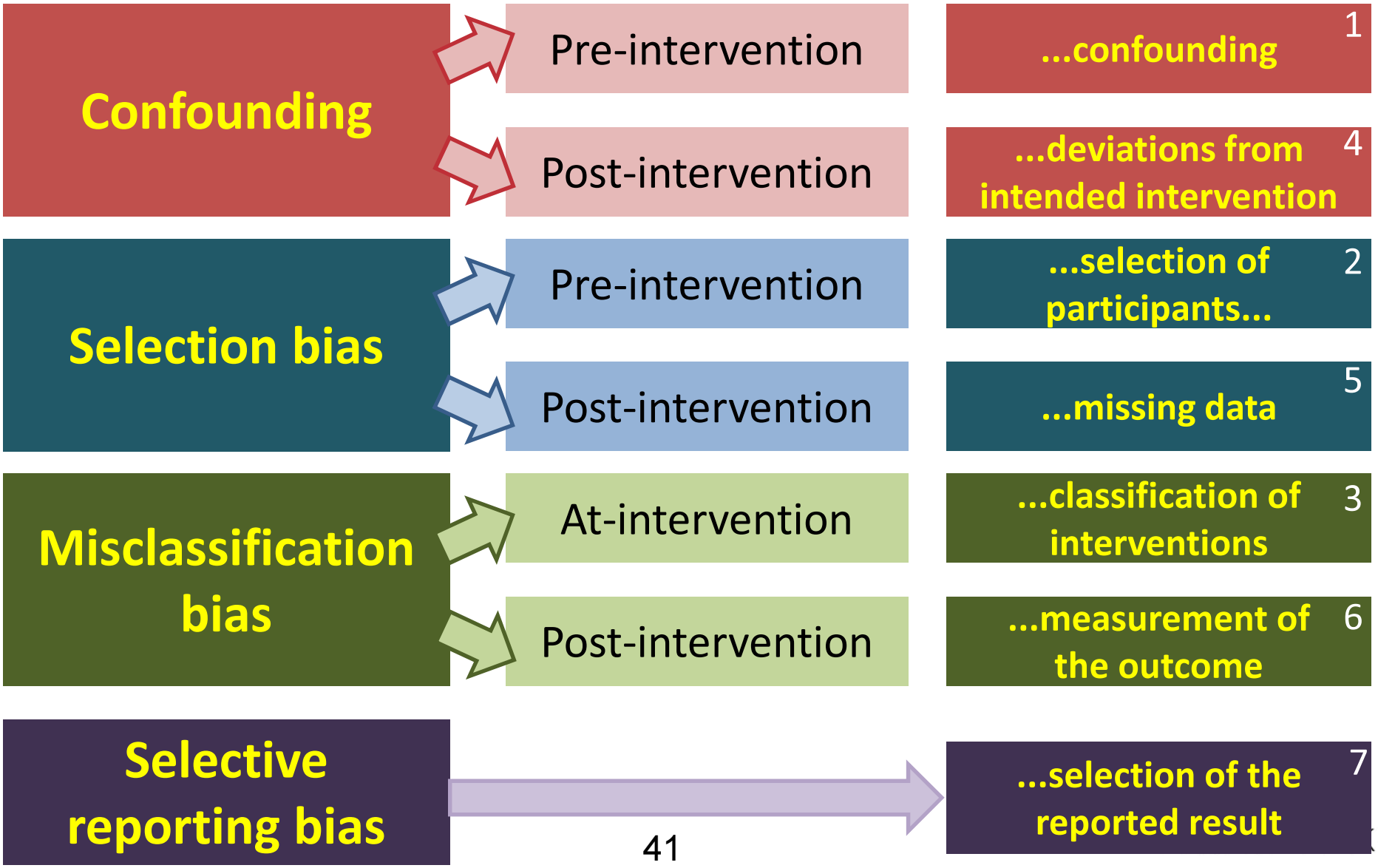
University of BRISTOL

# Dimensions of bias

| Bias dimension | Note |
|---|---|
| Bias due to confounding | *Pre-treatment* features, for which considerations of bias in observational studies are mainly distinct from those in RCTs |
| Bias in selection of participants into the study | |
| Bias in classification of interventions | |
| Bias due to deviations from intended interventions | *Post-treatment* features, for which many considerations of bias in observational studies are similar to those in RCTs |
| Bias due to missing data | |
| Bias in measurement of the outcome | |
| Bias in selection of the reported result | |

