Association Between Risk-of-Bias Assessments and Results of Randomized Trials in Cochrane Reviews: the ROBES Meta-Epidemiologic Study

Jelena Savović, Rebecca M. Turner, David Mawdsley, Hayley E. Jones, Rebecca Beynon, Julian P.T. Higgins, and Jonathan A.C. Sterne

Correspondence to Dr Jelena Savović, NIHR CLAHRC West, University Hospitals Bristol NHS Foundation Trust, 9th Floor, Whitefriars, Lewins Mead, Bristol BS1 2NT, United Kingdom (e-mail: J.Savovic@bristol.ac.uk, telephone: +44 117 3421243, fax: not available).

Jelena Savović will be the main point of contact during editing.

Author affiliations: Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom (Jelena Savović, David Mawdsley, Hayley E. Jones, Rebecca Beynon, Julian P.T. Higgins and Jonathan A.C. Sterne); The National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West (NIHR CLAHRC West), University Hospitals Bristol National Health Service Foundation Trust, Bristol, United Kingdom (Jelena Savović, Julian P.T. Higgins and Jonathan A.C. Sterne); Medical Research Council (MRC) Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom (Rebecca M. Turner); MRC Clinical Trials Unit, University College London, London, United Kingdom (Rebecca M. Turner).

This work was supported by the MRC Fellowship (Savović; G0701659/1) and by the MRC Methodology Research Panel grant (Turner, MR/K014587/1). Savović's time was partly supported by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West at University Hospitals Bristol National Health Service Foundation Trust. Turner was

supported by the MRC (MC-U105260558 and MC-UU12023/24). Sterne is funded by a National Institute for Health Research Senior Investigator award (NF-SI-0611-10168). The views expressed are those of the authors and not necessarily those of the MRC, the National Health Service, the National Institute for Health Research or the Department of Health.

Conflict of interest: Mawdsley was employed at the University of Bristol on an unrelated project part funded by Pfizer between January 2015 and November 2016. The other authors report no financial or commercial conflicts of interest.

Declaration of non-commercial interests: Savović, Sterne, Jones and Beynon are the co-authors of the BRANDO study referred to in the manuscript. Sterne, Higgins and Savović were involved in the development or evaluation of the Cochrane risk of rias tool. The other authors report no academic conflicts of interest.

Running head: Risk of Bias in Evidence Syntheses of Randomized Trials

ABSTRACT

Flaws in trial design may bias intervention effect estimates and increase between-trial heterogeneity. Empirical evidence suggests that these problems are greatest for subjectively assessed outcomes. For the ROBES study, we extracted risk-of-bias judgements (for sequence generation, allocation concealment, blinding and incomplete data) from a large collection of meta-analyses published in the Cochrane Library, issue 4, 2011. We categorized outcome measures as mortality, other objective or subjective, and estimated associations of bias judgements with intervention effect estimates using Bayesian hierarchical models. Among 2443 trials in 228 meta-analyses, intervention effect estimates were on average exaggerated in trials with high or unclear risk-of-bias judgements (versus low) for sequence generation (ratio of odds ratio 0.91 [95% credible interval 0.86, 0.98]), allocation concealment (0.92 [0.86, 0.98]) and blinding (0.87 [0.80, 0.93]). In contrast to previous work, we did not observe consistently different bias for subjective outcomes compared with mortality. However, we found an increase in between-trial heterogeneity associated with lack of blinding in meta-analyses with subjective outcomes. Inconsistency in criteria for risk-of-bias judgments applied by individual reviewers is a likely limitation of routinely collected bias assessments. Inadequate randomization and lack of blinding may lead to exaggeration of intervention effect estimates in trials.

Keywords: meta-analysis, blinding, randomization, allocation concealment, missing data, bias, trials

List of abbreviations

(The) BRANDO study – Study name (BRANDO – Bias in Randomized and Observational studies)

Cr-I – credible interval

OR – Odds ratio

(The) ROBES study – Study name (ROBES - Risk of Bias in Evidence Synthesis)

ROR – ratio of odds ratio

SD – standard deviation

MRC - Medical Research Council (United Kingdom)

Meta-analyses of randomized trials are often more influential than single trials, and increasingly inform healthcare decisions made by clinicians and health authorities. For their results to be valid, trials should employ rigorous methods that can achieve and preserve comparability of the intervention and control groups.(1) For example, concealment of randomized allocation prevents an influence of patient characteristics on allocation to intervention and control groups; blinding of participants and trial personnel prevents differences in patient management between groups; and blinding of outcome assessors prevents knowledge of the assigned intervention group influencing outcome measurement. Randomized trials vary in methodological rigour, and flaws in trial conduct can lead to biased estimation of the intervention effect.(2) Systematic reviewers should therefore assess the risk of bias in intervention effect estimates from each included trial.

Meta-epidemiologic studies analyse collections of meta-analyses to provide empirical evidence about the influence of trial design characteristics on trial results.(3) Such studies have, however, reached differing conclusions about which trial design characteristics most influence their results.(4-8) For example, four studies found that lack of adequate allocation concealment was associated with overestimation of treatment effect,(9-12) while several other studies did not find evidence for this.(4, 5, 13-15) In a previous study we explored reasons for these discrepancies by combining data from seven meta-epidemiologic studies.(16, 17) This was the first study to explore the effects of bias on between and within meta-analysis heterogeneity using Bayesian hierarchical models. The results suggested that trial results based on subjectively assessed outcomes are more susceptible to bias and that the effect of bias is unpredictable, leading to increased heterogeneity in meta-analyses assessing subjective outcomes. (16, 17) Further investigation of the effects of trial characteristics across different interventions, settings and outcomes in larger collections of meta-analyses (not previously used) may provide more clarity and resolve inconsistencies between previous empirical studies.

