Empirical evidence on sources of bias in randomised controlled trials: methods of and results from the BRANDO study

Jonathan Sterne

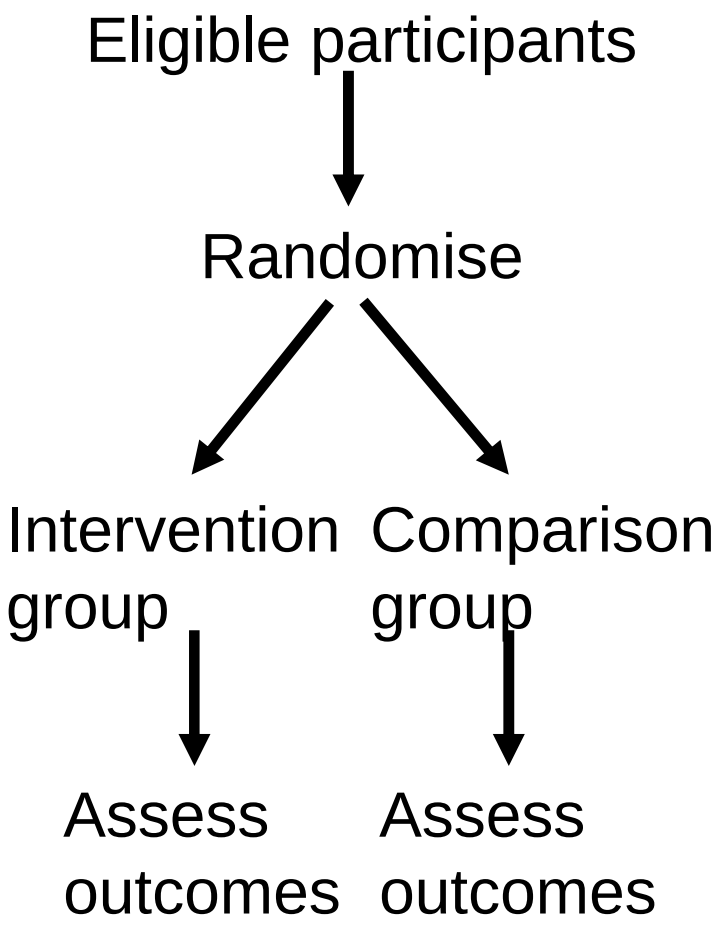
School of Social and Community Medicine,
University of Bristol, UK

Acknowledgements: Tony Ades, Bodil Als-Nielsen, Douglas Altman, Ethan Balk, Rebecca Beynon, John Carlin, Jon Deeks, Matthias Egger, Ross Harris, Julian Higgins, Lise Lotte Gluud, Christian Gluud, John Ioannidis, Peter Juni, Julie Pildal, Jelena Savovic, Ken Schulz, Nicky Welton, Lesley Wood

Randomised Controlled Trials (RCTs)

- **Deceptively simple idea**

Bias can be introduced at all stages of the conduct of RCTs



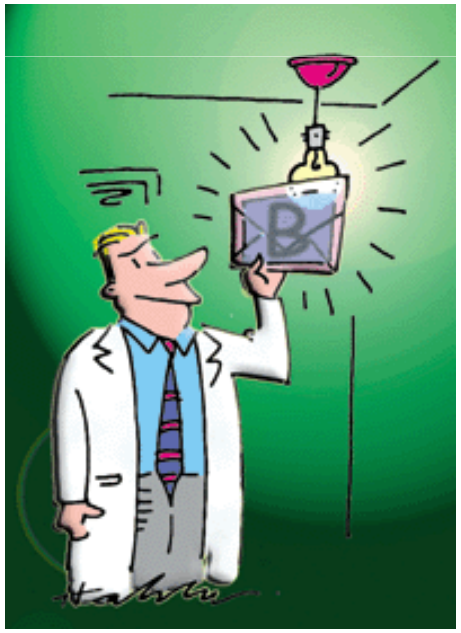
Flaws in the conduct of RCTs

- Trials provide causal inferences about the effect of the intervention if we randomise sufficient individuals and avoid selection and performance biases
- This can be undermined by:
  - Inadequate generation of randomisation sequence

Flaws in the conduct of RCTs

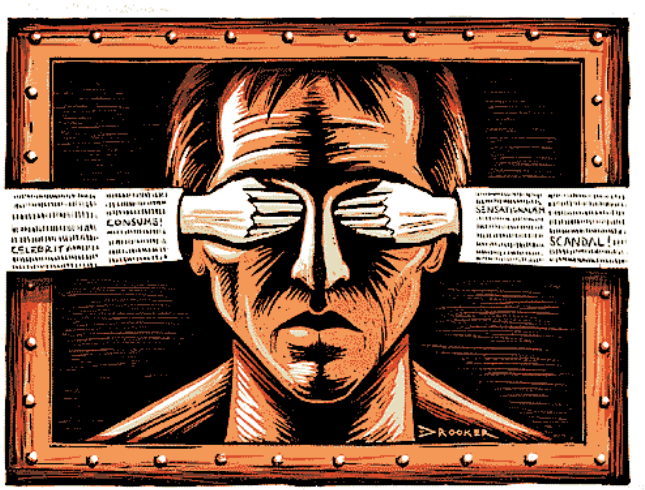
- Trials provide causal inferences about the effect of the intervention if we randomise sufficient individuals and avoid selection and performance biases
- This can be undermined by:
  - Inadequate generation of randomisation sequence
  - Inadequate concealment of allocation

Problems with randomisation may cause selection bias, if participants or healthcare providers can predict treatment allocation



Flaws in the conduct of RCTs

- Trials provide causal inferences about the effect of the intervention if we randomise sufficient individuals and avoid selection and performance biases
- This can be undermined by:
  - Inadequate generation of randomisation sequence
  - Inadequate concealment of allocation
  - Inadequate blinding



- **Performance bias**
  - Care of intervention and control groups not comparable
- **Detection bias**
  - Measurement of outcomes not comparable

Flaws in the conduct of RCTs

- Trials provide causal inferences about the effect of the intervention if we randomise sufficient individuals and avoid selection and performance biases
- This can be undermined by:
  - Inadequate generation of randomisation sequence
  - Inadequate concealment of allocation
  - Inadequate blinding
  - Excluding patients, or analysing them in the wrong group

