Recent developments for combining evidence within evidence streams: bias-adjusted meta-analysis

Julian Higgins

University of Bristol

Outline

- Introduction to concepts
- Standard approaches to dealing with bias in meta-analysis
  - assessment tools
  - narrative summary of study limitations
  - stratification / sensitivity analysis
- Approaches to bias adjustment
  - weighting
  - regression
  - direct adjustments
  - prior distributions for bias
  - triangulation
- Concluding remarks

Concepts (1)

- It is important to determine the extent to which results of the included studies can be believed
- We do this by assessing **risk of bias**, which is not the same as...

Imprecision

- random error due to sampling variation
- reflected in the confidence interval

Quality

- bias can occur in well-conducted studies
- not all methodological flaws introduce bias

Reporting

- good methods may have been used but not well reported

University of BRISTOL

Concepts (2)

- RoB assessment facilitated by considering each study as an attempt to mimic a high quality **hypothetical experiment** examining the exposures of interest
  - “Target experiment”
  - Need not be feasible or ethical

Traditional approaches to addressing risk of bias

Assessing the Quality of Randomized Controlled Trials: An Annotated Bibliography of Sca

David M
Graham
Peter Tu
Clinical Epid
Department
Relief Unit, U
of Medicine,

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2007;**36**:666–676
doi:10.1093/ije/dym018

Advance Access publication 30 April 2007

Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography

Simon Sanderson,[1]* Iain D Tatt[2,4] and Julian PT Higgins[3]

University of BRISTOL

Example: narrative summary

Table 3: Methodological assessment of eligible studies

CAN_039 Pesticides

8.1.1. Leukemias

Overall, 26 studies (and 2 abstracts) examined associations between pesticide exposure and various forms of leukaemia. Fourteen out of these 26 studies were reports from the AHS with some overlapping results and examination of different pesticide groups. Only 2 studies, both on DDE (ID CAN_063, ID CAN_064) examined residential exposure and all the remaining studies examined occupation exposure to pesticides. Twelve out of 99 different analyses were statistically significant with effect sizes across all studies ranging between 6.1 and 0.2. Statistically significant results come from 7 different studies; with the exception of the AHS all were of modest to low quality. Table 7 shows summarised results across studies that reported information on the same pesticide class. The vast majority of results are non-significant and of small effect sizes. Figure 8 shows random effect meta-analyses keeping analyses with largest sample size form each study. The meta-analysis resulted in a non-significant pooled effect (OR 1.26, 95% CI 0.93, 1.71) and had modest heterogeneity. Previous meta-analyses on occupational exposure to pesticides and leukaemia were published in 2008 and 2007 (Merhi 2007, Van Maele-Fabry 2008). The overall summary effect estimates from previous meta-analyses suggested that there is a significantly positive, albeit weak, association between occupational exposure to pesticides and all hematopoietic cancers. But both reports acknowledged a wide range of limitations including the lack of sufficient data about exposure information and other risk factors for hematopoietic cancer and unclear definition of exposure and of leukemia type.

University of BRISTOL

Modern generation of tools to assess risk of bias

BMJ

RESEARCH METHODS & REPORTING

The Cochrane Collaboration's tool for assessing risk of bias in randomised trials

Julian P T Higg
Jelena Savović
Cochrane Stat

RESEARCH METHODS AND REPORTING

ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions

thebmj

Jonathan AC Sterne,[1] Miguel A Hernán,[2] Barnaby C Reeves,[3] Jelena Savović,[1,4] Nancy D Berkman,[5] Meera Viswanathan,[6] David Henry,[7] Douglas G Altman,[8] Mohammed T Ansari,[9] Isabelle Boutron,[10] James R Carpenter,[11] An-Wen Chan,[12] Rachel Churchill,[13] Jonathan J Deeks,[14] Asbjørn Hróbjartsson,[15] Jamie Kirkham,[16] Peter Jüni,[17] Yoon K Loke,[18] Theresa D Pigott,[19] Craig R Ramsay,[20] Deborah Regidor,[21] Hannah R Rothstein,[22] Lakhbir Sandhu,[23] Pasqualina L Santaguida,[24] Holger J Schünemann,[25] Beverly Shea,[26] Ian Shrier,[27] Peter Tugwell,[28] Lucy Turner,[29] Jeffrey C Valentine,[30] Hugh Waddington,[31] Elizabeth Waters,[32] George A Wells,[33] Penny F Whiting,[34] Julian PT Higgins[35]

