

ORIGINAL RESEARCH

Empirical evidence of observer bias in randomized clinical trials: updated and expanded analysis of trials with both blinded and non-blinded outcome assessors

Josefina Salazar^{a,b,*}, Helene Moustgaard^{a,b}, Javier Bracchiglione^{c,d}, Asbjørn Hróbjartsson^{a,b}^a*Cochrane Denmark and Centre for Evidence-Based Medicine Odense (CEBMO), Department of Clinical Research, University of Southern Denmark, Campusvej 55, Odense 5230, Denmark*^b*Open Patient data Exploratory Network (OPEN), Odense University Hospital, Odense, Denmark*^c*Iberoamerican Cochrane Centre, Institut de Recerca Sant Pau (IR Sant Pau), CIBERESP, Barcelona, Spain*^d*Interdisciplinary Centre for Health Studies (CIESAL), Universidad de Valparaíso, Viña del Mar, Chile*

Accepted 10 April 2025; Published online 19 April 2025

Abstract**Objectives:** To study the impact of lack of blinding of outcome assessors on estimated treatment effects of randomized clinical trials.**Study Design and Setting:** Meta-epidemiological study. We included randomized trials with binary or measurement scale outcomes that (1) allocated patients to subtrials with and without blinded outcome assessment, or (2) had both blinded and non-blinded assessments of the same outcome. We identified trials from previous meta-epidemiological studies and searched six databases from 2013 to 2024. We calculated a ratio of odds ratios (ROR) for each trial. A ROR < 1 indicated a more favorable effect estimate by the non-blinded assessor. We pooled RORs using random effects meta-analysis and conducted meta-regression, subgroup, and sensitivity analyses.**Results:** We included 66 trials (9451 patients) across 18 clinical specialties. The pooled RORs in 43 trials (7055 patients) was 0.71 (0.55–0.92). Thirty of the 43 trials assessed highly subjective outcomes. Meta-regression showed no statistically significant association between ROR and scores for outcome subjectivity ($P = .53$), vulnerability ($P = .91$), and involvement ($P = .99$). Heterogeneity was partly explained by a larger impact in non-drug trials, ROR 0.62 (0.46–0.84), and industry-funded trials, ROR 0.57 (0.37–0.88). Sensitivity analyses, including imputed data for 23 trials (2396 patients randomized), did not modify the observed impact importantly.**Conclusion:** We provide empirical evidence of considerable bias in effect estimates of randomized trials with non-blinded assessors of subjective binary and measurement scale outcomes. Non-blinded assessors exaggerated effect estimates, expressed as odds ratios, by 29% (8%–45%) on average. This strongly supports blinding outcome assessors of subjective outcomes. © 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).**Keywords:** Bias; Blinding; Observer bias; Randomized controlled trials as topic; Meta-epidemiological studies; Empirical evidence

1. Introduction

The randomized clinical trial is the most trustworthy research design for assessing the effects of interventions and the primary source of evidence for systematic reviews of interventions, clinical practice guidelines, and decisions made by regulatory agencies [1]. However, results from randomized trials may be biased due to limitations in

design, execution, and reporting. One source of bias derives from the expectations and preferences of the persons responsible for assessing trial outcomes, often called “observers” or “outcome assessors.” Predispositions of observers may result in systematic differences in how outcomes are assessed between experimental and control groups, so-called observer bias. For example, expectations about a promising new treatment may lead to a more optimistic evaluation of the overall improvement of a patient in the experimental compared to the control group.

Blinding outcome assessors in a trial, that is, keeping them unaware of the allocated intervention of individual patients, is intended to prevent observer bias and is an established methodological principle. Still, many trials do

Funding: This study received no specific funding.

* Corresponding author. Cochrane Denmark and Centre for Evidence-Based Medicine Odense (CEBMO), Department of Clinical Research, University of Southern Denmark, Campusvej 55, Odense 5230, Denmark.

E-mail address: jsalazar@health.sdu.dk (J. Salazar).

Plain Language Summary

In a randomized clinical trial, the person evaluating the results (assessor) may be either unaware of the intervention received by participants (blinded assessor) or aware of it (non-blinded assessor). Knowing which treatment a patient received can influence the assessor's evaluation of the effect; for example, if an assessor has high expectations for a new experimental intervention, they may rate a patient's improvement more favorably in the group that received the intervention compared to the group that did not. We call this observer bias. In this study, we compared the results obtained from blinded assessors to those from non-blinded assessors within the same trials to estimate the impact of observer bias in randomized trials. We found that non-blinded assessors exaggerated the experimental intervention effect by approximately 29%, on average, compared to blinded assessors. Our results indicate that when an evaluation of a patient in a randomized trial requires judgment, there is potential for substantial bias if assessors are not blinded. To ensure more reliable results, randomized clinical trials should blind assessors whenever possible.

not use blinded outcome assessors. A review of randomized trials published in general medicine journals found that among 225 trials that required assessment of the primary outcome, (ie, not all-cause mortality), 79 (35%) had blinded outcome assessors [2]. Furthermore, 46 of 100 trials reported as “double blind” had identified the outcome assessor as blind [2].

The empirical evaluation of the impact of lack of blinding of outcome assessors provides an estimate of the average degree of observer bias, its range, direction, and heterogeneity. Thus, it can inform an evidence-based interpretation of individual trial results. In the context of systematic reviews, such evaluations support the risk of bias assessment in the included trials, which will directly influence the analysis and conclusions drawn from the review.

Unfortunately, empirical evaluation of observer bias has so far provided inconsistent results. Classic meta-epidemiological studies that compare effect estimates from subgroups of trials with and without blinded outcome assessors within meta-analyses, have reported estimates with wide confidence intervals consistent with some degree of bias in either direction (ie, over- and under-estimation of the treatment effect; Table 1) [3–11]. While some studies reported a tendency to exaggerate treatment effect estimates in trials without blinded outcome assessors [3,6,8,11], the largest study, MetaBLIND, notably found no evidence for a difference between trials with and without blinded outcome assessors [7]. However, the observational nature of these comparisons carries a risk of confounding, as any true impact of lack of blinding may have been nullified or modified through other systematic differences between the trials in the included meta-analyses.

Between 2012 and 2014, we published three meta-epidemiological studies based on randomized trials using both blinded and non-blinded assessments of the same outcome and patients [12–14]. This design considerably reduces the risk of confounding compared with observational between-trial analyses. We found that lack of blinding of outcome assessors in trials with subjective outcomes was associated with substantially biased effect estimates:

non-blinded assessors exaggerated odds ratios (ORs) by 36% (4%–57%) [12], standardized mean differences by 68% (14%–230%) [13], and hazard ratios by 27% (7%–43%), on average, compared with blinded assessors [14].

The discrepancy between the results of MetaBLIND and other similar meta-epidemiological studies and the results based on analysis of trials with both kinds of assessors is disconcerting and challenging to interpret. Classic meta-epidemiological studies are easier to conduct and include a large number of trials. In contrast, trials with both types of assessors are rare, and the three atypical meta-epidemiological studies each have a modest number of included trials within a limited range of clinical specialties [12–14]. Yet, they are methodologically more reliable.

We, therefore, found it relevant to update and expand our analysis of trials with both blinded and non-blinded outcome assessors with the aim of re-estimating the impact of lack of blinding of outcome assessors on the estimated treatment effects of randomized clinical trials.

2. Methods

We conducted a meta-epidemiological study of trials with blinded and non-blinded outcome assessors [12,13]. The main innovations are the (1) inclusion of trials that directly randomize patients to blinded and non-blinded subtrials, (2) identification of additional randomized trials with blinded and non-blinded assessors, mostly published since 2013, and (3) conversion of measurement scale outcomes to binary outcomes, resulting in a larger sample size.

2.1. Protocol and amendments

We prespecified the methods in a protocol available in [Appendix A](#). We also report all the amendments made to the protocol in the [Appendix](#).