Including biased trials will cause meta-analyses to be biased

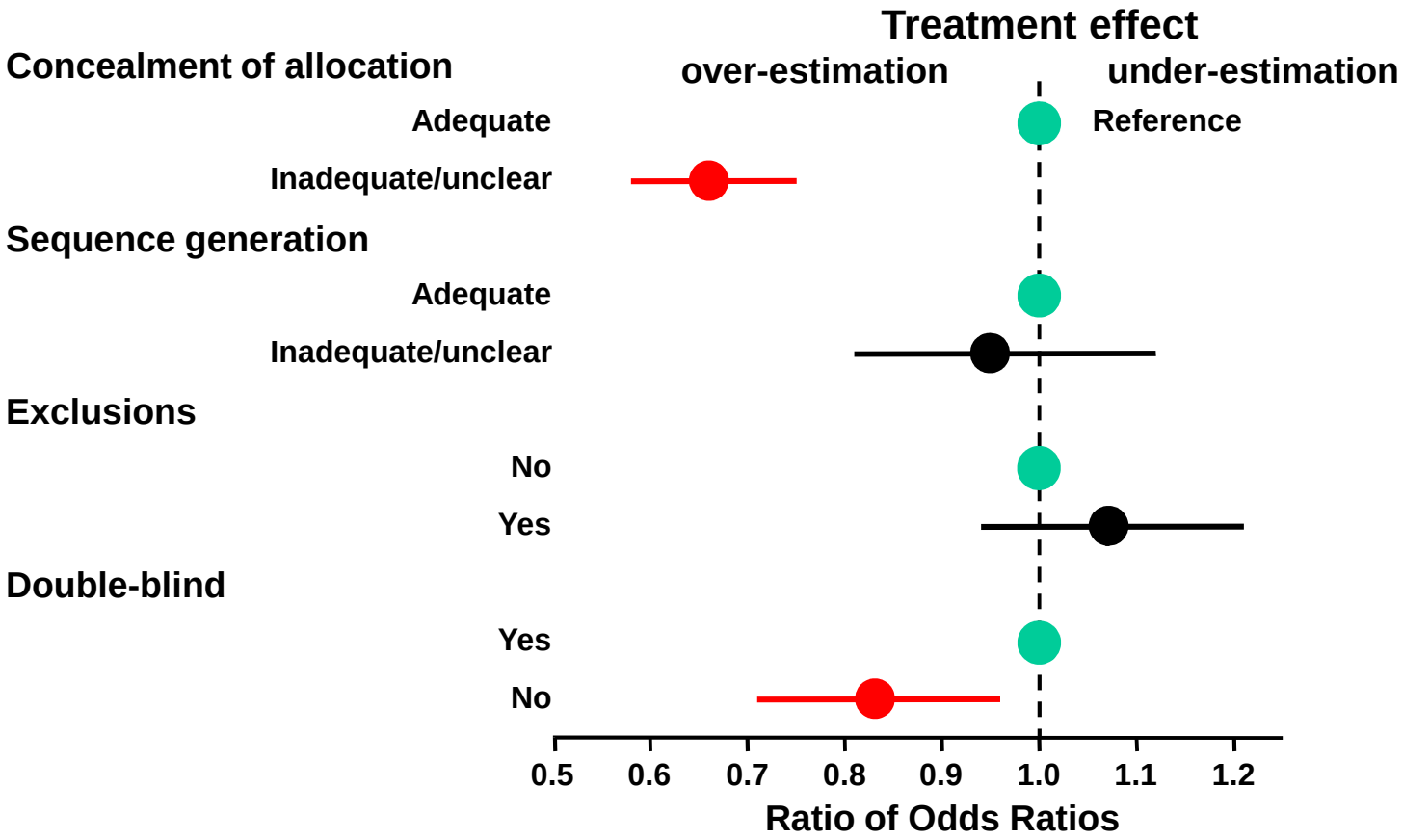
Meta-epidemiology

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

- Identify a large number of meta-analyses
- Record characteristics of individual studies (allocation concealment, blinding, type of publication, language etc.)
- Compare treatment effects *within* each meta-analysis (for example not double blind vs. double blind)
- Estimate ratio of odds ratios

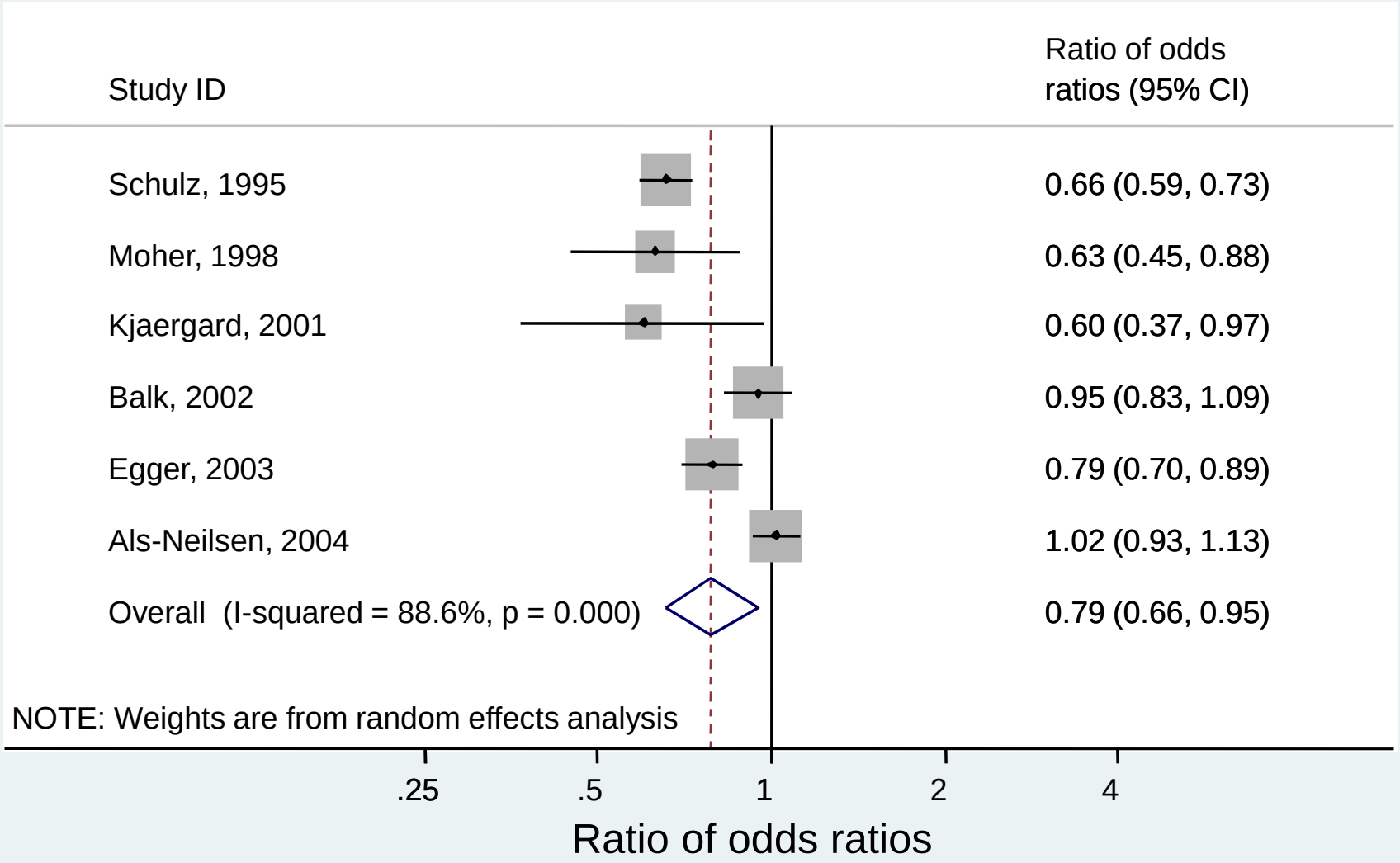
Empirical evidence of bias

33 meta-analyses, 250 RCTs



Schulz KF, Chalmers I, Hayes RJ, Altman DG. (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* **273**: 408-412.