Key tools for exposure-outcome studies

- ROBINS-E
  - development ongoing

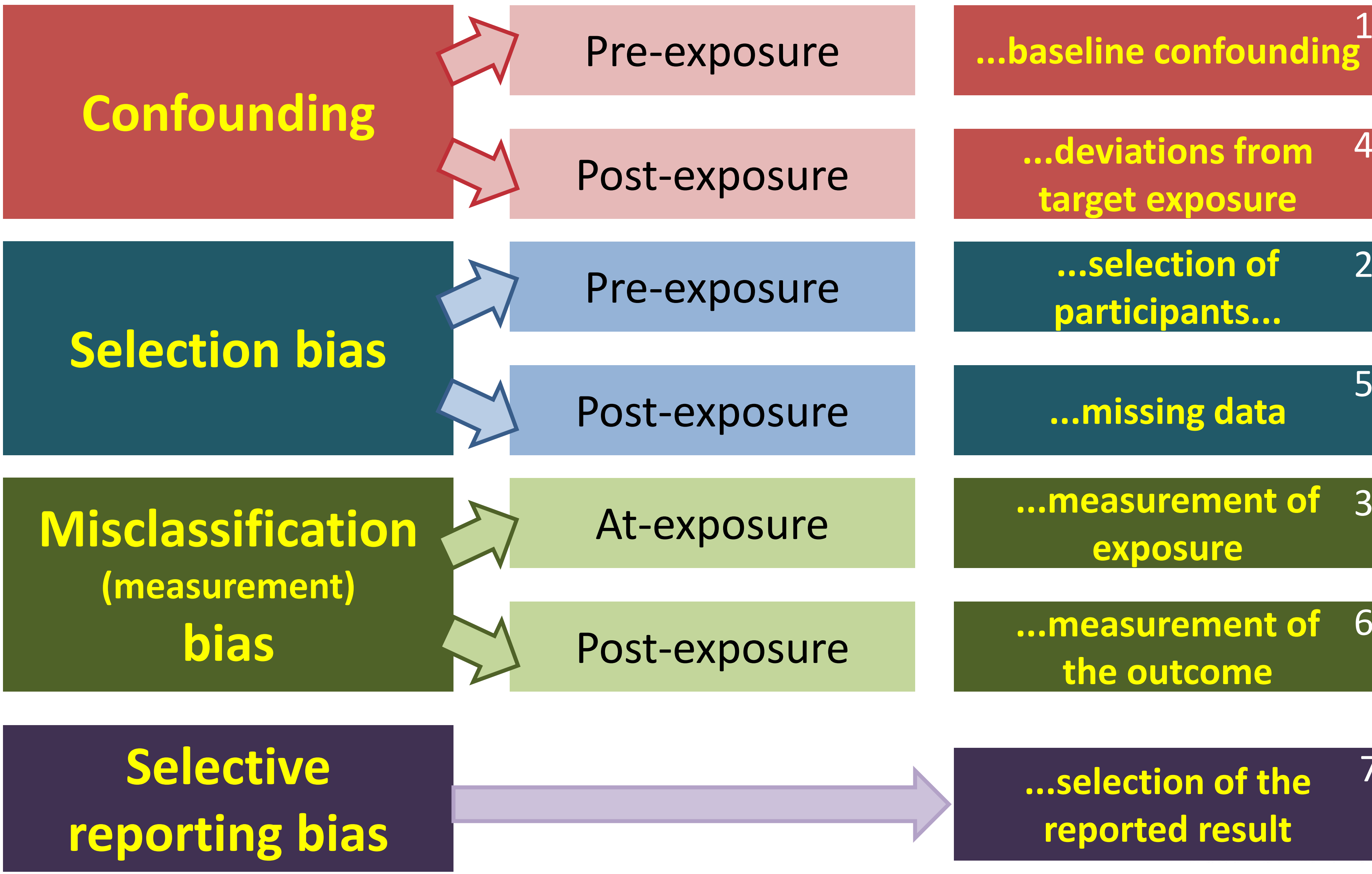
- OHAT/NTP integrated tool

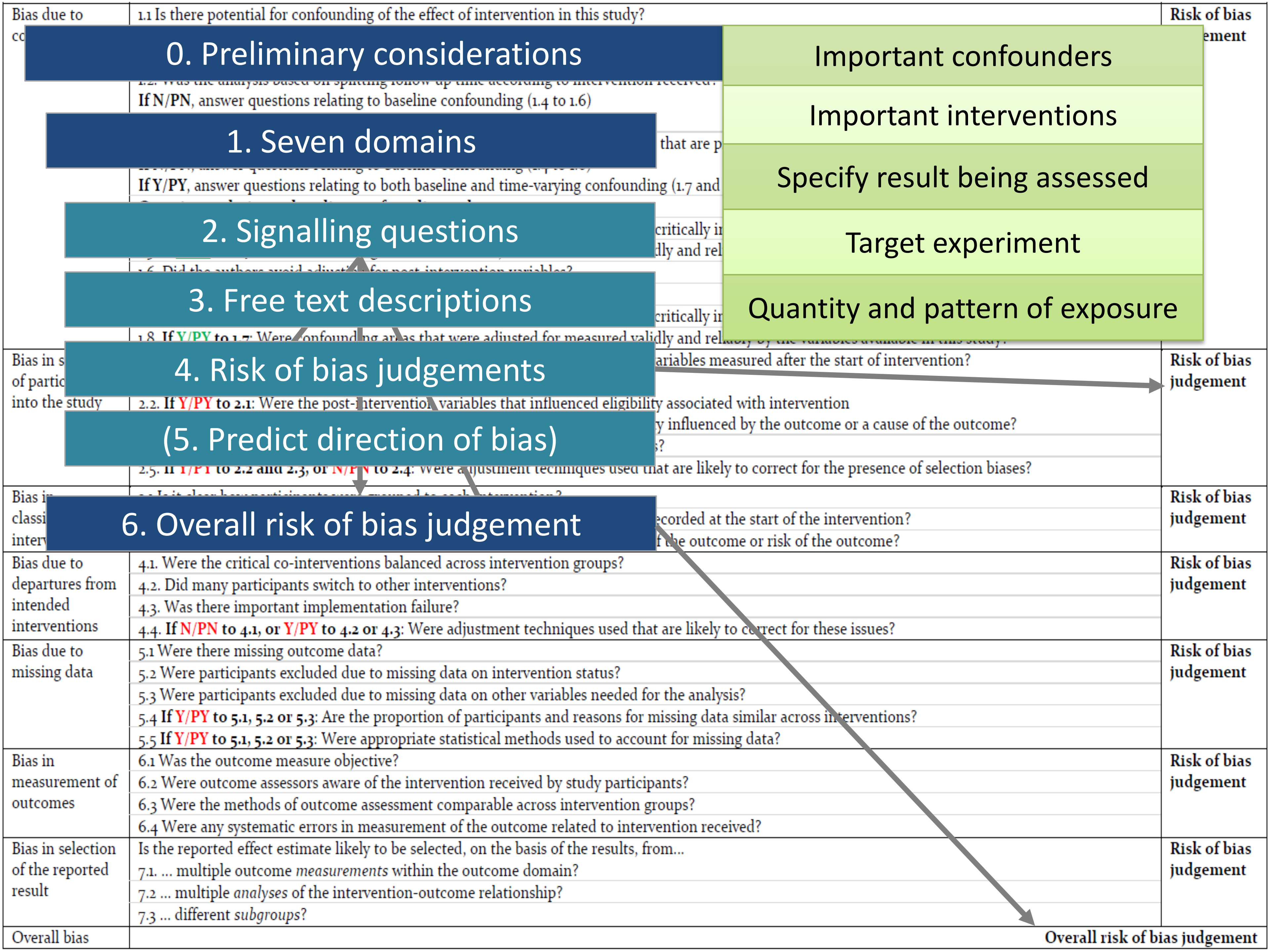
National Toxicology Program
U.S. Department of Health and Human Services

Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration

University of BRISTOL

Issues covered by ROBINS



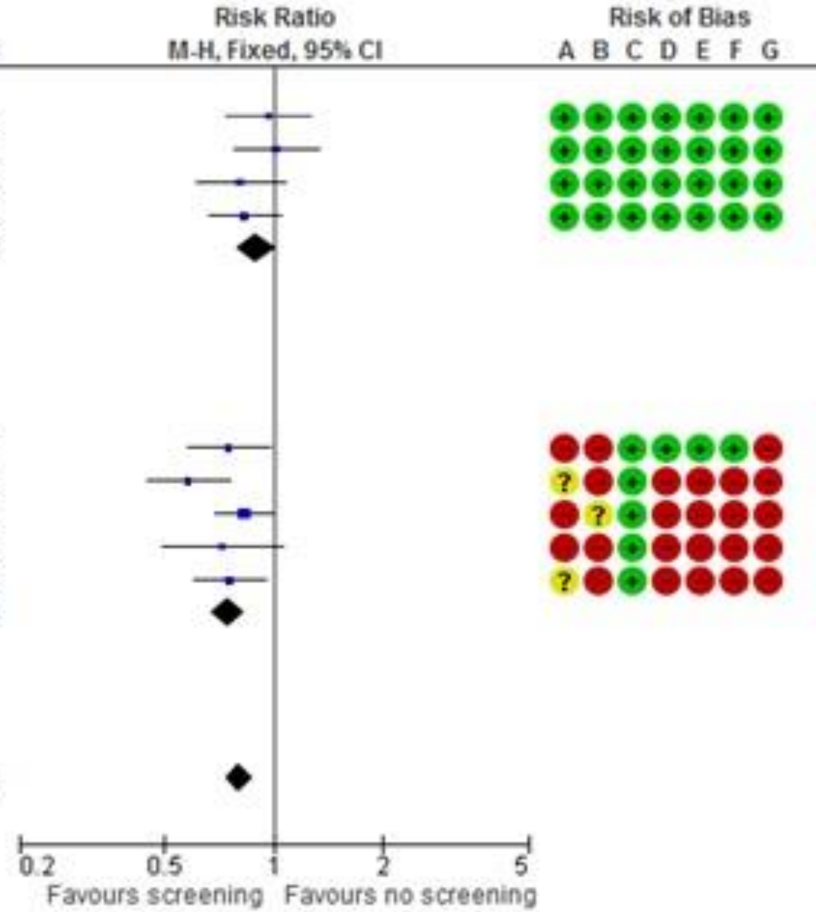


0. Preliminary considerations
Important confounders
Important interventions
Specify result being assessed
Target experiment
Quantity and pattern of exposure
1. Seven domains
2. Signalling questions
3. Free text descriptions
4. Risk of bias judgements
(5. Predict direction of bias)
6. Overall risk of bias judgement
Bias due to
1.1 Is there potential for confounding of the effect of intervention in this study?
If N/PN, answer questions relating to baseline confounding (1.4 to 1.6)
If Y/PY, answer questions relating to both baseline and time-varying confounding (1.7 and
Risk of bias judgement
ariables measured after the start of intervention?
2.2. If Y/PY to 2.1: Were the post-intervention variables that influenced eligibility associated with intervention
y influenced by the outcome or a cause of the outcome?
ecorded at the start of the intervention?
e outcome or risk of the outcome?
Bias due to departures from intended interventions
4.1. Were the critical co-interventions balanced across intervention groups?
4.2. Did many participants switch to other interventions?
4.3. Was there important implementation failure?
4.4. If N/PN to 4.1, or Y/PY to 4.2 or 4.3: Were adjustment techniques used that are likely to correct for these issues?
Bias due to missing data
5.1 Were there missing outcome data?
5.2 Were participants excluded due to missing data on intervention status?
5.3 Were participants excluded due to missing data on other variables needed for the analysis?
5.4 If Y/PY to 5.1, 5.2 or 5.3: Are the proportion of participants and reasons for missing data similar across interventions?
5.5 If Y/PY to 5.1, 5.2 or 5.3: Were appropriate statistical methods used to account for missing data?
Bias in measurement of outcomes
6.1 Was the outcome measure objective?
6.2 Were outcome assessors aware of the intervention received by study participants?
6.3 Were the methods of outcome assessment comparable across intervention groups?
6.4 Were any systematic errors in measurement of the outcome related to intervention received?
Bias in selection of the reported result
Is the reported effect estimate likely to be selected, on the basis of the results, from...
7.1. ... multiple outcome measurements within the outcome domain?
7.2 ... multiple analyses of the intervention-outcome relationship?
7.3 ... different subgroups?
Overall bias
Overall risk of bias judgement

University of BRISTOL

Example

Study or Subgroup	Screening Events	Screening Total	No screening Events	No screening Total	Weight	Risk Ratio M-H, Fixed, 95% CI
1.2.1 Adequately randomised trials						
Canada 1980a	105	25214	108	25216	8.6%	0.97 [0.74, 1.27]
Canada 1980b	107	19711	105	19694	8.3%	1.02 [0.78, 1.33]
Malmö 1976	87	20695	108	20783	8.5%	0.81 [0.61, 1.07]
UK age trial 1991	105	53884	251	106956	13.3%	0.83 [0.66, 1.04]
Subtotal (95% CI)		**119504**		**172649**	**38.7%**	**0.90 [0.79, 1.02]**
Total events	404		572			
Heterogeneity: Chi² = 2.16, df = 3 (P = 0.54); I² = 0%						
Test for overall effect: Z = 1.64 (P = 0.10)						
1.2.2 Suboptimally randomised trials						
Göteborg 1982	88	21650	162	29961	10.8%	0.75 [0.58, 0.97]
Kopparberg 1977	126	38589	104	18582	11.1%	0.58 [0.45, 0.76]
New York 1963	218	31000	262	31000	20.7%	0.83 [0.70, 1.00]
Stockholm 1981	66	40318	45	19943	4.8%	0.73 [0.50, 1.06]
Östergötland 1978	135	38491	173	37403	13.9%	0.76 [0.61, 0.95]
Subtotal (95% CI)		**170048**		**136889**	**61.3%**	**0.75 [0.67, 0.83]**
Total events	633		746			
Heterogeneity: Chi² = 4.94, df = 4 (P = 0.29); I² = 19%						
Test for overall effect: Z = 5.34 (P < 0.00001)						
Total (95% CI)		**289552**		**309538**	**100.0%**	**0.81 [0.74, 0.87]**
Total events	1037		1318			
Heterogeneity: Chi² = 11.82, df = 8 (P = 0.16); I² = 32%						
Test for overall effect: Z = 5.15 (P < 0.00001)						
Test for subgroup differences: Chi² = 4.55, df = 1 (P = 0.03), I² = 78.0%						



Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Bias-adjustment approaches

- **Weighting by quality**
- Regression approaches
- Direct adjustment
- Prior distributions for bias
- Triangulation approaches

University of BRISTOL

Historical arguments against weighting by quality

American Journal of Epidemiology
Copyright © 1994 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved
Vol 140, No 3
Printed in U S.A

Quality Scores Are Useless and Potentially Misleading

Reply to "Re: A Crit

Sander Greenland

Biostatistics (2001), **2**, 4, *pp.* 463–471
Printed in Great Britain

On the bias produced by quality scores in meta-analysis,

Department of Epidemiolog
College of Letters

Department of Surgery an

ORIGINAL CONTRIBUTION

The Hazards of Scoring the Quality of Clinical Trials for Meta-analysis

Peter Jüni, MD
Anne Witschi, MD
Ralph Bloch, MD, PhD
Matthias Egger, MD, MSc

Context Although it is widely recommended that clinical trials undergo some type of quality review, the number and variety of quality assessment scales that exist make it unclear how to achieve the best assessment.

