Name that bias: lessons learned from empirical studies of bias in clinical research

Lisa Bero, Professor, Department of Clinical Pharmacy and Institute for Health Policy Studies, UCSF and Director SF Cochrane Center
Risk of Bias

- Methodological characteristics of a study that can introduce a systematic error in the magnitude or direction of the results (Higgins and Green 2008).
The objective of pharmacological, toxicological, chemical studies is to determine if an intervention or an exposure is causally related to an outcome.

Risk of bias criteria are applicable to all these types of studies as they are adherent to the same scientific principles.

As the starting assumption is that there is no effect on outcome, the methodological criteria are independent of the “expected” or “desired” direction of the effect. In the case of drug efficacy, this may be a “positive” outcome (to observe an effect or larger effect size), in the case of drug harm this may be a “negative” outcome (to observe no effect or smaller effect size).

The methodological issues are independent of the data stream (animal or human).
Industry sponsored drug studies are more likely to have favorable efficacy results than non-industry sponsored studies.

Industry sponsored drug studies are more likely to have favorable harm results than non-industry sponsored studies.

Risk of Bias

- In human clinical trials of drug efficacy, studies with a high risk of bias, such as those lacking randomization, allocation concealment, or blinding of participants, personnel and outcome assessors produce larger treatment effect sizes, thus falsely inflating the efficacy of the drugs, compared to studies that have these design features (Schulz et al. 1995; Schulz and Grimes 2002a, b).

- Biased human studies assessing the harms of drugs are less likely to report statistically significant adverse effects (Nieto et al. 2007).
Risk of Bias is not........

- An assessment of how the study is conducted (e.g., in compliance with human subjects guidelines)
- An assessment of how the study is reported (e.g., study population described, abstract is structured).
- The same as imprecision. While bias refers to systematic error, imprecision refers to random error. Although smaller studies are less precise, they may not be more biased.
Types of bias
Selection bias

- Introduces systematic differences between baseline characteristics in comparison groups
- Minimized by randomization and concealment of allocation
Empirical evidence of bias

- Analysis of 250 randomized controlled trials
- Inadequate concealment of allocation
  - Estimates of effect 30-41% greater than studies with adequate concealment
- Inadequate generation of allocation schedule
  - Small change in estimate of treatment effect

Effect sizes of studies with adequate concealment of allocation differ from those with inadequate concealment of allocation.
Performance / detection biases

- **Performance** - systematic difference between treatment and control groups with regard to care or other exposure besides the intervention / treatment.

- Blinding of investigators can protect against performance bias

- **Detection** - systematic differences between treatment and control groups with regards to how outcomes are assessed.

- Blinding of outcome assessors is a primary way of reducing detection bias.
Empirical evidence of bias

- **No double blinding**
  - Analysis of 250 randomized controlled trials
  - Estimates of effect 17% greater than studies with double blinding

Who is blinded and how?

“double blinding” not informative.

Need to know:
- Participants
- Investigators
- Data collectors
- Outcome assessors
- Data analysts

Ways to achieve blinding in animal studies – having coded data analyzed by a statistician who is independent of the study team
Attrition / Exclusion bias

- The systematic difference between treatment and control groups in the number of subjects that were included in and completed the study.
- Data on whether all subjects in the study are accounted for and use of intention-to-treat analysis can reduce exclusion bias.
Attrition / exclusion bias

- Participants lost to follow-up may differ from those who complete the study
- Data could be differentially excluded from comparison groups
- Intention to treat analysis
  - Means analyzing all patients in the group to which they were assigned, regardless of whether they actually received the treatment / exposure and how else they may differ
Methods specific to the research question

- For drug studies, choice of dose / duration of treatment
- Also applies to toxicology studies
- “Intensity of intervention” applies to all studies
Gaming the dose

- High dose of drug A compared to low dose of drug B to “demonstrate” efficacy of A
- Low dose of drug A compared to high dose of drug B to “demonstrate” drug A has less adverse effects
- = favorable outcome for drug A
How did they get this result for efficacy?