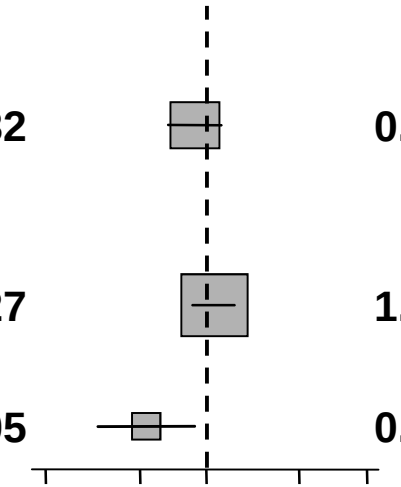
Allocation concealment: combined evidence



Combined analysis of three empirical studies: blinding

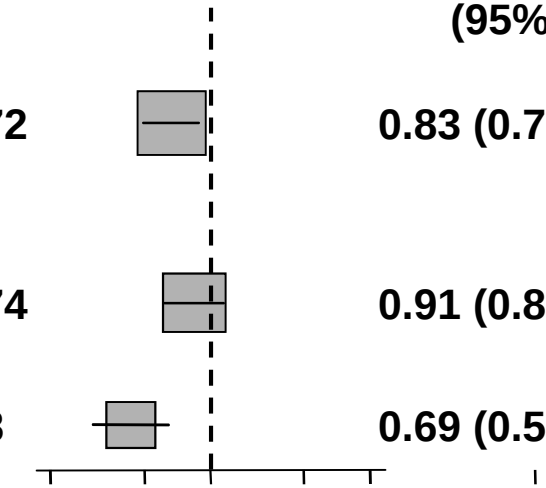
Comparison (No. of meta-analyses)	No. of trials*	Ratio of odds ratios (95% CI)	Variability in bias (P value)
Overall (76)	314 *vs.* 432	0.93 (0.83, 1.04)	0.11 (p<0.001)
Objective outcomes (44)	210 *vs.* 227	1.01 (0.92, 1.10)	0.08 (p<0.001)
Subjective outcomes (32)	104 *vs.* 205	0.75 (0.61, 0.93)	0.14 (p=0.001)

* Non blinded *vs.* blinded



Combined analysis of three empirical studies: allocation concealment

Comparison (No. of meta-analyses)	No. of trials*	Ratio of odds ratios (95% CI)	Variability in bias (P value)
Overall (102)	532 *vs* 272	0.83 (0.74, 0.93)	0.11 (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/unclearly concealed *vs*. adequately concealed

Wood, L., Egger, M., Gluud, L.L., Schulz, K., Jüni, P., Altman, D.G., Gluud, C., Martin, R.M., Wood, A.J.G. and Sterne, J.A.C. (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*, **336**: 601-605.

Emprical evidence on the effect of flaws in trial conduct will always be limited by imperfect reporting of trials

Journal of the Royal Statistical Society
SERIES A Statistics in Society

J. R. Statist. Soc. A (2009)
172, *Part* 1, *pp.* 119–136

Models for potentially biased evidence in meta-analysis using empirically based priors

N. J. Welton and A. E. Ades,

University of Bristol, UK

J. B. Carlin,

Murdoch Children's Research Institute and University of Melbourne, Australia

D. G. Altman

Centre for Statistics in Medicine, Oxford, UK

and J. A. C. Sterne

University of Bristol, UK

[Received April 2007. Revised February 2008]

Summary. We present models for the combined analysis of evidence from randomized controlled trials categorized as being at either low or high risk of bias due to a flaw in their conduct.

Effects of flaws in the conduct of trials

- Change in average intervention effect (bias), δ_0
  - the focus of most previous research, (measured as log ROR)
- Between-meta-analysis variability in average effect of bias, ϕ^2
- Increases in between-trial heterogeneity, κ^2
- In addition, there is uncertainty in the mean bias δ_0: $\delta_0 \sim N(D_0, \sigma^2_{D_0})$
- We can estimate δ_0 and κ^2 using data from a single meta-analysis, but D_0 or φ^2 can only be estimated using collections of meta-analyses

Possible analysis models

1. Fixed effect of intervention within meta-analyses (among trials at low risk of bias
2. Extends model 1 to allow κ^2 to vary between meta-analyses
3. **Extend model 1 to allow random effects of intervention within meta-analyses**
4. Combine models 2 and 3

How might we use evidence about the effects of flaws in trial conduct, from meta-epidemiological studies, to combine data from studies at high and low risk of bias in meta-analyses?

Consequences for a single study

- Given values for D_0, κ^2, φ^2 and $\sigma^2_{D_0}$, we can obtain the posterior distribution of the true intervention effect in a single study at high risk of bias from a new meta-analysis *m**:

$$E(\mu_{j,m^*} \mid \{\hat{\beta}_{j,m^*}, \sigma^2_{j,m^*}\}) = \hat{\beta}_{j,m^*} - D_0,$$

$$\mathrm{Var}(\mu_{j,m^*} \mid \{\hat{\beta}_{j,m^*}, \sigma^2_{j,m^*}\}) = \sigma^2_{j,m^*} + \kappa^2 + \varphi^2 + \sigma^2_{D_0}$$

Fixed-effect meta-analysis combining studies at high (*H*) and low (*L*) risk of bias

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

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

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

Implications

- 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 $\kappa^2 / n_H + \sigma_{D_0}^2 + \varphi^2$
3. Conducting large meta-epidemiological studies could in principle reduce $\sigma_{D_0}^2$, but φ^2 is a characteristic of the bias
4. However, φ^2 may be lower in certain situations (eg when outcomes are objectively assessed)

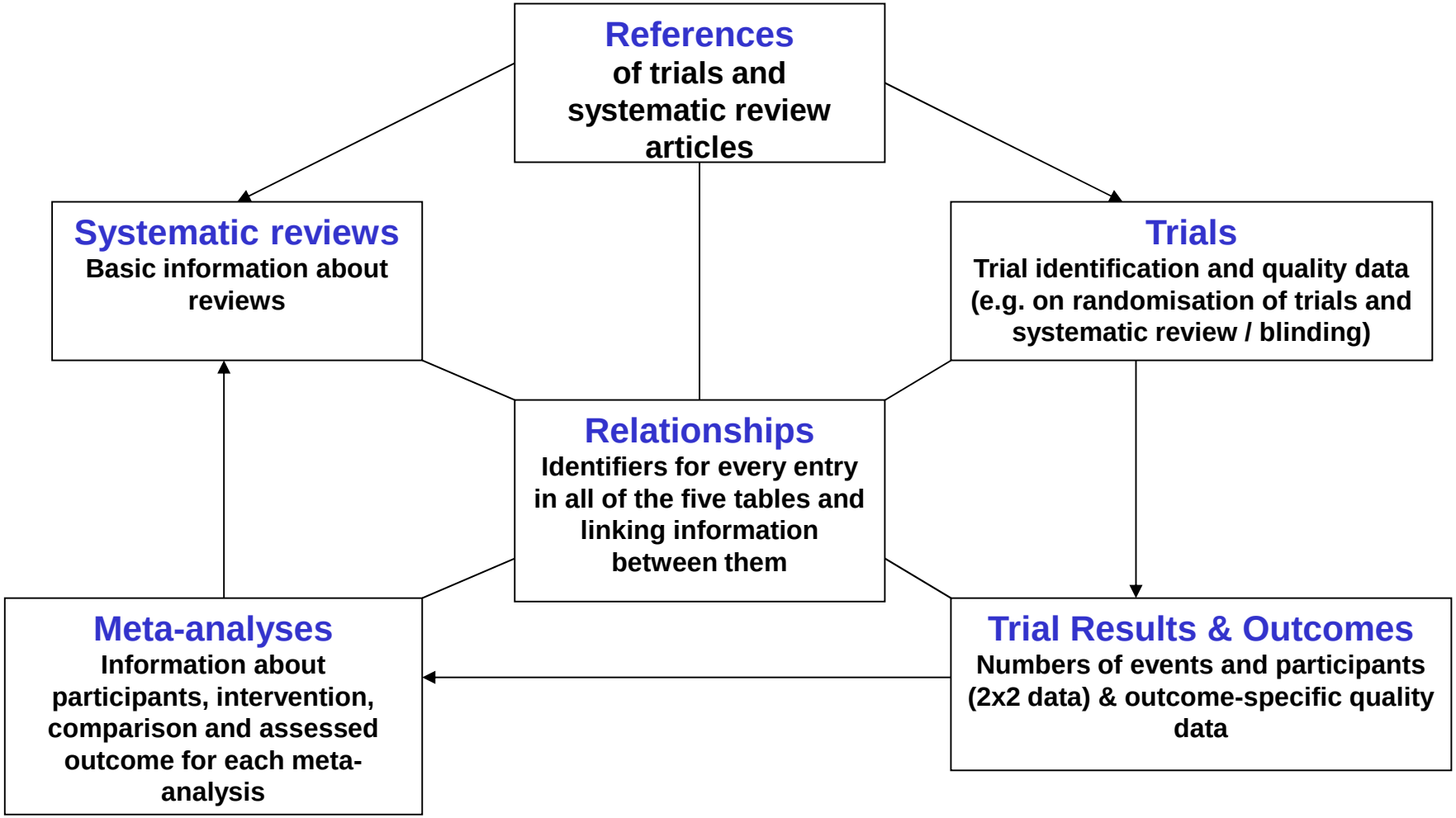
The BRANDO study
(Bias in Randomised and Observational Studies)