2.2. Eligibility criteria

We included randomized clinical trials in which patients were allocated to two subtrials that differed only in one

What is new?**Key findings**

- Our results show substantial bias in the effect estimates of randomized trials with non-blinded assessors of subjective outcomes, with an exaggeration of effect estimates, expressed as odds ratios, of 29% (95% CI 8%–45%) on average. There was an apparent larger impact on non-drug and industry-funded trials.

What this adds to what is known?

- Our study provides updated empirical evidence based on comparisons within trials with both blinded and non-blinded assessors. This result is more reliable than results based on between-trial comparisons from classic meta-epidemiological studies.

What is the implication and what should change now?

- Our results should be considered when designing randomized trials and when interpreting findings of individual trials, for example, when conducting systematic reviews.
- Outcome assessors of randomized trials with subjective outcomes should be blinded whenever possible.

having blinded assessors and the other non-blinded assessors. We also included randomized clinical trials in which blinded and non-blinded assessors evaluated the same outcome for all patients or a random subset of patients. We refer to the first type as “randomized between-trial comparison” and the second as “observational within-trial comparison.”

We excluded trials where it was not clear which group was experimental, since these do not allow for a non-arbitrary evaluation of the direction of any bias; trials in which blinded and non-blinded assessors had access to each other’s results when assessing the outcome, or where it was obvious that blinded assessors became aware of the patient’s treatment group (eg, radiographs showed ceramic material indicative of the experimental intervention); trials with an adjudication committee as the blinded assessor, since adjudication can involve knowledge of the local non-blinded assessment, and the information available for the blinded and non-blinded assessors could be systematically different (eg, restricted to events only or no further examinations after locally determined event); trials with patient-reported outcomes (ie, patients as the blinded or non-blinded “assessors” of their outcomes); trials that only

reported blinded and non-blinded assessment of time-to-event outcomes (see [Appendix A](#)); trials where the blinded and non-blinded assessors used different scales for the same outcome; and trials that only investigated outcomes other than clinical therapeutic effects of interventions.

2.3. Information sources and search strategy

We identified all trials included in the previous meta-epidemiological studies of within-trial comparisons [12–14]. We searched for trials published from September 15, 2013 (final search for previous studies, covering all outcomes) to February 16, 2024, in the bibliographic databases PubMed, Embase, PsycINFO, CINAHL, Cochrane Central Register of Controlled Trials, and the full-text database Google Scholar. We used the search strategy developed in the previous meta-epidemiological studies ([Appendix B](#)). For each search string used in Google Scholar, we screened the first 200 hits.

We searched for potentially eligible trials in the reference lists and the first 50 studies that cited the newly included trials (via Google Scholar).

2.4. Selection process

One author read all abstracts from standard databases and all text fragments from full-text databases. When a study was deemed potentially eligible, one author retrieved the full trial publication and excluded clearly ineligible studies. Two researchers decided on eligibility for the remaining studies. Disagreements were resolved by discussion.

2.5. Data collection process

We obtained the data sets from the previous meta-epidemiological studies [12,13]. We checked the soundness of the data by replicating the descriptive tables and main analyses of the previous studies. We report the minor adjustments to the data sets with reasons in [Appendix C](#).

Two researchers independently extracted information from newly included trials using a previously tested MS Excel data collection form. Disagreements were resolved by discussion. When trials had incomplete or unclear data (eg, no SDs), we contacted the corresponding author via e-mail and sent at least two reminders. We asked trial authors for either summary results or individual patient data. We also searched for data on the Food and Drug Administration website.

2.6. Data items

We extracted the following trial characteristics: design, main objective, number of study groups, type of experimental intervention, control group and outcomes examined, medical specialty, funding, and type of journal (specialty or general medicine). We also extracted methodological

Table 1. Meta-epidemiological studies comparing trials with blinded outcome assessors to trials without, within meta-analysis (between-trial comparison)

Study	Clinical area	No of MA in the analysis	No of trials in the analysis	Outcome assessed (as reported by the study)	Ratio of odds ratio (95% CI)
Studies not including meta-analyses of all-cause mortality					
Chaimani 2013 [3]	Multiple specialties	NS	254	Non-mortality outcomes (NS)	0.87 (0.63–1.20) ^{a,b}
Hartling 2014 [4]	Child health	NS	NS	Incl. acute otitis media, chronic suppurative otitis media, clinical malaria, acute respiratory tract infection, emergency department visits, failure to rehydrate, diarrhea, aggression, nocturnal enuresis, depression, anxiety, DMFS index	1.00 (0.82–1.22) ^c
Armijo-Olivo 2017 [5]	Physical therapy	17	165	Incl. Chalder fatigue scale, fasting plasma glucose, physical activity, range of motion, disability, gait speed, pain, reduction of depression symptoms, treadmill walking distance	1.14 (0.86–1.49) ^{b,c}
Haring 2020 [6]	Testosterone therapy	5	36	Incl. prostate-specific antigen levels, hemoglobin A1C levels, body weight, and cardiovascular-related events	0.55 (0.20–1.54)
Moustgaard 2020—MetaBLIND [7]	Multiple specialties	46	397	Incl. cause-specific mortality, clinician-assessed outcomes (eg, lung function), global improvement, laboratory-reported outcomes, lifestyle outcomes (eg, diet), major morbidity event, mental health outcomes, pain, perinatal and pregnancy outcomes, quality of life, radiological outcomes, and surgical and device-related outcomes	1.01 (0.86–1.18) ^a
Amer 2021 [8]	Surgery	NS	161	Length of hospital stay	0.39 (0.15–1.00) ^c
Studies including meta-analyses of all-cause mortality					
Bialy 2014 [9]	Neonatal	NS	NS	Incl. pneumothorax, all-cause mortality, improvement in oxygenation, need for mechanical ventilation, need for retreatment, chronic lung disease, failure of permanent patent ductus arteriosus closure	1.08 (0.85–1.33) ^b
Wang 2021 [10]	Chronic medical conditions	NS	NS	Incl. all-cause and cause-specific mortality, disease progression, adverse events, others (NS)	1.04 (0.93–1.16) ^b
Stadelmaier 2022 [11]	Nutrition	12	NS	Incl. cardiovascular disease, all-cause mortality, cancer, pre-eclampsia, gestational diabetes, preterm birth, colorectal adenoma	0.88 (0.73–1.07)

MA, meta-analysis; NS, not specified.

^a Credible interval.^b Direction of the published estimate reversed, such that a ROR < 1 indicates that the non-blinded outcome assessor yielded a more favorable effect estimate than the blinded assessor.^c Converted from difference in standardized mean differences = Hartling 2014—0.00 (–0.11 to 0.11); Armijo-Olivo 2017—0.07 (–0.22 to 0.08); Amer 2021—0.52 (–1.04 to 0.00).

characteristics: random sequence generation and allocation concealment, blinding of patients and treatment providers, and handling of attrition.

From each trial, we selected one outcome with blinded and non-blinded assessments. If more than one outcome was available, we selected the primary outcome. If no outcome was defined as primary, we selected the most clinically relevant (see [Appendix A](#) for more details on outcome prioritization).

For binary outcomes, we extracted the total number of failures and successes in each group resulting from blinded and non-blinded assessments. For measurement scale outcomes, we extracted the posttreatment mean (or change from baseline), its respective SD, and number of patients in each group for blinded and non-blinded assessments. When necessary, we inverted the scale by multiplying the mean values by -1 to ensure that higher values represented worse states.

2.7. Data assessment

We classified each outcome according to whether it was measured based on a hypothesis of benefit (eg, the experimental intervention will reduce pain compared to the control intervention) or based on another type of hypothesis (eg, a hypothesis of harm when measuring side effects or a hypothesis of equivalence).

