Investigating and dealing with bias in randomised trials and meta-analyses

Jonathan Sterne
Department of Social Medicine, University of Bristol, UK

Acknowledgements: Tony Ades, Douglas Altman, Rebecca Beynon, John Carlin, Jon Deeks, Matthias Egger, Ross Harris, Lise Lotte Gluud, Christian Gluud, Peter Juni, Jelena Savovic, Ken Schulz, Nicky Welton, Lesley Wood
Outline

• The epidemiology of bias in systematic reviews
• Variability in the effect of trial quality, and its implications
• Bayesian framework for including results of potentially biased studies in meta-analyses
• Discussion and possible developments
Meta-epidemiology
(Naylor, *BMJ* 1997; **315**: 617-619)

- Identify a large number of meta-analyses
- Record characteristics of individual studies (quality, type of publication, language etc.)
- Compare treatment effects *within* each meta-analysis (for example high-quality vs. low-quality according to some dimension of trial quality)
- Estimate *ratio of odds ratios* comparing high quality and low quality trials
Allocation concealment: combined evidence

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Ratio of odds ratios (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schulz, 1995</td>
<td>0.66 (0.59, 0.73)</td>
</tr>
<tr>
<td>Moher, 1998</td>
<td>0.63 (0.45, 0.88)</td>
</tr>
<tr>
<td>Kjaergard, 2001</td>
<td>0.60 (0.37, 0.97)</td>
</tr>
<tr>
<td>Balk, 2002</td>
<td>0.95 (0.83, 1.09)</td>
</tr>
<tr>
<td>Egger, 2003</td>
<td>0.79 (0.70, 0.89)</td>
</tr>
<tr>
<td>Als-Neilsen, 2004</td>
<td>1.02 (0.93, 1.13)</td>
</tr>
<tr>
<td>Overall (I-squared = 88.6%, p = 0.000)</td>
<td>0.79 (0.66, 0.95)</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis
Blinding: combined evidence

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Ratio of odds ratios (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schulz, 1995</td>
<td>0.83 (0.71, 0.96)</td>
</tr>
<tr>
<td>Moher, 1998</td>
<td>1.11 (0.76, 1.63)</td>
</tr>
<tr>
<td>Kjaergard, 2001</td>
<td>0.56 (0.33, 0.98)</td>
</tr>
<tr>
<td>Balk, 2002</td>
<td>0.98 (0.81, 1.27)</td>
</tr>
<tr>
<td>Egger, 2003</td>
<td>0.88 (0.75, 1.04)</td>
</tr>
<tr>
<td>Overall (I-squared = 27.6%, p = 0.237)</td>
<td>0.88 (0.78, 0.99)</td>
</tr>
</tbody>
</table>

Blinding: combined evidence

NOTE: Weights are from random effects analysis
The death of quality scores

• 25 known checklists
• Between 3 and 34 components
• Frequently no definitions of quality
• Most components said to be based on “accepted criteria”

(Moher et al. Controlled Clinical Trials 1995; 16: 62-73)
“Quality scores are useless and potentially misleading”

“perhaps the most insidious form of subjectivity masquerading as objectivity is ‘quality scoring’. This practice subjectively merges objective information with arbitrary judgements in a manner that can obscure important sources of heterogeneity among study results”

Greenland Am.J.Epidemiol. 1994;140:290-296
Relative risk with 95% CI

Re-analysis using 25 different scales to assess trial quality

**Endpoint:** Deep vein thrombosis

Jüni et al. *JAMA* 1999 282: 1054-1060
Bias assessment in Cochrane reviews

• “Risk of Bias” project led by Julian Higgins and Doug Altman

• Cochrane reviewers are now explicitly advised not to use quality scores

• Instead, they will be asked to record details of key aspects of trial conduct (allocation concealment, blinding of patients, therapists and outcome assessors, etc.) in a standardised way

• For each of these, reviewers will be asked to judge whether there is a risk of bias in the results of the trial because of the way that the trial was done
  – “yes”: high risk of bias
  – “no”: low risk of bias
Bias assessment in Cochrane reviews