Since January 2008, authors of Cochrane reviews have used a 'Risk of Bias' tool for assessing included trials.(18) The assessors make judgments in relation to “sequence generation”, “allocation concealment”, “blinding of participants, personnel and outcome assessors”, “incomplete outcome data”, “selective outcome reporting”, and a general category of “other potential threats to validity”. For each of these areas, review authors record whether there was a judgement of low, high or unclear risk of bias for each trial, together with comments or quotes to justify each judgement. Accumulated standardized risk-of-bias assessments are a potentially a useful resource for meta-epidemiologic research.

This paper describes and reports the main results from the ROBES study (Risk Of Bias in Evidence Synthesis). The ROBES study is a new, large empirical study investigating the associations of risk-of-bias judgements for sequence generation, allocation concealment, blinding and incomplete outcome data with treatment effect estimates. Our aims were to examine whether routinely collected risk of bias assessments relating to methodological characteristics are associated with effect estimates, to compare these associations with findings from our previous study,(17) and to examine further the effect of outcome types in a new collection of meta-analyses.

METHODS

Data source

The April 2011 issue of the Cochrane Database of Systematic Reviews included 4371 intervention reviews (excluding protocols), of which 1399 had at least two completed domains in the Risk of Bias tables. The complete 1399 reviews were supplied by the Cochrane Informatics & Knowledge Management Department in the format of Review Manager (rm5) files.(19) We converted these to a customized Microsoft Access database using bespoke software which we commissioned from Riskaware Limited. (Bristol, United Kingdom).

Data selection and categorization

We selected meta-analyses that fulfilled the following criteria: (i) address a binary outcome; (ii) include at least five trials, each with at least one event across the two trial arms; (iii) accompanied by risk of bias assessments, with all five core domains of the tool having been assessed (sequence generation, allocation concealment, blinding, incomplete outcome data and selective outcome reporting); (iv) compare an active intervention with a control or 'older' intervention; (v) include no trials that overlapped with another meta-analysis in the data set. Details of the process for selecting eligible meta-analyses are provided in Web Appendix 1. Meta-analyses can inform estimation of the bias associated with a particular domain only if they contain at least one trial at 'low risk' of bias and one at 'high or unclear' risk of bias. We refer to these as *informative* meta-analyses for that bias domain.

We categorized each meta-analysis according to objectivity of the outcome measure (see below); direction of outcome (whether adverse or favorable),(16, 17) type of intervention (pharmacological, surgical, psychosocial and behavioral, care pathways, and other), clinical area (based on the International Classification of Diseases 10th Revision, ICD-10),(20) and whether the comparator was an active intervention (i.e. not a placebo, untreated or standard care). Classification of outcome measure objectivity followed Savović et al. (16, 17): we categorized outcome measures as (a) all-cause mortality; (b) other objectively assessed (including live births, non-cephalic births, low birthweight, miscarriage, pregnancy and all automated laboratory outcomes); (c) semi-objective (where the outcome event is considered to be measured accurately but the decision behind it influenced by a clinician's or patient's judgement, e.g. hospital admissions or re-admissions, total dropouts/withdrawals for any reason, treatment completers, caesarean section, spontaneous vaginal birth, operative/assisted delivery, conversion to open surgery, additional treatments administered); and (d) subjectively assessed (e.g. clinician assessed outcomes, symptoms and symptom scores, pain, mental health outcomes, cause-specific mortality). Too few meta-analyses had outcomes in the objective and semi-objective categories (b and c) for separate analyses to be possible, so we combined these categories as "other objective". When both objective and subjective methods of outcome assessment were used in different trials contributing to the same meta-analysis, the meta-

analysis was categorized as having a subjectively assessed outcome (e.g. some trials in meta-analyses examining smoking cessation used a laboratory measure, while others used patient self-reporting).

Statistical analysis

To explore correlations between bias domains, we computed odds ratios for the association between risk of bias judgements for pairs of domains using logistic regression in Stata 14. For the main analyses, we modelled intervention effects as log odds ratios (ORs) with outcomes coded so that ORs less than 1 corresponded to beneficial intervention effects in all meta-analyses. In the main analysis, 'high risk' and 'unclear risk' bias judgments were grouped together. The underlying idea of the analysis is described in Web Appendix 2 and illustrated in Web Figure 1. We fitted Bayesian hierarchical bias models, assuming a binomial likelihood ('Model 3' by Welton *et al.*(21)). This assumes random intervention effects (between-trial heterogeneity) within meta-analyses, which allows us to assess whether individual bias domains are associated with increased heterogeneity. The model includes parameters for average bias in intervention effects (log odds ratios comparing trials at 'high or unclear' with 'low' risk of bias, averaged across all meta-analyses) and two sources of variation in bias. Variation in bias among trials within meta-analyses was quantified using a term κ^2, representing the average increase in between-trial heterogeneity in trials at 'high or unclear risk' of bias (vs. 'low' risk of bias) for each bias domain. Variation in mean bias across meta-analyses was quantified by a between–meta-analysis variance φ^2. Posterior means for average bias were exponentiated and reported as ratios of odds ratios (RORs); posterior medians for κ and φ are reported on the log odds ratio scale; all are presented with 95% credible intervals (CrIs). Meta-analyses containing fewer than two studies at 'low risk' of bias and at 'high or unclear' risk of bias are uninformative for κ, and were prevented from influencing the estimation of this parameter. Additional statistical analysis information and analysis code is provided in Web Appendix 2.

