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

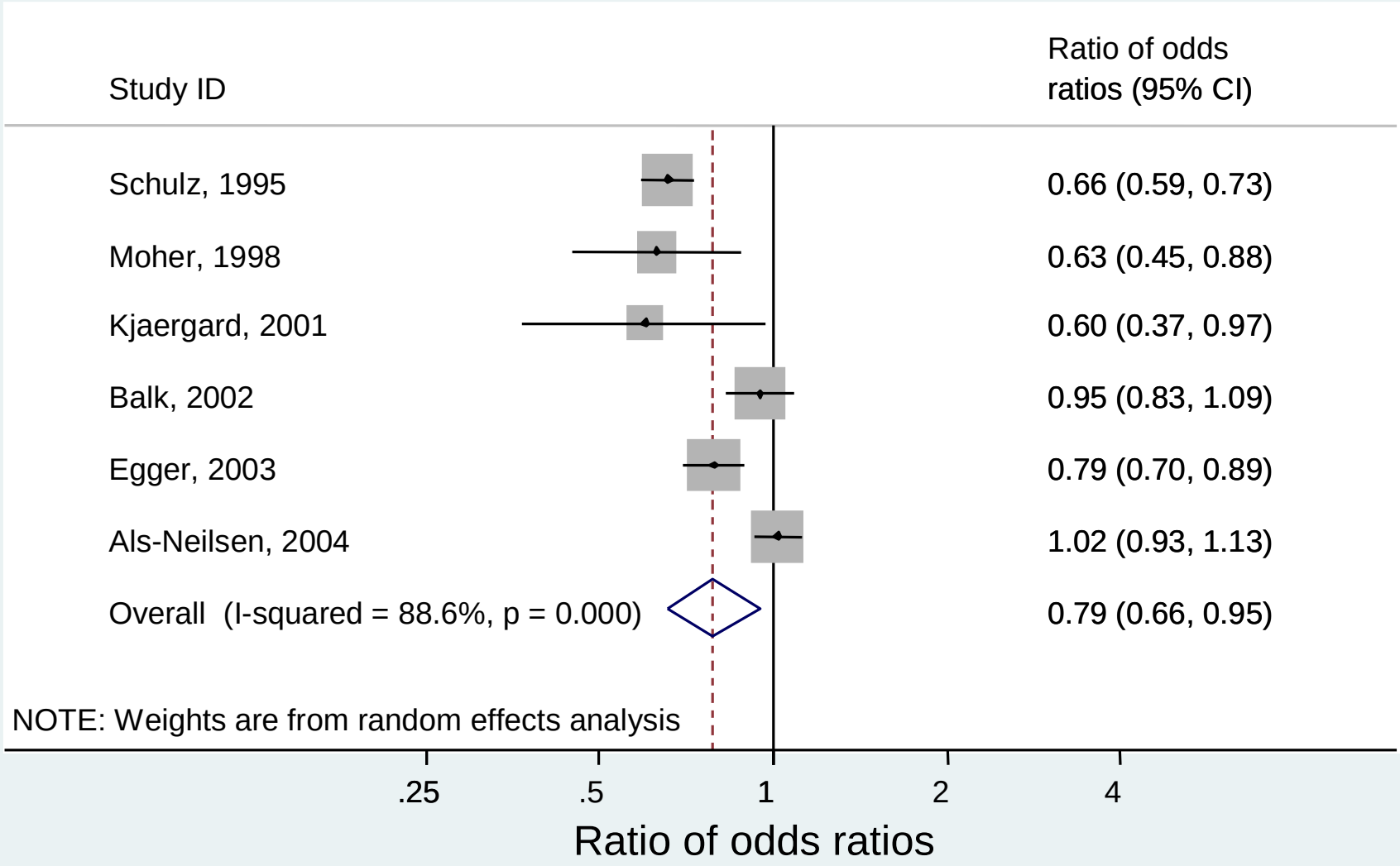
- The epidemiology of bias in systematic reviews
- Variability in the effect of trial quality, and its implications
- Bayesian framework for including results of potentially biased studies in meta-analyses
- Discussion and possible developments

Meta-epidemiology

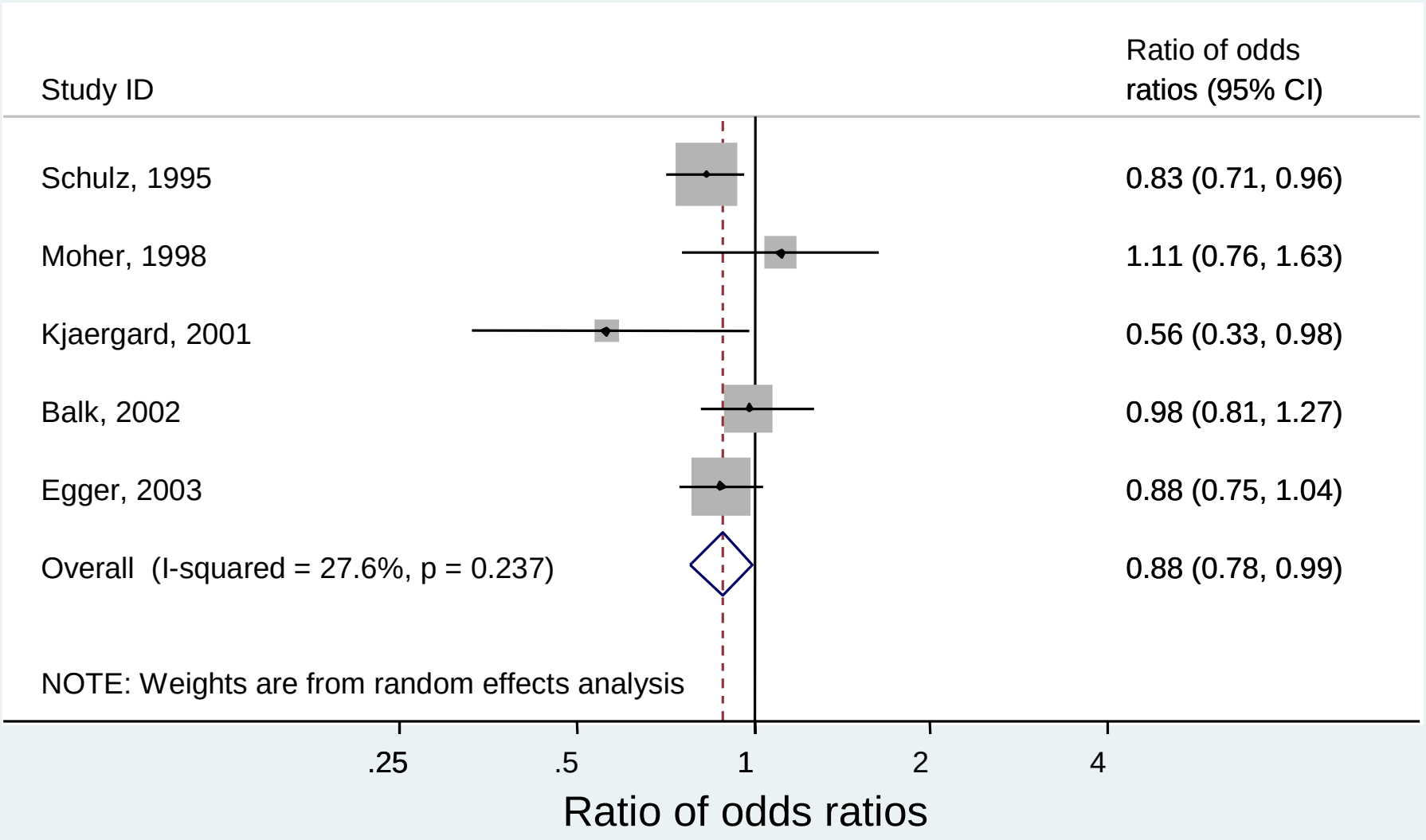
(Naylor, *BMJ* 1997; **315**: 617-619)