- Combine data from all existing empirical studies in a single database
- Final database contains data on 234 meta-analyses that included 1973 unique trials
- Aim to quantify effects of flaws in trial conduct on:
  - Average intervention effects
  - Uncertainty about intervention effects

Contributing studies

Contributing study	No. of meta-analyses (trials)	Clinical areas / types of interventions	No. of meta-analyses (trials) in final database
Als-Nielsen et al.	48 (523)	Various	46 (506)
Balk et al.	26 (276)	Circulatory, Paediatrics, Infection, Surgery	23 (251)
Contopoulos-Ioannidis et al.	16 (133)	Mental health	11 (94)
Egger et al.	165 (1776)	Various	121 (1115)
Kjaergard et al.	14 (190)	Various	8 (72)
Pildal et al.	68 (474)	Various	67 (460)
Schulz et al.	33 (250)	Pregnancy & childbirth	27 (210)

Database structure



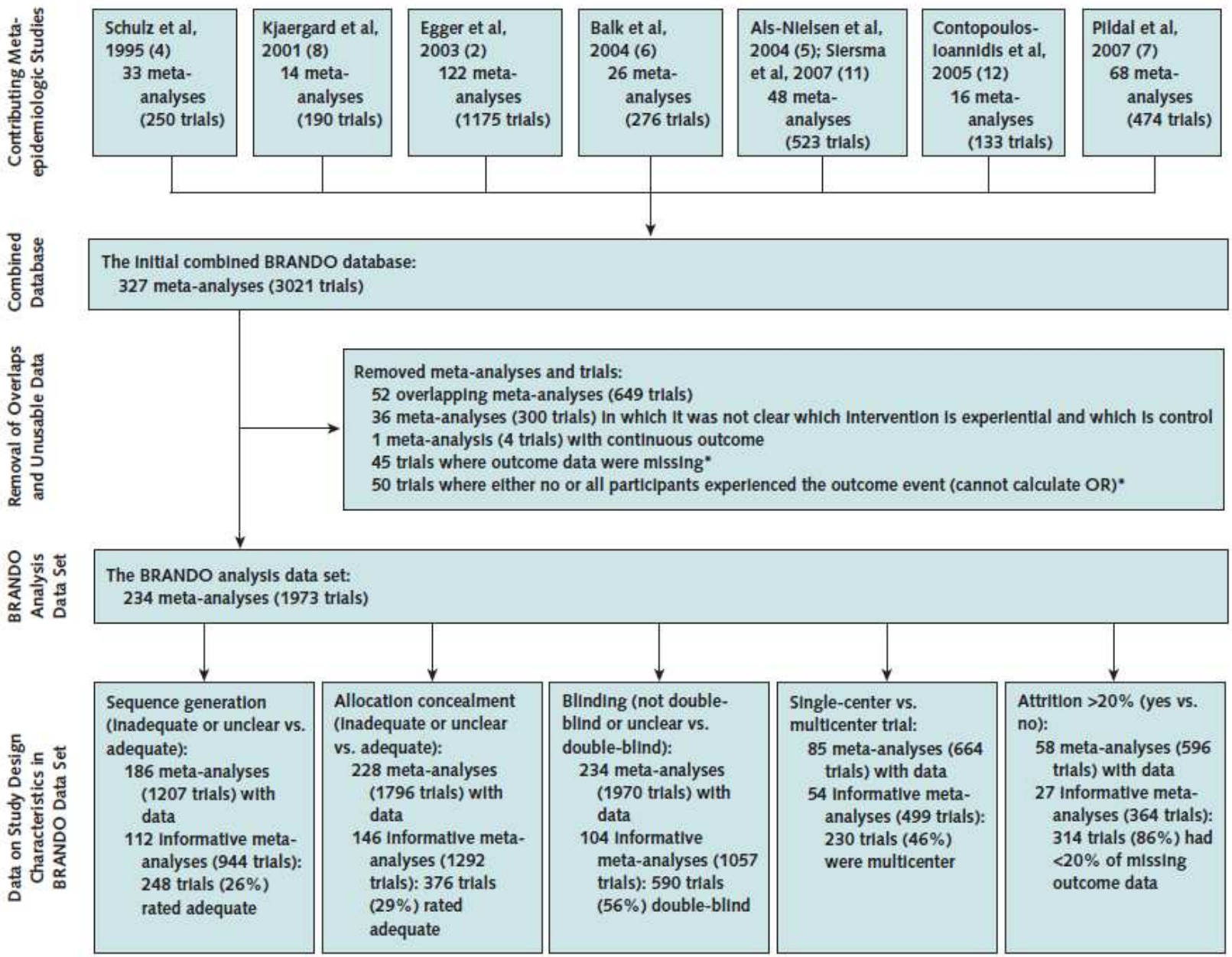
Removing overlaps

1. Assign unique ID (MEDLINE, EMBASE, or ISI Web of Science) to each trial and meta-analysis
2. Identify sets of meta-analyses containing any trial in common
3. Exclude meta-analyses from each set until there was minimal overlap between those remaining
   - Exclude those with fewer assessed trial characteristics in preference to those with more
   - Exclude those from less recently reviews, in preference to more recently published
   - Exclude those including fewer trials in preference to those including more
4. Randomly remove trials from meta-analyses with minimal overlap