University of BRISTOL

# An epidemiological perspective



University of BRISTOL

# An example of the complexity in considering risk of bias in non-randomized studies

(... if there is time)

Skip 

# POSTMENOPAUSAL ESTROGEN THERAPY AND CARDIOVASCULAR DISEASE

## Ten-Year Follow-up from the Nurses' Health Study

MEIR J. STAMPFER, M.D., GRAHAM A. COLDITZ, M.B., B.S., WALTER C. WILLETT, M.D., JOANN E. MANSON, M.D., BERNARD ROSNER, PH.D., FRANK E. SPEIZER, M.D., AND CHARLES H. HENNEKENS, M.D.

**Abstract** *Background.* The effect of postmenopausal estrogen therapy on the risk of cardiovascular disease remains controversial. Our 1985 report in the *Journal,* based on four years of follow-up, suggested that estrogen therapy reduced the risk of coronary heart disease, but a report published simultaneously from the Framingham Study suggested that the risk was increased. In addition, studies of the effect of estrogens on stroke have yielded conflicting results.

*Methods.* We followed 48,470 postmenopausal women, 30 to 63 years old, who were participants in the Nurses' Health Study and who did not have a history of cancer or cardiovascular disease at base line. During up to 10 years of follow-up (337,854 person-years), we documented 224 strokes, 405 cases of major coronary disease (nonfatal myocardial infarctions or deaths from coronary causes), and 1263 deaths from all causes.

*Results.* After adjustment for age and other risk factors, the overall relative risk of major coronary disease in women currently taking estrogen was 0.56 (95 percent confidence interval, 0.40 to 0.80); the risk was significantly reduced among women with either natural or surgical menopause. We observed no effect of the duration of estrogen use independent of age. The findings were similar in analyses limited to women who had recently visited their physicians (relative risk, 0.45; 95 percent confidence interval, 0.31 to 0.66) and in a low-risk group that excluded women reporting current cigarette smoking, diabetes, hypertension, hypercholesterolemia, or a Quetelet index above the 90th percentile (relative risk, 0.53; 95 percent confidence interval, 0.31 to 0.91). The relative risk for current and former users of estrogen as compared with those who had never used it was 0.89 (95 percent confidence interval, 0.78 to 1.00) for total mortality and 0.72 (95 percent confidence interval, 0.55 to 0.95) for mortality from cardiovascular disease. The relative risk of stroke when current users were compared with those who had never used estrogen was 0.97 (95 percent confidence interval, 0.65 to 1.45), with no marked differences according to type of stroke.

*Conclusions.* Current estrogen use is associated with a reduction in the incidence of coronary heart disease as well as in mortality from cardiovascular disease, but it is not associated with any change in the risk of stroke. (N Engl J Med 1991; 325:756-62.)

# Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women

## Principal Results From the Women's Health Initiative Randomized Controlled Trial

Writing Group for the Women's Health Initiative Investigators

**Context** Despite decades of accumulated observational evidence, the balance of risks and benefits for hormone use in healthy postmenopausal women remains uncertain.

**Objective** To assess the major health benefits and risks of the most commonly used combined hormone preparation in the United States.

**Design** Estrogen plus progestin component of the Women's Health Initiative, a randomized controlled primary prevention trial (planned duration, 8.5 years) in which 16608 postmenopausal women aged 50-79 years with an intact uterus at baseline were recruited by 40 US clinical centers in 1993-1998.