- 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



Blinding: combined evidence



The death of quality scores

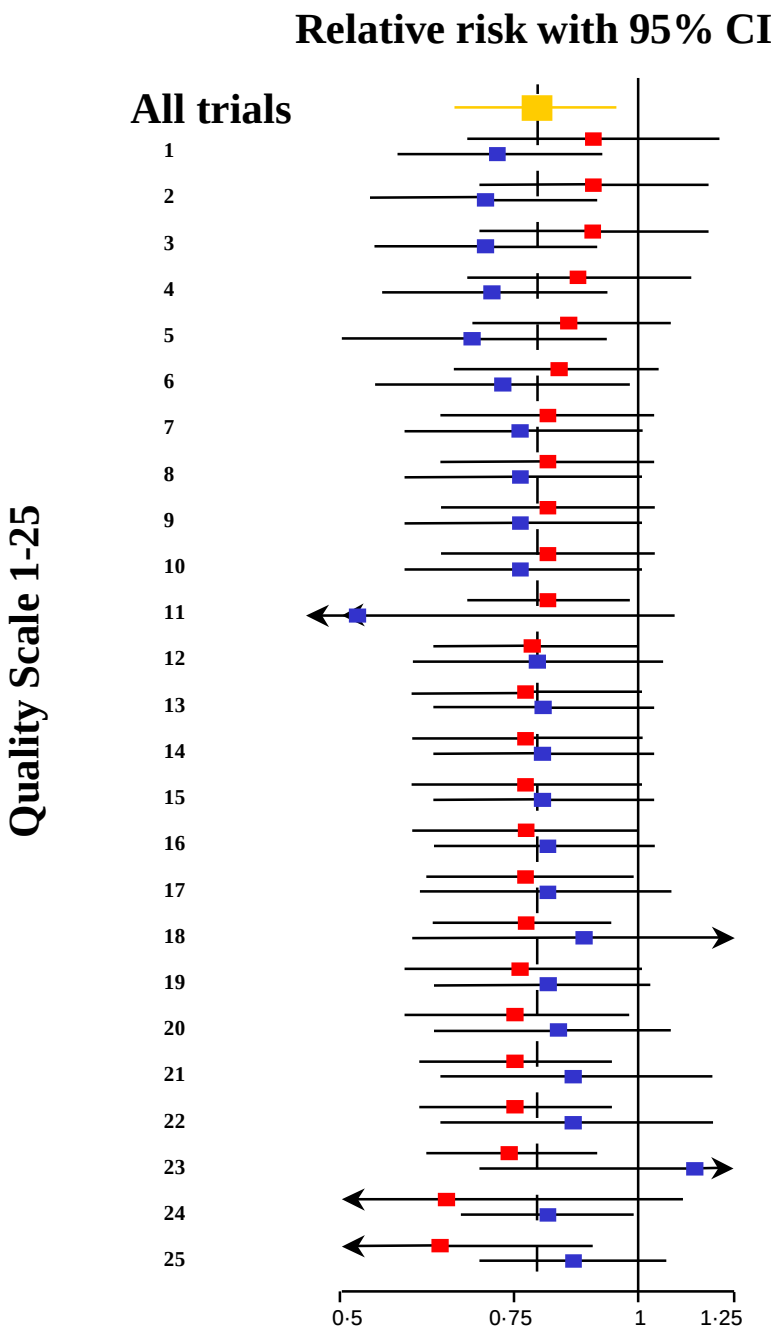
- 25 known checklists
- Between 3 and 34 components
- Frequently no definitions of quality
- Most components said to be based on “accepted criteria”

(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



Re-analysis using 25 different scales to assess trial quality

Endpoint: Deep vein thrombosis

High quality

Low quality

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:

$$\text{logit}(p) = b_0 + b_1 I_t + b_2 I_{tc} \sum_{i=2}^{M} g_i I_{tm_i} + \sum_{j=2}^{S} d_j I_{s_j}$$

where:

- p = Pr(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)

