

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

J Savović,<sup>1</sup> HE Jones,<sup>1</sup> DG Altman,<sup>2</sup> RJ Harris,<sup>3</sup>  
P Jüni,<sup>4</sup> J Pildal,<sup>5</sup> B Als-Nielsen,<sup>6</sup> EM Balk,<sup>7</sup>  
C Gluud,<sup>8</sup> LL Gluud,<sup>9</sup> JPA Ioannidis,<sup>10</sup>  
KF Schulz,<sup>11</sup> R Beynon,<sup>1</sup> N Welton,<sup>1</sup> L Wood,<sup>12</sup>  
D Moher,<sup>13</sup> JJ Deeks<sup>14</sup> and JAC Sterne<sup>1\*</sup>

<sup>1</sup>School of Social and Community Medicine, University of Bristol, Bristol, UK

<sup>2</sup>Centre for Statistics in Medicine, University of Oxford, Wolfson College, Oxford, UK

<sup>3</sup>Centre for Infections, Health Protection Agency, Colindale, UK

<sup>4</sup>Division of Clinical Epidemiology and Biostatistics, Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland; CTU Bern, Bern University Hospital, Bern, Switzerland

<sup>5</sup>The Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

<sup>6</sup>Copenhagen Trial Unit, Centre for Clinical Intervention Research and Department of Paediatrics, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark



## Executive summary

*Health Technology Assessment* 2012; Vol. 16: No. 35  
DOI: 10.3310/hta16350

Health Technology Assessment  
NIHR HTA programme  
[www.hta.ac.uk](http://www.hta.ac.uk)



<sup>7</sup>Center for Clinical Evidence Synthesis, Tufts Clinical and Translational Science Institute, Tufts Medical Center, Boston, MA, USA

<sup>8</sup>Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

<sup>9</sup>Department of Internal Medicine, Copenhagen University Hospital Gentofte, Copenhagen, Denmark

<sup>10</sup>Stanford Prevention Research Center, Department of Medicine and Department of Health Research and Policy, Stanford University School of Medicine and Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA

<sup>11</sup>Quantitative Sciences, Family Health International, Research Triangle Park, Durham, NC, USA

<sup>12</sup>School Food Trust, Sheffield, UK

<sup>13</sup>Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

<sup>14</sup>Public Health Epidemiology and Biostatistics, School of Health and Population Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

\*Corresponding author

# Executive summary

## Background

Systematic reviews of randomised controlled trials (RCTs) provide the best evidence for clinical and policy decision-making about medical interventions. The design of RCTs should incorporate characteristics (such as concealment of randomised allocation and blinding of participants and personnel) that avoid biases resulting from lack of comparability of the intervention and control groups. Empirical evidence suggests that absence of such characteristics, as reported in trial publications, leads to biased estimates of intervention effects. Collections of meta-analyses assembled in meta-epidemiological studies are used to study associations of study design characteristics with intervention effect estimates, but findings vary between studies.

## Objectives

- To combine data from contributing meta-epidemiological studies into a single database, and derive a harmonised data set in which overlap between meta-analyses was removed.
- To examine agreement in the assessment of reported study design characteristics in the subset of trials that was assessed in two or more contributing meta-epidemiological studies.
- To examine the influence of inadequate or unclear (compared with adequate) random sequence generation and allocation concealment, and absent or unclear double blinding (compared with double blinding), on intervention effect estimates and between-trial heterogeneity.
- To examine whether or not these influences vary with the type of clinical area, intervention, comparison and outcome measure.
- To examine the effects of combinations of characteristics, and to estimate adjusted effects using multivariable models.
- To explore the implications of these findings for downweighting of trials whose characteristics are associated with bias in future meta-analyses.

## Methods

We combined data from 10 contributing meta-epidemiological studies into a single database containing 427 reviews, 454 meta-analyses and 4874 trial results, whose design allowed trials to be contained in different meta-analyses, multiple meta-analyses in systematic reviews, overlapping meta-analyses between systematic reviews and multiple references to the same trial or review. Unique identifiers were used to identify sets of meta-analyses with overlapping trials: 258 meta-analyses were unique, whereas for 196 at least one trial overlapped with another meta-analysis. Overlapping meta-analyses and trials were removed according to a pre-specified protocol. The final database contained 363 meta-analyses and 3477 unique trial results. Overlapping trials were used to estimate kappa statistics for agreement between assessments of reported study design characteristics.

Information on outcome measures, interventions and comparisons was extracted from the included systematic reviews. Interventions were classified as pharmacological; surgical; psychosocial and behavioural; and all other interventions. Comparison interventions were classified as inactive (placebo, no intervention, standard care) or active. Outcome measures were grouped as all-cause mortality; other objectively assessed; and subjectively assessed.