Cochrane risk of bias project: mandatory items to address

1. Sequence generation (randomisation)
2. Allocation concealment
3. Blinding of participants, personnel and outcomes
4. Incomplete outcome data (attrition and exclusions)
5. Selective outcome reporting
6. Other (including topic-specific, design-specific)
Analysis of meta-epidemiological studies (1)

• Suppose we have data from $M$ meta-analyses, containing a total of $S$ studies

• To estimate the effect of a binary study characteristic $C$ on estimated treatment effects we fit the model:

$$\text{logit}(\pi) = \beta_0 + \beta_1 I_t + \beta_2 I_{tc} \sum_{i=2}^{M} \gamma_i I_{tm_i} + \sum_{j=2}^{S} \delta_j I_{s_j}$$

where:

– $\pi = \text{Pr}(\text{adverse outcome event})$

– $I_t, I_{tc}, \{I_{tm}\}$ and $\{I_s\}$ are indicator variables denoting, respectively, the effects of treatment, the treatment-characteristic interaction, the treatment-meta-analysis interactions and study number

(Sterne et al. Statistics in Medicine 2002; 21: 1513-1524)
Analysis of meta-epidemiological studies (2)

• Two-stage approach:
  – Estimate the effect of publication status and language of publication separately in each meta-analysis
  – Combine estimates across meta-analyses

• Analyses using fixed effects within and between meta-analyses are equivalent to the logistic regression analyses used in previous studies

• For the effects of well-known predictors of treatment effects (components of trial quality, publication status, language of publication) there is clear evidence of between-meta-analysis heterogeneity

(Sterne et al. Statistics in Medicine 2002; 21: 1513-1524)
Data from Schulz et al. (JAMA 1995)

The effects of components of trial quality are usually imprecisely estimated in a single meta-analysis.

Little hope of adjusting for the effects of trial quality using only the information available in the meta-analysis.
Data from Schulz et al. (JAMA 1995)

The ratio of odds ratios (ROR) comparing studies that were not and were adequately concealed was 0.67 (95% CI 0.57 to 0.78)

The between-meta-analysis variance in the log ROR was 0.065
Variability in effects of trial conduct

• Good a priori reasons to expect this:
  – Blinding more important when outcomes are subjectively assessed?
  – Differences between placebo-controlled trials and comparative trials?
  – Differences between areas of medicine?
Effect of inadequate/unclear versus adequate allocation concealment

<table>
<thead>
<tr>
<th>Comparison (number of meta-analyses)</th>
<th>Number of trials</th>
<th>ROR (95% CI)</th>
<th>Heterogeneity variance (p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall (102)</td>
<td>532 vs. 272</td>
<td>0.83 (0.74, 0.93)</td>
<td>0.11 (p&lt;0.001)</td>
</tr>
<tr>
<td>All-cause mortality (23)</td>
<td>117 vs. 90</td>
<td>1.01 (0.90, 1.15)</td>
<td>0.02 (p=0.235)</td>
</tr>
<tr>
<td>Other outcomes (79)</td>
<td>415 vs. 182</td>
<td>0.76 (0.66, 0.87)</td>
<td>0.14 (p&lt;0.001)</td>
</tr>
<tr>
<td>Objective outcomes (62)</td>
<td>310 vs. 174</td>
<td>0.91 (0.80, 1.03)</td>
<td>0.11 (p&lt;0.001)</td>
</tr>
<tr>
<td>Subjective outcomes (40)</td>
<td>222 vs. 98</td>
<td>0.69 (0.59, 0.82)</td>
<td>0.07 (p=0.011)</td>
</tr>
</tbody>
</table>

Wood et al., BMJ, nearly in press
Effect of inadequate/unclear versus adequate blinding