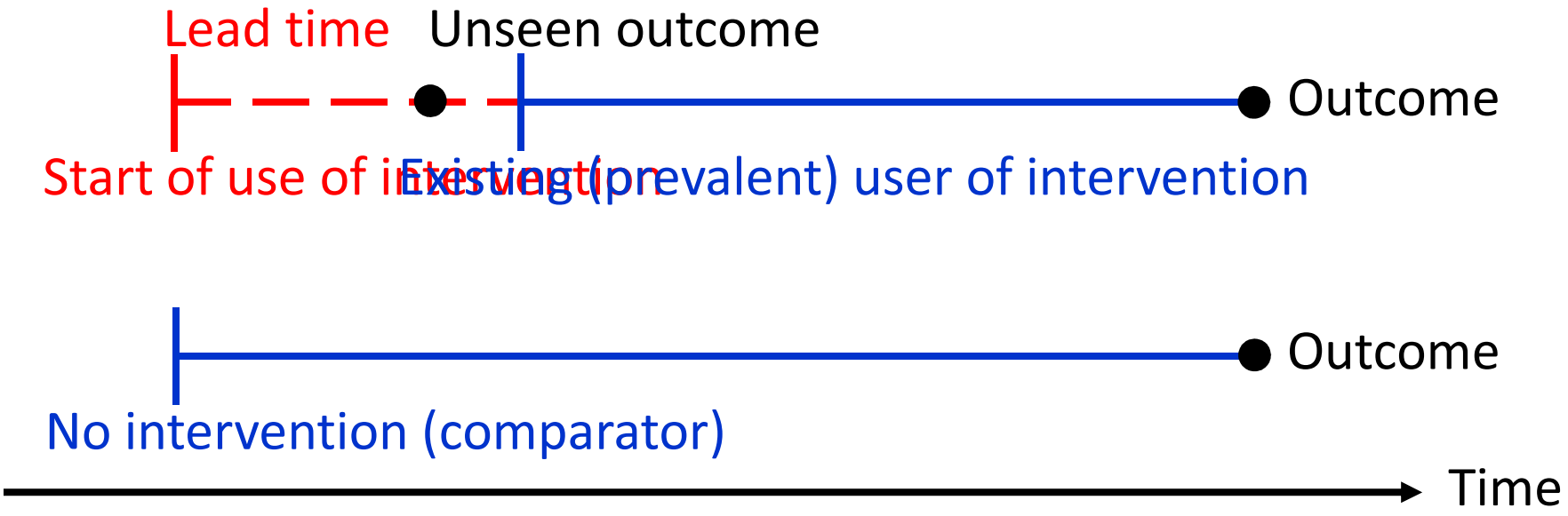
**Interventions** Participants received conjugated equine estrogens, 0.625 mg/d, plus medroxyprogesterone acetate, 2.5 mg/d, in 1 tablet (n=8506) or placebo (n=8102).

**Main Outcomes Measures** The primary outcome was coronary heart disease (CHD) (nonfatal myocardial infarction and CHD death), with invasive breast cancer as the primary adverse outcome. A global index summarizing the balance of risks and benefits included the 2 primary outcomes plus stroke, pulmonary embolism (PE), endometrial cancer, colorectal cancer, hip fracture, and death due to other causes.

**Results** On May 31, 2002, after a mean of 5.2 years of follow-up, the data and safety monitoring board recommended stopping the trial of estrogen plus progestin vs placebo because the test statistic for invasive breast cancer exceeded the stopping boundary for this adverse effect and the global index statistic supported risks exceeding benefits. This report includes data on the major clinical outcomes through April 30, 2002. Estimated hazard ratios (HRs) (nominal 95% confidence intervals [CIs]) were as follows: CHD, 1.29 (1.02-1.63) with 286 cases; breast cancer, 1.26 (1.00-1.59) with 290 cases; stroke, 1.41 (1.07-1.85) with 212 cases; PE, 2.13 (1.39-3.25) with 101 cases; colorectal cancer, 0.63 (0.43-0.92) with 112 cases; endometrial cancer, 0.83 (0.47-1.47) with 47 cases; hip fracture, 0.66 (0.45-0.98) with 106 cases; and death due to other causes, 0.92 (0.74-1.14)