We conducted univariable analyses for each of four risk of bias domains (sequence generation, allocation concealment, blinding and incomplete outcome data) using all informative meta-analyses

for that domain (Model A in Web Appendix 2). We did not explore the association between the selective outcome reporting domain and intervention effect estimates. This domain currently addresses the non-reporting of outcomes rather than bias in the results available for meta-analysis, so is not directly relevant to bias in the observed results. Analyses were also stratified according to type of outcome measure (all-cause mortality, other objectively assessed, and subjectively assessed). Multivariable analyses were based on an extended model assuming distinct variance components associated with each bias domain (Model B in Web Appendix 2), described in Savović et al.(16) We also fitted multivariable analyses that allowed interactions between sequence generation and allocation concealment; allocation concealment and blinding; and sequence generation and blinding (Model C in Web Appendix 2). We conducted a univariable sensitivity analysis combining trials with an 'unclear risk' of bias judgement with those with 'low risk' of bias (rather than with 'high risk'). We also conducted separate analyses for objective and semi-objective outcomes.

RESULTS

Following our selection process the final ROBES study dataset consisted of 228 meta-analyses containing 2443 trials (**Error! Reference source not found.**). The full list of included reviews and meta-analysis is provided in the Web Appendix 3. The median year of publication of included reviews was 2008 (interquartile range 2005 to 2010, range 1996 to 2011), and for trials 1999 (inter-quartile range 1992 to 2005, range 1950 to 2011). The median sample size was 1290 (inter-quartile range 676 to 3403, range 110 to 341,351) for meta-analyses and 114 (interquartile range 60 to 256, range 8 to 182,000) for trials. Based on the categorization of clinical areas according to World Health Organization International Classification of Diseases 10th revision, the most frequently assessed conditions were related to pregnancy and childbirth (28 meta-analyses, 12.3 %) and mental health (27, 11.8 %), followed by circulatory (21, 9.2 %), and respiratory (20, 8.8 %) system conditions. Subjectively assessed outcomes were reported most frequently, in 127 (55.7%) meta-analyses, followed by all-cause mortality (42, 18.4 %) (Table 1).

The proportion of trials judged as at low risk of bias was highest for the incomplete outcome data domain (1493 trials, 61.1%), followed by sequence generation (1143, 46.8%), blinding (1119, 45.8%) and allocation concealment (1033, 42.3%). The proportion of trials with unclear risk of bias was highest for allocation concealment (1267, 51.9%) and sequence generation (1226, 50.2%), and was markedly lower for blinding (641, 26.2%) and incomplete outcome data (580, 23.7%). The proportion of trials rated as high risk of bias was highest for blinding (683, 28.0%), followed by incomplete outcome data (370, 15.2%), with low proportions rated as high risk for allocation concealment (143, 5.9%), and sequence generation (74, 3.0%) (Table 2). Numbers of trials with each combination of the four risk of bias domain judgements by type of outcome are shown in Web Figure 2, Web Appendix 4.

For sequence generation, 2158 trials were included in 189 (82.9%) informative meta-analyses, of which 1006 (46.6%) were judged as low, 1081 (50.1%) unclear, and 71 (3.3%) high risk of bias. For allocation concealment, 2121 trials were included in 188 (82.5%) informative meta-analyses, of which 933 (44.0%) were judged as low, 1068 (50.3%) as unclear and 120 (5.7%) as high risk of bias. Only 144 (63.2%) meta-analyses (1678 trials) were informative for blinding: 854 (50.9%) trials were judged as low, 437 (26.0%) as unclear, and 387 (23.1%) as high risk of bias. For incomplete outcome data, 1956 trials were included in 167 (73.2%) informative meta-analyses: 1156 (59.1%) were judged as low, 475 (24.3%) unclear, and 325 (16.6%) high risk of bias.

There was a strong association between judgements of low risk of bias for sequence generation and allocation concealment (OR 10.4 [95% confidence interval 8.6, 12.5]) (Table 3). Odds ratios for this association were consistent across types of outcome variable. Associations between low risk of bias judgements for the other 5 pairs of domains were of smaller magnitude, with odds ratios across all trials varied between 1.8 and 2.9 (Table 3).

Table 4 and Web Figure 3 show results from univariable analyses (based on Model A). Intervention effect estimates were exaggerated by an average 9% in trials judged as at high or unclear risk of bias for sequence generation (ROR 0.91 [95% Cr-I 0.86, 0.98]). There was only a modest increase in between-trial heterogeneity among such trials (standard deviation (SD) increase 0.09 [95% Cr-I 0.02,

0.21]). Mean bias varied between meta-analyses, although this variability was imprecisely estimated (SD 0.10 [95% Cr-I 0.02, 0.20], Table 4). There was no convincing evidence that the magnitude of average bias differed according to the type of outcome. Meta-analyses with subjective outcomes contributed the most data to the analysis, and the average bias among these studies was similar to the overall result (ROR 0.90 [95% Cr-I 0.83, 0.98]). In multivariable analyses (based on Model B) the association between risk of bias judgement and intervention effect estimate was attenuated after adjusting for risk of bias judgements for allocation concealment, blinding and incomplete outcome data (ROR 0.95 [95% Cr-I 0.89, 1.03]). The average bias was similar across all outcome types (Table 5, Web Figure 4).