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Number of trials</th>
<th>ROR (95% CI)</th>
<th>Heterogeneity variance (p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall (76)</td>
<td>314 vs. 432</td>
<td>0.93 (0.83, 1.04)</td>
<td>0.11 (p&lt;0.001)</td>
</tr>
<tr>
<td>All-cause mortality (18)</td>
<td>79 vs. 121</td>
<td>1.04 (0.95, 1.14)</td>
<td>0.01 (p=0.265)</td>
</tr>
<tr>
<td>Other outcomes (58)</td>
<td>235 vs. 311</td>
<td>0.83 (0.70, 0.98)</td>
<td>0.18 (p&lt;0.001)</td>
</tr>
<tr>
<td>Objective outcomes (44)</td>
<td>210 vs. 227</td>
<td>1.01 (0.92, 1.10)</td>
<td>0.08 (p&lt;0.001)</td>
</tr>
<tr>
<td>Subjective outcomes (32)</td>
<td>104 vs. 205</td>
<td>0.75 (0.61, 0.93)</td>
<td>0.14 (p&lt;0.001)</td>
</tr>
</tbody>
</table>

Wood et al., BMJ, nearly in press
Effects of flaws in the conduct of trials

• Change in average intervention effect (bias)
  – the focus of most previous research
• Variability in average effect of bias between-meta-analyses
•Increases in between-trial variability (heterogeneity)
Why do meta-epidemiology?

• Improve the quality of future trials, by identifying important dimensions of trial quality
  – CONSORT statement

• Provide empirical evidence on how to combine evidence from trials of differing methodological quality?
Notation and bias model (1)

• Two types of studies, denoted by $L$ (low risk of bias) and $H$ (high risk of bias) due to a specific flaw in their conduct, e.g.
  – RCTs in which randomisation was ($L$) and was not ($H$) adequately concealed
  – RCTs that did ($L$) and did not ($H$) blind physicians, participants and outcome assessors

• We consider a new meta-analysis (indexed by $m^*$) in which $n_L$ studies at low risk and $n_H$ at high risk of bias are identified

• Estimates of intervention effect from study $j$ are denoted $\hat{\beta}_{j,m^*}$ $(j = 1, \ldots, n_L + n_H)$, with variances $\sigma_{j,m^*}^2$
Notation and bias model (2)

• For study $j$, we assume: $\hat{\beta}_{j,m^*} \sim N(\mu_{j,m^*} + \delta_{j,m^*}, \sigma_{j,m^*}^2)$

• The true intervention effect in trial $j$ is $\mu_{j,m^*}$

• $\delta_{j,m^*} = 0$ is the bias in this estimate, assumed zero in type $L$ studies

Unless the bias is known, we can’t use the type $H$ studies
Notation and bias model (3)

- We distinguish two types of variation in $\delta_{j,m^*}$
  - *within* meta-analysis variation
  - *between* meta-analysis variation

\[
\delta_{j,m^*} = 0 \quad j = 1, \ldots, n_L
\]

\[
\delta_{j,m^*} \sim N(d_{m^*}, \kappa^2), \quad d_{m^*} \sim N(\delta_0, \phi^2) \quad j = (n_L + 1), \ldots, (n_L + n_H)
\]

- In addition, there is uncertainty in the mean bias $\delta_0$:

\[
\delta_0 \sim N(D_0, \sigma_{D_0}^2)
\]

- We can estimate $d_m$ and $\kappa^2$ using data from a single meta-analysis, but information about $\delta_0$ or $\phi^2$ can only be estimated using collections of meta-analyses
Consequences for a single study

- For a single study at low risk of bias:
  \[
  E(\mu_{j,m^*} \mid \{\hat{\beta}_{j,m^*}, \sigma_{j,m^*}^2\}) = \hat{\beta}_{j,m^*}
  \]
  \[
  \text{Var}(\mu_{j,m^*} \mid \{\hat{\beta}_{j,m^*}, \sigma_{j,m^*}^2\}) = \sigma_{j,m^*}^2
  \]

- Given values for \(D_0, \kappa^2, \varphi^2\) and \(\sigma_{D_0}^2\), we can obtain the posterior distribution of the true intervention effect in a single study at high risk of bias:
  \[
  E(\mu_{j,m^*} \mid \{\hat{\beta}_{j,m^*}, \sigma_{j,m^*}^2\}) = \hat{\beta}_{j,m^*} - D_0,
  \]
  \[
  \text{Var}(\mu_{j,m^*} \mid \{\hat{\beta}_{j,m^*}, \sigma_{j,m^*}^2\}) = \sigma_{j,m^*}^2 + \kappa^2 + \varphi^2 + \sigma_{D_0}^2
  \]

- Information from type \(H\) studies will be limited!