Classification of outcomes

Type of outcome	No of meta-analyses	Outcome group*
Adverse events (as adverse effects of the treatment)	6	Subjective
All-cause mortality	64	Mortality
Cause-specific mortality	2	Subjective
Clinician-assessed outcomes (e.g. BMI, blood pressure, lung function)	51	Mostly subjective
Composite endpoint including mortality and/or major morbidity	9	Mix
Global improvement	4	Subjective
Health perceptions (person's own view of general health)	0	Subjective
Laboratory-reported outcomes (e.g. blood components, tissue analysis, urinalysis)	29	Mostly objective
Lifestyle outcomes (including diet, exercise, smoking)	12	Mostly subjective
Major morbidity event (including myocardial infarction, stroke, hemorrhage)	6	Subjective
Mental health outcomes (including cognitive function, depression, anxiety)	16	Subjective
Other outcomes (not classified elsewhere)	7	Mix
Pain (extent of pain a patient is experiencing)	13	Subjective
Perinatal outcomes	32	Objective (some potentially influenced)
Pregnancy outcomes	11	Objective
Quality of life (including ability to perform physical, daily and social activities)	0	Subjective
Radiological outcomes (including abnormalities, ultrasound, MRI results)	12	Subjective
Resource use (including cost, hospital stay duration, no. of procedures)	4	Objective, potentially influenced by judgement
Satisfaction with care (including patient views and clinician assessments)	0	Subjective
Surgical and device-related outcomes	16	Mostly subjective
Symptoms or signs of illness or condition	35	Subjective
Withdrawals/dropouts/compliance	16	Objective, potentially influence by judgement

Summarised as mortality, other objective, subjective/mixed


Contributing Meta-epidemiologic Studies
Schulz et al, 1995 (4) 33 meta-analyses (250 trials)
Kjaergard et al, 2001 (8) 14 meta-analyses (190 trials)
Egger et al, 2003 (2) 122 meta-analyses (1175 trials)
Balk et al, 2004 (6) 26 meta-analyses (276 trials)
Als-Nielsen et al, 2004 (5); Siersma et al, 2007 (11) 48 meta-analyses (523 trials)
Contopoulos-Ioannidis et al, 2005 (12) 16 meta-analyses (133 trials)
Pildal et al, 2007 (7) 68 meta-analyses (474 trials)
Combined Database
The initial combined BRANDO database: 327 meta-analyses (3021 trials)
Removal of Overlaps and Unusable Data
Removed meta-analyses and trials:
52 overlapping meta-analyses (649 trials)
36 meta-analyses (300 trials) in which it was not clear which intervention is experiential and which is control
1 meta-analysis (4 trials) with continuous outcome
45 trials where outcome data were missing*
50 trials where either no or all participants experienced the outcome event (cannot calculate OR)*
BRANDO Analysis Data Set
The BRANDO analysis data set: 234 meta-analyses (1973 trials)
Data on Study Design Characteristics in BRANDO Data Set
Sequence generation (inadequate or unclear vs. adequate):
186 meta-analyses (1207 trials) with data
112 informative meta-analyses (944 trials): 248 trials (26%) rated adequate
Allocation concealment (inadequate or unclear vs. adequate):
228 meta-analyses (1796 trials) with data
146 informative meta-analyses (1292 trials): 376 trials (29%) rated adequate
Blinding (not double-blind or unclear vs. double-blind):
234 meta-analyses (1970 trials) with data
104 informative meta-analyses (1057 trials): 590 trials (56%) double-blind
Single-center vs. multicenter trial:
85 meta-analyses (664 trials) with data
54 informative meta-analyses (499 trials): 230 trials (46%) were multicenter
Attrition >20% (yes vs. no):
58 meta-analyses (596 trials) with data
27 informative meta-analyses (364 trials): 314 trials (86%) had <20% of missing outcome data

Numbers of trials

Sequence generation	Allocation concealment	Blinding	All	Mortality	Objective	Subjective
Adequate	Adequate	Double blind	60	12	22	26
Adequate	Adequate	Not double blind*	31	4	10	17
Adequate	Inadequate*	Double blind	67	9	22	36
Adequate	Inadequate*	Not double blind*	115	16	42	57
Inadequate*	Adequate	Double blind	95	9	35	51
Inadequate*	Adequate	Not double blind*	33	6	10	17
Inadequate*	Inadequate*	Double blind	317	44	92	181
Inadequate*	Inadequate*	Not double blind*	453	57	135	261
Total			**1171**	**157**	**368**	**646**

Issues in analysis and interpretation

- Different numbers of trials and included meta-analyses in different analyses
  - Fewer studies assessed sequence generation than allocation concealment or blinding
  - Only meta-analyses containing trials with and without characteristic contribute to analyses
- Bias models were fitted using WinBUGS assuming vague prior distributions.
  - Two parallel MCMC chains, with a burn-in of 50,000 iterations followed by at least a further 500,000 iterations, with a thinning of 5.
- Many variance components are imprecisely estimated
  - Analyses were substantially delayed because of sensitivity to priors on variances
  - Based on a simulation study, we chose a modified Inverse Gamma(0.001, 0.001) prior with increased weight on small values
- Comparisons are of presence versus absence of methodological flaw (high versus low risk of bias)

Annals of Internal Medicine | Research and Reporting Methods

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; John P.A. Ioannidis, MD, DSc; Kenneth F. Schulz, PhD, MBA; Rebecca Beynon, MA; Nicky J. Welton, PhD; Lesley Wood, PhD; David Moher, PhD; Jonathan J. Deeks, PhD; and Jonathan A.C. Sterne, PhD

Published evidence suggests that aspects of trial design lead to biased intervention effect estimates, but findings from different studies are inconsistent. This study combined data from 7 meta-epidemiologic studies and removed overlaps to derive a final data set of 234 unique meta-analyses containing 1973 trials. Outcome measures were classified as "mortality," "other objective," "or subjective," and Bayesian hierarchical models were used to estimate associations of trial characteristics with average bias and between-trial heterogeneity. Intervention effect estimates seemed to be exaggerated in trials with inadequate or unclear (vs. adequate) random-sequence generation (ratio of odds ratios, 0.89 [95% credible interval {CrI}, 0.82 to 0.96]) and with inadequate or unclear (vs. adequate) allocation concealment (ratio of odds ratios, 0.93 [CrI, 0.87 to 0.99]). Lack of or unclear double-blinding (vs. double-blinding) was associated with an average of 13% exaggeration of intervention effects (ratio of odds ratios, 0.87 [CrI, 0.79 to 0.96]), and between-trial heterogeneity was increased for such studies (SD increase in heterogeneity, 0.14 [CrI, 0.02 to 0.30]). For each characteristic, average bias and increases in between-trial heterogeneity were driven primarily by trials with subjective outcomes, with little evidence of bias in trials with objective and mortality outcomes. This study is limited by incomplete trial reporting, and findings may be confounded by other study design characteristics. Bias associated with study design characteristics may lead to exaggeration of intervention effect estimates and increases in between-trial heterogeneity in trials reporting subjectively assessed outcomes.

Ann Intern Med. 2012;157:429-438. www.annals.org

For author affiliations, see end of text.

This article was published at www.annals.org on 4 September 2012.