Dosage of Manufacturer's Drug vs. Comparitor Drug

Selective reporting bias

- The selective reporting of entire studies (publication bias) or outcomes from studies (selective outcome reporting).
- Publication bias minimized by registries, published protocols.
- Selective reporting minimized by full access to study protocols, results, reports.
Selective reporting of drug studies

Identify and characterize discrepancies, if any, between clinical trial data submitted to the Food and Drug Administration (FDA) in approved new drug applications (NDAs) and the corresponding published trials


Some trials are not published

Of 164 Trials submitted in NDAs...

PUBLISHED within 5 years: 78% (128)

OF 33 NDAs...

All trials published: 52% (17)

NO trials published: 2 (with a total of 5 trials)
## Predictors of Publication

<table>
<thead>
<tr>
<th>Trial Characteristic</th>
<th>OR (95% CI)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable primary outcome(s)</td>
<td>4.77 (1.33-17.06)</td>
<td>0.018</td>
</tr>
<tr>
<td>Active control (vs. placebo only)</td>
<td>3.37 (1.02-11.22)</td>
<td>0.047</td>
</tr>
</tbody>
</table>
Papers include more outcomes favoring the test drug

- 179 primary outcomes reported in NDAs
  - 41 were *omitted* from the papers
- Papers had 138 outcomes also reported in NDAs (77%)
  - PLUS 15 additional outcomes that favored the test drug
  - PLUS 2 other neutral outcomes
43 outcomes in the NDAs did not favor the test drug
- 20 were not included in the papers
- 5 changed statistical significance, with 4 changing to favor test drug in the paper

* changes in outcomes occurred in 36 (22%) trials found in 19 (58%) NDAs
How do we measure “quality”? 