As there was a strong association between sequence generation and allocation concealment, the estimates of average bias for these two domains may be expected to be similar. Intervention effect estimates were exaggerated by an average 8% (ROR 0.92 [95% Cr-I 0.86, 0.98]) in trials judged to be at high or unclear risk of bias for allocation concealment, but there was very little evidence of an increase in between-trial heterogeneity (SD increase 0.05 [95% Cr-I 0.01, 0.15]). The variability in average bias across meta-analyses was small (SD 0.05 [95% Cr-I 0.01, 0.17]). There was little evidence that the average bias varied according to type of outcome. Estimates of both between-trial and between meta-analysis heterogeneity in bias were low for all outcome types. As for sequence generation, the analysis adjusted for the other three domains (Model B) produced an attenuated estimate of average bias (ROR 0.96 [95% Cr-I 0.88, 1.03]) and the estimates were very similar across all outcome types (Table 5, Web Figure 4).

Intervention effect estimates were exaggerated by an average 13% (ROR 0.87 [95% Cr-I 0.80, 0.93]) in trials judged to be at high or unclear risk of bias for blinding. Between-trial heterogeneity was modestly increased for such studies (SD increase 0.10 [95% Cr-I 0.02, 0.25]), and average bias varied between meta-analyses (SD 0.12 [95% Cr-I 0.02, 0.24]). There was little evidence that intervention effects differed according to type of outcome. Increases in between-trial heterogeneity (SD increase 0.22 [95% Cr-I 0.04, 0.36]), and between-meta-analysis heterogeneity in average bias (SD 0.19 [95% Cr-I 0.03, 0.34]) appeared greater in meta-analyses assessing subjective outcomes

than for all-cause mortality or other objective outcomes. In adjusted analysis (Model B), the estimated effect of high or unclear risk of bias due to blinding was similar to the unadjusted estimate (ROR 0.88 [95% Cr-I 0.81, 0.94]).

There was little evidence that intervention effects were exaggerated in trials judged to be at high or unclear risk of bias for incomplete outcome data (ROR 0.98 [95% Cr-I 0.92, 1.05]). The corresponding estimated increase in between-trial heterogeneity was small (SD 0.05 [95% Cr-I 0.01, 0.15]). There was little evidence that average bias or increases in between-trial heterogeneity varied according to type of outcome. The adjusted estimates were very similar to unadjusted (Table 5, Web Figure 4).

The results of the sensitivity analysis (Model A) in which trials with an unclear risk of bias judgement were combined with those at low risk of bias are shown in Web Table 1. The average intervention effects in meta-analyses with high risk of bias for blinding compared with those with low or unclear risk of bias were exaggerated on average by 13 % (ROR 0.87 [95% Cr-I 0.79, 0.95]), consistent with the main analysis. For the other three bias domains the credible intervals for estimates of average bias included the null. These analyses included fewer informative meta-analyses, especially for sequence generation and allocation concealment, and consequently have wider credible intervals. Estimated increases in between-trial heterogeneity were larger for sequence generation, compared with that observed in the main analysis.

The separate estimates for subgroups of meta-analyses with ‘other objective’ and ‘semi-objective’ outcomes (which were analysed together in the main analysis) were similar to each other for allocation concealment and blinding. They differed somewhat for sequence generation (RORs 0.85 [95% Cr-I 0.67, 1.09] for other objective and 1.08 [95% Cr-I 0.91, 1.34] for semi-objective outcomes) and incomplete outcome data (RORs 0.94 [95% Cr-I 0.72, 1.22] for other objective and 1.11 [95% Cr-I 0.93, 1.30] for semi-objective), but the credible intervals are wide and overlapping (Web Table 2).

In multivariable models with interaction terms (Model C), an interaction was observed between allocation concealment and blinding (ROR 0.84 [95% Cr-I 0.74, 0.96]), and between sequence generation and blinding (ROR 0.77 [95% Cr-I 0.66, 0.91]) (Web Table 3). This means that lack of

blinding may introduce greater bias in estimation of intervention effects within studies with inadequate randomization than within studies with adequate randomization.

DISCUSSION

Using a collection of 2443 trials included in 228 meta-analyses, our estimates of the association between average intervention effect estimates and routinely collected risk of bias judgements for sequence generation, allocation concealment, blinding and incomplete outcome data confirm that problems with randomization and a lack of blinding are on average associated with a modest (around 10%) exaggeration of treatment effect estimates. Lack of blinding appears to have the largest influence on treatment effect estimates: this remains after adjusting for other domains. There was little evidence that these biases varied according to the type of outcome measure assessed: although there were some differences in RORs for different outcome types in univariable analyses, the credible intervals overlapped and the differences were attenuated or disappeared in adjusted analyses. We found little evidence that trials assessed as at high risk of bias for incomplete outcome data produced systematically different estimates compared with trials at low risk of bias for this domain, for all types of outcome measures. Variability of treatment effects was higher in trials that lacked blinding and had subjective outcomes, suggesting that for such trials the direction and magnitude of bias is unpredictable. Such variability in bias was observed both between trials within a meta-analysis and across meta-analyses. There was little evidence of such variation in bias for other bias domains or for objectively determined outcomes. Multivariable analyses suggested that effects of individual risk of bias domain judgements were less than additive, in that estimated effects of two bias domain judgements together were less than the combined individual effects.