Two authors masked to the trial results independently assigned each outcome a score of 1–3 (low to high) on three factors: outcome subjectivity, that is, the degree of assessor's judgment involved in assessing the outcome (eg, low for reading a laboratory sheet and high when evaluating overall aesthetic improvement); vulnerability of the outcome to the reporting and behavior of non-blinded patients (eg, low for a radiographic assessment and high for evaluations using questionnaires); and non-blinded outcome assessor's overall involvement in the trial (eg, high for an assessor who is also the lead investigator and low for an independent outcome assessor). Disagreements were resolved by discussion.

Two authors independently evaluated the presence (yes/no) of four factors that could systematically distort the measured differences between blinded and non-blinded assessments: considerable difference in timing of blinded and non-blinded assessment, difference in type of assessor (eg, pulmonologist *v* radiologist), difference in method (eg, direct visual assessment *v* photograph), and whether both blinded and non-blinded assessors evaluated identical (or randomly sampled) groups of patients. We evaluated the first three factors in a masked manner. Disagreements were resolved by discussion.

2.8. Data conversion

We treated results from cross-over and split-body trials as results from parallel group trials. We calculated ORs

for failures for blinded and non-blinded assessments for each trial with binary outcomes. When there were zero events in one group, we used the “treatment arm” continuity correction, adding a factor of the reciprocal of the size of the opposite treatment arm [15]. An OR < 1 indicates a beneficial effect of the experimental intervention.

We expressed the impact of lack of blinding of outcome assessors for each trial as a ratio of OR (ROR = OR_{non-blinded}/OR_{blinded}). An ROR < 1 indicates that the non-blinded outcome assessor yielded a more favorable effect estimate than the blinded assessor (details in Appendix D).

For trials that only reported measurement scale outcomes, we calculated standardized mean differences (SMDs), correcting for small sample bias (ie, Hedges' *g*), for blinded and non-blinded assessments using the pooled SD of the blinded assessment as the common standardizing unit. We then calculated a difference in SMD (details in [Appendix D](#)) and re-expressed it as ROR, assuming an underlying logistic distribution [16].

2.9. Data analysis

For our main analysis, we pooled individual RORs of trials with complete data for outcomes measured based on a hypothesis of benefit using random effects meta-analysis with inverse variance weights (DerSimonian-Laird) [17]. Our estimation of the impact of lack of blinding of outcome assessors was based solely on SEs that disregarded the dependence of blinded and non-blinded assessments within trials. We stratified the main analysis by type of comparison (randomized between-trial or observational within-trial) but conducted all other analyses irrespective of the type of comparison. We evaluated inconsistency using the I^2 statistic [18], and we present prediction intervals to aid the interpretation of heterogeneity [19].

We conducted random effects meta-regressions on the log(ROR) of the main analysis to explore the association with our scores for the degree of (1) outcome subjectivity, (2) vulnerability to the reporting and behavior of non-blinded patients, and (3) non-blinded assessor's overall involvement in the trial, separately and as continuous variables.

To investigate the impact of the factors that could systematically distort the observed differences, we conducted four subgroup analyses based on differences in timing, type of assessor, method, and group of patients assessed by blinded and non-blinded assessors. In the latter, we did not include randomized between-trial comparisons.

To investigate other sources of heterogeneity, we conducted five subgroup analyses by type of experimental intervention (drug *v* surgery/psychosocial/other), source of funding (industry *v* nonindustry), trial design (parallel arm *v* not parallel), publication objective (observer bias main objective *v* not), and type of non-blinded assessor (panel consensus *v* single assessor). We considered a *P*-

value $<.1$ for the test for subgroup differences to indicate a statistically significant effect [20].

We tested the robustness of our main analysis with five sensitivity analyses. In the preplanned sensitivity analyses, we first gave all trials equal weight. Then, we calculated the impact of lack of blinding of outcome assessors for binary and measurement scale outcomes separately. We conducted three post hoc sensitivity analyses: (1) excluding trials with zero events in one group, (2) reversing the direction of bias of one trial, and (3) including trials with incomplete or unclear data. For the latter, we first included the trials with tentative ROR, calculated with available but incomplete, unclear, or discrepant data, and in the second stage, we included trials with no data (ie, only measures of agreement) assuming two scenarios: (1) non-blinded assessors yielded more favorable effect estimates (ROR 0.90) and (2) blinded assessors yielded more favorable effect estimates (ROR 1.11; details in Appendix E).

We conducted all analyses using Stata 18 (StataCorp LLC).

3. Results

We identified 11622 records after removing duplicates (Fig 1). We assessed 163 full reports for eligibility and excluded 97: 11 from the previous meta-epidemiological studies as they reported only time-to-event outcomes, and for the remaining 86, the reasons for exclusion were mainly that they lacked a non-blinded assessment, the experimental group was not clearly specified, or the assessors had access to each other's results (confirmation rather than an independent outcome assessment; Appendix F). Thus, we included 66 trials (9451 patients randomized) with blinded and non-blinded outcome assessments [21–86].

One trial published in 1966 aimed to investigate observer bias and randomized 32 patients receiving psychiatric interventions to two subtrials with either blinded or non-blinded outcome assessment, thus allowing for a randomized between-trial comparison [36].

Sixty-five trials published from 1983 to 2023 studied the effect of interventions across 18 clinical specialties and had

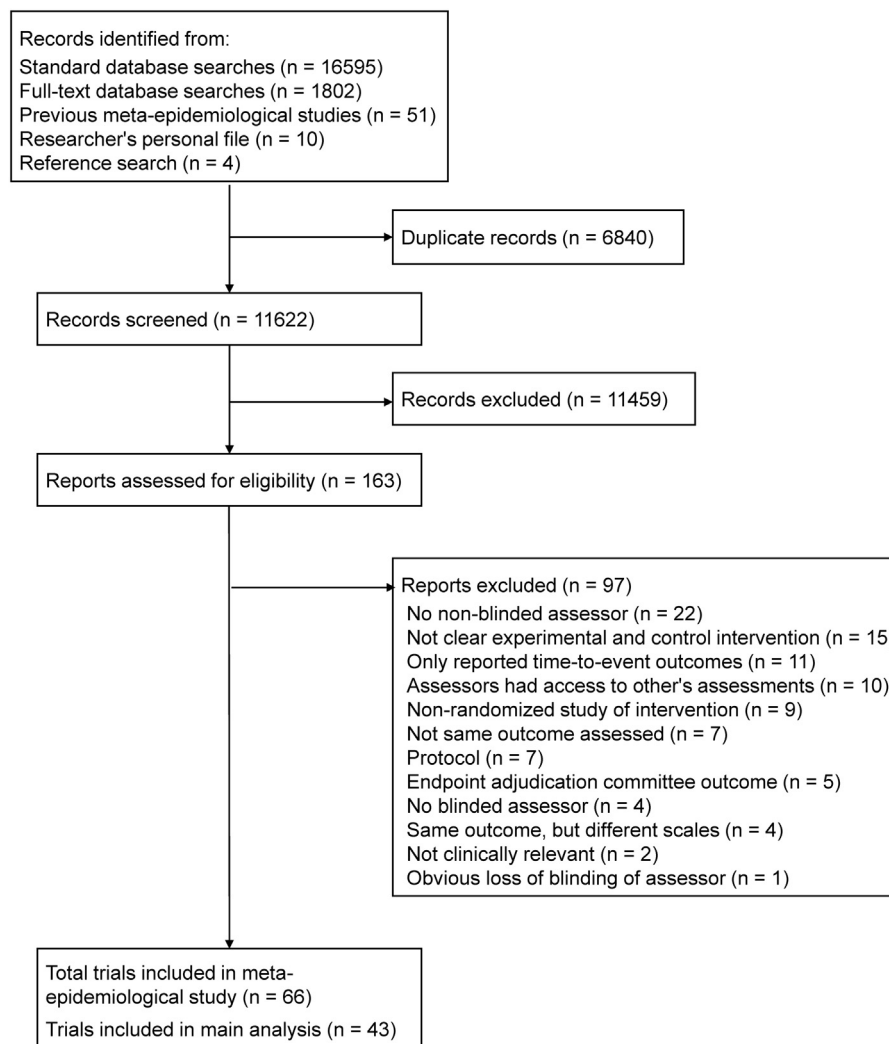


Figure 1. Flow diagram for identification of eligible trials.