25 scales, 9 checklists and more........
<table>
<thead>
<tr>
<th>Scale Name</th>
<th>Type of Scale</th>
<th>Quality Defined</th>
<th>Type of Quality Assessed</th>
<th>Items Selected</th>
<th>Patient Assignment</th>
<th>Masking</th>
<th>Patient Follow-up</th>
<th>Statistical Analysis</th>
<th>Number of Items</th>
<th>Scale Development</th>
<th>Inter-rater Reliability</th>
<th>Time to Complete</th>
<th>Scoring Range</th>
<th>Detailed Instructions for Scoring Items</th>
<th>Meta-analysis Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrew</td>
<td>s</td>
<td>n</td>
<td>r</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>11</td>
<td>nr</td>
<td>0.95</td>
<td>10</td>
<td>0–22</td>
<td>y</td>
<td>56</td>
</tr>
<tr>
<td>Annals</td>
<td>g</td>
<td>y</td>
<td>r</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>34</td>
<td>nr</td>
<td>0.12</td>
<td>15</td>
<td>34–170</td>
<td>y</td>
<td>75</td>
</tr>
<tr>
<td>Beckerman</td>
<td>g</td>
<td>y</td>
<td>m</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>25</td>
<td>nr</td>
<td>0.89</td>
<td>10</td>
<td>0–25</td>
<td>n</td>
<td>35</td>
</tr>
<tr>
<td>Brown</td>
<td>s</td>
<td>n</td>
<td>m</td>
<td>ac</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>6</td>
<td>nr</td>
<td>0.89</td>
<td>10</td>
<td>0–21</td>
<td>y</td>
<td>55</td>
</tr>
<tr>
<td>Chalmers, F</td>
<td>g</td>
<td>y</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>3</td>
<td>nr</td>
<td>&lt;10</td>
<td>10</td>
<td>0–9</td>
<td>y</td>
<td>58</td>
</tr>
<tr>
<td>Chalmers, TC</td>
<td>g</td>
<td>n</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>27</td>
<td>nr</td>
<td>0.89</td>
<td>45</td>
<td>0–100</td>
<td>y</td>
<td>45</td>
</tr>
<tr>
<td>Cho</td>
<td>g</td>
<td>y</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>24</td>
<td>nr</td>
<td>0.89</td>
<td>30</td>
<td>0–1</td>
<td>y</td>
<td>60</td>
</tr>
<tr>
<td>Colditz</td>
<td>g</td>
<td>n</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>8</td>
<td>nr</td>
<td>10</td>
<td>10</td>
<td>0–8</td>
<td>n</td>
<td>56</td>
</tr>
<tr>
<td>Detrsky</td>
<td>g</td>
<td>n</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>5</td>
<td>nr</td>
<td>10</td>
<td>10</td>
<td>0–15</td>
<td>n</td>
<td>56</td>
</tr>
<tr>
<td>Evans</td>
<td>g</td>
<td>n</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>33</td>
<td>nr</td>
<td>15</td>
<td>15</td>
<td>0–8</td>
<td>n</td>
<td>63</td>
</tr>
<tr>
<td>Gotzsche</td>
<td>s</td>
<td>n</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>8.8</td>
<td>nr</td>
<td>15</td>
<td>25,38</td>
<td>0–8</td>
<td>n</td>
<td>25</td>
</tr>
<tr>
<td>Imperiale</td>
<td>g</td>
<td>n</td>
<td>m</td>
<td>ac</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>5</td>
<td>nr</td>
<td>0.79</td>
<td>&lt;10</td>
<td>0–5</td>
<td>n</td>
<td>82</td>
</tr>
<tr>
<td>Jadad</td>
<td>g</td>
<td>y</td>
<td>r</td>
<td>pool</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>6</td>
<td>y</td>
<td>0.65,0.75</td>
<td>&lt;10</td>
<td>0–8</td>
<td>y</td>
<td>56</td>
</tr>
<tr>
<td>Jonas</td>
<td>g</td>
<td>n</td>
<td>m</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>20</td>
<td>nr</td>
<td>0.6</td>
<td>20</td>
<td>0–100</td>
<td>y</td>
<td>54</td>
</tr>
<tr>
<td>Kleijnen</td>
<td>g</td>
<td>p</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>7</td>
<td>nr</td>
<td>0.87</td>
<td>15</td>
<td>0–100</td>
<td>y</td>
<td>42</td>
</tr>
<tr>
<td>Koeps</td>
<td>s</td>
<td>p</td>
<td>m</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>17</td>
<td>nr</td>
<td>0.8</td>
<td>15</td>
<td>0–100</td>
<td>y</td>
<td>37</td>
</tr>
<tr>
<td>Linde</td>
<td>g</td>
<td>y</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>24</td>
<td>nr</td>
<td>30</td>
<td>0–100</td>
<td>0–8</td>
<td>y</td>
<td>55</td>
</tr>
<tr>
<td>Nourmohamed</td>
<td>s</td>
<td>p</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>8</td>
<td>nr</td>
<td>10</td>
<td>15</td>
<td>0–10</td>
<td>y</td>
<td>55</td>
</tr>
<tr>
<td>Onghena</td>
<td>s</td>
<td>n</td>
<td>m&amp;c</td>
<td>ac</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>14</td>
<td>nr</td>
<td>&gt;0.6</td>
<td>15</td>
<td>0–2–26</td>
<td>y</td>
<td>13</td>
</tr>
<tr>
<td>Pouyward</td>
<td>g</td>
<td>p</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>34</td>
<td>nr</td>
<td>0.99,0.71</td>
<td>30</td>
<td>0–34</td>
<td>y</td>
<td>45</td>
</tr>
<tr>
<td>Reisch</td>
<td>g</td>
<td>n</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>34</td>
<td>nr</td>
<td>0.99,0.71</td>
<td>30</td>
<td>0–34</td>
<td>y</td>
<td>45</td>
</tr>
<tr>
<td>Smith</td>
<td>s</td>
<td>n</td>
<td>m</td>
<td>ac</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>8</td>
<td>nr</td>
<td>nr</td>
<td>10</td>
<td>0–40</td>
<td>n</td>
<td>62</td>
</tr>
<tr>
<td>Spitzer</td>
<td>s</td>
<td>n</td>
<td>m</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>5</td>
<td>nr</td>
<td>nr</td>
<td>25</td>
<td>0–5</td>
<td>n</td>
<td>nr</td>
</tr>
<tr>
<td>Taps</td>
<td>g</td>
<td>n</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>29</td>
<td>nr</td>
<td>nr</td>
<td>30</td>
<td>0–100</td>
<td>y</td>
<td>nr</td>
</tr>
<tr>
<td>Ter Riette</td>
<td>s</td>
<td>p</td>
<td>m&amp;c</td>
<td>ac</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>18</td>
<td>nr</td>
<td>0.93</td>
<td>15</td>
<td>1–100</td>
<td>y</td>
<td>47</td>
</tr>
</tbody>
</table>
Several problems with all measures