To our knowledge, this study represents the most comprehensive attempt to date to quantify the influence of four bias domains on intervention effect estimates from randomized controlled trials using routinely collected risk of bias assessments from published Cochrane reviews. Our findings indicate that assessments are associated with effect sizes, on average, for three of the four domains,

providing some degree of validation of the risk of bias tool. However, to interpret our findings as evidence of bias due to the methods implemented in the trials, it is important to consider the accuracy and reliability of these risk of bias assessments. The assessments were made by a large number of Cochrane review authors with varying degrees of experience and training, and we did not replicate assessments to determine how appropriate they were. Although detailed guidance on how to assess risk of bias in trials included in Cochrane reviews is available in chapter 8 of the Cochrane Handbook,(18) review authors have reported that they find aspects of the assessment difficult.(22) Indeed, some studies have reported that the assessor agreement and inter-rater reliability of the risk of bias tool is suboptimal.(23, 24) Specifically, individual reviewers have different criteria for judging a study to be 'low risk' of bias: some may be more confident to make a judgment with less information, while others would opt for 'unclear risk'. Standard advice is that two assessors independently assess risk of bias and resolve disagreements through discussion. We presume that this advice was followed. As a safeguard that recommended assessment methods were followed at least to some extent, we restricted eligibility to reviews that had completed all five prescribed bias domains. It is possible that individual review teams had their own criteria for rating study 'low risk' for each of the domains, which may have differed from those described in the handbook.

In our main analyses, risk of bias judgements were dichotomized so that "high" and "unclear" risk were considered together. This allows for like-for-like comparisons with results from most of the previous empirical studies, including our previous study.(17) Furthermore, there were few "high" risk of bias judgements, so analyses with the alternative dichotomization of "high" versus "low" or "unclear" risk of bias were not informative. For domains of sequence generation and allocation concealment, a "high" risk of bias judgement was recorded in only 3% and 6% trials, respectively (Table 2). We demonstrated that Cochrane assessors frequently reach a judgement of "unclear" risk of bias (Table 2). Inadequate reporting of key features of trial design is a likely explanation of this high rate of uncertainty, particularly for methods of sequence generation and allocation concealment. This observation is consistent with findings from Turner et al. that allocation concealment was reported in sufficient detail in 30% (722/2396) of published randomized trials in their study.(25)

Our adjusted results for sequence generation and allocation concealment are largely consistent with recent meta-analyses of all previous meta-epidemiologic studies reported in a recent systematic review.(26) Blinding and incomplete data in studies included in this review were not assessed in the same way as in our study and cannot be meaningfully compared with our results. Our results for average bias are slightly smaller than those from our previous study (the BRANDO study – Bias in Randomized and Observational Studies).(17) This may be dilution due to measurement error, arising because the risk of bias assessments in the current study were conducted by a heterogeneous group of Cochrane reviewers. In contrast, assessments used in the BRANDO study were done by teams of trained methodologists, and data were only combined in the BRANDO analyses where the definitions for adequate versus inadequate study method were consistent across studies. Our finding that the lack of blinding in trials with subjective outcomes can lead to biased effect estimates, but the direction and magnitude of such bias is unpredictable, also confirms a finding from the BRANDO study. The main difference between findings from the current study and the BRANDO study is that we do not see a clear difference in the magnitude of bias according to type of outcome.

In summary, our results confirm that some aspects of the conduct of randomized trials, particularly blinding, are associated with a modest exaggeration of treatment effects on average, but there is little evidence that the average bias differs according to whether the outcome was subjectively or objectively assessed. However, lack of blinding in trials with subjective outcomes leads to increased heterogeneity and hence unpredictable bias in effect estimates. As far as possible, clinical and policy decisions should be cautious when they are based on trials in which blinding was not reported or not feasible and outcome measures were subjectively assessed. Future developments of tools for assessing risk of bias in randomized trials (27, 28) should reflect this observation and collect information on the subjectivity of an outcome. Facilities for capture of detailed routine assessments of risk of bias in randomized trials should be made available for future meta-epidemiological research and could contribute to further improvements in methods of risk of bias assessments.

Acknowledgments

Author affiliations: Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom (Jelena Savović, David Mawdsley, Hayley E. Jones, Rebecca Beynon, Julian P.T. Higgins and Jonathan A.C. Sterne); The National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West (NIHR CLAHRC West), University Hospitals Bristol National Health Service Foundation Trust, Bristol, United Kingdom (Jelena Savović, Julian P.T. Higgins and Jonathan A.C. Sterne); Medical Research Council (MRC) Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom (Rebecca M. Turner); MRC Clinical Trials Unit, University College London, London, United Kingdom (Rebecca M. Turner).

Author contributions: Savović contributed to the conception of the study, designed the study, obtained data, collected additional data and categorized it, carried out most of the analyses, and wrote the first draft. Turner contributed to data categorization, adapted the WinBUGS models and carried out some of the analyses. Mawdsley carried out some of the analyses. Beynon carried out independent categorizations of meta-analyses. Jones helped adapt the WinBUGS models and provided statistical and methodological advice throughout the study. Higgins provided methodological and statistical advice throughout the study. Sterne conceived the study, provided methodological and statistical advice throughout the study. All co-authors critically revised the manuscript for important intellectual content.