- Note that the specific value of the prior average bias \(D_0\) makes no difference to the informational value of the type \(H\) evidence.
Fixed-effect meta-analysis combining type $H$ and $L$ studies

\[
E(\mu_{m*} \mid \text{all evidence}) = \left[ \sum_{j=1}^{n_L} \frac{\hat{\beta}_{j,m*}}{\sigma^2_{j,m*}} + \sum_{j=(n_L+1)}^{n_L+n_H} \frac{\hat{\beta}_{j,m*} - D_0}{\sigma^2_{j,m*} + \kappa^2} \right] w
\]

where \( w = \left[ 1 + \sum_{j=(n_L+1)}^{n_L+n_H} \left( \frac{\sigma^2_{D_0} + \phi^2}{\sigma^2_{j,m*} + \kappa^2} \right) \right]^{-1} \)

\[
V(\mu_{m_{new}} \mid \text{all evidence}) = \left[ \sum_{H \text{ Studies}} \frac{1}{\sigma^2_{j,m_{new}}} + \sum_{L} \left( \frac{1}{\sigma^2_{j,m_{new}} + \kappa^2} \right) w \right]^{-1}
\]
Implications (1)

• **Standard sensitivity analyses are special cases:**

1. Standard inverse-variance-weighted meta-analysis: $D_0 = 0$, $\kappa^2 = 0$, $\varphi^2 = 0$ and $\sigma^2_{D_0} = 0$

2. Omit the type $L$ studies: $\kappa^2$, $\varphi^2$ or $\sigma^2_{D_0} = \infty$
Implications (2)

• Informational value of studies at high risk of bias:

1. Intervention effect from a large type $H$ study has minimum variance $\kappa^2 + \sigma_{D_0}^2 + \varphi^2$

2. A meta-analysis of $n_H$ large type $H$ studies has minimum variance $\kappa^2 / n_L + \sigma_{D_0}^2 + \varphi^2$

3. Conducting large meta-epidemiological studies could in principle reduce $\sigma_{D_0}^2$, but $\varphi^2$ is a characteristic of the bias

4. However, $\varphi^2$ may be lower in certain situations (eg when outcomes are objectively assessed)

5. A new meta-analysis including both type $L$ and $H$ studies can identify both the underlying intervention effect $\mu_{m*}$ and the expectation of the meta-analysis-specific bias $d_{m*}$
Estimation of bias parameters (1)

- Data from Schulz et al.: 250 trials in 33 meta-analyses classified as adequately (H) or inadequately/unclearly (L) concealed

- Random-effects logistic regression using WinBUGS, using the bias model defined earlier:

\[
\text{logit}(p_{j,m}) = \begin{cases} 
\alpha_{j,m} & \text{Control Arm} \\
\alpha_{j,m} + \mu_m + \delta_{j,m} & \text{Treatment Arm}
\end{cases}
\]

\[
\delta_{j,m} = 0 \quad \text{L studies}
\]

\[
\delta_{j,m} \sim \text{N}(d_m, \kappa^2), \quad d_m \sim \text{N}(\delta_0, \phi^2) \quad \text{H studies}
\]
### Estimation of bias parameters (2)

- **Results:**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>sd</th>
<th>Median</th>
<th>95% credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_0$</td>
<td>-0.47</td>
<td>0.095</td>
<td>-0.46</td>
<td>(-0.65, -0.28)</td>
</tr>
<tr>
<td>$\kappa^2$</td>
<td>0.25</td>
<td>0.063</td>
<td>0.25</td>
<td>(0.15, 0.39)</td>
</tr>
<tr>
<td>$\varphi^2$</td>
<td>0.08</td>
<td>0.072</td>
<td>0.06</td>
<td>(0.00, 0.26)</td>
</tr>
</tbody>
</table>

\[
D_0 = -0.47 \\
\sigma_{D_0}^2 = 0.095^2 = 0.009 \\
\kappa^2 = 0.25 \\
\varphi^2 = 0.08
\]
Examples