- Mix reporting and actual study design
- No empirical evidence for weighting the scores
- Reliability and validity not measured
- Does not typically assess: the question asked, how the study is conducted, whether the study is reported (publication bias, selective outcome reporting)

- What does a quality “score” mean? How can you use them?
Cochrane Criteria for RCTs

- Sequence generation
- Allocation concealment
- Blinding of participants, personnel and outcome assessors
- Incomplete outcome data
- Selective outcome reporting
- Other threats to validity
How to use Risk of Bias Assessment

In meta-analysis

Exclude studies
Report descriptively
Sensitivity analyses
Figure 1. Methodological quality summary: review authors' judgements about each methodological quality item for each included study.
Figure 2. Methodological quality graph: review authors' judgements about each methodological quality item presented as percentages across all included studies.
Sensitivity analysis

Odds Ratio = 1.0

Favors treatment vs. Favors control

LOW RoB studies

All studies

LOW RoB studies ONLY

Odds Ratio

REMOVE HIGH RoB studies
Sensitivity analysis

Odds Ratio 1.0

HIGH RoB studies ONLY

All studies

Odds Ratio

Favors treatment Favors control

REMOVE LOW RoB studies
Reporting of clinical research has improved as risk of bias assessments for systematic reviews and other purposes became more prevalent and standards for reporting were implemented by journals.
Special Communication

Improving the Quality of Reporting of Randomized Controlled Trials

The CONSORT Statement

Colin Begg, PhD; Mildred Cho, PhD; Susan Eastwood, ELS(D); Richard Horton, MB; David Moher, MSc; Ingram Olkin, PhD; Roy Pitkin, MD; Drummond Rennie, MD; Kenneth F. Schulz, PhD; David Simel, PhD; Donna F. Stroup, PhD

THE RANDOMIZED controlled trial (RCT), more than any other methodology, can have a powerful and immediate impact on patient care. Ideally, the report of such an evaluation needs to convey to the reader relevant information concerning the design, conduct, analysis, and generalizability of the trial. This information should provide the reader with the ability to make informed judgments regarding the internal and external validity of the trial. Adequate and complete reporting also benefits editors and reviewers in their deliberations regarding submitted manuscripts. For RCTs to ultimately benefit patients, the published report should be of the highest possible standard.

For editorial comment see p 649.

Evidence produced repeatedly over the past 30 years indicates a wide chasm between what a trial should report and what is actually published in the literature.

In a review of 71 RCTs in the New England Journal of Medicine, we found that failure to report at least 40% of the important results (both positive and negative) published between 1960 and 1979. The authors reported that the vast majority of those with too few patients to observe moderate or large differences. Twenty years later, THE JOURNAL reported research indicating few improvements in this situation and expressed a concern about the reporting of RCTs in general.

In an effort to correct these and other problems, the Standards of Reporting Trials (SORT) group met on October 7 and 8, 1994. At the conclusion of the 2-day meeting, the group presented a new proposal for the reporting of RCTs: structured reporting. The proposal includes criteria for the items that should be included in the report of a trial, provided empirical evidence as to why the items should be included, and provided a format showing how they should be included.