Health Technology Assessment 2012; Vol. 16: No. 35
ISSN 1366-5278

Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies

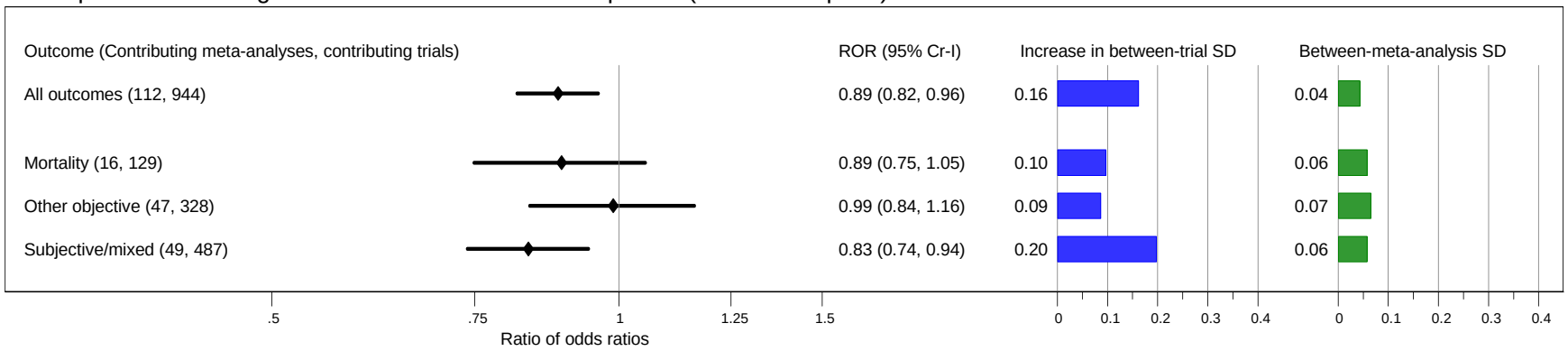
J Savović, HE Jones, DG Altman, RJ Harris,
P Jűni, J Pildal, B Als-Nielsen, EM Balk,
C Gluud, LL Gluud, JPA Ioannidis, KF Schulz,
R Beynon, N Welton, L Wood, D Moher,
JJ Deeks and JAC Sterne

Associations between design characteristics

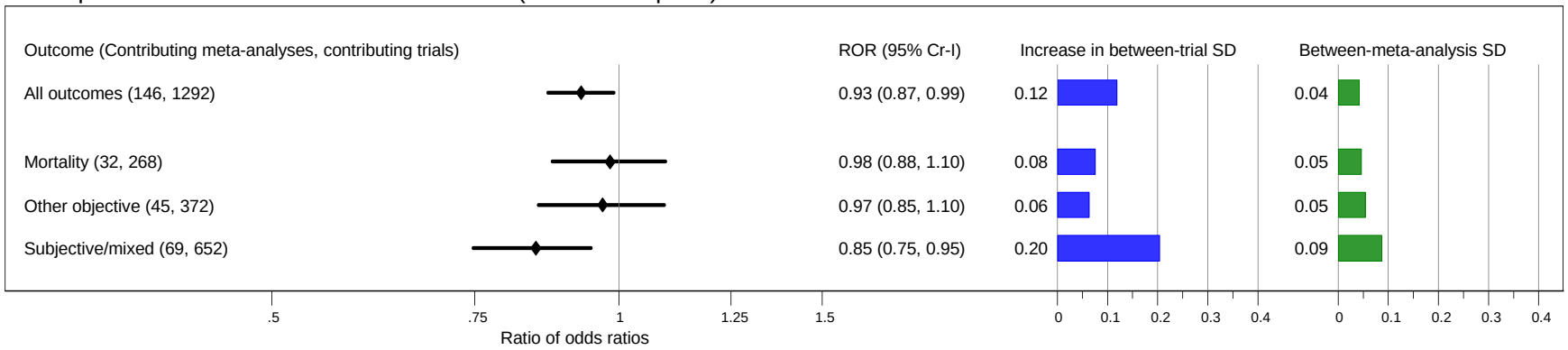
		No. (%) of trials			
Study characteristic 1	Study characteristic 2	All trials	Mortality outcome	Objective outcome	Subjective outcome
Sequence generation	*Allocation concealment*	1171	157	368	646
Adequate	Adequate	91 (7.8)	16 (10.2)	32 (8.7)	43 (6.7)
Adequate	Inadequate/unclear	182 (15.5)	25 (15.9)	64 (17.4)	93 (14.4)
Inadequate/unclear	Adequate	128 (10.9)	15 (9.6)	45 (12.2)	68 (10.5)
Inadequate/unclear	Inadequate/unclear	770 (65.8)	101 (64.3)	227 (61.7)	442 (68.4)
OR (95% CI)		3.01 (2.20 to 4.12)	4.31 (1.88 to 9.88)	2.52 (1.48 to 4.29)	3.01 (1.93 to 4.68)
Sequence generation	*Blinding*	1171	157	368	646
Adequate	Double blind	127 (10.8)	21 (13.4)	44 (12.0)	62 (9.6)
Adequate	Not double blind/unclear	146 (12.5)	20 (12.7)	52 (14.1)	74 (11.5)
Inadequate/unclear	Double blind	412 (35.2)	53 (33.8)	127 (34.5)	232 (35.9)
Inadequate/unclear	Not double blind/unclear	486 (41.5)	63 (40.1)	145 (39.4)	278 (43.0)
OR (95% CI)		1.03 (0.78 to 1.35)	1.25 (0.61 to 2.55)	0.97 (0.61 to 1.54)	1.00 (0.69 to 1.47)
Allocation concealment	*Blinding*	1793	328	550	915
Adequate	Double blind	283 (15.8)	65 (19.8)	93 (16.9)	125 (13.7)
Adequate	Not double blind/unclear	133 (7.4)	30 (9.1)	45 (8.2)	58 (6.3)
Inadequate/unclear	Double blind	556 (31.0)	108 (32.9)	159 (28.9)	289 (31.6)
Inadequate/unclear	Not double blind/unclear	821 (45.8)	125 (38.1)	253 (46.0)	443 (48.4)
OR (95% CI)		3.14 (2.49 to 3.96)	2.51 (1.52 to 4.15)	3.29 (2.19 to 4.94)	3.30 (2.34 to 4.66)

Univariable analyses

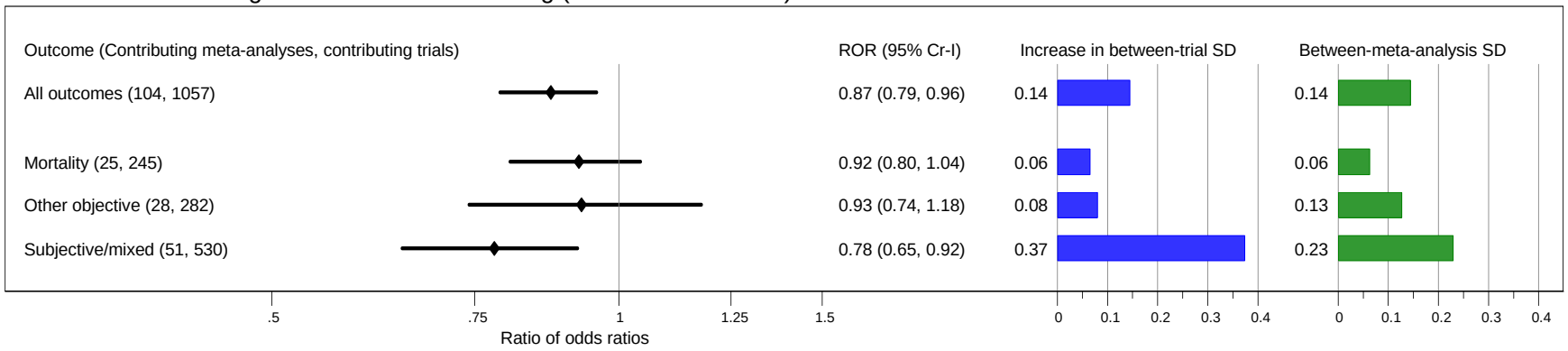
Inadequate or unclear generation of randomization sequence (versus adequate)