Table 2. Characteristics of 66 randomized clinical trials with both blinded and nonblinded outcome assessment

Characteristic	All trials 66 (100)	Comparison type		Trials included in the main analysis 43 (65)
		Randomized between-trial 1 (2)	Observational within-trial 65 (98)	
General				
Parallel group design	54 (82)	1 (100)	53 (82)	37 (86)
Two study groups	62 (94)	1 (100)	61 (94)	6 (14)
Experimental intervention				
Drug	19 (29)	1 (100)	18 (28)	12 (28)
Surgery	25 (38)	0 (0)	25 (38)	18 (42)
Psychosocial/behavioral/education	3 (5)	0 (0)	3 (5)	2 (5)
Other	19 (29)	0 (0)	19 (29)	11 (26)
Control group				
Placebo	5 (8)	0 (0)	5 (8)	5 (12)
Standard care	41 (62)	0 (0)	41 (63)	29 (67)
No treatment	6 (9)	0 (0)	6 (9)	3 (7)
Active comparator	14 (21)	1 (100)	13 (20)	6 (14)
Primary outcome defined	52 (79)	0 (0)	52 (80)	34 (79)
Publication objective: observer bias	10 (15)	1 (100)	9 (14)	7 (16)
Published in specialty journal	48 (73)	0 (0)	48 (74)	29 (67)
Published in general medicine journal	15 (23)	1 (100)	14 (22)	12 (28)
Funding: industry ^a	31 (47)	1 (100)	30 (46)	42 (53)
Trial methods ^b				
Random allocation sequence adequately generated	22 (33)	0 (0)	22 (34)	15 (35)
Random allocation sequence adequately concealed	21 (32)	0 (0)	21 (32)	17 (39)
Patients blinded	16 (24)	0 (0)	16 (25)	13 (30)
All treatment providers blinded	2 (3)	0 (0)	2 (3)	1 (2)
Attrition: ITT and dropouts <15%	27 (41)	1 (100)	26 (40)	19 (44)
Medical specialty				
Cosmetic surgery	11 (17)	0 (0)	11 (17)	9 (21)
Neurology	8 (12)	0 (0)	8 (12)	6 (14)
Dermatology	7 (11)	0 (0)	7 (11)	3 (7)
Vascular surgery	7 (11)	0 (0)	7 (11)	4 (9)
Psychiatry	6 (9)	1 (100)	5 (8)	4 (9)
Orthopedic surgery	4 (6)	0 (0)	4 (6)	4 (9)
Anesthesiology	3 (5)	0 (0)	3 (5)	1 (2)
Cardiology	3 (5)	0 (0)	3 (5)	2 (5)
Infectious diseases	3 (5)	0 (0)	3 (5)	2 (5)
Surgery	3 (5)	0 (0)	3 (5)	2 (5)
Ophthalmology	2 (3)	0 (0)	2 (3)	2 (5)
Otolaryngology	2 (3)	0 (0)	2 (3)	2 (5)
Plastic surgery	2 (3)	0 (0)	2 (3)	0 (0)
Specialties with one trial ^c	5 (8)	0 (0)	5 (8)	2 (5)

ITT, intention to treat.

Data are number (%).

^a We considered trials as industry-funded if any of the funding sources was from the drug and device industry.^b We considered the random allocation sequence to be adequately generated if the trial described it as computer generated sequence, random number tables, or coin tosses. We considered the sequence adequately concealed if the trial described the allocation as central or using sealed envelopes. We considered patients or treatment/care providers to be blinded if the trial explicitly stated it for each group separately.^c Dentistry, gynecology, physiotherapy, pulmonary medicine, and traumatology.

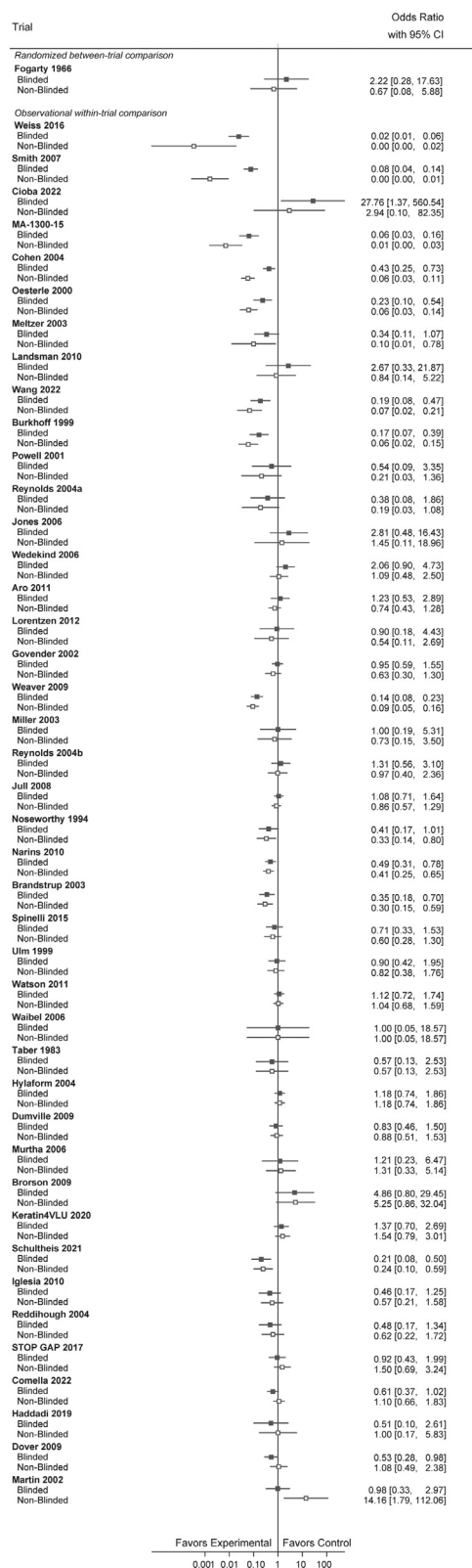


Figure 2. Estimated effects according to blinded and nonblinded outcome assessor.

both types of outcome assessors, allowing for observational within-trial comparisons. Most studies (86%) aimed to investigate clinical therapeutic effects of interventions rather than observer bias and included a second assessor to validate the other's assessment or secure that data were available for all patients. The median sample size was 90 (10th–90th centile 21–299), and 53 trials (82%) had a parallel group design. All outcomes were evaluated based on a hypothesis of benefit. Table 2 presents a summary of the characteristics of the included trials, and Appendix Table G.1 presents full details.

Nine (35%) of the 26 newly included trials published data for both blinded and non-blinded assessments. After contacting authors, the number of such trials with data increased to 13 (50%). Of the 40 trials included in the previous meta-epidemiological studies, 30 (75%) had data for blinded and non-blinded assessments. Thus, we included 43 trials in the main analysis (7055 patients randomized), 10 of which provided individual patient data [22,28,35,37,38,44,46,58,66,67].

The effect estimates were more favorable when based on non-blinded assessment in the single randomized between-trial comparison, and 27 of the 42 trials (64%) in the observational within-trial comparison (Fig 2). The ROR in the 43 trials ranged from 0.01 to 14.39. The pooled ROR was 0.71 (95% CI 0.55–0.92), indicating that, on average, the OR based on non-blinded assessments was exaggerated by 29% (8%–45%) compared with the OR based on blinded assessments. There was moderate inconsistency ($I^2 = 44\%$), and the 95% prediction interval ranged from 0.25 to 2.07 (Fig 3).