Independently, approximately 5 months later (March 14 to 16, 1994), another group, the Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature, met to discuss similar challenges regarding the reporting of clinical trials. Their proposal consisted of a checklist of items that should be included when reporting a clinical trial, along with a suggestion that editors add it to the Instructions for Authors.

A subsequent Editorial urged both groups to meet and decide which recommendations from each group's proposal should be retained. Besides being pragmatic, this suggestion had the potential for increasing consensus, which in turn might afford a greater chance of meaningful change in the reporting of clinical trials to a wider audience.

On September 20, 1995, a total of 9 members including editors, epidemiologists, and statisticians of the SORT group and the Asilomar Working Group met in Chicago, Ill. Two other people participated in the meeting: a journal editor (Dr Simel) and an operations manager (Dr Simel).

METHODS

We started the day by reviewing both the SORT and Asilomar checklists in order to ascertain which items covered similar content areas and which ones were unique. Those items having similar content areas were then reviewed individually. We decided, a priori, to keep only the items that were emphasized in both checklists, and provide empirical evidence, when available, that not reporting them resulted in bias in the estimates of the effects of interventions. We used common sense for those items included for which there was no empirical evidence. The selection of items was achieved using a modified Delphi process.

We also emphasized the need to keep the number of items to a minimum, while maintaining adequate standards of reporting RCTs. We used a similar approach in deciding which of the unique items should remain in the resulting checklist. The day ended with a discussion on the use of the flow diagram proposed by the SORT group and the format of a trial report. Within a week of the meeting, a draft report was circulated to the entire group for further refinement. This process was continued until we felt the report accurately represented what had gone on during the meeting.

RESULTS

This meeting resulted in the Consolidated Standards of Reporting Trials (CONSORT) statement—a check list (Table) and a flow diagram (Figure). The checklist consists of 21 items that pertain mainly to the methods, results, and discussion of an RCT report and identify key pieces of information necessary to verify the internal and external validity of the report. We have included at least 1 reference for each item, when appropriate (Table). The flow diagram provides information about the progress of patients throughout a 2-group parallel-
Improvements in Reporting

- 50 evaluations of reporting in CONSORT-endorsing vs. Non-CONSORT endorsing journals
- 25 of 27 reporting elements improved post-CONSORT
- Largest improvements:
  - Allocation concealment (RR 1.81, CI 1.25, 2.61)
  - Sequence generation (RR 1.59, CI 1.38, 1.84)
  - Any mention of blinding (RR 1.23, CI 0.98, 1.55)
  - Sample size (1.61, CI 1.13, 2.29)
Improvements in Reporting

Figure 4. Pooled risk ratios across assessed 2001 CONSORT checklist items with 99% confidence intervals for primary comparison, adherence of RCTs published in CONSORT-endorsing journals versus RCTs published in CONSORT non-endorsing journals. Plot generated in Comprehensive Meta-analysis Version 2.0 (CMA).