- Three Cochrane reviews
- Allocation concealment categorised as adequate or inadequate/unclear
- Parameters of prior from analysis of Schulz data:
  1. $D_0 = -0.47$ (ROR=0.63)
  2. $\sigma_{D_0}^2 = 0.009$
  3. $\kappa^2 = 0.25$
  4. $\varphi^2 = 0.06$
# Example 1. Clozapine versus neuroleptic medication for schizophrenia

## Concealment inadequate ($H$)

<table>
<thead>
<tr>
<th>Location</th>
<th>Treatment Events</th>
<th>Control Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia 1976</td>
<td>11/33</td>
<td>17/31</td>
</tr>
<tr>
<td>Europe 1974</td>
<td>6/110</td>
<td>8/113</td>
</tr>
<tr>
<td>Europe 1984</td>
<td>8/39</td>
<td>15/40</td>
</tr>
<tr>
<td>Germany 1989</td>
<td>2/16</td>
<td>2/16</td>
</tr>
<tr>
<td>Germany 1994</td>
<td>1/18</td>
<td>1/18</td>
</tr>
<tr>
<td>Hong Kong 1974</td>
<td>1/20</td>
<td>1/20</td>
</tr>
<tr>
<td>Japan 1977</td>
<td>4/47</td>
<td>0/41</td>
</tr>
<tr>
<td>Switzerland 1975</td>
<td>1/18</td>
<td>3/22</td>
</tr>
<tr>
<td>Taiwan 1997</td>
<td>2/21</td>
<td>2/19</td>
</tr>
<tr>
<td>USA 1979a</td>
<td>6/16</td>
<td>11/15</td>
</tr>
<tr>
<td>USA 1979b</td>
<td>4/7</td>
<td>5/8</td>
</tr>
<tr>
<td>USA 1987</td>
<td>27/75</td>
<td>36/76</td>
</tr>
<tr>
<td>USA 1994a</td>
<td>2/21</td>
<td>0/20</td>
</tr>
<tr>
<td>USA 1994d</td>
<td>6/25</td>
<td>1/14</td>
</tr>
<tr>
<td>USA 1996a</td>
<td>46/136</td>
<td>58/89</td>
</tr>
<tr>
<td>USA 1997</td>
<td>88/205157/218</td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>215/807317/760</td>
<td></td>
</tr>
</tbody>
</table>

## Concealment adequate ($L$)

<table>
<thead>
<tr>
<th>Location</th>
<th>Treatment Events</th>
<th>Control Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada 1977</td>
<td>6/22</td>
<td>9/28</td>
</tr>
<tr>
<td>Romania 1976</td>
<td>1/20</td>
<td>0/20</td>
</tr>
<tr>
<td>USA 1988</td>
<td>15/126</td>
<td>18/142</td>
</tr>
<tr>
<td>USA 1994b</td>
<td>12/37</td>
<td>21/34</td>
</tr>
<tr>
<td>USA 1996b</td>
<td>3/10</td>
<td>1/11</td>
</tr>
<tr>
<td>Subtotal</td>
<td>37/215</td>
<td>49/235</td>
</tr>
</tbody>
</table>

Overall

<table>
<thead>
<tr>
<th>Treatment Events</th>
<th>Control Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>252/1022366/995</td>
<td></td>
</tr>
</tbody>
</table>

## Odds Ratio

The odds ratio is calculated as the ratio of the treatment group's events to the control group's events. The graph shows the distribution of odds ratios across different studies, with the data points indicating the number of events in the treatment and control groups.
Example 1. Clozapine versus neuroleptic medication for schizophrenia

Concealment inadequate ($H$)

Concealment adequate ($L$)

All

Bias-adjusted

ROR: 0.66 (0.31, 1.41)

Relative weight (type $H$ studies) = 81%

Relative weight (type $H$ studies) = 37%

$w = 0.41$
Example 2. Vaginal misoprostol versus prostaglandin for induction of labour

Concealment inadequate ($H$)
- Chang 1997: 3/30 2/30
- Chuck 1995: 10/49 10/50
- Fletcher 1994: 1/32 3/31
- Herabutya 1997: 19/60 16/50
- Kadanali 1996: 12/112 22/112
- Tabor 1995: 17/68 15/59
- Subtotal: 62/351 68/332