Objective To determine whether the type of quality assessment scale used affects the conclusions of meta-analytic studies.

University of BRISTOL

ORIGINAL ARTICLE

A Quality-Effects Model for Meta-Analysis

Suhail A. R. Doi† and Lukman Thalib‡*

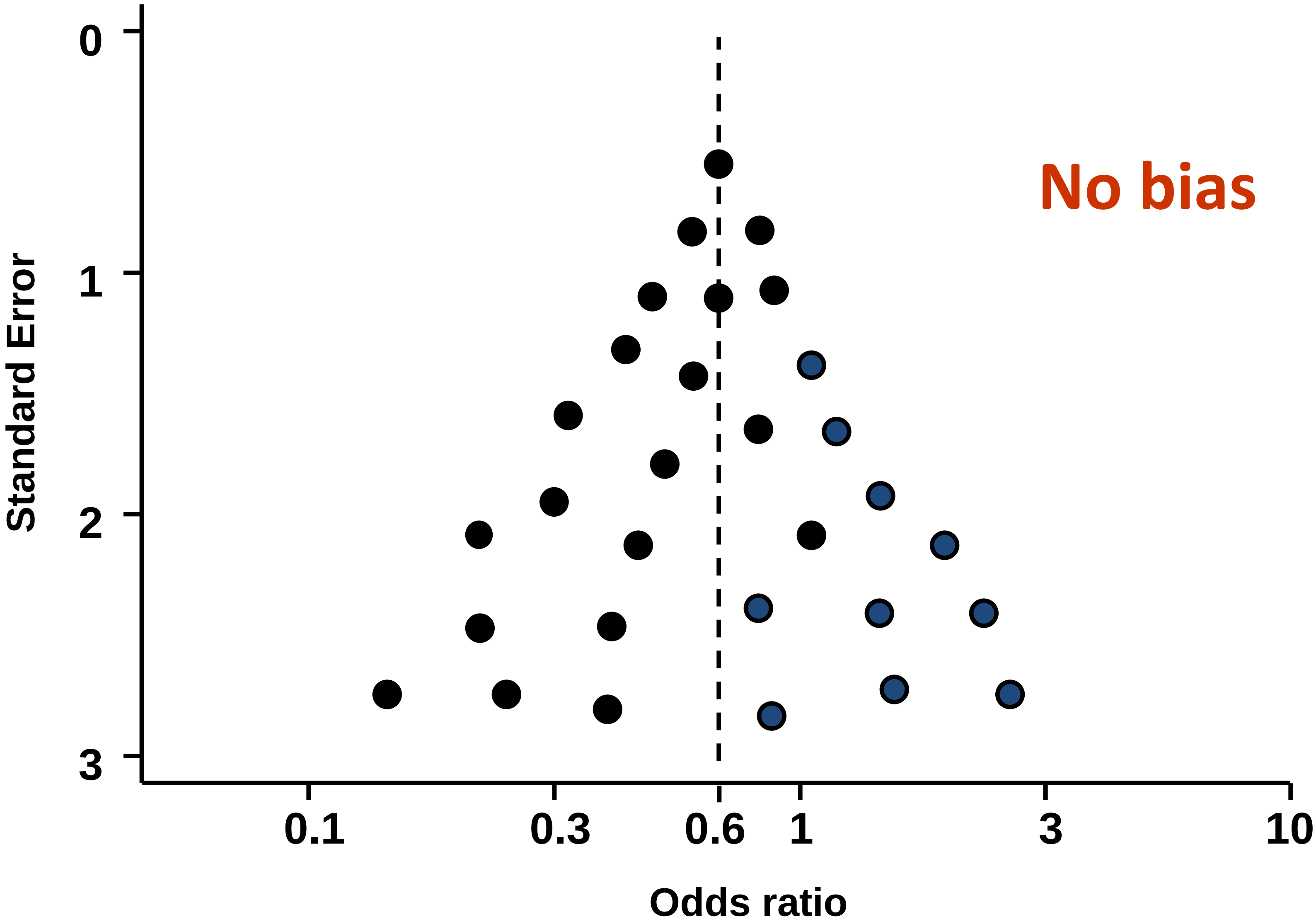
"For the QE model, the weighted estimator… has weights that are adjusted from inverse variance weights based on the additional variance contribution from internal study biases"

- Weighting by 'quality' or 'risk of bias' features indirectly adjusts for bias by shifting centre of mass towards the results of the 'better' studies

Bias-adjustment approaches

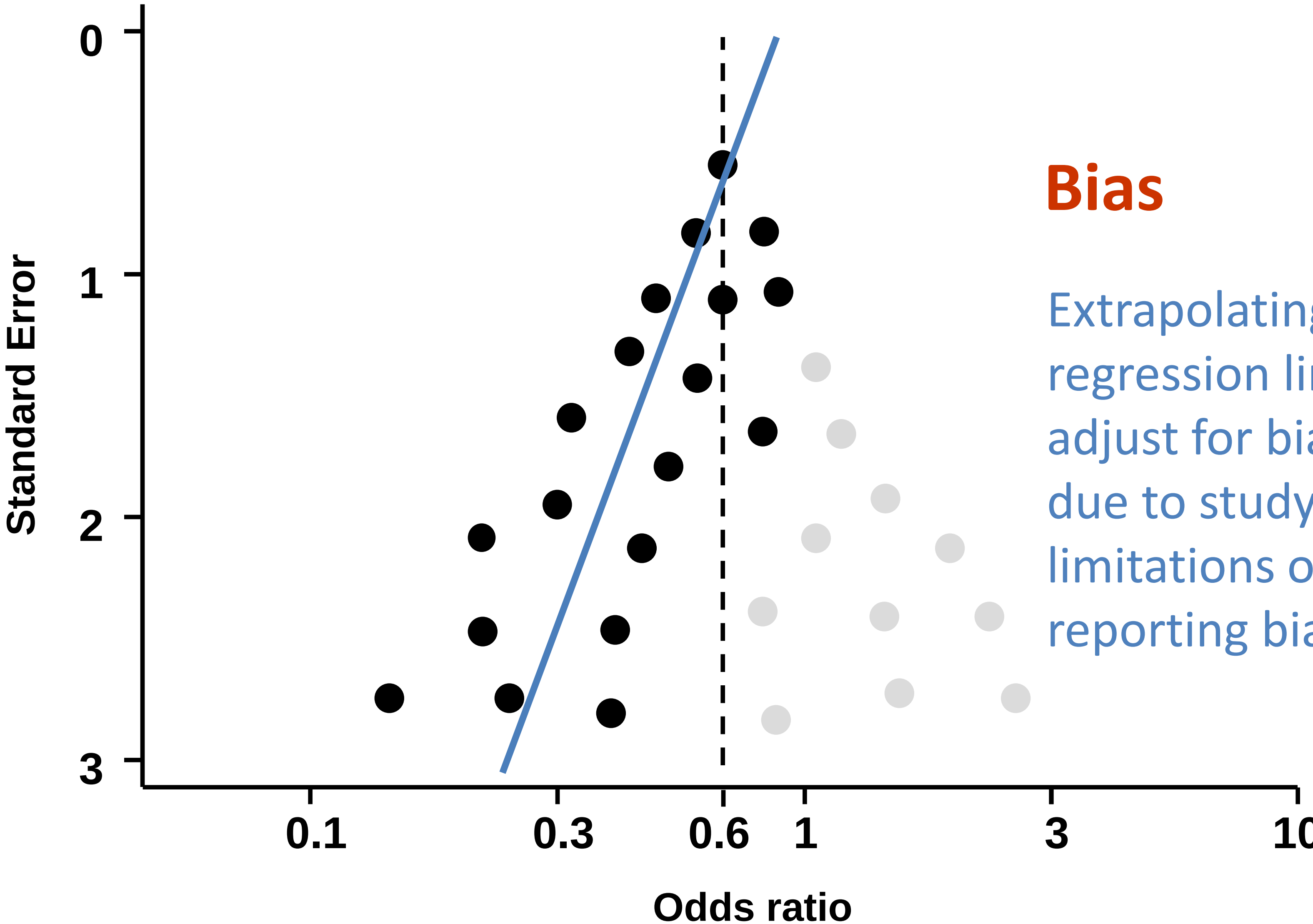
- Weighting by quality
- **Regression approaches**
- Direct adjustment
- Prior distributions for bias
- Triangulation approaches

Funnel plot: symmetrical



University of BRISTOL

Funnel plot: asymmetrical



Bias

Extrapolating this regression line may adjust for bias due to study limitations or reporting bias

Regression approaches

- Regression approaches may be used to extrapolate to various types of limit
  - Very large study (as previous slide)
  - Lowest risk-of-bias profile
  - Highest quality score
  - etc

Bias-adjustment approaches

- Weighting by quality
- Regression approaches
- **Direct adjustment**
- Prior distributions for bias
- Triangulation approaches

University of BRISTOL

Direct adjustments

- Bespoke adjustment according to type of bias
- Adjustment for missing data (e.g. “informative missingness parameters”)
- Example of adjustment for healthy worker effects

Pedeli *et al*. *Environmental Health* 2011, **10**:30

008; 5: 225–239

Review

Table 4 Summary results of the meta-analysis of benzene exposure and non-Hodgkins lymphoma (NHL) and meta-analysis of refinery work and NHL