<table>
<thead>
<tr>
<th>CONSORT Checklist Item</th>
<th># of Evaluations</th>
<th># of RCTs</th>
<th>RR</th>
<th>99% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title and Abstract</td>
<td>7</td>
<td>1,233</td>
<td>1.13</td>
<td>(0.96, 1.33)</td>
</tr>
<tr>
<td>Introduction</td>
<td>5</td>
<td>513</td>
<td>1.07</td>
<td>(1.01, 1.14)</td>
</tr>
<tr>
<td>Participants</td>
<td>6</td>
<td>683</td>
<td>0.95</td>
<td>(0.96, 1.05)</td>
</tr>
<tr>
<td>Interventions</td>
<td>6</td>
<td>638</td>
<td>1.00</td>
<td>(0.95, 1.05)</td>
</tr>
<tr>
<td>Objectives</td>
<td>5</td>
<td>540</td>
<td>1.01</td>
<td>(0.96, 1.06)</td>
</tr>
<tr>
<td>Outcomes</td>
<td>8</td>
<td>1,302</td>
<td>1.17</td>
<td>(0.95, 1.44)</td>
</tr>
<tr>
<td>Sample Size</td>
<td>11</td>
<td>1,843</td>
<td>1.61</td>
<td>(1.13, 2.29)</td>
</tr>
<tr>
<td>Sequence Generation</td>
<td>14</td>
<td>2,231</td>
<td>1.59</td>
<td>(1.30, 1.94)</td>
</tr>
<tr>
<td>Allocation Concealment</td>
<td>16</td>
<td>2,396</td>
<td>1.81</td>
<td>(1.26, 2.52)</td>
</tr>
<tr>
<td>Implementation</td>
<td>5</td>
<td>498</td>
<td>1.47</td>
<td>(0.65, 3.32)</td>
</tr>
<tr>
<td>Blinding of Participants</td>
<td>5</td>
<td>711</td>
<td>1.39</td>
<td>(0.97, 1.92)</td>
</tr>
<tr>
<td>Blinding of Intervention</td>
<td>5</td>
<td>710</td>
<td>1.26</td>
<td>(0.74, 2.12)</td>
</tr>
<tr>
<td>Blinding of Outcome Assessor</td>
<td>5</td>
<td>719</td>
<td>1.72</td>
<td>(0.59, 4.30)</td>
</tr>
<tr>
<td>Blinding of Data Analyst</td>
<td>3</td>
<td>497</td>
<td>3.56</td>
<td>(0.40, 31.6)</td>
</tr>
<tr>
<td>Blinding Any Description</td>
<td>8</td>
<td>1,851</td>
<td>1.23</td>
<td>(0.93, 1.62)</td>
</tr>
<tr>
<td>Statistical Methods</td>
<td>9</td>
<td>894</td>
<td>1.03</td>
<td>(0.90, 1.18)</td>
</tr>
<tr>
<td>Participant Flow</td>
<td>16</td>
<td>2,461</td>
<td>1.16</td>
<td>(0.94, 1.44)</td>
</tr>
<tr>
<td>Recruitment</td>
<td>6</td>
<td>959</td>
<td>1.03</td>
<td>(0.75, 1.41)</td>
</tr>
<tr>
<td>Baseline Data</td>
<td>5</td>
<td>529</td>
<td>1.07</td>
<td>(0.94, 1.22)</td>
</tr>
<tr>
<td>Numbers Analysed</td>
<td>13</td>
<td>2,145</td>
<td>1.22</td>
<td>(0.90, 1.55)</td>
</tr>
<tr>
<td>Outcomes and Estimation</td>
<td>6</td>
<td>617</td>
<td>1.00</td>
<td>(0.95, 1.05)</td>
</tr>
<tr>
<td>Ancillary Analyses</td>
<td>4</td>
<td>378</td>
<td>1.31</td>
<td>(0.48, 3.58)</td>
</tr>
<tr>
<td>Adverse Events</td>
<td>8</td>
<td>911</td>
<td>1.14</td>
<td>(0.86, 1.52)</td>
</tr>
<tr>
<td>Interpretation</td>
<td>5</td>
<td>540</td>
<td>1.01</td>
<td>(0.96, 1.06)</td>
</tr>
<tr>
<td>Generalisability</td>
<td>5</td>
<td>540</td>
<td>1.22</td>
<td>(0.88, 1.70)</td>
</tr>
<tr>
<td>Overall Evidence</td>
<td>4</td>
<td>317</td>
<td>1.03</td>
<td>(0.91, 1.17)</td>
</tr>
</tbody>
</table>

Bottom line on bias

- Assess specific risk of bias criteria relevant to the study design
- Don’t confuse risk of bias with reporting criteria, imprecision
- Don’t let a lack of reporting be an excuse for not assessing risk of bias
- Don’t use “quality” or “methodology” scores
- Decide *a priori* what to do with the risk of bias assessment: exclude studies, report descriptively, sensitivity analysis