THE WOMEN'S HEALTH INITIATIVE (WHI) focuses on defining the risks and benefits of strategies that could potentially reduce the incidence of heart disease, breast and colorectal cancer, and fractures in postmenopausal women. Between 1993 and 1998, the WHI enrolled 161 809 postmenopausal women in the age range of 50 to 79 years into a set of clinical trials (trials of low-fat dietary pattern, calcium and vitamin D supplementation, and 2 trials of postmenopausal hormone use) and an observational study at 40 clinical centers in the United States.[1] This article reports principal results for the trial of combined estrogen and progestin in women with a uterus. The trial was stopped early based on health risks that exceeded health benefits over an aver-

# Bias due to selection of follow up time




Observational Studies Analyzed Like Randomized Experiments

*An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease*

*Miguel A. Hernán,*[a,b] *Alvaro Alonso,*[c] *Roger Logan,*[a] *Francine Grodstein,*[a,d] *Karin B. Michels,*[a,d,e] *Walter C. Willett,*[a,d,f] *JoAnn E. Manson,*[a,d,g] *and James M. Robins*[a,h]

University of BRISTOL

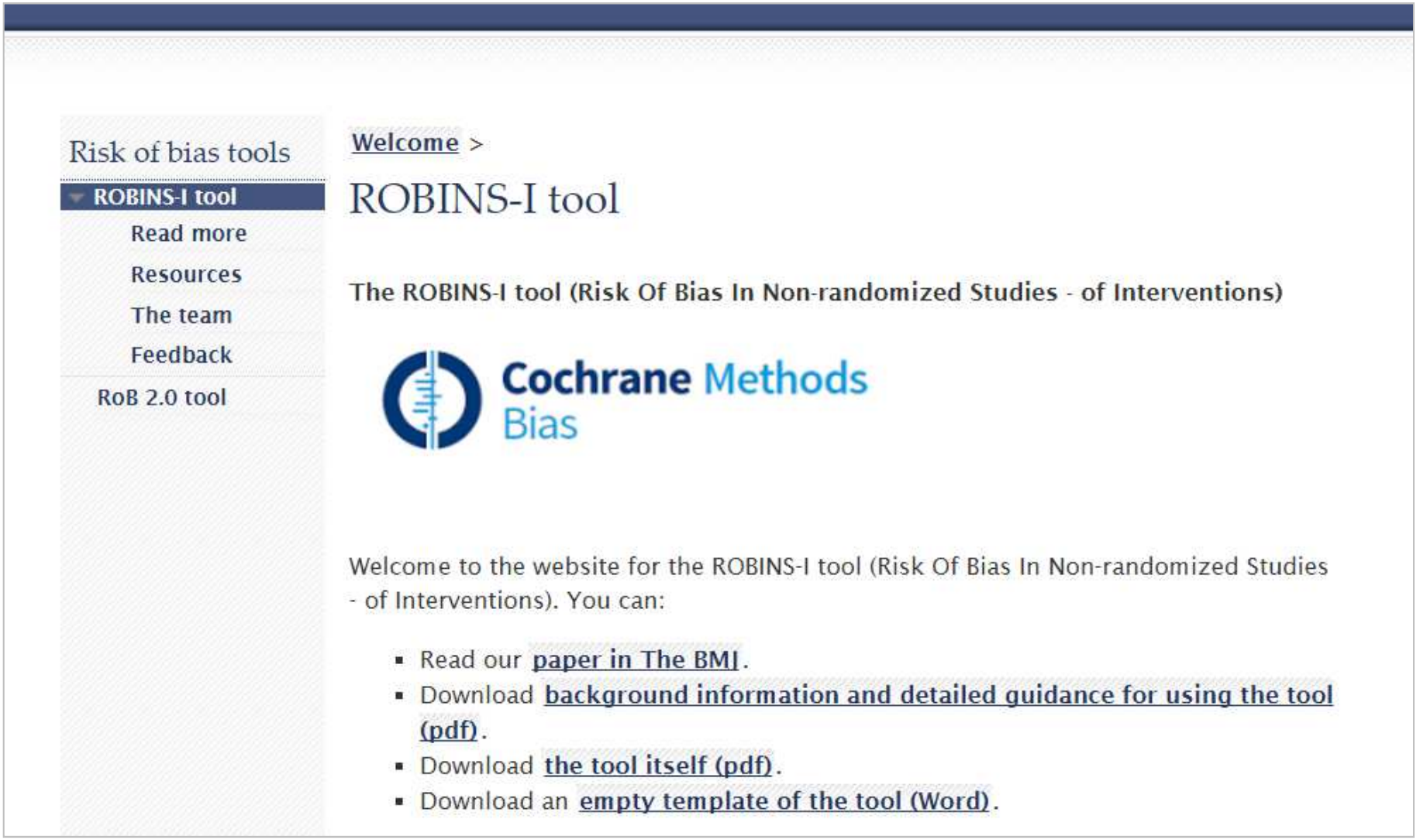
# Risk of bias judgements

| Response option | Interpretation |
|---|---|
| **Low risk of bias** | The study is comparable to a well-performed randomized trial with regard to this bias dimension. |
| **Moderate risk of bias** | The study is sound for a non-randomized study with regard to this bias dimension but cannot be considered comparable to a well-performed randomized trial. |
| **Serious risk of bias** | The study has some important problems in this dimension of bias. |
| **Critical risk of bias** | The study is too problematic in this dimension of bias to provide any useful evidence. |
| **No information** | No information on which to base a judgement about risk of bias for this dimension. |

It is usually impossible to exclude bias due to residual or unmeasured confounding of the results of an non-randomized study. **Therefore we expect very few NRSI to be assessed as at low risk of bias due to confounding**

