RoB 2 Domain IV
Bias in measurement of the outcome

Asbjørn Hróbjartsson\textsuperscript{1,2} and Isabelle Boutron\textsuperscript{1,3,4}

Bias Methods Group\textsuperscript{1}, University of Southern Denmark\textsuperscript{2}, Cochrane France\textsuperscript{3}, Université de Paris\textsuperscript{4}

Trusted evidence.
Informed decisions.
Better health.
• Introduction
• Bias mechanisms and empirical evidence
• Assessing the risk of bias in measurement of the outcome: signalling questions 1-2
• Assessing the risk of bias in measurement of the outcome: Signalling questions 3-5
• Questions
Risk of bias in randomized trials

Bias arising from the randomization process

Bias due to deviations from intended intervention

Bias due to missing outcome data

Experimental

Comparator

Outcome

Bias in measurement of the outcome

Bias in selection of the reported result

<table>
<thead>
<tr>
<th>1.02</th>
<th>3.87</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.20</td>
<td>4.32</td>
</tr>
<tr>
<td>1.38</td>
<td>5.44</td>
</tr>
</tbody>
</table>
Bias mechanisms

Errors in measuring outcome variables

The measured value ≠ the true value of the outcome.

Terminology

- Measurement error (continuous outcome)
- Misclassification (dichotomous outcome, categorical outcome)
- Under/over-ascertainment (event)

Errors

- Non-differential
- Differential
Bias mechanisms
Errors in measuring outcome variables

Non differential errors:

• Errors occurs similarly in both groups
• Errors are not related to the treatment allocated
• **Example:**
  - blood test to measure haemoglobin level
  - Blood pressure measurement
Double blind randomized trials

The outcome of composites of death and myocardial infarction with or without refractory angina

**Figure 2** A figure summarizing the changes in the number of patients with end-points at 30 days after randomization as calculated before and after the judgements and decisions of the End Point Committee (EPC). AMI = acute myocardial infarction. Index MI = myocardial infarction at inclusion in the study (not an end-point).
Bias mechanisms
Errors in measuring outcome variables

Non differential errors:

• **Continuous outcome (mean difference)**
  – Usually **no bias**

• **Dichotomous/categorical outcome (OR, RR, HR)**:
  – **Bias toward the null**

• Situations where non-differential error can bias effect estimates away from the null are **unlikely** to occur in randomized trials
Bias mechanisms
Errors in measuring outcome variables

Differential errors

• Errors are related to the treatment allocated
• Example: assessment of the pain level on a VAS systematically assessed as lower in the intervention arm

⇒ Bias
⇒ Essential role of blinding
Blinding terminology

- "Single" blind
- "Double" blind
- "Triple" blind

<table>
<thead>
<tr>
<th>Physicians</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single blind</td>
<td>17</td>
</tr>
<tr>
<td>Double blind</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Textbooks</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

Devereaux, JAMA, 2001
Empirical evidence of bias

Influence of Reported Study Design Characteristics on Intervention Effect Estimates From Randomized, Controlled Trials

Jelena Savović, PhD; Hayley E. Jones, PhD; Douglas G. Altman, DSc; Ross J. Harris, MSc; Peter Jüni, MD; Julie Pildal, MD, PhD; Bodil Als-Nielsen, MD, PhD; Ethan M. Balk, MD, MPH; Christian Gluud, DrSciMed; Lise Lotte Gluud, DrSciMed;

Lack of Double-Blinding or Unclear Double-Blinding (vs. Double-Blind)

Outcome (Contributing Meta-analyses/Contributing Trials) | ROR (95% Crl)
--- | ---
All outcomes (104/1057) | 0.87 (0.79–0.96)
Mortality (25/245) | 0.92 (0.80–1.04)
Other objective (28/282) | 0.93 (0.74–1.18)
Subjective or mixed (51/530) | 0.78 (0.65–0.92)

Estimated intervention effect according to blinded or non-blinded outcome assessor