This work was supported by the MRC Fellowship (Savović; G0701659/1) and by the MRC Methodology Research Panel grant (Turner, MR/K014587/1). Savović's time was partly supported by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West at University Hospitals Bristol National Health Service Foundation Trust. Turner was supported by the MRC (MC-U105260558 and MC-UU12023/24). Sterne is funded by a National Institute for Health Research Senior Investigator award (NF-SI-0611-10168). The views expressed are those of the authors and not necessarily those of the MRC, the National Health Service, the National Institute for Health Research or the Department of Health.

The authors thank Professor Nicky J. Welton for advice about statistical models.

Conflict of interest: Mawdsley was employed at the University of Bristol on an unrelated project part funded by Pfizer between January 2015 and November 2016. The other authors report no financial or commercial conflicts of interest.

Declaration of non-commercial interests: Savović, Sterne, Jones and Beynon are the co-authors of the BRANDO study referred to in the manuscript. Sterne, Higgins and Savović were involved in the development or evaluation of the Cochrane risk of rias tool. The other authors report no academic conflicts of interest.

REFERENCES

1. Gluud LL. Bias in clinical intervention research. *Am J Epidemiol* 2006;163(6):493-501.
2. Sterne JA, Juni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;21(11):1513-1524.
3. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *BMJ* 1997;315(7109):617-619.
4. Als-Nielsen B, Chen W, Gluud LL, et al. Are trial size and reported methodological quality associated with treatment effects? Observational study of 523 randomised trials [abstract]. Presented at 12th Cochrane Colloquium: Bridging the Gaps, Ottawa, Ontario, Canada, October 2-6, 2004.
5. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287(22):2973-2982.
6. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;135(11):982-989.
7. Nuesch E, Reichenbach S, Trelle S, et al. The importance of allocation concealment and patient blinding in osteoarthritis trials: a meta-epidemiologic study. *Arthritis Rheum* 2009;61(12):1633-1641.
8. Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007;36(4):847-857.
9. Egger M, Juni P, Bartlett C, et al. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;7(1):1-76.
10. Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. *J Clin Epidemiol* 2006;59(12):1249-1256.

11. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352(9128):609-613.
12. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273(5):408-412.
13. Bialy L, Vandermeer B, Lacaze-Masmonteil T, et al. A meta-epidemiological study to examine the association between bias and treatment effects in neonatal trials. *Evid Based Child Health* 2014;9(4):1052-1059.
14. Chaimani A, Vasiliadis HS, Pandis N, et al. Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study. *Int J Epidemiol* 2013;42(4):1120-1131.
15. Hartling L, Hamm MP, Fernandes RM, et al. Quantifying bias in randomized controlled trials in child health: a meta-epidemiological study. *PLoS One* 2014;9(2):e88008.
16. Savović J, Jones H, Altman D, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess* 2012;16(35):1-82.
17. Savović J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157(6):429-438.
18. Higgins JPT, Altman DG. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, England: John Wiley & Sons Ltd; 2008:187-241.
19. Review Manager (RevMan) software, version 5.0. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2008.
20. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10), Version: 2010. World Health Organization; 2010. (http://apps.who.int/classifications/icd10/browse/2010/en#/). (Accessed 30/03/2017).

21. Welton NJ, Ades AE, Carlin JB, et al. Models for potentially biased evidence in meta-analysis using empirically based priors. *J R Stat Soc Ser A Stat Soc* 2009;172:119-136.

22. Savović J, Weeks L, Sterne JA, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. *Syst Rev* 2014;3:37.

23. Armijo-Olivo S, Ospina M, da Costa BR, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One* 2014;9(5):e96920.

24. Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 2013;66(9):973-981.

25. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev* 2012;11:MR000030.

26. Page MJ, Higgins JP, Clayton G, et al. Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. *PLoS One* 2016;11(7):e0159267.

27. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

28. Higgins JPT, Sterne JC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials In: Chandler J, McKenzie J, Boutron I, et al., eds. *Cochrane Methods*. *Cochrane Database Syst Rev* 2016; Issue 10(Suppl 1):29-31 (dx.doi.org/10.1002/14651858.CD201601).

Table 1. Characteristics of Included Meta-analyses and Trials

Characteristics of meta-analyses and trials	Meta-analyses (N=228)		Trials (N=2443)	
	n	%	n	%
Clinical area according to ICD-10 chapters				
Pregnancy and childbirth (O)	28	12.3	387	15.8
Mental & behavioural (F)	27	11.8	286	11.7
Circulatory (I)	21	9.2	259	10.6
Respiratory (J)	20	8.8	196	8.0
Genitourinary (N)	19	8.3	214	8.8
Perinatal (P)	18	7.9	155	6.3
Digestive (K)	17	7.5	193	7.9
Infectious & parasitic (A-B)	11	4.8	113	4.6
Neoplasms (C-D)	11	4.8	103	4.2
Nervous system (G)	10	4.4	102	4.2
Injury & poisoning (S-T)	10	4.4	98	4.0
Other ICD-10 chapters	34	14.9	319	13.1
Unclassified	2	0.9	18	0.7
Type of experimental intervention				
Pharmacological	151	66.2	1688	69.1
Provision of care	14	6.1	111	4.5
Surgical interventions or procedures	12	5.3	126	5.2
Psychosocial & behavioural	11	4.8	125	5.1
Other	40	17.5	393	16.1
Type of comparison intervention				
Pharmacological	26	11.4	251	10.3
Surgical interventions or procedures	8	3.5	99	4.1
Other active interventions	4	1.8	33	1.4
Placebo / no treatment [a]	58	25.4	677	27.7
Placebo	51	22.4	560	22.9
Standard/usual care	32	14.0	307	12.6
No treatment	25	11.0	233	9.5
Standard care / placebo / no treatment [a]	24	10.5	283	11.6
Type of outcome measure [b]				
All-cause mortality	42	18.4	429	17.6
Other objective	20	8.8	197	8.1
Subjective	127	55.7	1356	55.5
Mixture of objective and subjective [a]	2	0.9	70	2.9
Semi-objective	37	16.2	391	16.0

ICD-10 = International Classification of Diseases, 10th edition (World Health Organization).