Concealment adequate ($L$)
- Buser 1997: 27/76 17/79
- Howarth 1996: 6/36 15/36
- Mundle 1996: 15/111 12/111
- Varalakis 1995: 8/36 3/33
- Wing 1995a: 10/68 13/67
- Wing 1995b: 28/138 38/137
- Wing 1997: 18/99 20/98
- Subtotal: 112/564 118/561

Overall: 174/915 186/893
Example 2. Vaginal misoprostol versus prostaglandin for induction of labour

Concealment adequate (L)

Concealment inadequate (H)

All

Bias-adjusted

ROR: 0.81 (0.47, 1.40)

Relative weight (type H studies)=36%

Relative weight (type H studies)=13%

Treatment odds ratio (log scale)

w=0.57
Example 3. Ovulation suppression compared to Danazol for endometriosis

### Concealment inadequate (H)

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dmowski 1989</td>
<td>10/18</td>
<td>3/8</td>
</tr>
<tr>
<td>Fedele 1989a</td>
<td>14/20</td>
<td>12/19</td>
</tr>
<tr>
<td>Fedele 1989b</td>
<td>18/30</td>
<td>19/32</td>
</tr>
<tr>
<td>NEET 1992</td>
<td>67/100</td>
<td>45/57</td>
</tr>
<tr>
<td>Noble 1979</td>
<td>6/10</td>
<td>5/12</td>
</tr>
<tr>
<td>Shaw 1992</td>
<td>80/113</td>
<td>41/54</td>
</tr>
<tr>
<td>Subtotal</td>
<td>195/291</td>
<td>125/182</td>
</tr>
</tbody>
</table>

### Concealment adequate (L)

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Henzl 1988</td>
<td>62/104</td>
<td>29/45</td>
</tr>
<tr>
<td>Subtotal</td>
<td>62/104</td>
<td>29/45</td>
</tr>
</tbody>
</table>

### Overall

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtotal</td>
<td>257/395</td>
<td>154/227</td>
</tr>
</tbody>
</table>
Example 3. Ovulation suppression compared to Danazol for endometriosis

ROR: 1.06 (0.46, 2.45)

Relative weight (type H studies) = 75%

Relative weight (type H studies) = 45%

Bias-adjusted: w = 0.58
Meta-confounding?

• Components of trial quality are likely to be associated with each other and with other trial characteristics.
• It follows that crude estimates of the association of individual components of trial quality with treatment effect estimates may be biased.
• Only one published paper (Siersma et al., *Statistics in Medicine* 2007 **26**: 2745-58) addresses this issue.
Multiple dimensions of quality

- Sensitivity analyses rapidly become impossible, or uninterpretable
  - allocation concealment
  - blinding of patient
  - blinding of outcome assessor
  - intention to treat analysis

- Could generalise the approach to correct for a number of dimensions of quality:
  - estimated intervention effect = \( \log \text{OR} - \delta_{AC} - \delta_{BP} - \delta_{BOA} - \delta_{ITT} \)
  - weight in meta-analysis = \( \frac{1}{\sigma^2 + \kappa_{AC}^2 + \kappa_{BP}^2 + \kappa_{BOA}^2 + \kappa_{ITT}^2} \)
Future work

• We need further meta-epidemiological research in order to derive evidence-based priors
  – The BRANDO study (Bias in Randomised AND Observational studies), has combined data from all (except one) published meta-epidemiological studies
  – Combined database (after removing overlapping meta-analyses) has data from around 2500 trials with quality assessment, from around 270 meta-analyses
  – Potential for analyses specific to clinical area, type of control group, type of outcome variable….

• Need to correct for multiple dimensions of quality
  – methods to allow for meta-confounding require both development and application
  – If data become available, how should they be used to formulate multivariate priors?
Conclusions

• If we want to include flawed evidence in a systematic review, then we should downweight and correct for bias, based on evidence from meta-epidemiological studies
  – How far should priors be based on evidence, and how far on other factors?
  – e.g. should we further increase $\kappa^2$ to account for our uncertainty about the relevance of past evidence to future trials?
  – Need to convince meta-analysts that they are using priors even if these are not explicitly acknowledged