<table>
<thead>
<tr>
<th>Study</th>
<th>Odds ratio (95% CI)</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith 2007²⁷</td>
<td>0.08 (0.04 to 0.14)</td>
<td>0.00 (0.00 to 0.01)</td>
</tr>
<tr>
<td>Non-blinded</td>
<td>0.06 (0.03 to 0.16)</td>
<td>0.01 (0.00 to 0.03)</td>
</tr>
<tr>
<td>MA-1300-15²⁸</td>
<td>0.23 (0.10 to 0.54)</td>
<td>0.06 (0.03 to 0.14)</td>
</tr>
<tr>
<td>Non-blinded</td>
<td>0.34 (0.11 to 1.07)</td>
<td>0.10 (0.01 to 0.78)</td>
</tr>
<tr>
<td>Oesterle 2000²²</td>
<td>2.67 (0.33 to 21.87)</td>
<td>0.84 (0.14 to 5.22)</td>
</tr>
<tr>
<td>Blinded</td>
<td>0.17 (0.07 to 0.39)</td>
<td>0.06 (0.03 to 0.15)</td>
</tr>
<tr>
<td>Non-blinded</td>
<td>0.38 (0.08 to 1.86)</td>
<td>0.19 (0.03 to 1.08)</td>
</tr>
<tr>
<td>Landsman 2010¹⁶</td>
<td>2.81 (0.48 to 16.43)</td>
<td>1.46 (0.11 to 18.96)</td>
</tr>
<tr>
<td>Non-blinded</td>
<td>1.23 (0.53 to 2.89)</td>
<td>0.74 (0.43 to 1.28)</td>
</tr>
<tr>
<td>Meltzer 2003¹⁸</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-blinded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reynolds 2004α²⁴</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-blinded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burkhoff 1999¹⁰</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-blinded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jones 2006¹³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-blinded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aro 2011¹⁴</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-blinded</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hróbjartsson A et al. BMJ 2012;344:bmj.e1119
Is blinding always feasible?

Assessment of zinc treatment for common cold\(^1,\)\(^2\)

- Specific taste and aftertaste of zinc
- Hunches: « anything tasting as bad as zinc and with as much as aftertaste as zinc must be a good medicine »
- Success of blinding was questionable

1) Desbiens et al, Annals of Internal Medicine, 2000
2) Fair, J et al.. Chronic Dis., 1987
Is blinding always feasible?
Blinding of outcome assessment

Centralized blinded assessment

- Radiography
- Video
- Audiotape
- Photography
- Blinded adjudication committee

Not always possible

- Patient reported outcome?
Assessing the risk of bias in measurement of the outcome

Signalling questions
4.1 Was the method of measuring the outcome inappropriate

4.2 Measurement or ascertainment of outcome differ between groups?

4.1 Method of measuring the outcome inappropriate?

4.3 Outcome assessors aware of intervention received?

4.3 Outcome assessors aware of intervention received?

4.4 Could assessment have been influenced by knowledge of intervention?

4.5 Likely that assessment was influenced by knowledge of intervention?

Low risk

Some concerns

High risk
To download the file

Assessment at the outcome level
Assessing the risk of bias in measurement of the outcome

Signalling questions 1-2.
4.1 Was the method of measuring the outcome inappropriate?

- Poor validity of the methods
  - The methods **does not measure** what it is intended to measure
  - Example 1:
    - Event: severe hypoglycemia
    - Measurement: portable blood glucose machine used by patients
    - **Issue**: does not reliably measure glycemia <3.1 mmol/l

- Poor reliability of the methods
  - Example: Four-point rating scale for assessing pain level is less reliable than VAS or numeric rating scale
4.1 Was the method of measuring the outcome inappropriate?

⚠️ The question does **NOT** aim to assess whether the choice of the outcome is relevant

- **NO/Probably NO**
  - In **most trials**, for pre-specified outcomes

- **Yes/probably yes:**
  - Measurement unlikely to identify plausible intervention effect
  - Measurement has been **demonstrated** to have poor validity
4.1 Was the method of measuring the outcome inappropriate

4.2 Measurement or ascertainment of outcome differ between groups?

4.3 Outcome assessors aware of intervention received?

4.4 Could assessment have been influenced by knowledge of intervention?

4.5 Likely that assessment was influenced by knowledge of intervention?

Low risk

Some concerns

High risk
In randomized trials outcome measurement is usually performed similarly in both group.