[a] Combined at meta-analysis level;

[b] **Other objective**: automated or semi-automated laboratory measures including biochemical measurements and serological tests; birthweight, live births, preterm birth, clinical pregnancy, un-intended pregnancy, non-cephalic births; **Semi-objective** (outcomes for which ascertainment is accurate but their occurrence was influenced by a patient's or care-provider's subjective judgment): blood transfusion given, prescribed anti-platelet medication, caesarean section, spontaneous vaginal birth, preterm birth, oxytocin augmentation, failure of extubation, surgical evacuation, conversion to open surgery, need for further surgery, radical resection, hospital admissions, admissions to neonatal intensive care unit, hospital readmissions, presentations at emergency department, compliance with intervention, completion of study, withdrawals or dropouts from the study, discontinuation of treatment; not remaining in contact with psychiatric services; **Subjective**: signs and symptoms of disease and improvement thereof; symptom scales and scores; mental health outcomes, imaging and radiological outcomes, pain, quality of life, adverse events of treatment, other patient reported outcomes or those relying on a diagnosis by a physician, cause-specific deaths. **Mixture**: meta-analyses where some trials used laboratory validation, while others used self-report for smoking cessation.

Table 2. Number and Percentage of Trials by Risk-of-Bias Judgment

Risk of bias domain judgements	*Low risk*		*High risk*		*Unclear risk*	
	n	%	n	%	n	%
Sequence generation	1143	46.8	74	3.0	1226	50.2
Allocation concealment	1033	42.3	143	5.9	1267	51.9
Blinding	1119	45.8	683	28.0	641	26.2
Incomplete outcome data	1493	61.1	370	15.2	580	23.7

Table 3. Associations Between Risk of Bias Domains

Risk of bias domain pairs	All trials (N=2443)		All-cause mortality (N=429)		Other objective outcome (N=197)		[a] 'Semi-objective' outcome (N=391)		[b]Subjective outcome (N=1426)	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Sequence generation, Allocation concealment	10.4	8.6, 12.5	11.3	7.1, 17.9	16.7	7.9, 34.9	9.7	6.1, 15.4	9.5	7.4, 12.2
Sequence generation, Blinding	2.5	2.2, 3.0	3.1	2.1, 4.6	2.0	1.0, 3.8	2.2	1.5, 3.3	2.8	2.2, 3.4
Sequence generation, Incomplete outcome data	2.1	1.8, 2.4	2.7	1.8, 4.0	5.3	2.8, 9.8	1.7	1.1, 2.6	1.8	1.4, 2.2
Allocation concealment, Blinding	2.9	2.4, 3.4	4.0	2.7, 6.0	6.0	3.0, 12.1	1.3	0.8, 1.9	3.2	2.6, 4.1
Allocation concealment, Incomplete outcome data	2.2	1.8, 2.6	2.9	1.9, 4.4	4.4	2.4, 8.3	1.3	0.9, 2.0	2.0	1.6, 2.5
Blinding, Incomplete outcome data	1.8	1.5, 2.1	1.8	1.2, 2.6	1.4	0.7, 2.6	2.1	1.4, 3.2	1.8	1.5, 2.3

OR = Odds ratio; CI = confidence interval.

[a] Outcomes for which ascertainment is accurate but their occurrence was influenced by a patient's or care-provider's subjective judgment (e.g. duration of hospital stay, admissions, withdrawals, caesarean section).

[b] Includes meta-analyses in which some trials had subjective measures and some objective (e.g. self-report and laboratory measures).

Table 4. Estimated Ratios of Odds Ratios and Between-Meta-Analysis Heterogeneity in Mean Bias Associated With Risk of Bias Judgements, According to Type of Outcome Measure: Univariable Analyses (Model A)