University of BRISTOL

# www.riskofbias.info



University of BRISTOL

# Closing remarks

University of BRISTOL

# Piloting and implementation

- Both tools have undergone multiple phases of piloting
  - informed development and refinement
  - More is always welcome
- Formal studies of inter-rater agreement not yet performed

- Full guidance for both tools available at riskofbias.info
  - ROBINS-I is official version 1 (BMJ paper)
  - RoB 2.0 is initial draft, subject to minor refinements

- Implementation
  - We are implementing ROBINS-I in an interactive online system
  - RoB 2.0 is very new; implementation options yet to be discussed in detail

University of BRISTOL

# Some unresolved issues

- How many results to assess per study?
- How to integrate into data collection process
- How to present assessments in a review?

- Ongoing work on ROBINS-I adaptations to case-control studies, before-after studies, interrupted time series, instrumental variables, regression discontinuities, ...

- RoB 2.0 available for parallel group trials, crossover trials and cluster-randomized trials
  - what else is needed?

| RoB 1.0 | RoB 2.0 | ROBINS-I |
|---|---|---|
| Random sequence generation (*selection bias*) | Bias arising from the randomization process | Bias due to confounding |
| Allocation concealment (*selection bias*) | | |
| N/A | N/A | Bias in selection of participants into the study |
| N/A | N/A | Bias in measurement of interventions |
| Blinding of participants and personnel (*performance bias*) | Bias due to deviations from intended interventions | Bias due to deviations from intended interventions |
| Incomplete outcome data (*attrition bias*) | Bias due to missing outcome data | Bias due to missing data |
| Blinding of outcome assessment (*detection bias*) | Bias in measurement of the outcome | Bias in measurement of the outcome |
| Selective reporting (*reporting bias*) | Bias in selection of the reported result | Bias in selection of the reported result |
| Other bias | N/A | N/A |
| N/A | Overall bias | Overall bias |