However, specific situations may arise

- **different outcome assessors**
  - surgeons/GP

- ‘Diagnostic detection bias’
  - Number of visits differ because of the intervention evaluated -> increasing opportunities to detect outcome events
  - Treatment adverse event → complementary tests more frequently performed on one arm
  - Ex: headache -> MRI -> tumor more likely to be detected
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?

- **NO/Probably NO**
  - Comparable methods of measurement (same methods, definition, time points, assessors)

- **Yes/probably yes**
  - Context of passive collection of outcome data (adverse event) or additional visits in one group
4.1 Method of measuring the outcome inappropriate?

4.2 Measurement or ascertainment of outcome differ between groups?

4.3 Outcome assessors aware of intervention received?

4.4 Could assessment have been influenced by knowledge of intervention?

4.5 Likely that assessment was influenced by knowledge of intervention?

Low risk

Some concerns

High risk
Assessing the risk of bias in measurement of the outcome

Signalling questions 3-5
Bias in measurement of the outcome

The role of blinding

- who is assessing the outcome
- whether outcome assessor is blinded to intervention assignment
- whether assessment of outcome is likely to be influenced by knowledge of intervention assignment
A person measuring, ascertaining or recording the outcome is an ‘outcome assessor’:

i. an observer not directly involved in the intervention provided to the participant, such as an adjudication committee, a biologist performing an automated test, or a health professional recording outcomes from health records or disease registries.

ii. the participant when the outcome is participant-reported: for example pain, quality of life, or self-completed questionnaire evaluating depression, anxiety or function.

iii. the intervention provider when the outcome is the result of a therapeutic decision such as a decision to offer a surgical intervention or to discharge the patient.
Reporting
Often inadequate in trial reports.

‘26% of journal articles reported no information on blinding whatsoever beyond the trial being ‘double blind’.

More details in protocols

<table>
<thead>
<tr>
<th></th>
<th>Double blind</th>
<th>Single blind</th>
<th>Not DB/SB</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>156</td>
<td>12</td>
<td>32</td>
</tr>
<tr>
<td>%</td>
<td>78</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>Reporting of blinding status of key trial persons</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete (all categories)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Partial (patients and health care providers and data collectors)</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Minimal (at least one category)</td>
<td>65</td>
<td>42</td>
<td>24</td>
</tr>
<tr>
<td>None (no explicit information)</td>
<td>88</td>
<td>50</td>
<td>19</td>
</tr>
<tr>
<td>No information at all beyond trial being ‘blind’, eg, ‘double blind’</td>
<td>41</td>
<td>26</td>
<td>13</td>
</tr>
<tr>
<td>Experimental and control treatments appear ‘similar’</td>
<td>72</td>
<td>46</td>
<td>10</td>
</tr>
<tr>
<td>Time of unblinding described</td>
<td>14</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Blinding mentioned in discussion</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

1 Trials not described as ‘single-blind’ or ‘double-blind’. Typically such trials described blinding with other words (eg, ‘assessor-blind’).
2 Patients, health care providers, data collectors, assessors of outcome, data analysts, manuscript writers.
3 Excluding trials with ‘partial reporting’ and ‘complete reporting’.
4 Including analogue terms, eg, ‘identical’ or ‘indistinguishable’.

Haahr Clin Trials 2006, Hróbjartsson et al. JCE 2009
Whether outcome assessors were aware of the intervention received by study participants?

It is important to determine whether outcome assessments were made blinded to intervention assignment. If blinding was successfully implemented, then the risk of bias due to differential measurement error is low.

**Component 1:** were outcome assessors intended to be blind?

**Component 2:** was intention of blinding successful?
When is blinding of outcome assessors intended?

**Green flag**
- "Outcome assessors were blinded"
- "Non-blind participants and blind outcome assessor"
- "Double-blind drug trial with no indication of lack of blinding of outcome assessor"

**Red flag**
- "single blind" or "double-blind" only information
- external assessors not involved in patient care (but blinding not mentioned explicit)
- "Blind assessors interviewed non-blind patients"
Successfully implemented blinding of outcome assessor

When is blinding of outcome assessors successful?