Risk of bias domain and outcome	No. Contributing		Average bias		Meta-analyses contributing to kappa estimation	Within meta-analysis heterogeneity		Between meta-analysis heterogeneity	
	Meta-analyses	Trials	ROR	95% Cr-I		*Kappa*	95% Cr-I	*Phi*	95% Cr-I
Sequence generation: High/unclear risk of bias vs. low risk of bias									
All	189	2,158	0.91	0.86, 0.98	142	0.09	0.02, 0.21	0.10	0.02, 0.20
Mortality	34	363	0.84	0.71, 1.01	27	0.13	0.01, 0.39	0.09	0.01, 0.37
Other objective/Semi-objective	47	523	0.99	0.87, 1.16	38	0.10	0.01, 0.31	0.14	0.01, 0.41
Subjective/Mixed	108	1,272	0.90	0.83, 0.98	77	0.08	0.01, 0.21	0.08	0.01, 0.22
Allocation concealment: High/unclear risk of bias vs. low risk of bias									
All	188	2,121	0.92	0.86, 0.98	139	0.05	0.01, 0.15	0.05	0.01, 0.17
Mortality	35	358	0.84	0.71, 1.01	27	0.07	0.01, 0.30	0.12	0.01, 0.42
Other objective/ Semi-objective	49	524	0.96	0.86, 1.07	40	0.04	0.01, 0.14	0.05	0.01, 0.19
Subjective/Mixed	104	1,239	0.91	0.83, 0.99	72	0.08	0.01, 0.25	0.06	0.01, 0.20
Blinding: High/unclear risk of bias vs. low risk of bias									
All	144	1,678	0.87	0.80, 0.93	105	0.10	0.02, 0.25	0.12	0.02, 0.24
Mortality	31	327	0.83	0.72, 0.97	25	0.06	0.01, 0.26	0.06	0.01, 0.25
Other objective/ Semi-objective	32	334	0.94	0.81, 1.10	24	0.06	0.01, 0.21	0.06	0.01, 0.28
Subjective/Mixed	81	1,017	0.83	0.73, 0.93	56	0.22	0.04, 0.36	0.19	0.03, 0.34
Incomplete outcome data: High/unclear risk of bias vs. low risk of bias									
All	167	1,956	0.98	0.92, 1.05	112	0.05	0.01, 0.16	0.05	0.01, 0.15
Mortality	29	303	0.92	0.79, 1.08	19	0.08	0.01, 0.32	0.06	0.01, 0.24
Other objective/ Semi-objective	43	471	1.03	0.90, 1.19	28	0.07	0.01, 0.25	0.06	0.01, 0.25
Subjective/Mixed	95	1,182	0.97	0.88, 1.07	65	0.06	0.01, 0.17	0.10	0.01, 0.30

ROR = ratio of odds ratios; Cr-I = credible interval; *Kappa* - measure of within meta-analysis heterogeneity; *Phi* - measure of between meta-analysis heterogeneity. For graphical representation of these results, see Web Figure 3.

Table 5. Estimated Ratios of Odds Ratios and Between-Meta-Analysis Heterogeneity in Mean Bias Associated Risk of Bias Judgements, According to Type of Outcome Measure: Multivariable Analyses (Model B)

Risk of bias domain and outcome	No. Contributing		Average bias		Meta-analyses contributing to kappa estimation	Within meta-analysis heterogeneity		Between meta-analysis heterogeneity	
	Meta-analyses	Trials	ROR	95% Cr-I		*Kappa*	95% Cr-I	*Phi*	95% Cr-I
Sequence generation: High/unclear risk of bias vs. low risk of bias									
All	189	2,158	0.95	0.88, 1.03	142	0.08	0.02, 0.18	0.11	0.03, 0.22
Mortality	34	363	0.92	0.75, 1.18	27	0.14	0.02, 0.36	0.14	0.03, 0.42
Other objective/Semi-objective	47	523	1.06	0.90, 1.28	38	0.14	0.03, 0.33	0.20	0.04, 0.44
Subjective/Mixed	108	1,272	0.94	0.84, 1.04	77	0.08	0.02, 0.18	0.11	0.02, 0.24
Allocation concealment: High/unclear risk of bias vs. low risk of bias									
All	188	2,121	0.96	0.88, 1.03	139	0.06	0.01, 0.15	0.07	0.02, 0.16
Mortality	35	358	0.92	0.74, 1.13	27	0.11	0.03, 0.29	0.15	0.03, 0.42
Other objective/ Semi-objective	49	524	0.94	0.81, 1.08	40	0.07	0.01, 0.18	0.09	0.02, 0.25
Subjective/Mixed	104	1,239	0.95	0.86, 1.07	72	0.10	0.02, 0.23	0.08	0.02, 0.20
Blinding: High/unclear risk of bias vs. low risk of bias									
All	144	1,678	0.88	0.81, 0.94	105	0.10	0.02, 0.22	0.12	0.03, 0.23
Mortality	31	327	0.87	0.73, 1.03	25	0.10	0.02, 0.26	0.10	0.02, 0.28
Other objective/ Semi-objective	32	334	0.95	0.79, 1.12	24	0.09	0.02, 0.24	0.10	0.02, 0.34
Subjective/Mixed	81	1017	0.84	0.75, 0.95	56	0.17	0.04, 0.33	0.19	0.05, 0.35
Incomplete outcome data: High/unclear risk of bias vs. low risk of bias									
All	167	1,956	1.01	0.94, 1.09	112	0.07	0.01, 0.16	0.07	0.02, 0.16
Mortality	29	303	0.99	0.82, 1.18	19	0.11	0.02, 0.31	0.10	0.02, 0.30
Other objective/ Semi-objective	43	471	1.04	0.90, 1.21	28	0.11	0.02, 0.30	0.09	0.02, 0.26
Subjective/Mixed	95	1,182	1.00	0.90, 1.12	65	0.07	0.01, 0.17	0.11	0.03, 0.27

Analyses for each bias domain were adjusted for risk of bias judgements for the other three domains (Model B in Web Appendix 2). For graphical representation of these results, see Web Figure 4. ROR = ratio of odds ratios; Cr-I = credible interval; *Kappa* - measure of within meta-analysis heterogeneity; Phi – measure of between meta-analysis heterogeneity.

Figure

Figure 1. Study Selection Flow Char