	Fixed effects				Shore CI			Random effects			Heterogeneity		
	N	RR	CI_{low}	CI_{up}	RR	CI_{low}	CI_{up}	RR	CI_{low}	CI_{up}	χ^2	p	I^2 (%)
Benzene and NHL													
All studies	22	1.22	1.03	1.46	1.22	1.02	1.47	1.23	1.02	1.48	22.8	0.36	8
Case-control studies	16	1.23	1.00	1.50	1.23	0.99	1.52	1.21	0.97	1.51	16.8	0.33	12
Cohort studies	6	1.21	0.86	1.71	1.21	0.83	1.77	1.34	0.86	2.09	5.9	0.31	16
High exposure studies													
All	13	1.49	1.15	1.92	1.49	1.12	1.97	1.49	1.09	2.04	14.9	0.25	20
No self-reported data	6	2.12	1.11	4.02	na*	na	na	na	na	na	2.1	0.83	0
Healthy worker effect adjusted													
Cohort studies	6	1.22	0.89	1.67	1.22	0.80	1.85	1.54	0.92	2.59	8.6	0.13	42
All studies (cohort and case control)	22	1.22	1.03	1.45	1.22	1.02	1.48	1.24	1.01	1.51	25.4	0.23	17
All high exposure studies	13	1.53	1.19	1.96	1.53	1.15	2.03	1.55	1.14	2.12	15.8	0.20	24
High exposure, no self-reported data	6	2.26	1.29	3.97	na*	na	na	na	na	na	2.1	0.83	0

Bias-adjustment approaches

- Weighting by quality
- Regression approaches
- Direct adjustment
- **Prior distributions for bias**
- Triangulation approaches

University of BRISTOL

Sources of prior distributions

Journal of the Royal Statistical S[...]
SERIES A Statistics [...]
J. R. Statist. Soc. A (2009)
172, *Part* 1, *pp.* 21–47

Bias modell[...]

Rebecca M. Turne[...]
Medical Research C[...]
Gordon C. S. Smit[...]
University of Cambr[...]
and Simon G. Tho[...]
Medical Research C[...]

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

Meta-epidemiology

Review

Empirical Evidence of Bias

Dimensions of Methodological Quality Associated With Estimates of Treatment Effects in Controlled Trials

Kenneth F. Schulz, PhD, MBA; Iain Chalmers, MBBS, MSc; Richard J. Hayes, MSc; Douglas G. Altman

Objective.—To determine if inadequate approaches to randomized controlled trial design and execution are associated with evidence of bias in estimating treatment effects.

Design.—An observational study in which we assessed the methodological quality of 250 controlled trials from 33 meta-analyses and then analyzed, using multiple logistic regression models, the associations between those assessments and estimated treatment effects.

Data Sources.—Meta-analyses from the Cochrane Pregnancy and Childbirth Database.

Main Outcome Measures.—The associations between estimates of treatment effects and inadequate allocation concealment, exclusions after randomization, and lack of double-blinding.

Results.—Compared with trials in which authors reported adequately concealed treatment allocation, trials in which concealment was either inadequate or unclear (did not report or incompletely reported a concealment approach) yielded larger estimates of treatment effects ($P<.001$). Odds ratios were exaggerated by 41% for inadequately concealed trials and by 30% for unclearly concealed trials (adjusted for other aspects of quality). Trials in which participants had been excluded after randomization did not yield larger estimates of effects, but that lack of association may be due to incomplete reporting. Trials that were not double-blind also yielded larger estimates of effects ($P=.01$), with odds ratios being exaggerated by 17%.

Conclusions.—This study provides empirical evidence that inadequate methodological approaches in controlled trials, particularly those representing poor allocation concealment, are associated with bias. Readers of trial reports should be wary of these pitfalls, and investigators must improve their design, execution, and reporting of trials.

(*JAMA*. 1995;273:408-412)

ditionally, they suspected that methodologically inferior trials might produce bias in both directions, thereby causing greater variability in estimates of treatment effects. In neither analysis, however, did they detect a relationship.

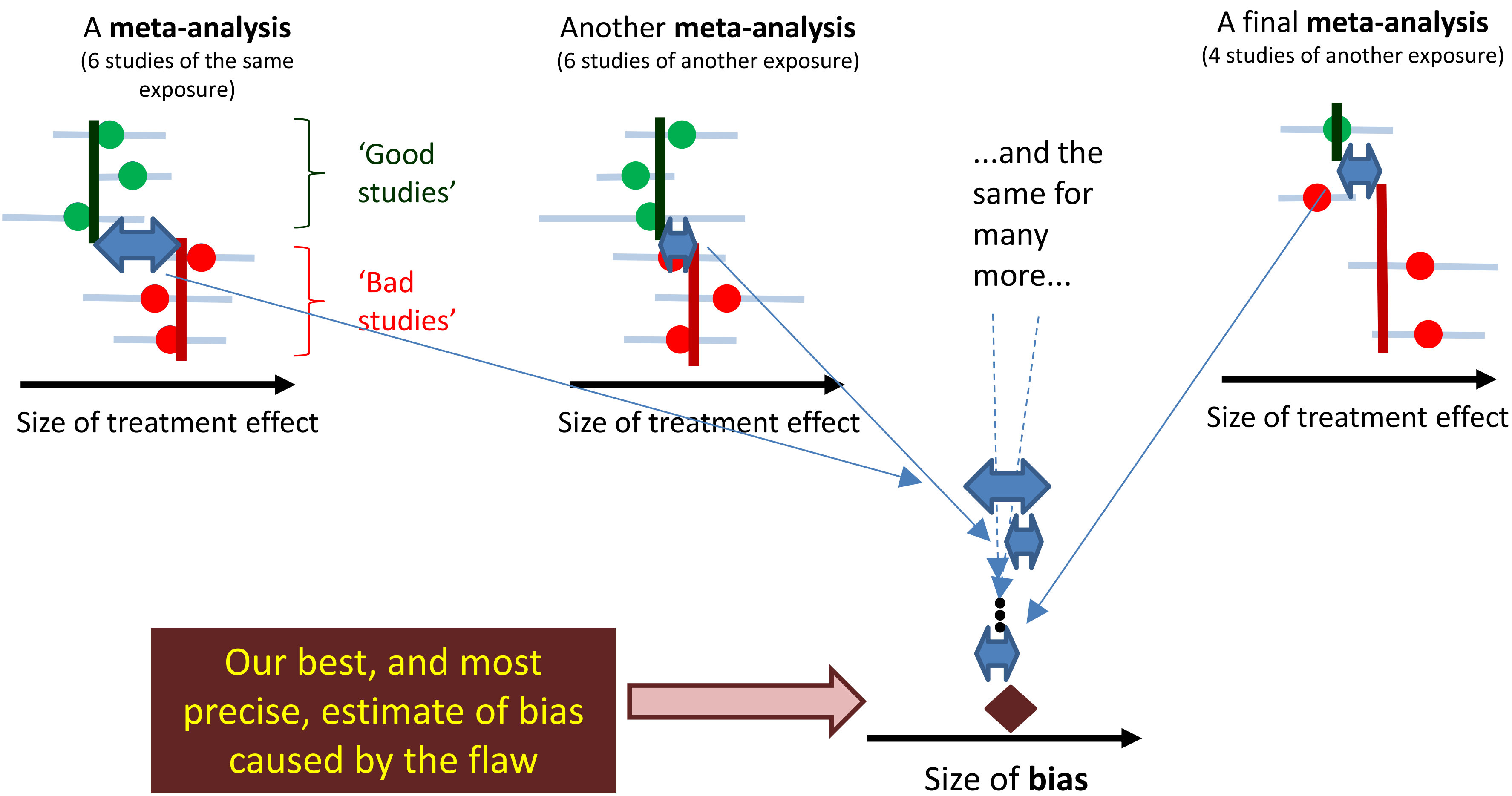
Using a database of systematic reviews of controlled trials in pregnancy and childbirth,[12] we sought evidence of bias related to use of inadequate methodological approaches to trial design and execution. Rather than using quality scores, we investigated specific aspects that we believed might be influential.[13] We hypothesized that estimates of treatment effects would be larger in trials in which (1) adequate measures had not been taken to conceal treatment allocation; (2) adequate measures had not been taken to generate the allocation schedule; (3) some allocated participants had been excluded from the analysis; and (4) measures had not been taken to implement double-blinding. Furthermore, we examined whether treatment effects varied more in trials in which allocation schedules had not been adequately concealed.