Green flag
Pre-trial testing of matching of compared interventions
Assessor interaction with non-blind patients and description of procedures to handle cases of accidental unblinding
No tell-tale effects

Red flags
Assessor interaction with non-blind patients and no procedures to handle the risk of unblinding
Tell-tale effects (taste of zinc)
Run-in periods (active or placebo)

Probably less important than if blinding was intended.
Signalling question 4.4

*Could assessment have been influenced by knowledge of intervention?*

The importance of lack of blinding of the outcome assessor will depend on the extent to which the assessment can be influenced by knowledge of the intervention assignment.

**Green Flag**

Objective outcomes: **all-cause mortality** and (some) automated test procedure, e.g. laboratory measurements

**Red flag**

Subjective outcomes
Subjective outcomes
involving judgement
moderate to high inter-observer variation

Objective outcomes
not involving judgement
no or low inter-observer variation

Other uses of subjective/objective not relevant for RoB2
Objective: observer-reported
Subjective: inherently private to a person

Moustgaard JCE 2014
A model for observer bias

Different outcomes

- Objective (e.g. all-cause mortality)
- **Subjective** (e.g. global improvement, clinical function score)

Different persons

- **Preconceptions:** none
- Preconceptions: some or strong.

Red flag

Person with **preconceptions** observing a **subjective outcome**
Was it likely that assessment was influenced by knowledge of intervention? When the outcome assessor could have been influenced by knowledge of intervention received, users should assess whether it is likely that such influence occurred.

Considerations: trial context
Preconceptions
Hope
Hunches

Conflicts of interest
Red flags: high risk of bias
Experimental intervention vs no-treatment or usual care control
Outcome assessors strongly engaged in other parts of the trial
Outcome with high degree of subjectivity

Click to add text

Green Flags: some concern
Active control group
External outcome assessor not otherwise engaged in the trial
Low degree of outcome subjectivity
4.1 Method of measuring the outcome inappropriate?

4.2 Measurement or ascertainment of outcome differ between groups?

4.3 Outcome assessors aware of intervention received?

4.4 Could assessment have been influenced by knowledge of intervention?

4.5 Likely that assessment was influenced by knowledge of intervention?

Low risk

Some concerns

High risk
Blinding terminology in flux
”double-blind” carry different meanings to different authors

*Look for direct descriptions*

Reporting of blinding often inadequate in publications

*Use supplemental sources of information*

Information on risk of unblinding often missing assessment informal, absent and not reported

*If suspected, contact authors*

RoB2 involves judgements based on imperfect information
277 patients randomised to usual care + rHBMP-2 vs. usual care

“This was a multicenter single-blinded randomized study conducted at twenty-eight European and South African sites”.

“The primary efficacy end point was the proportion of subjects with a healed fracture as demonstrated by radiographic and clinical assessment thirteen and twenty weeks after definitive wound closure.”

“This study was limited by its single-blind design. Given the nature of the intervention under study, it was not possible to blind the investigators to the study group.”

Outcome: radiographic union

Non-blinded surgeons, reported in paper:
OR 0.74 (0.43 to 1.23)

Blinded radiologists, not reported in paper:
OR 1.23 (0.53 to 2.89)
719 patients randomised to echinacea tablets vs placebo VS echinacea tablets vs no-treatment.

“Patients were assigned to 1 of 4 parallel groups: no pills, placebo pills (blinded), echinacea pills (blinded), or Echinacea pills (unblinded, open-label).”

“Placebo and echinacea tablets contained the same proportions of inert ingredients and were covered with identical digestible coatings”.

“The primary outcome was the area under the curve for global severity, with severity assessed twice daily by self-report using the Wisconsin Upper Respiratory Symptom Survey, short version”.

“Blinding seemed to be intact. Of the 363 participants who received pills and were blinded, 141 (39%) guessed their assignment correctly …”
Important to differentiate between trials with low and high risk of bias in measurement of the outcome

**Low risk of bias**
- blinding implemented, and unlikely that the blinding could have been broken
- no blinding, but measurement unlikely to be influenced by knowledge of the intervention assignment

**High risk of bias**
- no blinding or broken blinding, and measurement likely to be influenced
Acknowledgements

- Slides in part based on material from Miranda Cumpston, Julian Higgins and Jonathan Sterne, Cochrane Bias Methods Group and the Australasian Cochrane Centre
Questions