The main analyses excluded 87 meta-analyses (1093 trials) from four contributing studies that did not collect data on both study design characteristics and outcome events; 36 meta-analyses (300 trials) in which it was not possible to classify one intervention as experimental and the other as control; one meta-analysis (four trials) that had a continuous outcome measure; 45 trials with missing outcome data; and 50 trials in which either no or all subjects experienced the outcome event.

### **Statistical methods**

Intervention effects were modelled as log-odds ratios; odds ratios  $< 1$  corresponded to beneficial intervention effects. We fitted Bayesian hierarchical bias models that allowed for random intervention effects within meta-analyses, with meta-analysis-specific mean  $\mu$  and between-trial variance  $\tau^2$ . Three effects of study design characteristics were modelled. First, mean intervention effects among trials reported as having a particular study design characteristic may differ from those among trials without that characteristic: estimated mean differences were exponentiated and reported as ratios of odds ratios (RORs). Second, variation in bias between trials within meta-analyses was quantified by standard deviation  $\kappa$ ;  $\kappa^2$  corresponds to the average increase in between-trial heterogeneity in trials with a specified study design characteristic. Third, variation in mean bias between meta-analyses was quantified by between-meta-analysis standard deviation  $\phi$ . We derived 95% credible intervals (CrIs) for each parameter. Bias models were fitted using WinBUGS Version 1.4 (MRC Biostatistics Unit, Cambridge, UK), using vague prior distributions. The prior for variance parameters found to give the best overall performance was a modified Inverse Gamma(0.001, 0.001) prior with increased weight on small values. For location parameters (overall mean bias, baseline response rates, treatment effects), Normal(0, 1000) priors were assumed.

We first conducted univariable analyses for each bias domain separately using all informative meta-analyses (meta-analyses containing trials with and without the characteristic of interest) for each characteristic. The primary analysis used dichotomised variables for each characteristic (inadequate or unclear compared with adequate sequence generation and allocation concealment, and not double blind or unclearly blinded compared with double blind). Analyses were stratified according to type of outcome measure (all-cause mortality, other objectively assessed and subjectively assessed). Further univariable analyses (examining the influence of combinations of design characteristics) and multivariable analyses were conducted using two data subsets: trials with information on all three characteristics and trials with information on both allocation concealment and blinding.

## **Results**

There was good agreement between assessments of reported study design characteristics carried out in the different studies. For sequence generation (two comparisons), the percentages of studies in which the assessments were in agreement were 81% and 82% and kappa statistics were 0.56 and 0.64. For allocation concealment (12 comparisons), percentage agreement varied between 52% and 100% and kappa statistics between 0.19 and 1.00 (median 0.58). Assessments were most reliable for blinding (nine comparisons): percentage agreement ranged from 80% to 100% (in four comparisons), whereas kappa statistics ranged from 0.55 to 1.00 (median 0.87).

### ***Influence of reported study design characteristics: univariable analyses of individual characteristics***

The main analyses were based on a data set containing 1973 trials included in 234 meta-analyses. Based on 944 trials from 112 informative meta-analyses, intervention effect estimates were exaggerated by an average of 11% in trials with inadequate or unclear sequence generation

(ROR 0.89, 95% CrI 0.82 to 0.96), and between-trial heterogeneity was higher among such trials ( $\kappa=0.16$ , 95% CrI 0.03 to 0.27). When analyses were stratified according to type of outcome measure, the average effect of bias associated with inadequate or unclear sequence generation was greatest for subjective outcomes (ROR 0.83, 95% CrI 0.74 to 0.94) and the increase in between-trial heterogeneity was also greatest for such outcomes ( $\kappa=0.20$ , CrI 0.03 to 0.32). There was little evidence that inadequate or unclear sequence generation was associated with exaggeration of intervention effects for all-cause mortality (ROR 0.89, 95% CrI 0.75 to 1.05) or for other objective outcomes (ROR 0.99, 95% CrI 0.84 to 1.16). For all types of outcome, there was only limited between-meta-analysis heterogeneity in mean bias (estimated  $\varphi$  between 0.04 and 0.07).