MATERIALS AND METHODS

1995: First **meta-epidemiological** study, based on 250 clinical trials:

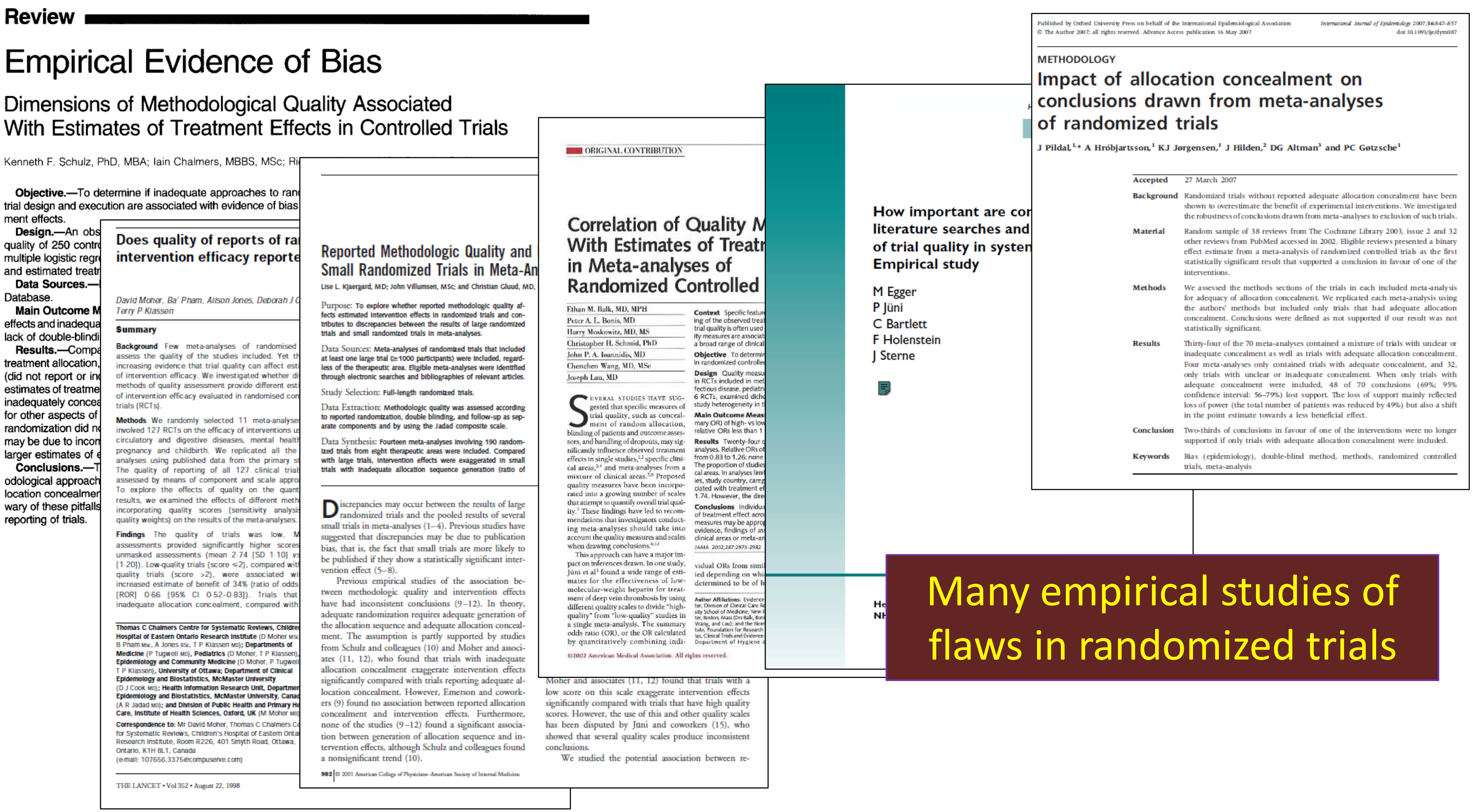
- Treatment effects exaggerated by **41%** in studies with inadequate concealment of allocation
- Treatment effects exaggerated by **17%** if studies not 'double-bind'

What is a meta-epidemiological study?



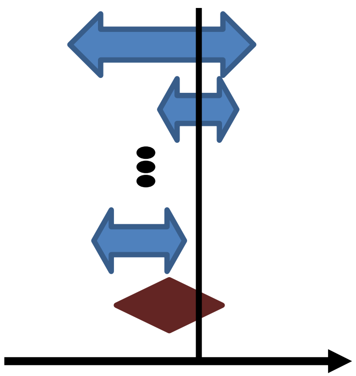
University of BRISTOL

And since that first study …

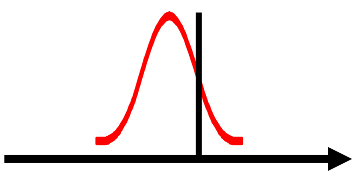


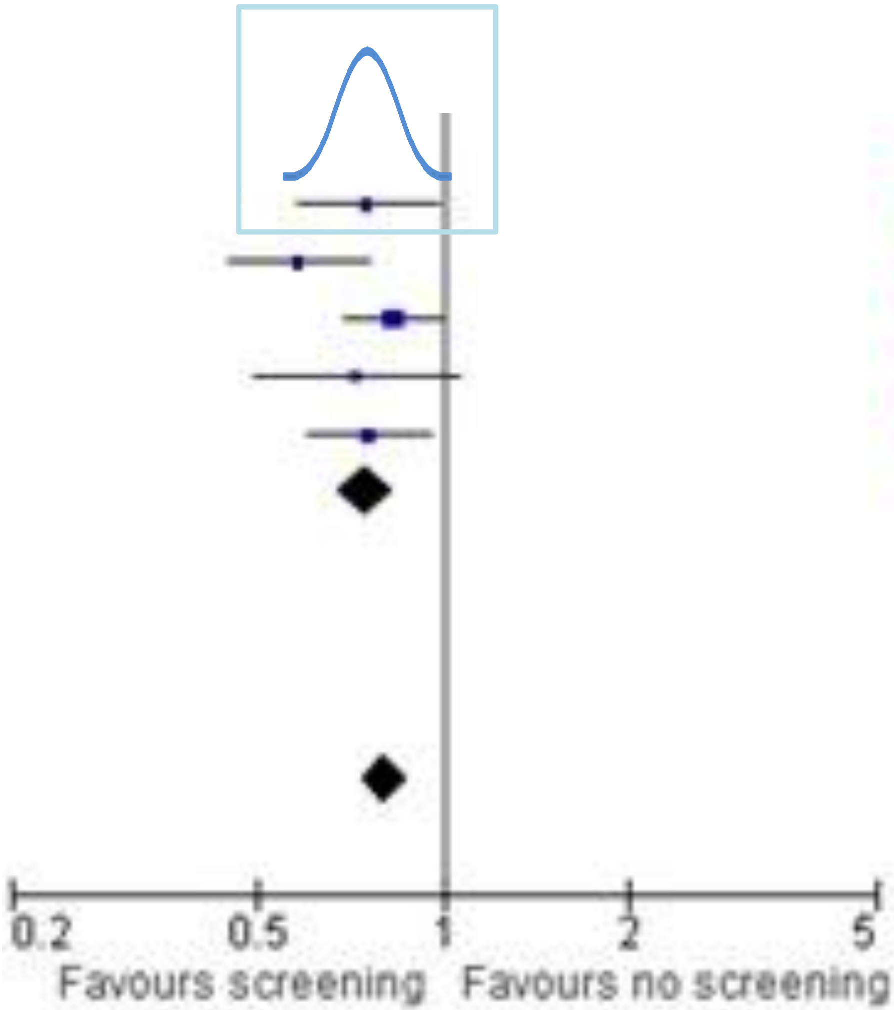
Many empirical studies of flaws in randomized trials

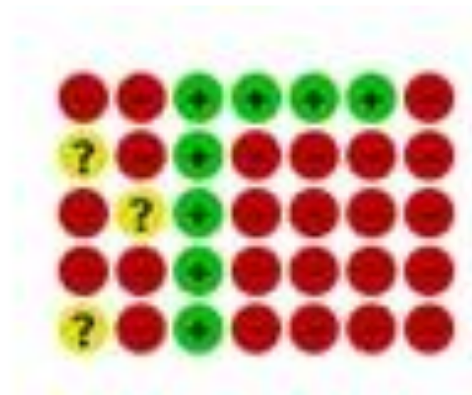
Using prior distributions for bias

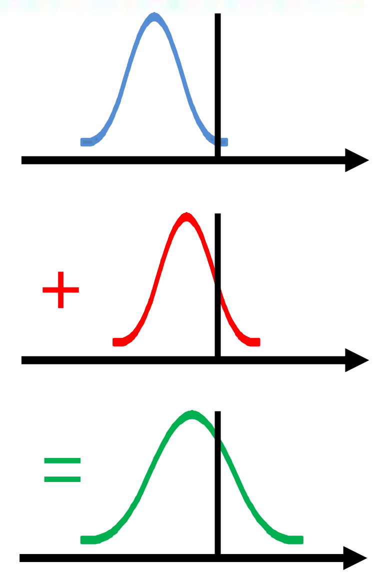


Size of **bias**



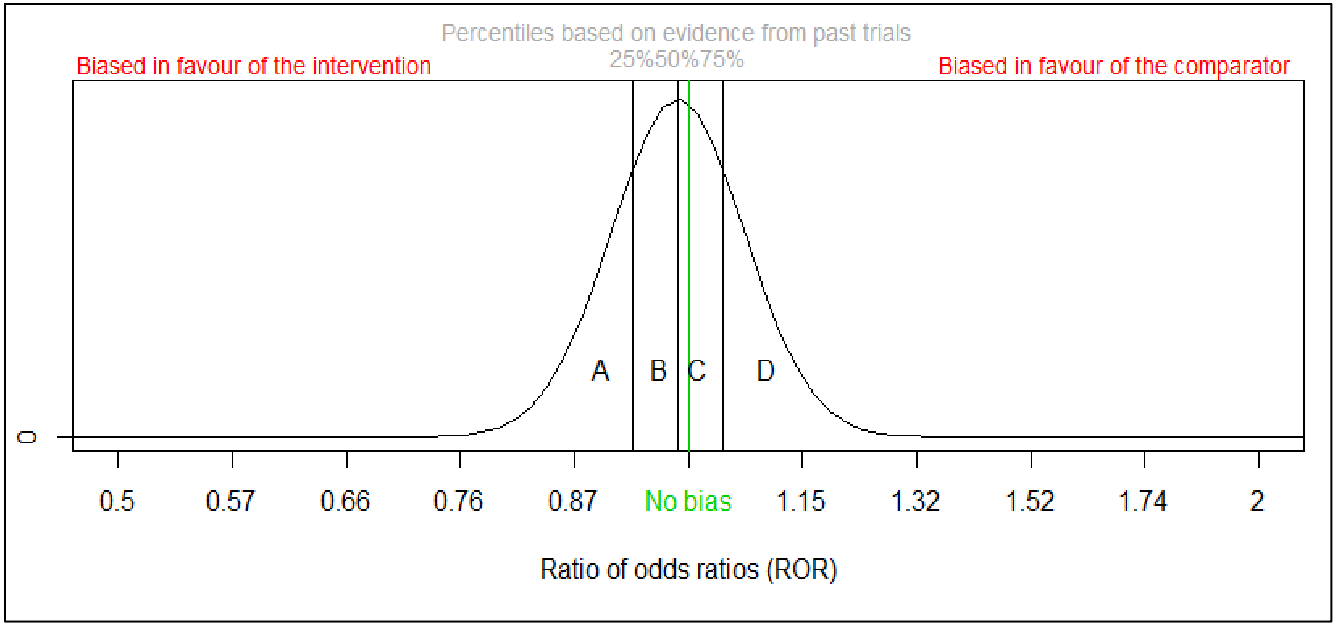






Combining the approaches

- An on-going MRC-funded project is exploring the combination of opinion-based and data-based priors
  - agreement between data and opinion is good for some domains
  - piloting data-informed elicitation proving successful