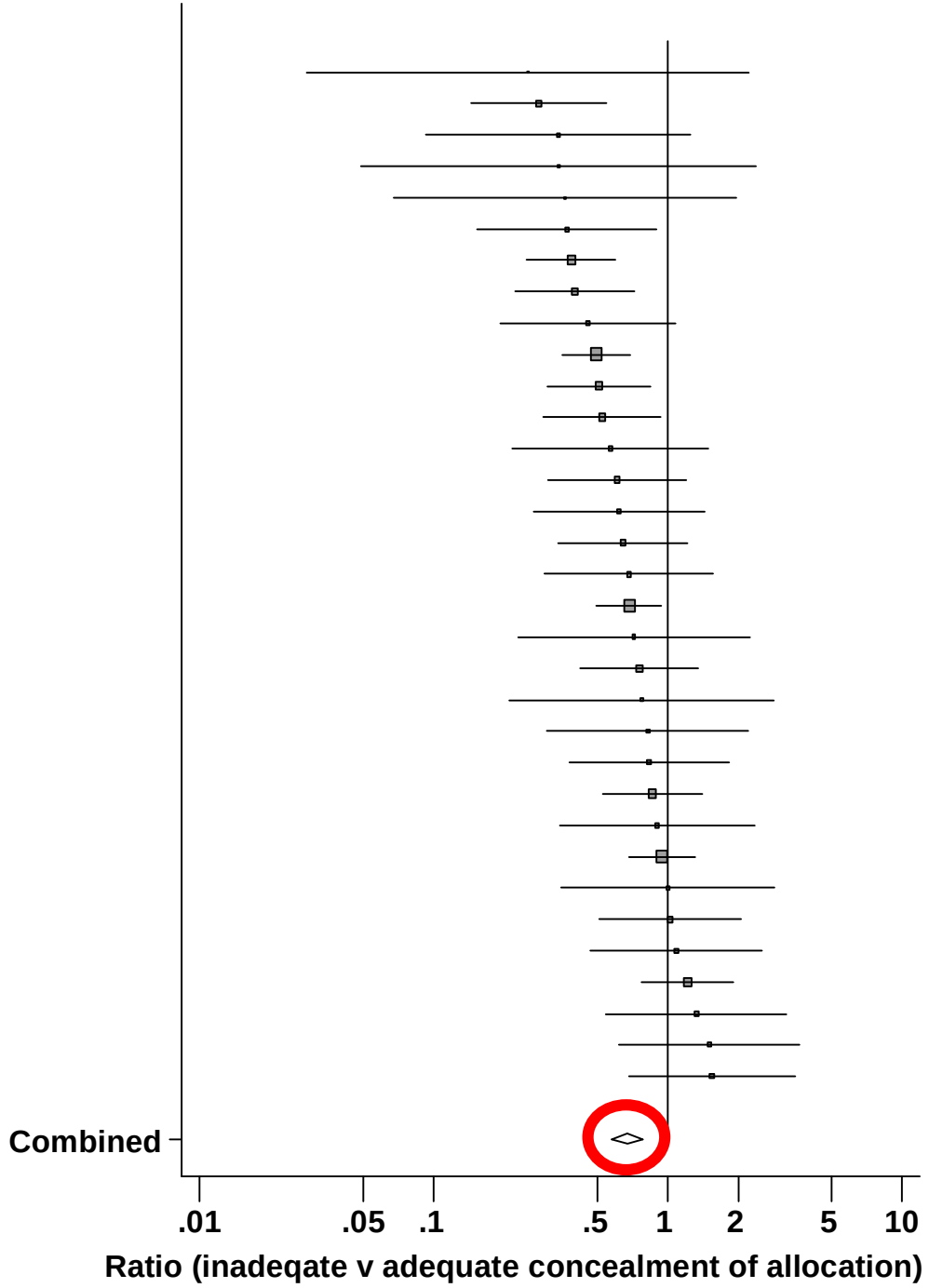
- Two-stage approach:
  - Estimate the effect of publication status and language of publication separately in each meta-analysis
  - Combine estimates across meta-analyses
- Analyses using fixed effects within and between meta-analyses are equivalent to the logistic regression analyses used in previous 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

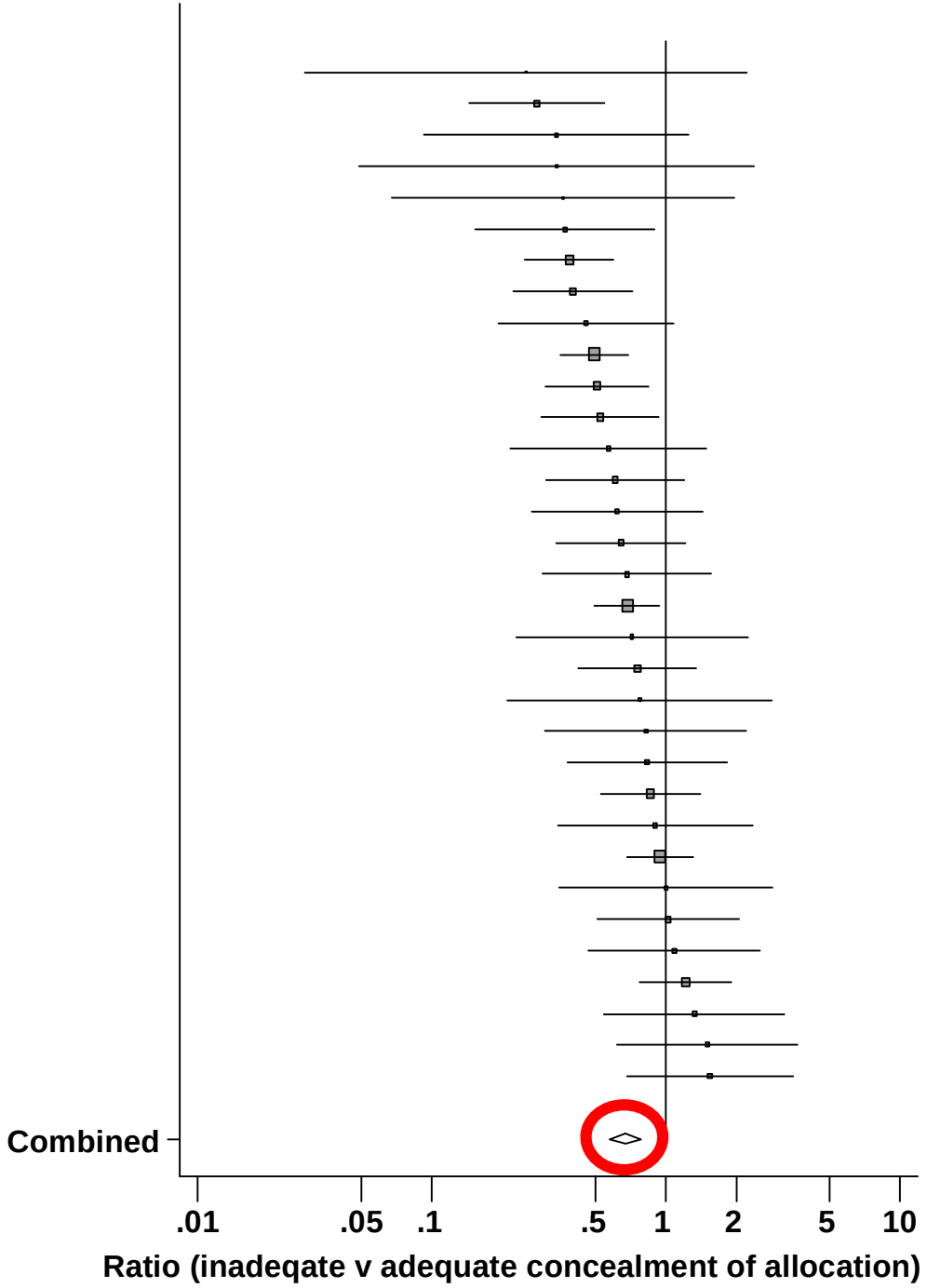
Little hope of adjusting for the effects of trial quality using only the information available in the meta-analysis