All trials evaluated subjective outcomes. Thirty trials (70%) included in the main analysis evaluated an outcome with high subjectivity (score 3 in a scale from 1 to 3). One trial (2%) reported an outcome with low subjectivity (score 1) [28]. The outcomes were highly vulnerable to the reporting and behavior of non-blinded patients in 22 trials (51%), and the non-blinded assessor had a high overall involvement in 34 trials (79%), for example, treating physician (Appendix Table G.2). Meta-regression analyses showed no statistically significant association between the ROR and the scores for outcome subjectivity ($P = .529$), outcome vulnerability to the reporting and behavior of non-blinded patients ($P = .908$), and non-blinded outcome assessors' involvement in the trial ($P = .992$). However, the slope of the regression line for outcome subjectivity was in the expected direction, with a greater impact of lack of blinding of outcome assessors on highly subjective outcomes (Appendix H).

The methods used by blinded and non-blinded assessors differed in 25 trials (58%), mainly non-blinded direct clinical assessment vs blinded photograph/video assessment. There was a difference in the type of assessor in 22 trials

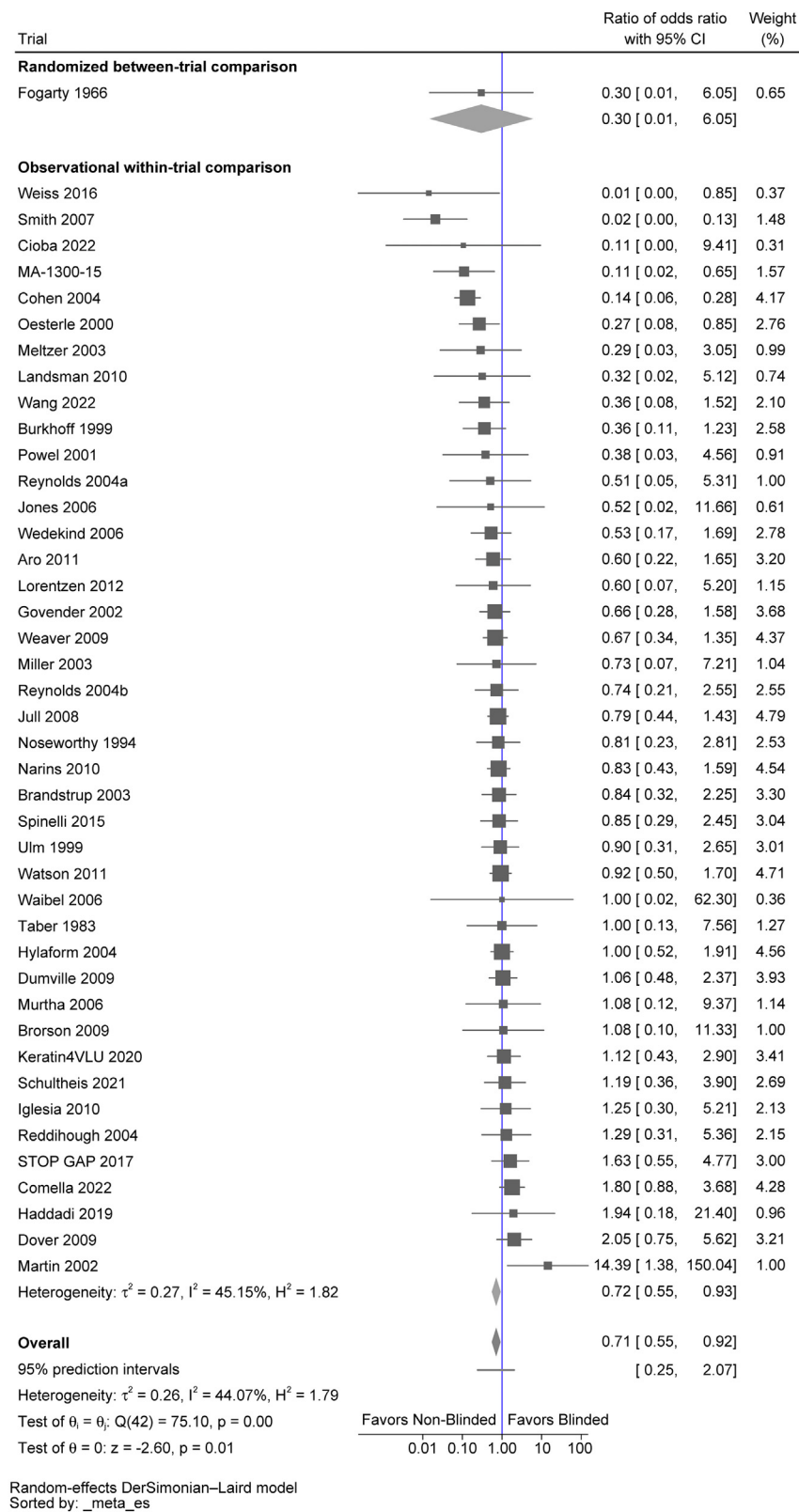


Figure 3. Impact of nonblinded outcome assessors on estimated effects in randomized clinical trials.

Table 3. Main, subgroup, and sensitivity analyses

Analysis	No of trials	<i>I</i> ² (%)	ROR (95%CI)	<i>P</i> -value for subgroup differences
Main analysis	43	44	0.71 (0.55–0.92)	
Randomized between-trial comparison	1	NA	0.30 (0.01–6.05)	
Observational within-trial comparison	42	45	0.72 (0.55–0.93)	
Stratified by factors that could systematically distort observed differences				
Timing of blinded and non-blinded assessments				
Same/similar	42	45	0.72 (0.56–0.93)	0.40
Not same/similar	1	NA	0.11 (0.00–9.41)	
Type of assessors				
Same type (such as dentist v dentist)	21	0	0.88 (0.68–1.15)	0.16
Not same type (such as radiologist v pulmonologist)	22	44	0.61 (0.40–0.94)	
Methods used				
Same procedure (such as radiographs v radiographs)	18	24	0.79 (0.53–1.19)	0.59
Not same procedure (such as live v photographs)	25	54	0.69 (0.50–0.95)	
Patients evaluated				
All evaluated by both assessors	22	0	0.81 (0.64–1.03)	0.37
Minority evaluated by one type of assessor	20	68	0.63 (0.38–1.03)	
Subgroup analyses				
Experimental intervention				
Drug	12	10	1.05 (0.71–1.55)	0.04
Surgery/psychosocial/other	31	48	0.62 (0.46–0.84)	
Funding ^a				
Industry	23	65	0.57 (0.37–0.88)	0.06
Non-industry	16	0	0.92 (0.71–1.19)	
Design				
Parallel group	37	44	0.74 (0.57–0.96)	0.48
Not parallel group (split-body or cross-over)	6	67	0.50 (0.17–1.45)	
Publication objective				
Observer bias main objective	7	0	0.87 (0.51–1.49)	0.46
Observer bias not main objective	36	52	0.69 (0.52–0.93)	
Nonblinded assessment				
Panel consensus	2	0	1.11 (0.55–2.24)	0.21
Single assessor	41	46	0.69 (0.53–0.90)	
Sensitivity analyses				
All trials equal weight	43		0.60 (NA)	
Only binary outcomes	28	38	0.66 (0.47–0.92)	
Only measurement scale outcomes expressed as ROR ^b	15	55	0.80 (0.53–1.20)	
Excluding trials with zero events in one arm	41	43	0.73 (0.57–0.94)	
Direction of bias reversed in one trial (Martin 2002) [53] ^c	43	42	0.68 (0.53–0.87)	

(Continued)

Table 3. Continued

Analysis	No of trials	I^2 (%)	ROR (95%CI)	P-value for subgroup differences
Missing data				
Including trials with tentative ROR ^d	53	36	0.71 (0.57–0.88)	
Also including trials with no data ^e				
Scenario A, ROR = 0.90	66	21	0.75 (0.63–0.89)	
Scenario B, ROR = 1.11	66	24	0.78 (0.66–0.93)	

NA, not applicable.