Bias-adjustment approaches

- Weighting by quality
- Regression approaches
- Direct adjustment
- Prior distributions for bias
- **Triangulation approaches**

Triangulation approaches

- General idea:
  - Use internal structure of data to estimate biases and adjust for them simultaneously
  - Network meta-analysis approach...
    - combines direct and indirect source of evidence on the same comparison

Journal of the Royal Statistical Society
SERIES A Statistics in Society

J. R. Statist. Soc. A (2010)
173, *Part* 4, *pp.*

Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis

S. Dias and N. J. Welton,
University of Bristol, UK

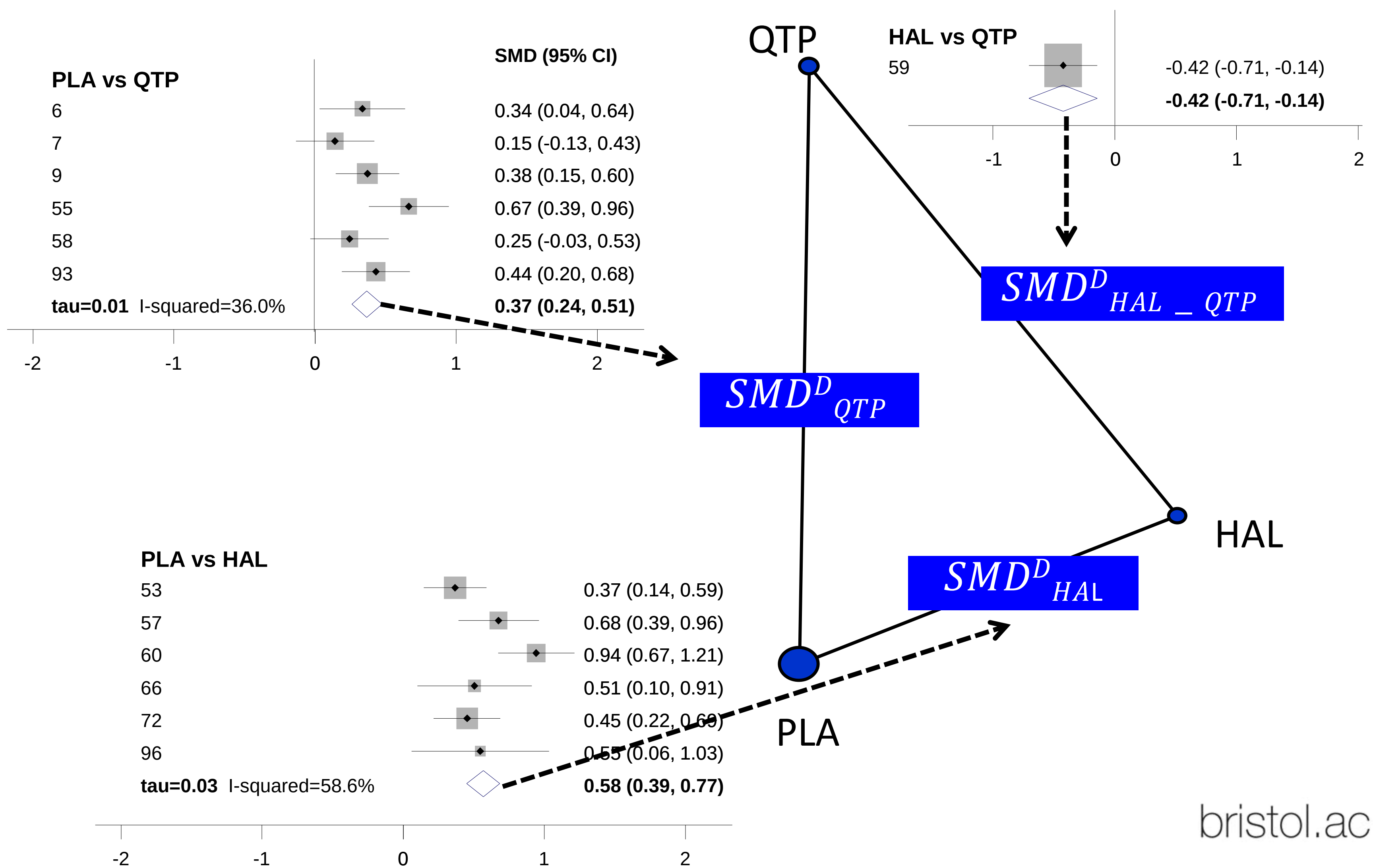
V. C. C. Marinho,
Queen Mary University of London, UK

G. Salanti,
University of Ioannina School of Medicine, Greece

J. P. T. Higgins
Medical Research Council Biostatistics Unit, Cambridge, UK

and A. E. Ades
University of Bristol, UK

Basic idea of network meta-analysis



Triangulation approaches

- General idea for network meta-analysis approach (ctd):
  - Assume bias is similar in all studies across the network
    - Triangle holds within ‘good’ and within ‘bad’ studies
  - Then can estimate bias as well as (adjusted) treatment effects

- Another possibility: multivariate meta-analysis to address missing results

Concluding remarks

- Numerous methods are available for attempting to adjust for bias in evidence synthesis
  - targeting each study individually
  - or targeting the body of evidence
- Informed by different things
  - assumptions
  - opinions
  - empirical evidence
- Bias-adjustment methods are appropriate also for
  - combining evidence **across evidence streams**
  - hazard characterization
- Some methods allow learning about biases; other don't

ROBINS-I

- Supplementary slides

Signalling questions

- e.g.
  - "Did the authors use an appropriate analysis method that controlled for all the important confounding domains?"
  - "Were outcome data available for all, or nearly all, participants?"

Yes
Probably yes
Probably no
No
No information

University of BRISTOL

Risk of bias judgements

Response option	Interpretation
Low risk of bias	The study is comparable to a well-performed randomized trial with regard to this bias domain.
Moderate risk of bias	The study is sound for a non-randomized study with regard to this bias domain but cannot be considered comparable to a well-performed randomized trial.
Serious risk of bias	The study has some important problems in this domain of bias.
Critical risk of bias	The study is too problematic in this domain of bias to provide any useful evidence.
No information	No information on which to base a judgement about risk of bias for this domain.

Overall risk of bias judgement

Low risk of bias	The study is judged to be at **low risk of bias for all domains** (for the result).
Moderate risk of bias	The study is judged to be at **low or moderate risk of bias for all domains** (for the result).
Serious risk of bias	The study is judged to be at **serious risk of bias** in at least one domain, but not at critical risk of bias in any domain.
Critical risk of bias	The study is judged to be at **critical risk of bias in at least one domain** (for the result).
No information	There is no clear indication that the study is at serious or critical risk of bias *and* there is a lack of information in one or more key domains of bias (*a judgement is required for this*).

University of BRISTOL

Bias due to confounding

Signalling Questions	Rationale
1.1 Is there potential for confounding of the effect of intervention in this study? If N/PN to 1.1: the study can be considered to be at low risk of bias due to confounding and no further signalling questions need be considered	In rare situations, such as when studying harms that are very unlikely to be related to factors that influence treatment decisions, no confounding is expected and the study can be considered to be at low risk of bias due to confounding, equivalent to a fully randomized trial. There is no NI (No information) option for this signalling question.
If Y/PY to 1.1: determine whether there is a need to assess time-varying confounding:	
1.2. Was the analysis based on splitting participants' follow up time according to intervention received? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, go to question 1.3.	If participants could switch between intervention groups then associations between intervention and outcome may be biased by time-varying confounding. This occurs when prognostic factors influence switches between intended interventions.
1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome? **If N/PN**, answer questions relating to baseline confounding (1.4 to 1.6) **If Y/PY**, answer questions relating to both baseline and time-varying confounding (1.7 and 1.8)	If intervention switches are unrelated to the outcome, for example when the outcome is an unexpected harm, then time-varying confounding will not be present and only control for baseline confounding is required.

Bias due to confounding

Signalling Questions	Rationale
Questions relating to baseline confounding only	
1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	Appropriate methods to control for measured confounders include stratification, regression, matching, standardization, and inverse probability weighting. They may control for individual variables or for the estimated propensity score. Inverse probability weighting is based on a function of the propensity score. Each method depends on the assumption that there is no unmeasured or residual confounding.
1.5. If Y/PY to 1.4: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Appropriate control of confounding requires that the variables adjusted for are valid and reliable measures of the confounding domains. For some topics, a list of valid and reliable measures of confounding domains will be specified in the review protocol but for others such a list may not be available. Study authors may cite references to support the use of a particular measure. If authors control for confounding variables with no indication of their validity or reliability pay attention to the subjectivity of the measure. Subjective measures (e.g. based on self-report) may have lower validity and reliability than objective measures such as lab findings.
1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	Controlling for post-intervention variables that are affected by intervention is not appropriate. Controlling for mediating variables estimates the direct effect of intervention and may introduce bias. Controlling for common effects of intervention and outcome introduces bias.