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

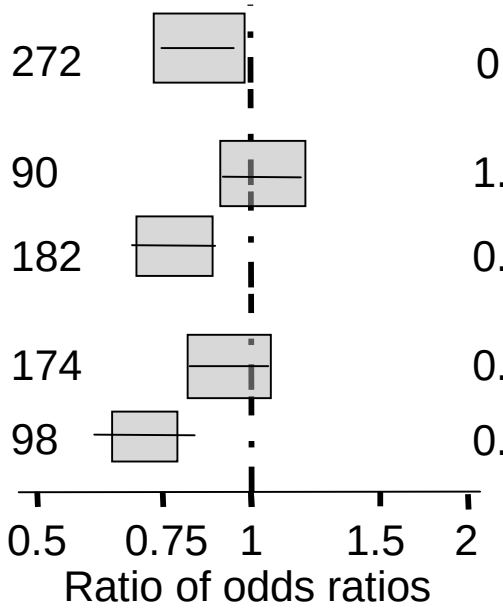


Variability in effects of trial conduct

- 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

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



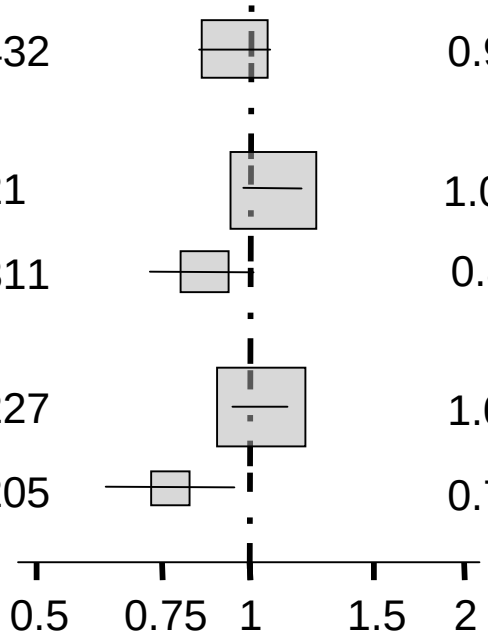
Inadequately concealed more beneficial

Inadequately concealed less beneficial

Wood et al., *BMJ*, nearly in press

Effect of inadequate/unclear versus adequate blinding

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



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 $\hat{b}_{j,m^*}$ $(j = 1, \ldots, n_L + n_H)$, with variances s^2_{j,m^*}

Notation and bias model (2)

- For study j, we assume: $\hat{b}_{j,m^*} \sim \mathrm{N}(m_{j,m^*} + d_{j,m^*}, s^2_{j,m^*})$
- The true intervention effect in trial j is m_{j,m^*}
- $d_{j,m^*} = 0$ 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 d_{j,m^*}
  - *within* meta-analysis variation
  - *between* meta-analysis variation

$$d_{j,m^*} = 0 \qquad j = 1,\ldots,n_L$$

$$d_{j,m^*} \sim \mathrm{N}(d_{m^*}, k^2),\ d_{m^*} \sim \mathrm{N}(d_0, j^2) \quad j = (n_L+1),\ldots,(n_L+n_H)$$

- In addition, there is uncertainty in the mean bias d_0:
  $d_0 \sim \mathrm{N}(\mathrm{D}_0, s^2_{\mathrm{D}_0})$
- We can estimate d_m and k^2 using data from a single meta-analysis, but information about d_0 or j^2 can only be estimated using collections of meta-analyses

Consequences for a single study

- For a single study at low risk of bias:

$$\mathrm{E}(m_{j,m^*} \mid \{\hat{b}_{j,m^*}, s^2_{j,m^*}\}) = \hat{b}_{j,m^*}$$

$$\mathrm{Var}(m_{j,m^*} \mid \{\hat{b}_{j,m^*}, s^2_{j,m^*}\}) = s^2_{j,m^*}$$

- Given values for D_0, k^2, j^2 and $s^2_{D_0}$, we can obtain the posterior distribution of the true intervention effect in a single study at high risk of bias:

$$E(m_{j,m^*} \mid \{\hat{b}_{j,m^*}, s^2_{j,m^*}\}) = \hat{b}_{j,m^*} - D_0,$$

$$\mathrm{Var}(m_{j,m^*} \mid \{\hat{b}_{j,m^*}, s^2_{j,m^*}\}) = s^2_{j,m^*} + k^2 + j^2 + s^2_{D_0}$$

- 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

$$\mathrm{E}(m_{m^*} \mid \text{all evidence}) = \frac{\left[\sum_{j=1}^{n_L} \frac{\hat{b}_{j,m^*}}{s^2_{j,m^*}} + \sum_{j=(n_L+1)}^{(n_L+n_H)} \left(\frac{\hat{b}_{j,m^*} - D_0}{s^2_{j,m^*} + k^2} \right) w \right]}{\left[\sum_{j=1}^{n_L} \frac{1}{s^2_{j,m^*}} + \sum_{j=(n_L+1)}^{(n_L+n_H)} \left(\frac{1}{s^2_{j,m^*} + k^2} \right) w \right]}$$

$$\text{where } w = \left[1 + \sum_{j=(n_L+1)}^{(n_L+n_H)} \left(\frac{s^2_{D_0} + j^2}{s^2_{j,m^*} + k^2} \right) \right]^{-1}$$

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

Implications (1)

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

1. Standard inverse-variance-weighted meta-analysis: $D_0 = 0$, $k^2 = 0$, $j^2 = 0$ and $s^2_{D_0} = 0$
2. Omit the type *L* studies: k^2, j^2 or $s^2_{D_0} = \infty$

Implications (2)

- Informational value of studies at high risk of bias:

1. Intervention effect from a large type *H* study has minimum variance $k^2 + s^2_{D_0} + j^2$
2. A meta-analysis of n_H large type *H* studies has minimum variance $k^2 / n_L + s^2_{D_0} + j^2$
3. Conducting large meta-epidemiological studies could in principle reduce $s^2_{D_0}$, but j^2 is a characteristic of the bias
4. However, j^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 m_{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:

$$logit(p_{j,m}) = \begin{cases} a_{j,m} & \text{Control Arm} \\ a_{j,m} + m_m + d_{j,m} & \text{Treatment Arm} \end{cases}$$

$$d_{j,m} = 0 \qquad L \text{ studies}$$

$$d_{j,m} \sim \mathrm{N}(d_m, k^2), \quad d_m \sim \mathrm{N}(d_0, j^2) \qquad H \text{ studies}$$

Estimation of bias parameters (2)

- Results:

Parameter	Mean	sd	Median	95% credible interval
d_0	-0.47	0.095	-0.46	(-0.65, -0.28)
k^2	0.25	0.063	0.25	(0.15, 0.39)
j^2	0.08	0.072	0.06	(0.00, 0.26)

$$D_0 = -0.47$$

$$s^2_{D_0} = 0.095^2 = 0.009$$

$$k^2 = 0.25$$

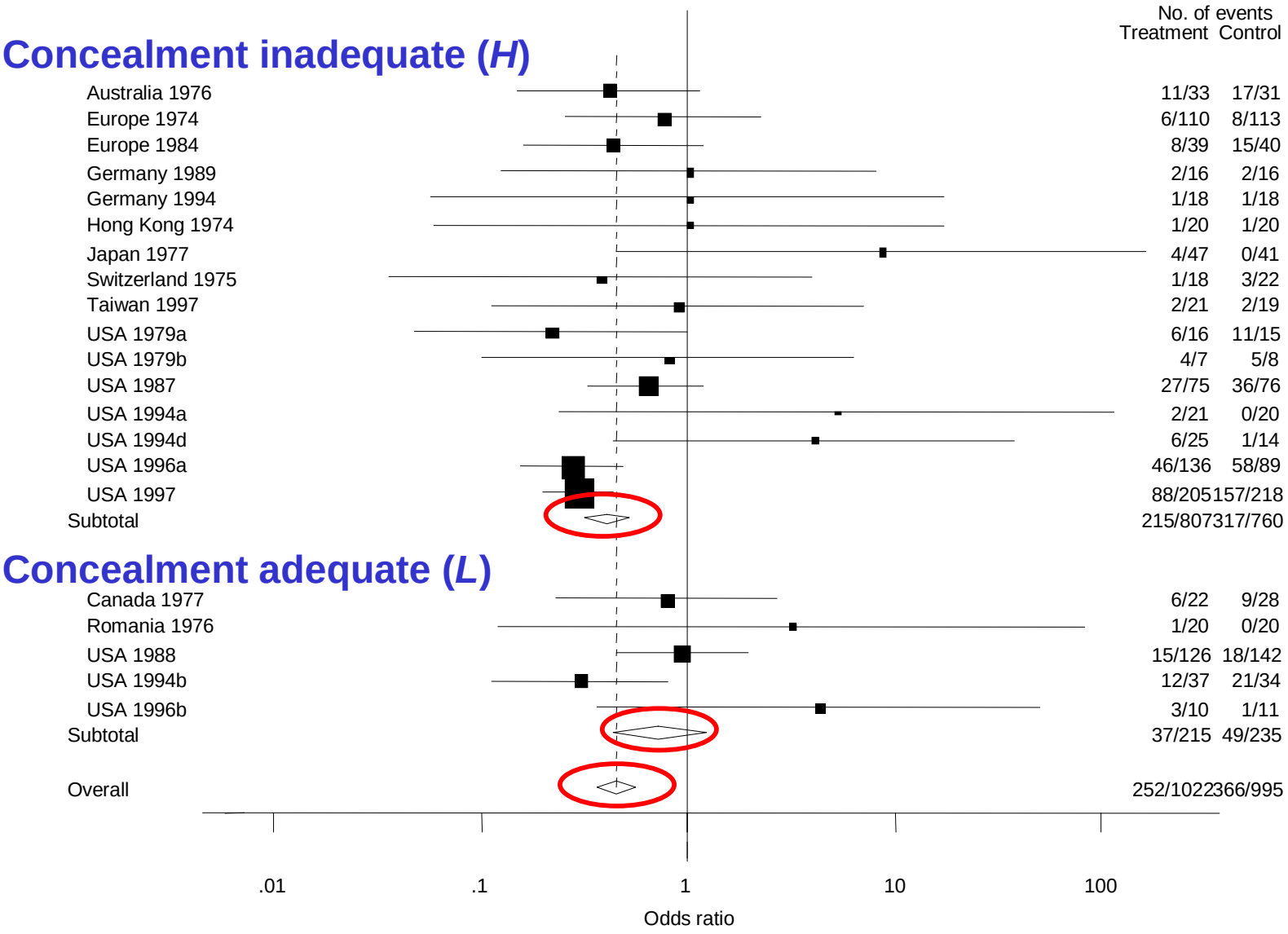
$$j^2 = 0.08$$

Examples

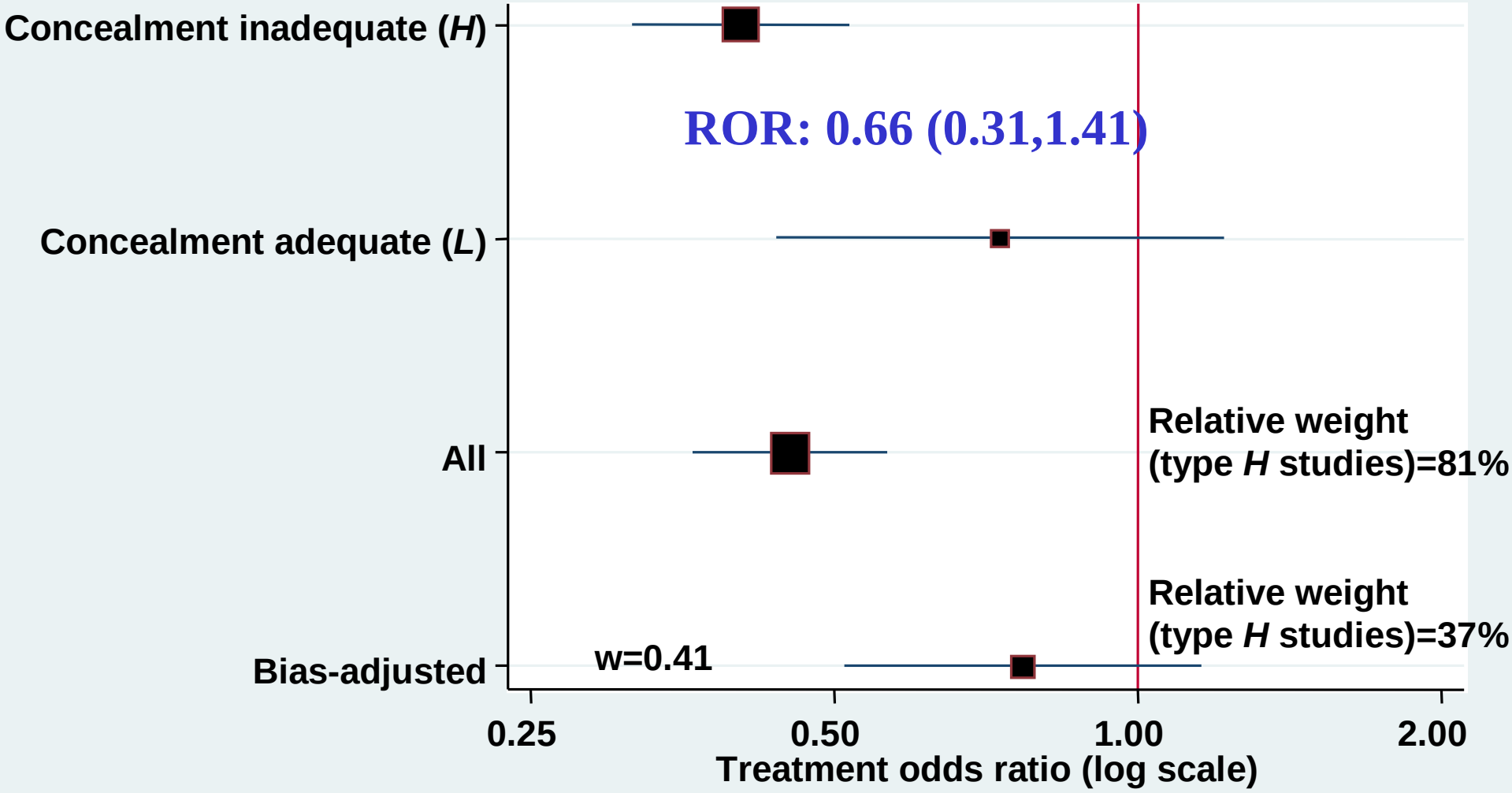
- Three Cochrane reviews
- Allocation concealment categorised as adequate or inadequate/unclear
- Parameters of prior from analysis of Schulz data:

1. D_0 = -0.47 (ROR=0.63)
2. $s^2_{D_0}$ = 0.009
3. k^2 = 0.25
4. j^2 = 0.06

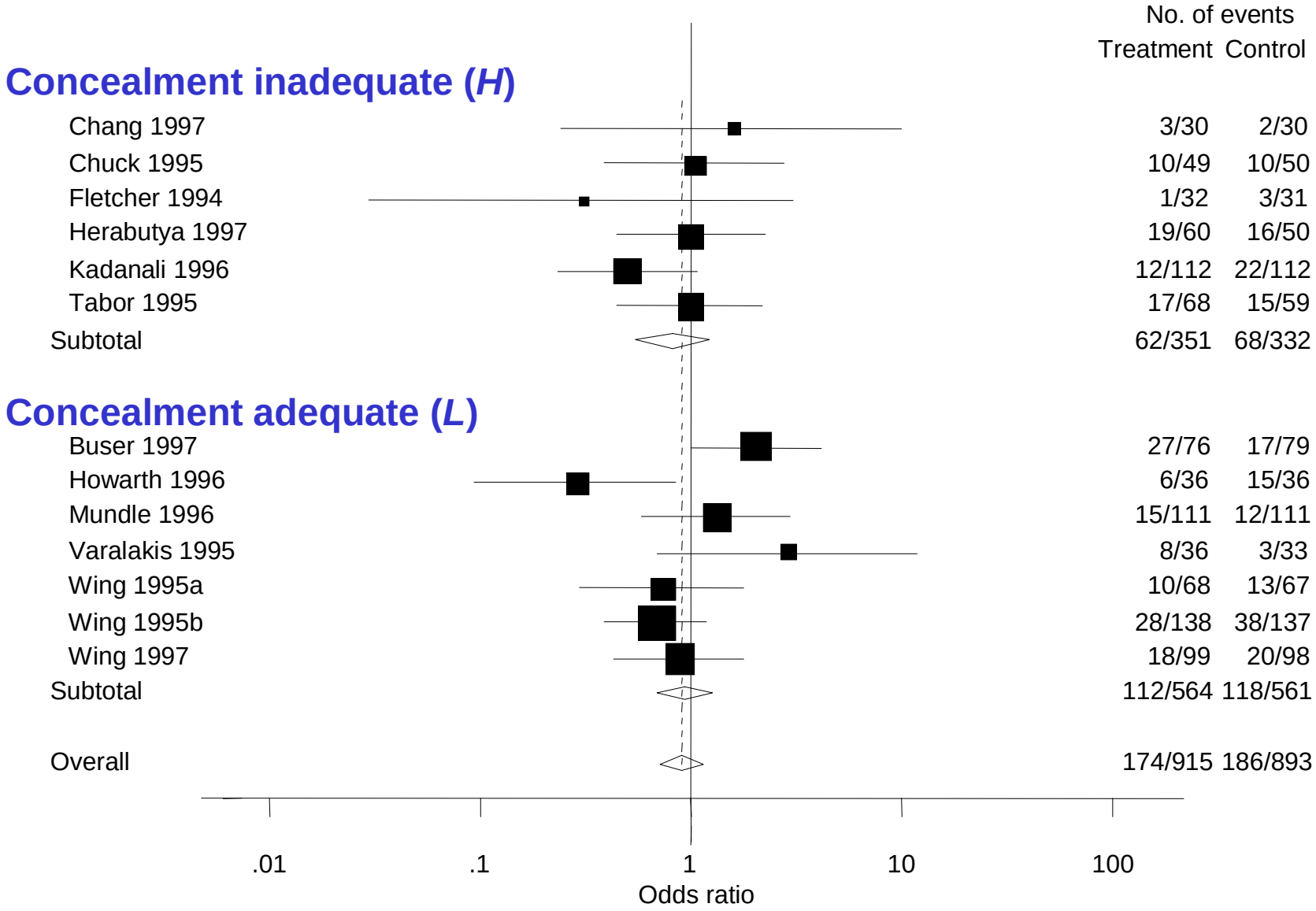
Example 1. Clozapine versus neuroleptic medication for schizophrenia