^a We excluded 4 trials because funding was not reported. Inclusion in the analysis as industry-funded trials results in an ROR (95% CI) of 0.56 (0.37–0.85), $I^2 = 61\%$, and a P -value of .05 for the test for subgroup differences.

^b dSMD (95% CI) –0.13 (–0.35 to 0.10). A dSMD <0 indicates that the non-blinded assessor yielded a more favorable effect estimate than the blinded assessor.

^c Martin 2002 compared oral (experimental) to intravenous (control) administration of the same drug for cytomegalovirus retinitis in patients with HIV. Although it was framed as expecting the oral administration to give better results (hypothesis of benefit), we find it plausible that the outcome assessors preferred the established intravenous treatment.

^d Based on available but incomplete, unclear, or discrepant data.

^e Only reported measures or descriptions of agreement between assessors. We included these trials following two scenarios: exaggeration of treatment effects by non-blinded (Scenario A) or blinded (Scenario B) assessors.

(51%) and a considerable difference in the timing between blinded and non-blinded assessments in 1 trial (2%). In 23 trials (53%), all patients were assessed by both blinded and non-blinded assessors (Appendix Table G.2). Four of the 43 trials (9%) were free of all differences and had a pooled ROR of 0.52 (95% CI 0.22–1.27; $I^2 = 0\%$) [36,54,74,81]. Overall, we found no significant subgroup effects for these factors (Table 3).

We explored sources of heterogeneity in the main analysis and found a statistically significant subgroup effect for type of experimental intervention ($P = .04$) and funding ($P = .06$). In trials with non-drug experimental interventions, non-blinded assessors yielded more optimistic effect estimates; however, in drug trials, there was a non-statistically significant effect in the opposite direction with low inconsistency ($I^2 = 10\%$). Regarding sources of funding, in both subgroups, the ROR indicated that, on average, non-blinded assessors yielded more optimistic effect estimates, but in industry-funded trials, the exaggeration was greater (43% vs 8%). However, there was still considerable inconsistency between trials within the industry-funded subgroup ($I^2 = 65\%$). We found no significant subgroup effects for the other analyses (Table 3).

Our results were largely consistent across sensitivity analyses (Table 3). We did not include 23 trials (35%) (2396 patients randomized) in the main analysis because of insufficient data. We imputed data and calculated a tentative ROR for 10 trials (1345 patients randomized). The remaining 13 trials (1051 patients randomized) reported only measures of agreement between blinded and non-blinded assessors, for example, kappa values, correlation coefficients, or narrative statements, such as “considerable agreement” (Appendix Table G.3). Sensitivity analysis including tentative ROR of 10 trials did not change the point estimate (ROR 0.71). Similarly, analyses including trials following two scenarios, where either non-blinded or blinded assessors yielded more favorable effect estimates, did not

modify the observed impact (ROR 0.75 and 0.78; Table 3 and Appendix H).

4. Discussion

We provide evidence of substantial bias in the effect estimates of randomized clinical trials with non-blinded assessors of subjective outcomes. The pooled ratio of ORs was 0.71 (0.55–0.92), indicating that nonblinded assessors of subjective outcomes exaggerated effect estimates, expressed as ORs, by an average of 29% (8%–45%). Notably, the 95% prediction interval ranged from 0.25 to 2.07, and the pooled result involved moderate inconsistency. Heterogeneity was partly explained by more pronounced observer bias in industry-funded trials ROR 0.57 (0.37–0.88) and non-drug trials ROR 0.62 (0.46–0.84), but some heterogeneity remains unexplained.

4.1. Strengths and weaknesses of the study

The main innovation and strengths of our study are its robust methodological design and the inclusion of a comparatively large number of trials. The included 66 trials covered 18 clinical areas, the main analysis was based on 43 trials with 7055 randomized patients, and individual patient data provided 24% of the weight of the main analysis (10 trials, 1781 randomized patients). We included one trial with direct randomization to blinded or non-blinded outcome assessor, which is a rare case of an experimental comparison of a methodological characteristic of trial design. The remaining 65 trials compared blinded and non-blinded assessments of the same outcome based on the same patients within the same trial, with a low risk of confounding.

We consider our analysis to have two main challenges. The first is missing data, as 23 trials (35%) did not provide

enough data to be included in the main analysis. In most cases, the separate results from the blinded and non-blinded assessments were not predefined trial outcomes. Therefore, we do not consider our results intrinsically at high risk of bias due to selective reporting. However, we did assess the impact of missing data by first imputing ratios of ORs for 10 trials with uncertain or discrepant data and then conducting analyses assuming an exaggeration of effect estimates by the non-blinded or blinded assessors for the remaining 13 trials. These two scenarios reasonably represent a “good agreement” between assessors. All analyses were consistent with the result of the main analysis.

The second challenge is external validity. None of the trials in our analysis assessed outcomes that were clearly objective. Furthermore, the inclusion of both types of assessors in a single trial may be interpreted as an indication that trial investigators considered the outcome to be particularly susceptible to bias. Still, the general and methodological characteristics of the trials included in our study mirror that reported by Hopewell et al. in their study of the epidemiology of clinical trials [87]. Our sample included 44 clinical conditions across 18 specialties, indicating fairly broad applicability across clinical settings. Overall, we consider our sample reasonably representative of trials with subjective outcomes, that is, outcomes that require some degree of judgment by the observer.

It is important to note that the blinded and non-blinded assessments of a trial outcome for a given patient are correlated. In our estimation of the impact observer bias, we used standard errors that disregarded the dependence of blinded and non-blinded assessments within trials. Disregarding the dependence, all other things being equal, tends to inflate the width of the CI. However, in our previous meta-epidemiological studies of observational within-trial comparisons, the secondary analyses that took this correlation into account yielded similar results to the main analysis [12,13].

4.2. Other studies

Our results are consistent with, and expand on, those from our previous meta-epidemiological studies of binary and measurement scale outcomes [12,13]. The meta-epidemiological study of trials with time-to-event outcomes reported a ratio of hazard ratio of 0.88 (0.69–1.12), also indicating an exaggeration of the treatment effects by the non-blinded assessors [14]. We excluded such trials from the present study, as conversion from time-to-event to binary outcomes involves a series of assumptions we were not confident were valid for the trials in question.

A recent review of meta-epidemiological studies [88] aimed to evaluate the impact of blinding of outcome assessors and included two other observational within-trial comparisons in addition to our three meta-epidemiological studies. Both studies included trials that allowed for a comparison between local outcome assessment and blinded

central review or adjudication committees [89,90]. We consider the trials included in these comparisons different from those in the present study. Even if adjudication applies both to events and nonevents, the methods for central assessment are not always clearly reported, and bias can be introduced through knowledge of the result of the local outcome assessment and informative censoring [91]. Still, their results are consistent with our findings in terms of the direction of bias. Dello Russo et al. reported a ratio of hazard ratios of 0.98 (0.91–1.03) [89], and Ndounga Diakou et al. reported a ratio of ORs of 0.76 (0.48–1.21) [90]. In contrast, a similar study not in the review found no evidence of a difference between central adjudication and local assessment (ratio of treatment effects 1.02, 0.95–1.09) [92].