Inadequate or unclear allocation concealment (versus adequate)

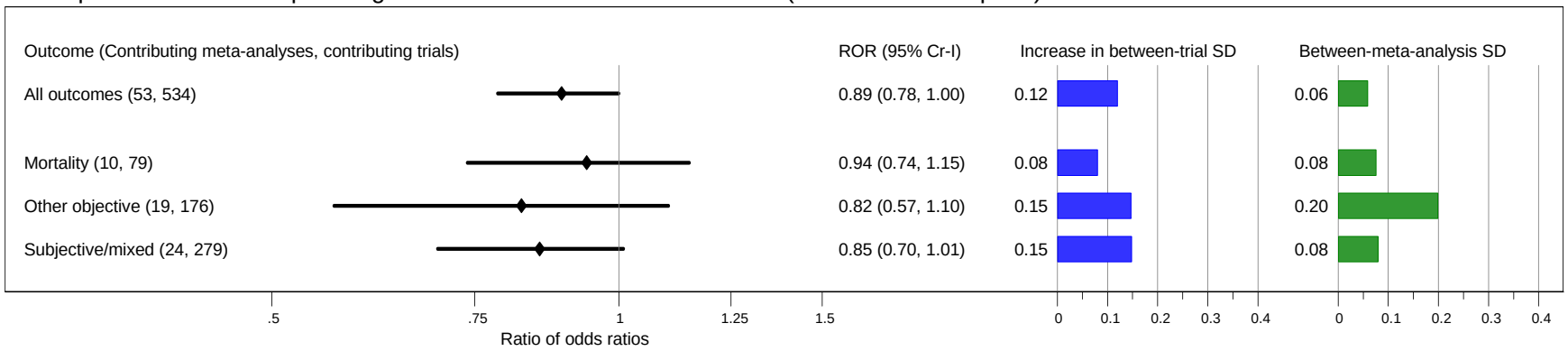


Lack of double blinding or unclear double blinding (versus double blind)

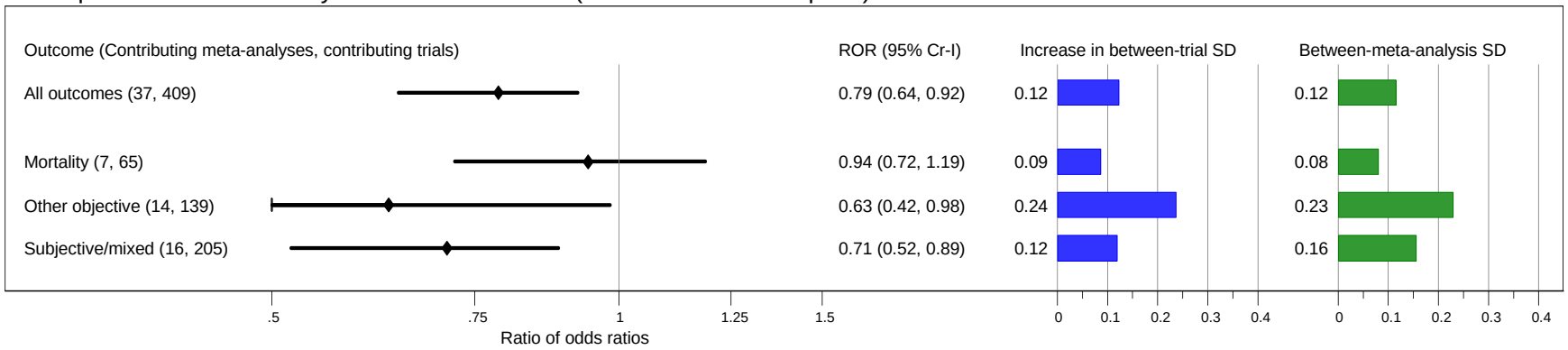


Combinations of characteristics

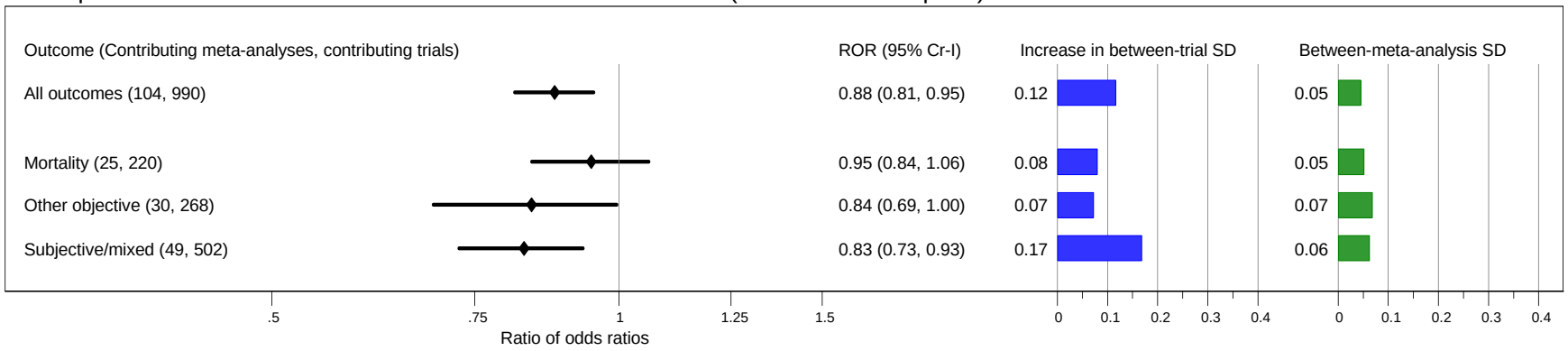
Inadequate or unclear sequence generation or allocation concealment (versus both adequate)



Inadequate or unclear for any of the three domains (versus all three adequate)

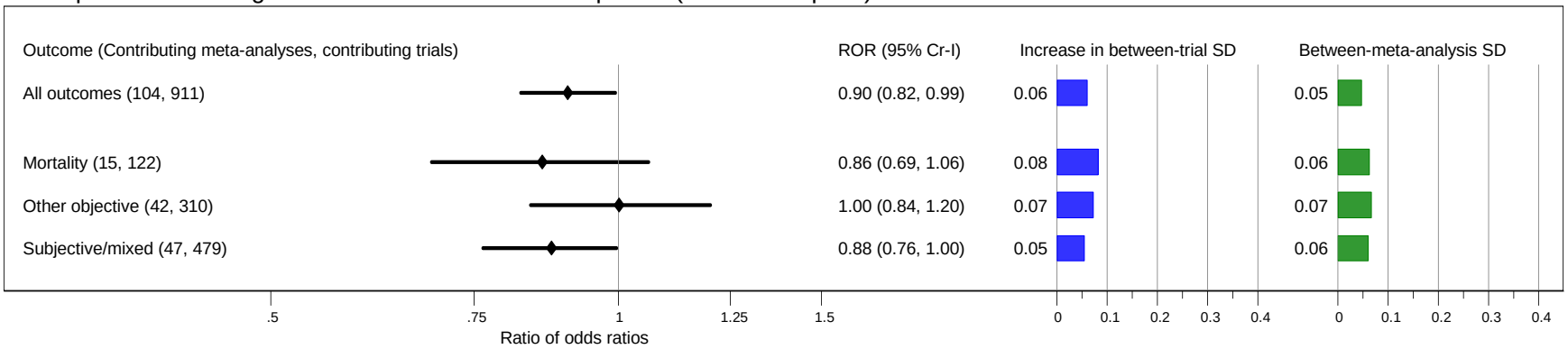


Inadequate or unclear allocation concealment or not double blind (versus both adequate)