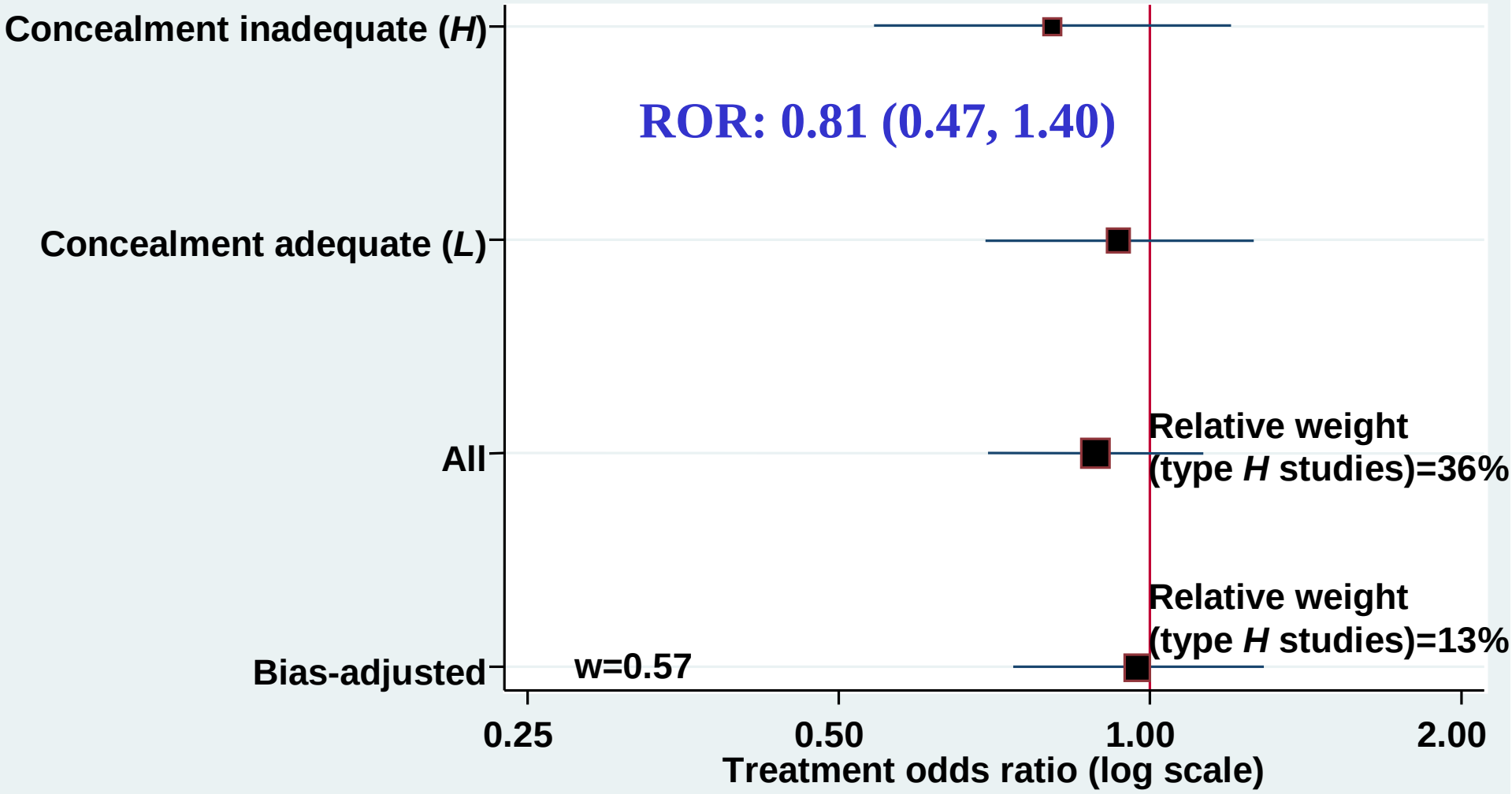
Example 1. Clozapine versus neuroleptic medication for schizophrenia