Previous meta-epidemiological studies of observational between-trial comparisons have seen a potential overestimation of treatment effects in trials without blinded outcome assessors [6]. However, the estimates of the average degree of bias had wide CIs, including considerable and no differences. MetaBLIND is the largest and most recent study including trials in multiple specialties and notoriously found no evidence of a difference between trials with and without blinded outcome assessment [7]. The difference between the impact observed in MetaBLIND and our present study can be partly explained by the risk of confounding that affects observational between-trial comparisons and differences in the characteristics of the trials included in the main analysis of both studies. For example, a larger proportion of trials in MetaBLIND evaluated experimental drug interventions (49% vs 27%), and only 1% evaluated surgical interventions (vs 41% in our present study). As suggested by our subgroup analysis, there may be a larger impact of lack of blinding of outcome assessors in non-drug trials. Further exploration of the characteristics of trials included in between- or within-trial comparisons can provide insights into the possible sources of variation in the estimated impact of lack of blinding of outcome assessors.

4.3. Mechanisms and implications

The impact of lack of blinding of outcome assessors, that is, observer bias, can be understood as the result of the interaction between assessors' predisposition and the subjectivity of the outcome. It is reasonable to expect that if assessors have little or no preconceptions, the impact of blinding would be less noticeable. However, it is challenging to evaluate reliably an assessor's preconceptions in any specific trial. Similarly, it is reasonable to expect that if an outcome is objective, that is, involving no judgment [93], blinding outcome assessors is irrelevant, as is the case for all-cause mortality and, possibly, non-repeatable laboratory outcomes.

Our findings and theoretical rationale suggest that for outcomes that require some degree of judgment by an

observer, there is a considerable risk that non-blinded assessors will give biased results, thus providing a solid argument in favor of the established methodological principle of blinding outcome assessors. Fortunately, it is possible to blind outcome assessors in most trials, including pragmatic effectiveness trials [94].

The difference between our results and those obtained from observational between-trial comparisons, such as the MetaBLIND study, emphasizes the limitations and risk of bias in the classic meta-epidemiological approach. For example, in addition to the risk of confounding, these studies have a comparatively higher risk of misclassifying the blinding status of the outcome assessor, which can bias the results [95]. Some studies rely on, sometimes reused, classifications based on risk of bias assessment, which considers additional trial aspects (eg, likelihood of outcome being affected by non-blinded assessor in specific contexts). Despite this, the MetaBLIND study has been cited as evidence against blinding outcome assessors. We suggest a cautious interpretation of the results of the increasing number of published observational between-trial comparisons and that our present study gives a more credible empirical assessment of observer bias.

This study provides additional empirical basis for the risk of bias assessment tools, for example, the Revised Cochrane risk of bias tool for randomized trials (RoB 2) [96]. Currently, RoB 2 guidance suggests that for trials with non-blinded outcome assessment, a higher risk of observer bias is expected if the control group receives no treatment or usual care, and if the outcome assessor is also involved in patient care. Our study adds that special attention should also be given to trials of non-drug interventions and trials funded by the drug and device industry.

In conclusion, we provide empirical evidence for the average degree of bias caused by the lack of blinding of outcome assessors in randomized clinical trials, as well as the range of bias, its general direction, and sources of heterogeneity. Non-blinded assessors of subjective outcomes exaggerate estimated effects, expressed as ORs, by an average of 29% (8%–45%). We suggest that this result, based on randomized between-trial comparisons and observational within-trial comparisons, is more trustworthy than that of meta-epidemiological studies based on observational between-trial comparisons. Outcome assessors in randomized clinical trials should be blinded whenever possible, except in trials with clearly objective outcomes.

Ethics statement

Not required.

Data statement

Data set available from the corresponding author.

CRedit authorship contribution statement

Josefina Salazar: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Helene Moustgaard:** Writing – review & editing, Validation, Methodology, Investigation, Data curation, Conceptualization. **Javier Bracchiglione:** Writing – review & editing, Validation, Investigation. **Asbjørn Hróbjartsson:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the following researchers for sharing unpublished outcome data with us: Andrew Jull, Andrea Nelson, Caroline Fairhurst, Yasser Haddadi, and Stig Brorson. We thank the authors from the 2 previous studies for sharing the data set: Ann Sofia Skou Thomsen, Frida Emanuelsson, Britta Tendal, Jørgen Hilden, Isabelle Boutron, Philippe Ravaud, and Stig Brorson. We also thank Lasse Adrup Benné Petersen and Mia Elkjær for their assistance with data extraction.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2025.111787>.

Data availability

Data will be made available on request.

References

- [1] Cumpston MFE, Thomas J, Higgins JPT, Deeks JJ, Clarke MJ. Chapter I: Introduction. In: Higgins JPT, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Cochrane. Available at: www.training.cochrane.org/handbook. Accessed June 17, 2024.
- [2] Kahan BC, Rehal S, Cro S. Blinded outcome assessment was infrequently used and poorly reported in open trials. *PLoS One* 2015; 10:e0131926. <https://doi.org/10.1371/journal.pone.0131926>.
- [3] Chaimani A, Vasiliadis HS, Pandis N, Schmid CH, Welton NJ, Salanti G. Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study. *Int J Epidemiol* 2013;42:1120–31. <https://doi.org/10.1093/ije/dyt074>.
- [4] Hartling L, Hamm MP, Fernandes RM, Dryden DM, Vandermeer B. Quantifying bias in randomized controlled trials in child health: a