Based on 1292 trials from 146 informative meta-analyses, intervention effect estimates were exaggerated by 7% in trials with inadequate or unclear allocation concealment (ROR 0.93, 95% CrI 0.87 to 0.99), and between-trial heterogeneity was increased for such studies ( $\kappa=0.12$ , 95% CrI 0.02 to 0.23). The effect of inadequate or unclear allocation concealment was greatest among meta-analyses with a subjectively assessed outcome measure (ROR 0.85, 95% CrI 0.75 to 0.95;  $\kappa=0.20$ , 95% CrI 0.02 to 0.33). In contrast, the average effect of inadequate or unclear allocation concealment was close to the null for meta-analyses with a mortality outcome (ROR 0.98, 95% CrI 0.88 to 1.10) and other objective outcomes (ROR 0.97, 95% CrI 0.85 to 1.10). Estimates of both between-trial and between-meta-analyses heterogeneity in bias were lower for such outcomes than for subjectively assessed outcomes.

Based on 1057 trials from 104 informative meta-analyses, lack of, or unclear, double blinding was associated with an average 13% exaggeration of intervention effects (ROR 0.87, 95% CrI 0.79 to 0.96). Between-trial heterogeneity was higher in such studies ( $\kappa=0.14$ , 95% CrI 0.02 to 0.30), and average bias varied between meta-analyses ( $\varphi=0.14$ , 95% CrI 0.03 to 0.28). Average bias (ROR 0.78, 95% CrI 0.65 to 0.92), between-trial heterogeneity ( $\kappa=0.37$ , 95% CrI 0.19 to 0.53) and between-meta-analysis heterogeneity in average bias ( $\varphi=0.23$ , 95% CrI 0.04 to 0.44) were all greatest for meta-analyses assessing subjective outcome measures. Among meta-analyses with subjectively assessed outcome measures, the effect of lack of blinding appeared greater than the effect of inadequate or unclear sequence generation or allocation concealment.

### ***Influence of reported study design characteristics: univariable analyses of combinations of characteristics***

Estimates of the influence of any risk of selection bias (inadequate or unclear sequence generation or allocation concealment, compared with other trials) on intervention effects and heterogeneity were based on 53 informative meta-analyses containing 534 trials, of which 89 (17%) were assessed as at low risk. Risk of selection bias was associated with an average 11% exaggeration of intervention effect estimates (ROR 0.89, 95% CrI 0.78 to 1.00) and with increased between-trial heterogeneity ( $\kappa=0.12$ , 95% CrI 0.02 to 0.27). Average effects did not differ substantially according to type of outcome measure.

Only 37 informative meta-analyses [409 trials, of which 58 (14%) were assessed as being at low risk of bias] contributed to the analysis of any risk of bias (inadequate or unclear sequence generation or allocation concealment, or lack of or unclear double blinding, compared with all other trials). Any risk of bias was associated with an average 21% exaggeration of intervention effect estimates (ROR 0.79, 95% CrI 0.64 to 0.92). The numbers of informative meta-analyses and trials included in analyses stratified by type of outcome measure were small, but the influence of any risk of bias appeared the smallest for all-cause mortality outcomes.

A total of 104 informative meta-analyses [990 trials, of which 259 (26%) were assessed as at low risk of bias] contributed to analyses of the influence of inadequate or unclear allocation concealment or lack of double blinding (compared with adequate allocation concealment and

presence of double blinding). Intervention effects from trials at high risk of bias according to this definition were exaggerated by an average 12% (ROR 0.88, 95% CrI 0.81 to 0.95). The ROR was smaller for all-cause mortality outcomes than for other objective or subjective outcome measures. The increase in between-trial heterogeneity appeared greatest for subjective outcome measures ( $\kappa=0.17$ , 95% CrI 0.02 to 0.31).

### ***Influence of reported study design characteristics: multivariable analyses***

Multivariable analyses of the influence of inadequate or unclear allocation concealment and lack of or unclear double blinding were based on 169 informative meta-analyses (1456 trials) in which both characteristics were assessed. Estimated RORs were similar to, or modestly attenuated compared with, those in the univariable analyses. Estimated influence on heterogeneity was also modestly attenuated.

Estimated RORs from multivariable analyses of the effects of all three characteristics were of similar magnitudes to those in the univariable analyses for each characteristic. For inadequate or unclear sequence generation or allocation concealment, estimated increases in between-trial heterogeneity (quantified by  $\kappa$ ) were smaller in multivariable analyses than in the corresponding univariable analyses. Estimates of between-meta-analysis variability in average bias were changed little compared with univariable analyses.

### ***Analyses according to type of intervention and clinical area***

For pregnancy and childbirth – the clinical area contributing most meta-analyses to the combined data set – RORs were further from 1 than in the analyses of the whole data set, whereas estimates of the influence of reported study design characteristics on heterogeneity were broadly consistent with analyses of the whole data set. For mental health and circulatory system conditions, RORs were attenuated towards 1. Only small numbers of meta-analyses contributed to estimation of  $\kappa$ , but estimated values of  $\kappa$  and  $\varphi$  were generally smaller for circulatory system meta-analyses than for the other two clinical areas.