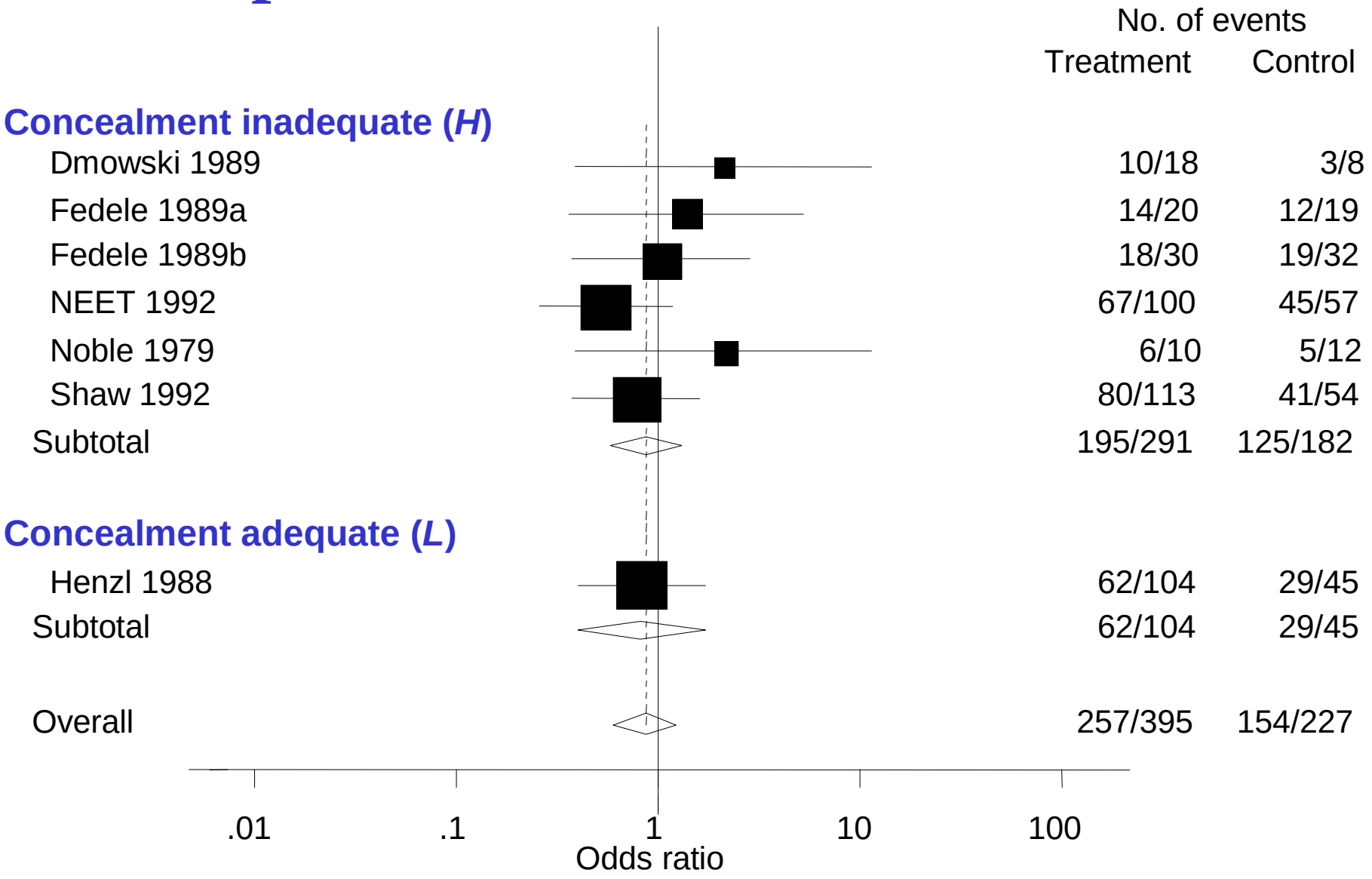
Example 2. Vaginal misoprostol versus prostaglandin for induction of labour



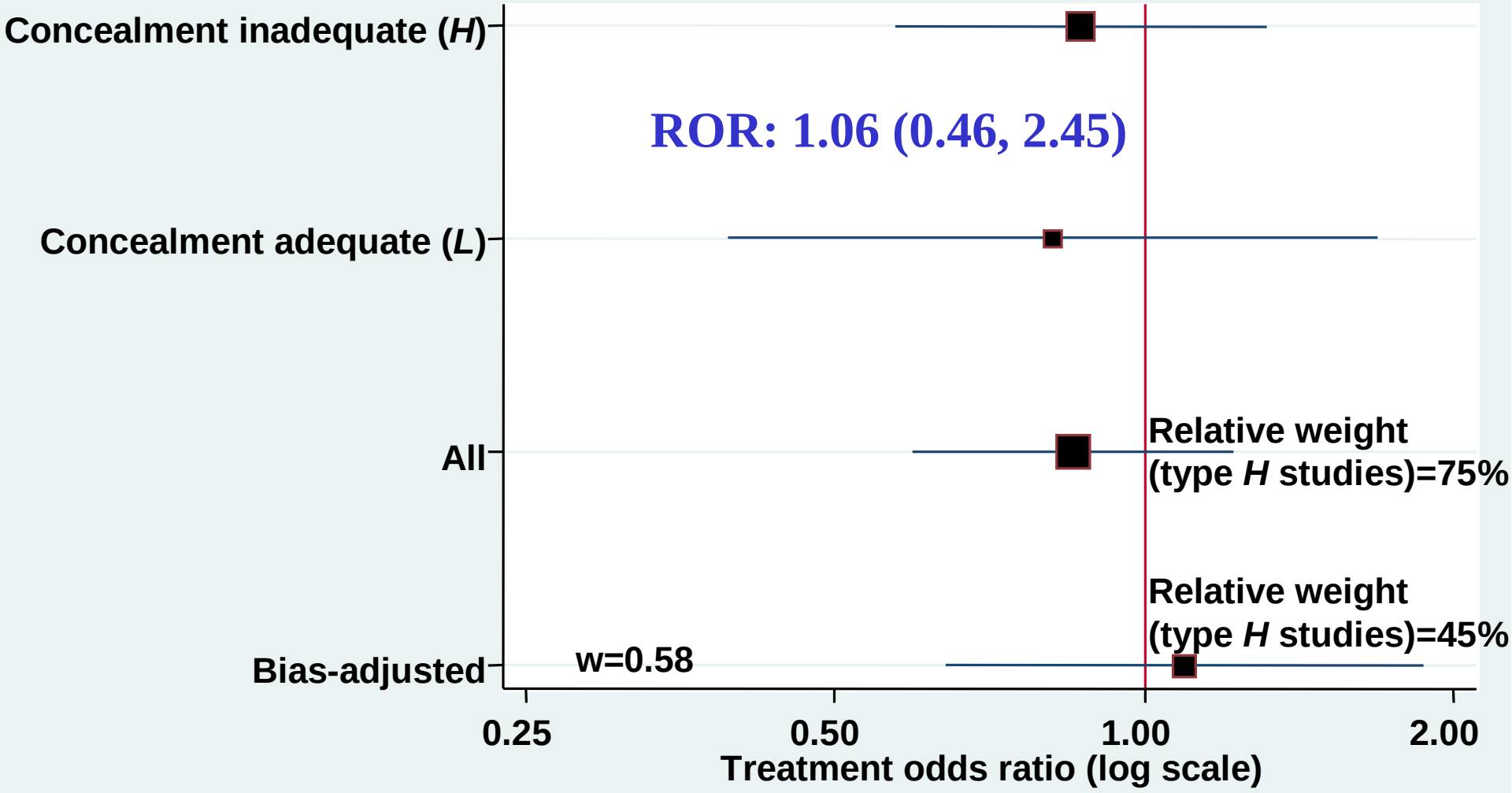
Example 2. Vaginal misoprostol versus prostaglandin for induction of labour



Example 3. Ovulation suppression compared to Danazol for endometriosis



Example 3. Ovulation suppression compared to Danazol for endometriosis



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 OR-δ_{AC}-δ_{BP}-δ_{BOA}-δ_{ITT}
  - weight in meta-analysis = $\frac{1}{s^2 + k_{AC}^2 + k_{BP}^2 + k_{BOA}^2 + k_{ITT}^2}$

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 k^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