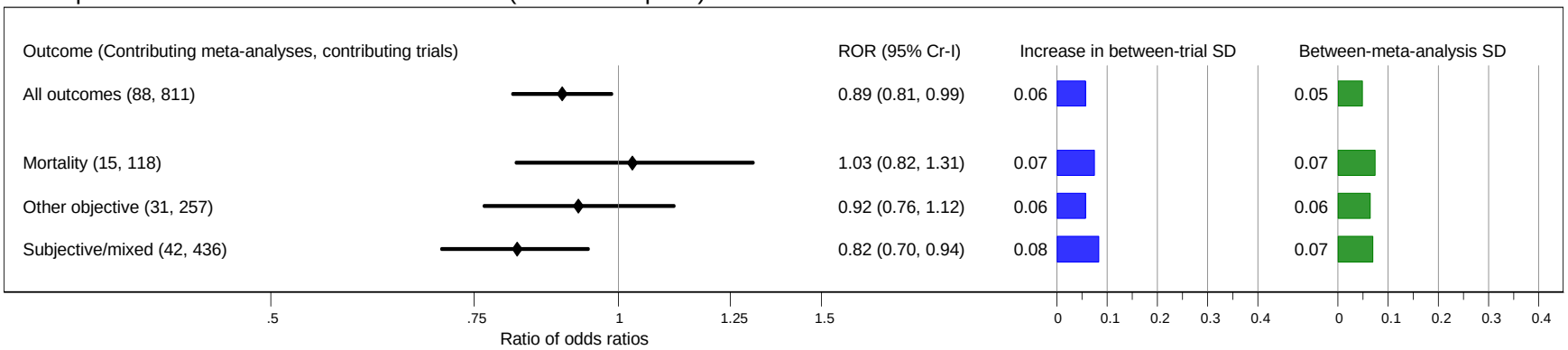


Multivariable analyses

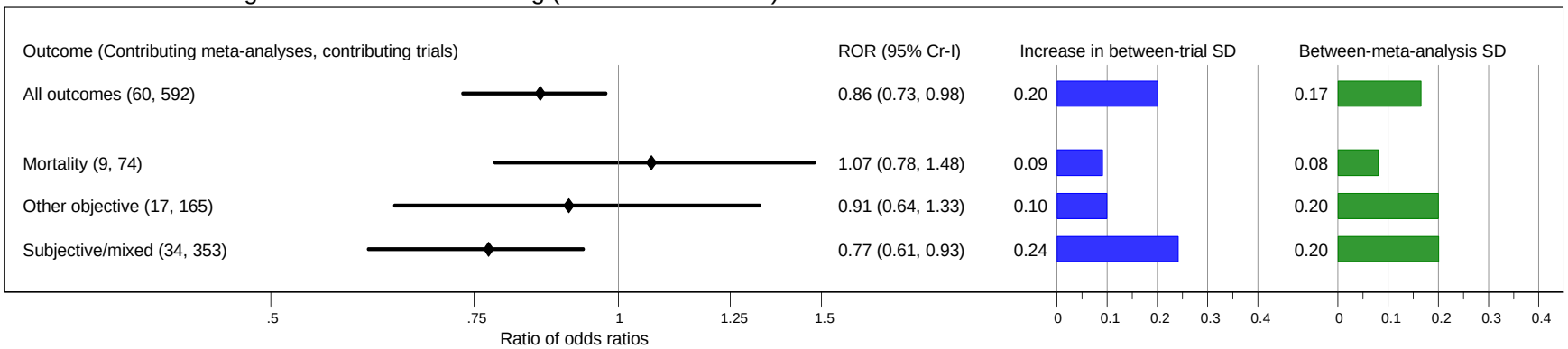
Inadequate or unclear generation of randomization sequence (versus adequate)



Inadequate or unclear allocation concealment (versus adequate)



Lack of double blinding or unclear double blinding (versus double blind)



Implications for downweighting of trials in a new meta-analysis

- For a new trial: $100 \times \frac{\sigma_i^2 + \kappa^2 + \varphi^2 + V_0}{\sigma_i^2}\%$
- For a new meta-analysis: using the formulae of Welton et al. using the distribution of trial and meta-analysis characteristics in the BRANDO database

Implications for downweighting

Study design characteristic and outcome	No. of high-risk trials	Minimum variance of trial at high risk of bias ($V_0+\kappa^2+\varphi^2$)	Median (IQR) increase in trial-level variance (%)	Median (IQR) increase in variance of summary intervention effect (%)		
				Downweighting all meta-analyses	Downweighting informative meta-analyses	Excluding all trials at high or unclear risk of bias
Inadequate or unclear sequence generation (vs adequate)						
All	901	0.030	10 (4 to 23)	12 (2 to 32)	13 (5 to 32)	217 (87 to 482)
Mortality	116	0.020	6 (3 to 14)	11 (1 to 25)	13 (6 to 36)	119 (70 to 336)
Objective	273	0.019	5 (3 to 11)	8 (1 to 19)	11 (2 to 32)	145 (62 to 559)
Subjective	512	0.046	20 (8 to 39)	31 (6 to 64)	31 (11 to 56)	282 (126 to 482)
Inadequate or unclear allocation concealment (vs adequate)						
All	1380	0.017	5 (2 to 12)	9 (3 to 23)	7 (3 to 20)	150 (49 to 411)
Mortality	233	0.011	4 (1 to 11)	8 (3 to 34)	8 (3 to 19)	121 (39 to 468)
Objective	413	0.011	3 (1 to 6)	9 (3 to 22)	6 (3 to 13)	175 (52 to 337)
Subjective	734	0.053	18 (7 to 40)	36 (8 to 73)	27 (7 to 59)	146 (55 to 411)
Lack of double blinding or unclear double blinding (vs double blind)						
All	1041	0.044	13 (6 to 31)	16 (0 to 62)	15 (3 to 48)	62 (19 to 143)
Mortality	170	0.013	4 (2 to 10)	5 (0 to 18)	5 (0 to 18)	46 (18 to 101)
Objective	336	0.036	11 (5 to 24)	22 (0 to 67)	16 (3 to 46)	79 (22 to 202)
Subjective	535	0.200	63 (22 to 138)	41 (1 to 175)	36 (8 to 72)	62 (19 to 140)

Future research

- Update to more recent meta-analyses and trials
- Investigate different aspects of blinding (separate blinding of participants and trial personnel from blinding of outcome assessors)
- Other trial characteristics (eg missing data)

Conclusions

- Flaws in the conduct of randomized controlled trials are important because they increase uncertainty
- Effects are most marked for subjectively assessed outcomes
  - Particularly for absence of blinding
- Patterns similar for different bias domains
  - Outcome-specific effects of flaws in the randomisation process were unexpected
- BRANDO results provide an empirical basis to include potentially biased evidence in a systematic review, by downweighting and correcting for bias