Bias due to confounding

Signalling Questions	Rationale
Questions relating to baseline and time-varying confounding	
1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Adjustment for time-varying confounding is necessary to estimate the effect of starting and adhering to intervention, in both randomized trials and NRSI. Appropriate methods include those based on inverse probability weighting. Standard regression models that include time-updated confounders may be problematic if time-varying confounding is present.
1.8. If Y/PY to 1.7: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	See 1.5 above.

Risk of bias judgements

- For each domain, there is guidance on how to judge risk of bias based on the answers to the signalling questions

Low risk of bias (the study is comparable to a well-performed randomized trial with regard to this domain)	No confounding expected.
Moderate risk of bias (the study is sound for a non-randomized study with regard to this domain but cannot be considered	(i) Confounding expected, all known important confounding domains appropriately measured and controlled for; *and*
	or (ii) Reliability or validity of measurement of an important domain was low enough that we expect serious residual confounding.
Critical risk of bias (the study is too problematic to provide any useful evidence on the effects of intervention)	(i) Confounding inherently not controllable *or* (ii) The use of negative controls strongly suggests unmeasured confounding.
No information on which to base a judgement about risk of bias for this domain	No information on whether confounding might be present.

It is usually impossible to exclude bias due to residual or unmeasured confounding of the results of an non-randomized study. **We expect few NRSI to be assessed as at low risk of bias due to confounding**

Bias in selection of participants into the study

<table>
<tr><th>Signalling Questions</th><th>Rationale</th></tr>
<tr><td>2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?
If <u>N/PN</u> to 2.1: go to 2.4</td><td rowspan="3">This domain is concerned only with selection into the study based on participant characteristics observed after the start of intervention. Selection based on characteristics observed before the start of intervention can be addressed by controlling for imbalances between experimental intervention and comparator groups in baseline characteristics that are prognostic for the outcome (baseline confounding).
Selection bias occurs when selection is related to an effect of either intervention or a cause of intervention and an effect of either the outcome or a cause of the outcome. Therefore, the result is at risk of selection bias if selection into the study is related to both the intervention and the outcome.</td></tr>
<tr><td>2.2. If Y/PY to 2.1.: Were the post-intervention variables that influenced selection likely to be associated with intervention?</td></tr>
<tr><td>2.3. If Y/PY to 2.2.: Were the post intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?</td></tr>
</table>

Bias in selection of participants into the study

Signalling Questions	Rationale
2.4. Do start of follow-up and start of intervention coincide for most participants?	If participants are not followed from the start of the intervention then a period of follow up has been excluded, and individuals who experienced the outcome soon after intervention will be missing from analyses. This problem may occur when prevalent, rather than new (incident), users of the intervention are included in analyses.
2.5. **If Y/PY to 2.2 and 2.3, or N/PN to 2.4**: Were adjustment techniques used that are likely to correct for the presence of selection biases?	It is in principle possible to correct for selection biases, for example by using inverse probability weights to create a pseudo-population in which the selection bias has been removed, or by modelling the distributions of the missing participants or follow up times and outcome events and including them using missing data methodology. However such methods are rarely used and the answer to this question will usually be "No".

University of BRISTOL

Risk of bias judgements

Low risk of bias	(i) All participants who would have been eligible for the target trial were included in the study; *and* (ii) For each participant, start of follow up and start of intervention coincided.
Moderate risk of bias	(i) Selection into the study may have been related to intervention and outcome; *and* The authors used appropriate methods to adjust for the selection bias; *or* (ii) Start of follow up and start of intervention do not coincide for all participants; *and* (a) the proportion of participants for which this was the case was too low to induce important bias; *or* (b) the authors used appropriate methods to adjust for the selection bias; *or* (c) the review authors are confident that the rate (hazard) ratio for the effect of intervention remains constant over time.

Risk of bias judgements

Serious risk of bias	(i) Selection into the study was related (but not very strongly) to intervention and outcome; *and* This could not be adjusted for in analyses; *or* (ii) Start of follow up and start of intervention do not coincide; *and* A potentially important amount of follow-up time is missing from analyses; *and* The rate ratio is not constant over time.
Critical risk of bias	(i) Selection into the study was very strongly related to intervention and outcome; *and* This could not be adjusted for in analyses; *or* (ii) A substantial amount of follow-up time is likely to be missing from analyses; *and* The rate ratio is not constant over time.
No information on which to base a judgement about risk of bias for this domain	No information is reported about selection of participants into the study or whether start of follow up and start of intervention coincide.

University of BRISTOL

Bias in classification of interventions

Signalling Questions	Rationale
3.1 Were intervention groups clearly defined?	A pre-requisite for an appropriate comparison of interventions is that the interventions are well defined. Ambiguity in the definition may lead to bias in the classification of participants. For individual-level interventions, criteria for considering individuals to have received each intervention should be clear and explicit, covering issues such as type, setting, dose, frequency, intensity and/or timing of intervention. For population-level interventions (e.g. measures to control air pollution), the question relates to whether the population is clearly defined, and the answer is likely to be 'Yes'.

University of BRISTOL

Bias in classification of interventions

Signalling Questions	Rationale
3.2 Was the information used to define intervention groups recorded at the start of the intervention?	In general, if information about interventions received is available from sources that could not have been affected by subsequent outcomes, then differential misclassification of intervention status is unlikely. Collection of the information at the time of the intervention makes it easier to avoid such misclassification. For population-level interventions (e.g. measures to control air pollution), the answer to this question is likely to be 'Yes'.
3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	Collection of the information at the time of the intervention may not be sufficient to avoid bias. The way in which the data are collected for the purposes of the NRSI should also avoid misclassification.

University of BRISTOL

Risk of bias judgements

Low risk of bias	(i) Intervention status is well defined; *and* (ii) Intervention definition is based solely on information collected at the time of intervention.
Moderate risk of bias	(i) Intervention status is well defined; *and* (ii) Some aspects of the assignments of intervention status were determined retrospectively.
Serious risk of bias	(i) Intervention status is not well defined; *or* (ii) Major aspects of the assignments of intervention status were determined in a way that could have been affected by knowledge of the outcome.
Critical risk of bias	(Unusual) An extremely high amount of misclassification of intervention status, e.g. because of unusually strong recall biases.
No information on which to base a judgement about risk of bias for this domain	No definition of the intervention or no explanation of the source of information about intervention status is reported.

University of BRISTOL

Bias due to deviations from intended interventions

	Signalling Questions	Rationale
	If your aim for this study is to assess the effect of assignment to intervention, answer questions 4.1 and 4.2	
	4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	Deviations that happen in usual practice following the intervention (for example, cessation of a drug intervention because of acute toxicity) are part of the intended intervention and therefore do not lead to bias in the effect of assignment to intervention. Deviations may arise due to expectations of a difference between intervention and comparator (for example because participants feel unlucky to have been assigned to the comparator group and therefore seek the active intervention, or components of it, or other interventions). Such deviations are not part of usual practice, so may lead to biased effect estimates. However these are not expected in observational studies of individuals in routine care.
	4.2. **If Y/PY to 4.1**: Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Deviations from intended interventions that do not reflect usual practice will be important if they affect the outcome, but not otherwise. Furthermore, bias will arise only if there is imbalance in the deviations across the two groups.

Bias due to deviations from intended interventions

	Signalling Questions	Rationale
	If your aim for this study is to assess the effect of starting and adhering to intervention, answer questions 4.3 to 4.6	
	4.3. Were important co-interventions balanced across intervention groups?	Risk of bias will be higher if unplanned co-interventions were implemented in a way that would bias the estimated effect of intervention. Co-interventions will be important if they affect the outcome, but not otherwise. Bias will arise only if there is imbalance in such co-interventions between the intervention groups. Consider the co-interventions, including any pre-specified co-interventions, that are likely to affect the outcome and to have been administered in this study. Consider whether these co-interventions are balanced between intervention groups.
	4.4. Was the intervention implemented successfully for most participants?	Risk of bias will be higher if the intervention was not implemented as intended by, for example, the health care professionals delivering care during the trial. Consider whether implementation of the intervention was successful for most participants.

Bias due to deviations from intended interventions

	Signalling Questions	Rationale
	4.5. Did study participants adhere to the assigned intervention regimen?	Risk of bias will be higher if participants did not adhere to the intervention as intended. Lack of adherence includes imperfect compliance, cessation of intervention, crossovers to the comparator intervention and switches to another active intervention. Consider available information on the proportion of study participants who continued with their assigned intervention throughout follow up, and answer 'No' or 'Probably No' if this proportion is high enough to raise concerns. Answer 'Yes' for studies of interventions that are administered once, so that imperfect adherence is not possible. We distinguish between analyses where follow-up time after interventions switches (including cessation of intervention) is assigned to (1) the new intervention or (2) the original intervention. (1) is addressed under time-varying confounding, and should not be considered further here.
	4.6. **If N/PN to 4.3, 4.4 or 4.5**: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	It is possible to conduct an analysis that corrects for some types of deviation from the intended intervention. Examples of appropriate analysis strategies include inverse probability weighting or instrumental variable estimation. It is possible that a paper reports such an analysis without reporting information on the deviations from intended intervention, but it would be hard to judge such an analysis to be appropriate in the absence of such information. Specialist advice may be needed to assess studies that used these approaches. If everyone in one group received a co-intervention, adjustments cannot be made to overcome this.