The majority of meta-analyses included in the full data set addressed pharmacological interventions; it was therefore unsurprising that overall results restricted to such interventions were consistent with those from the full data set. For surgical interventions, effects of inadequate or unclear sequence generation and inadequate or unclear allocation concealment were estimated from only six and nine meta-analyses respectively; confidence intervals were too wide to allow substantive conclusions to be drawn.

### ***Downweighting potentially biased evidence in future meta-analyses***

We investigated the implications of our results for downweighting of potentially biased evidence in future meta-analyses. Because estimated values of  $\kappa$  and  $\varphi$  were greatest for meta-analyses with subjectively assessed outcomes, the minimum variance of the estimated intervention effect for a trial at high or unclear risk of bias is greatest for such trials. Across all Bias in Randomised and Observational studies (BRANDO) trials with inadequate or unclear sequence generation, bias adjustment led to a median 10% [interquartile range (IQR) 4% to 23%] increase in trial-level variance. Downweighting based on results specific to type of outcome measure has the greatest effect in trials with subjectively assessed outcomes [median 20% (IQR 8% to 39%) increase in variance]. Results were broadly similar for downweighting based on inadequate or unclear allocation concealment. The median increase in variance for trials with subjectively measured outcomes that were not double blind or had unclear blinding status was 63% (IQR 22% to 138%).

Downweighting of all trials with inadequate or unclear sequence generation led to a median 13% (IQR 5% to 32%) increase in the variance of the summary (meta-analytic) intervention effect estimate among informative meta-analyses in the BRANDO database. This is in contrast to a

median increase of 217% (IQR 87% to 482%) when such trials are excluded from meta-analyses, because only 26% of trials were assessed to have adequate sequence generation. Bias adjustment for meta-analyses with subjectively assessed outcome measures led to a median 31% (IQR 11% to 56%) increase in the variance of the summary intervention effect estimate, which was again small compared with complete exclusion of such trials. Results were broadly similar for the other study design characteristics, although differences between the effects of downweighting and excluding trials at high or unclear risk of bias were smaller for double blinding, because 56% of trials from informative meta-analyses were double blind.

## Conclusions

Bias associated with specific reported study design characteristics of RCTs leads to exaggeration of intervention effect estimates and increases in between-trial heterogeneity. For each of the three characteristics assessed, these effects appeared greatest for subjectively assessed outcome measures. Assessments of the risk of bias in trial results should account for these findings. Downweighting trials at high risk of bias in future meta-analyses, based on these empirical findings, could be an alternative to completely excluding such trials from meta-analyses, resulting in a smaller loss of precision. The strategies used to combine data from several independent data sources and to remove overlap between meta-analyses may be of use for future empirical research.

## Recommendations for future research

- Tools for assessing risk of bias in results of RCTs should account for the findings of this study.
- Practical and acceptable methods for correcting and downweighting the results of trials at high risk of bias in new meta-analyses should be developed.
- The influence of further study design characteristics should be explored in new meta-epidemiological studies.
- As far as possible, clinical decisions should not be based on trials in which blinding is not feasible and outcome measures are subjectively assessed.

## Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

## Publication

Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, *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).

# NIHR Health Technology Assessment programme

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needed in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

## Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The research reported in this issue of the journal was commissioned by the National Coordinating Centre for Research Methodology (NCCRM), and was formally transferred to the HTA programme in April 2007 under the newly established NIHR Methodology Panel. The HTA programme project number is 06/91/10. The contractual start date was in October 2005. The draft report began editorial review in March 2011 and was accepted for publication in August 2011. The commissioning brief was devised by the NCCRM who specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA programme or the Department of Health.

Editor-in-Chief: Professor Tom Walley CBE

Series Editors: Dr Martin Ashton-Key, Professor Aileen Clarke, Dr Peter Davidson,  
Dr Tom Marshall, Professor William McGuire, Professor John Powell, Dr Rob Riemsmma,  
Professor Helen Snooks and Professor Ken Stein

Editorial Contact: edit@southampton.ac.uk

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

ISSN 2046-4932 (DVD)

**© Queen's Printer and Controller of HMSO 2012. This work was produced by Savović *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to NETSCC.**

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (<http://www.publicationethics.org/>).

This journal may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NETSCC, Health Technology Assessment, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by Prepress Projects Ltd, Perth, Scotland ([www.prepress-projects.co.uk](http://www.prepress-projects.co.uk)), on behalf of NETSCC, HTA.

Printed on acid-free paper in the UK by Charlesworth Press.