University of BRISTOL

Risk of bias judgements

Effect of assignment to intervention

Low risk of bias	(i) Any deviations from intended intervention reflected usual practice; *or* (ii) Any deviations from usual practice were unlikely to impact on the outcome.
Moderate risk of bias	There were deviations from usual practice, but their impact on the outcome is expected to be slight.
Serious risk of bias	There were deviations from usual practice that were unbalanced between the intervention groups and likely to have affected the outcome.
Critical risk of bias	There were substantial deviations from usual practice that were unbalanced between the intervention groups and likely to have affected the outcome.
No information on which to base a judgement about risk of bias for this domain	No information is reported on whether there is deviation from the intended intervention

Risk of bias judgements

Effect of starting and adhering to intervention

Low risk of bias	The important **co-interventions** were balanced across intervention groups, and there were no deviations from the intended interventions (in terms of **implementation or adherence**) that were likely to impact on the outcome.
Moderate risk of bias	(i) There were deviations from intended intervention, but their impact on the outcome is expected to be slight. *or* (ii) The important co-interventions were not balanced across intervention groups, or there were deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome; *and* The analysis was appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-intervention) that were likely to impact on the outcome.
Serious risk of bias	(i) The important co-interventions were not balanced across intervention groups, or there were deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome; *and* (ii) The analysis was not appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-intervention) that were likely to impact on the outcome.

Risk of bias judgements

Effect of starting and adhering to intervention

Critical risk of bias	(i) There were substantial imbalances in important co-interventions across intervention groups, or there were substantial deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome; *and* (ii) The analysis was not appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-intervention) that were likely to impact on the outcome.
No information on which to base a judgement about risk of bias for this domain	No information is reported on whether there is deviation from the intended intervention.

University of BRISTOL

Bias due to missing data

Signalling Questions	Rationale
5.1 Were outcome data available for all, or nearly all, participants?	"Nearly all" should be interpreted as "enough to be confident of the findings", and a suitable proportion depends on the context. In some situations, availability of data from 95% (or possibly 90%) of the participants may be sufficient, providing that events of interest are reasonably common in both intervention groups. One aspect of this is that review authors would ideally try and locate an analysis plan for the study.
5.2 Were participants excluded due to missing data on intervention status?	Missing intervention status may be a problem. This requires that the *intended* study sample is clear, which it may not be in practice.
5.3 Were participants excluded due to missing data on other variables needed for the analysis?	This question relates particularly to participants excluded from the analysis because of missing information on confounders that were controlled for in the analysis.

Bias due to missing data

Signalling Questions	Rationale
5.4 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Are the proportion of participants and reasons for missing data similar across interventions?	This aims to elicit whether either (i) differential proportion of missing observations or (ii) differences in reasons for missing observations could substantially impact on our ability to answer the question being addressed. "Similar" includes some minor degree of discrepancy across intervention groups as expected by chance.
5.5 **If PN/N to 5.1, or Y/PY to 5.2 or 5.3**: Is there evidence that results were robust to the presence of missing data?	Evidence for robustness may come from how missing data were handled in the analysis and whether sensitivity analyses were performed by the investigators, or occasionally from additional analyses performed by the systematic reviewers. It is important to assess whether assumptions employed in analyses are clear and plausible. Both content knowledge and statistical expertise will often be required for this. For instance, use of a statistical method such as multiple imputation does not guarantee an appropriate answer. Review authors should seek naïve (complete-case) analyses for comparison, and clear differences between complete-case and multiple imputation-based findings should lead to careful assessment of the validity of the methods used.

Risk of bias judgements

Low risk of bias	(i) Data were reasonably complete; *or* (ii) Proportions of and reasons for missing participants were similar across intervention groups; *or* (iii) The analysis addressed missing data and is likely to have removed any risk of bias.
Moderate risk of bias	(i) Proportions of and reasons for missing participants differ slightly across intervention groups; *and* (ii) The analysis is unlikely to have removed the risk of bias arising from the missing data.
Serious risk of bias	(i) Proportions of missing participants differ substantially across interventions; *or* Reasons for missingness differ substantially across interventions; *and* (ii) The analysis is unlikely to have removed the risk of bias arising from the missing data; *or* Missing data were addressed inappropriately in the analysis; *or* The nature of the missing data means that the risk of bias cannot be removed through appropriate analysis.

Risk of bias judgements

Critical risk of bias	(i) (Unusual) There were critical differences between interventions in participants with missing data; *and* (ii) Missing data were not, or could not, be addressed through appropriate analysis.
No information on which to base a judgement about risk of bias for this domain	No information is reported about missing data or the potential for data to be missing.

Bias in measurement of outcomes

Signalling Questions	Rationale
6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Some outcome measures involve negligible assessor judgment, e.g. all-cause mortality or non-repeatable automated laboratory assessments. Risk of bias due to measurement of these outcomes would be expected to be low.
6.2 Were outcome assessors aware of the intervention received by study participants?	If outcome assessors were blinded to intervention status, the answer to this question would be 'No'. In other situations, outcome assessors may be unaware of the interventions being received by participants despite there being no active blinding by the study investigators; the answer this question would then also be 'No'. In studies where participants report their outcomes themselves, for example in a questionnaire, the outcome assessor is the study participant. In an observational study, the answer to this question will usually be 'Yes' when the participants report their outcomes themselves.

Bias in measurement of outcomes

Signalling Questions	Rationale
6.3 Were the methods of outcome assessment comparable across intervention groups?	Comparable assessment methods (i.e. data collection) would involve the same outcome detection methods and thresholds, same time point, same definition, and same measurements.
6.4 Were any systematic errors in measurement of the outcome related to intervention received?	This question refers to differential misclassification of outcomes. Systematic errors in measuring the outcome, if present, could cause bias if they are related to intervention or to a confounder of the intervention-outcome relationship. This will usually be due either to outcome assessors being aware of the intervention received or to non-comparability of outcome assessment methods, but there are examples of differential misclassification arising despite these controls being in place.

University of BRISTOL

Risk of bias judgements

Low risk of bias	(i) The methods of outcome assessment were comparable across intervention groups; *and* (ii) The outcome measure was unlikely to be influenced by knowledge of the intervention received by study participants (i.e. is objective) or the outcome assessors were unaware of the intervention received by study participants; *and* (iii) Any error in measuring the outcome is unrelated to intervention status.
Moderate risk of bias	(i) The methods of outcome assessment were comparable across intervention groups; *and* (ii) The outcome measure is only minimally influenced by knowledge of the intervention received by study participants; *and* (iii) Any error in measuring the outcome is only minimally related to intervention status.

University of BRISTOL

Risk of bias judgements

Serious risk of bias	(i) The methods of outcome assessment were not comparable across intervention groups; *or* (ii) The outcome measure was subjective (i.e. vulnerable to influence by knowledge of the intervention received by study participants); *and* The outcome was assessed by assessors aware of the intervention received by study participants; *or* (iii) Error in measuring the outcome was related to intervention status.
Critical risk of bias	The methods of outcome assessment were so different that they cannot reasonably be compared across intervention groups.
No information on which to base a judgement about risk of bias for this domain	No information is reported about the methods of outcome assessment.

Bias in selection of the reported result

	Signalling Questions	Rationale
	Is the reported effect estimate likely to be selected, on the basis of the results, from...	
	7.1. ... multiple outcome *measurements* within the outcome domain?	For a specified outcome domain, it is possible to generate multiple effect estimates for different measurements. If multiple measurements were made, but only one or a subset is reported, there is a risk of selective reporting on the basis of results

Bias in selection of the reported result

Signalling Questions	Rationale
Is the reported effect estimate likely to be selected, on the basis of the results, from...	
7.2 ... multiple *analyses* of the intervention-outcome relationship?	Because of the limitations of using data from non-randomized studies for analyses of effectiveness (need to control confounding, substantial missing data, etc), analysts may implement different analytic methods to address these limitations. Examples include unadjusted and adjusted models; use of final value vs change from baseline vs analysis of covariance; different transformations of variables; a continuously scaled outcome converted to categorical data with different cut-points; different sets of covariates used for adjustment; and different analytic strategies for dealing with missing data. Application of such methods generates multiple estimates of the effect of the intervention versus the comparator on the outcome. If the analyst does not pre-specify the methods to be applied, and multiple estimates are generated but only one or a subset is reported, there is a risk of selective reporting on the basis of results.

University of BRISTOL

Bias in selection of the reported result

	Signalling Questions	Rationale
	Is the reported effect estimate likely to be selected, on the basis of the results, from...	
	7.3 ... different *subgroups*?	Particularly with large cohorts often available from routine data sources, it is possible to generate multiple effect estimates for different subgroups or simply to omit varying proportions of the original cohort. If multiple estimates are generated but only one or a subset is reported, there is a risk of selective reporting on the basis of results.

Risk of bias judgements

Low risk of bias	There is clear evidence (usually through examination of a pre-registered protocol or statistical analysis plan) that all reported results correspond to all intended outcomes, analyses and sub-cohorts.
Moderate risk of bias	(i) The outcome measurements and analyses are consistent with an *a priori* plan; or are clearly defined and both internally and externally consistent; *and* (ii) There is no indication of selection of the reported analysis from among multiple analyses; *and* (iii) There is no indication of selection of the cohort or subgroups for analysis and reporting on the basis of the results.
Serious risk of bias	(i) Outcomes are defined in different ways in the methods and results sections, or in different publications of the study; *or* (ii) There is a high risk of selective reporting from among multiple analyses; *or* (iii) The cohort or subgroup is selected from a larger study for analysis and appears to be reported on the basis of the results.

Risk of bias judgements

Critical risk of bias	(i) There is evidence or strong suspicion of selective reporting of results; *and* (ii) The unreported results are likely to be substantially different from the reported results.
No information on which to base a judgement about risk of bias for this domain	There is too little information to make a judgement (for example, if only an abstract is available for the study).