- meta-epidemiological study. *PLoS One* 2014;9:e88008. <https://doi.org/10.1371/journal.pone.0088008>.
- [5] Armijo-Olivo S, Fuentes J, da Costa BR, Saltaji H, Ha C, Cummings GG. Blinding in physical therapy trials and its association with treatment effects: a meta-epidemiological study. *Am J Phys Med Rehabil* 2017;96:34–44. <https://doi.org/10.1097/phm.0000000000000521>.
 - [6] Haring R, Ghannad M, Bertizzolo L, Page MJ. No evidence found for an association between trial characteristics and treatment effects in randomized trials of testosterone therapy in men: a meta-epidemiological study. *J Clin Epidemiol* 2020;122:12–9. <https://doi.org/10.1016/j.jclinepi.2020.02.004>.
 - [7] Moustgaard H, Clayton GL, Jones HE, Boutron I, Jørgensen L, Laursen DRT, et al. Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. *BMJ* 2020; 368:l6802. <https://doi.org/10.1136/bmj.l6802>.
 - [8] Amer MA, Herbison GP, Grainger SH, Khoo CH, Smith MD, McCall JL. A meta-epidemiological study of bias in randomized clinical trials of open and laparoscopic surgery. *Br J Surg* 2021;108: 477–83. <https://doi.org/10.1093/bjs/zna035>.
 - [9] Bialy L, Vandermeer B, Lacaze-Masmonteil T, Dryden DM, Hartling L. A meta-epidemiological study to examine the association between bias and treatment effects in neonatal trials. *Evid Based Child Health* 2014;9:1052–9. <https://doi.org/10.1002/ebch.1985>.
 - [10] Wang Z, Alahdab F, Farah M, Seisa M, Firwana M, Rajjoub R, et al. Association of study design features and treatment effects in trials of chronic medical conditions: a meta-epidemiological study. *BMJ Evid Based Med* 2022;27:104–8. <https://doi.org/10.1136/bmjebm-2021-111667>.
 - [11] Stadelmaier J, Roux I, Petropoulou M, Schwingshackl L. Empirical evidence of study design biases in nutrition randomised controlled trials: a meta-epidemiological study. *BMC Med* 2022;20:330. <https://doi.org/10.1186/s12916-022-02540-9>.
 - [12] Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 2012;344:e1119. <https://doi.org/10.1136/bmj.e1119>.
 - [13] Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ* 2013;185:E201–11. <https://doi.org/10.1503/cmaj.120744>.
 - [14] Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Rasmussen JV, Hilden J, et al. Observer bias in randomized clinical trials with time-to-event outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *Int J Epidemiol* 2014;43:937–48. <https://doi.org/10.1093/ije/dyt270>.
 - [15] Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* 2004;23:1351–75. <https://doi.org/10.1002/sim.1761>.
 - [16] Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;19:3127–31. [https://doi.org/10.1002/1097-0258\(20001130\)19:22<3127::aid-sim784>3.0.co;2-m](https://doi.org/10.1002/1097-0258(20001130)19:22<3127::aid-sim784>3.0.co;2-m).
 - [17] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2).
 - [18] Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of meta-analysis: I(2) is not an absolute measure of heterogeneity. *Res Synth Methods* 2017;8:5–18. <https://doi.org/10.1002/jrsm.1230>.
 - [19] Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009; 172:137–59. <https://doi.org/10.1111/j.1467-985X.2008.00552.x>.
 - [20] Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: a tutorial. *Clin Epidemiol Glob Health* 2019;7:192–8. <https://doi.org/10.1016/j.cegh.2018.05.005>.
 - [21] Alam M, Pon K, Van Laborde S, Kaminer MS, Arndt KA, Dover JS. Clinical effect of a single pulsed dye laser treatment of fresh surgical scars: randomized controlled trial. *Dermatol Surg* 2006;32:21–5. <https://doi.org/10.1111/1524-4725.2006.32029>.
 - [22] Aro HT, Govender S, Patel AD, Hernigou P, Perera de Gregorio A, Popescu GI, et al. Recombinant human bone morphogenetic protein-2: a randomized trial in open tibial fractures treated with reamed nail fixation. *J Bone Joint Surg Am* 2011;93:801–8. <https://doi.org/10.2106/jbjs.L01763>.
 - [23] Ash K, Lord J, Zukowski M, McDaniel DH. Comparison of topical therapy for striae alba (20% glycolic acid/0.05% tretinoin versus 20% glycolic acid/10% L-ascorbic acid). *Dermatol Surg* 1998;24: 849–56. <https://doi.org/10.1111/j.1524-4725.1998.tb04262.x>.
 - [24] Baumann LS, Shamban AT, Lupo MP, Monheit GD, Thomas JA, Murphy DK, et al. Comparison of smooth-gel hyaluronic acid dermal fillers with cross-linked bovine collagen: a multicenter, double-masked, randomized, within-subject study. *Dermatol Surg* 2007;33(Suppl 2):S128–35. <https://doi.org/10.1111/j.1524-4725.2007.33352.x>.
 - [25] Bayer A, Kuznik N, Langan EA, Recke A, Recke AL, Faerber G, et al. Clinical outcome of short-term compression after sclerotherapy for telangiectatic varicose veins. *J Vasc Surg Venous Lymphat Disord* 2021;9:435–43. <https://doi.org/10.1016/j.jvsv.2020.05.015>.
 - [26] Box RC, Reul-Hirche HM, Bullock-Saxton JE, Furnival CM. Shoulder movement after breast cancer surgery: results of a randomised controlled study of postoperative physiotherapy. *Breast Cancer Res Treat* 2002;75:35–50. <https://doi.org/10.1023/a:1016571204924>.
 - [27] Brandstrup B, Tønnesen H, Beier-Holgersen R, Hjortsø E, Ørding H, Lindorff-Larsen K, et al. Effects of intravenous fluid restriction on post-operative complications: comparison of two perioperative fluid regimens: a randomized assessor-blinded multicenter trial. *Ann Surg* 2003;238:641–8. <https://doi.org/10.1097/01.sla.0000094387.50865.23>.
 - [28] Brorson S, Olsen BS, Frich LH, Jensen SL, Johannsen HV, Sørensen AK, et al. Effect of osteosynthesis, primary hemiarthroplasty, and non-surgical management for displaced four-part fractures of the proximal humerus in elderly: a multi-centre, randomised clinical trial. *Trials* 2009;10:51. <https://doi.org/10.1186/1745-6215-10-51>.
 - [29] Burkhoff D, Schmidt S, Schulman SP, Myers J, Resar J, Becker LC, et al. Transmyocardial laser revascularisation compared with continued medical therapy for treatment of refractory angina pectoris: a prospective randomised trial. ATLANTIC Investigators. *Angina Treatments-Lasers and Normal Therapies in Comparison*. *Lancet* 1999;354:885–90. [https://doi.org/10.1016/s0140-6736\(99\)08113-1](https://doi.org/10.1016/s0140-6736(99)08113-1).
 - [30] Cioba C, Marafon SB, Fortes BGB, Cavalheiro MT, Fabris M, Michel G, et al. Autologous fibrin glue versus sutures for conjunctival autograft in primary pterygium: a randomized clinical trial. *Int Ophthalmol* 2023;43:2371–81. <https://doi.org/10.1007/s10792-023-02635-z>.
 - [31] Cohen SR, Holmes RE. Arteccl: a long-lasting injectable wrinkle filler material: report of a controlled, randomized, multicenter clinical trial of 251 subjects. *Plast Reconstr Surg* 2004;114:964–76. <https://doi.org/10.1097/01.prs.0000133169.16467.5f>; discussion 77–9.
 - [32] Comella C, Hauser RA, Isaacson SH, Truong D, Oguh O, Hui J, et al. Efficacy and safety of two incobotulinumtoxinA injection intervals in cervical dystonia patients with inadequate benefit from standard injection intervals of botulinum toxin: phase 4, open-label, randomized, noninferiority study. *Clin Park Relat Disord* 2022;6:100142. <https://doi.org/10.1016/j.prdoa.2022.100142>.
 - [33] Dover JS, Rubin MG, Bhatia AC. Review of the efficacy, durability, and safety data of two nonanimal stabilized hyaluronic acid fillers from a prospective, randomized, comparative, multicenter study.

- Dermatol Surg 2009;35(Suppl 1):322–30. <https://doi.org/10.1111/j.1524-4725.2008.01060.x>; discussion 30–1.
- [34] Duffley G, Szabo A, Lutz BJ, Mahoney-Rafferty EC, Hess CW, Ramirez-Zamora A, et al. Interactive mobile application for Parkinson's disease deep brain stimulation (MAP DBS): an open-label, multicenter, randomized, controlled clinical trial. *Parkinsonism Relat Disord* 2023;109:105346. <https://doi.org/10.1016/j.parkreldis.2023.105346>.
- [35] Dumville JC, Worthy G, Bland JM, Cullum N, Dowson C, Iglesias C, et al. Larval therapy for leg ulcers (VenUS II): randomised controlled trial. *BMJ* 2009;338:b773. <https://doi.org/10.1136/bmj.b773>.
- [36] Fogarty JP. Observer bias in drug trial. *J Ir Med Assoc* 1966;59:161–3.
- [37] Govender S, Csimma C, Genant HK, Valentin-Opran A, Amit Y, Arbel R, et al. Recombinant human bone morphogenetic protein-2 for treatment of open tibial fractures: a prospective, controlled, randomized study of four hundred and fifty patients. *J Bone Joint Surg Am* 2002;84:2123–34. <https://doi.org/10.2106/00004623-200212000-00001>.
- [38] Haddadi Y, Bahrami G, Isidor F. Accuracy of crowns based on digital intraoral scanning compared to conventional impression-a split-mouth randomised clinical study. *Clin Oral Invest* 2019;23:4043–50. <https://doi.org/10.1007/s00784-019-02840-0>.
- [39] Hariz GM, Fredricks A, Stenmark-Persson R, Hariz M, Forsgren L, Blomstedt P. Blinded versus unblinded evaluations of motor scores in patients with Parkinson's disease randomized to deep brain stimulation or best medical therapy. *Mov Disord Clin Pract* 2021;8:286–7. <https://doi.org/10.1002/mdc3.13141>.
- [40] Havel CJ Jr, Strait RT, Hennes H. A clinical trial of propofol vs midazolam for procedural sedation in a pediatric emergency department. *Acad Emerg Med* 1999;6:989–97. <https://doi.org/10.1111/j.1553-2712.1999.tb01180.x>.
- [41] Herberger K, Franzke N, Blome C, Kirsten N, Augustin M. Efficacy, tolerability and patient benefit of ultrasound-assisted wound treatment versus surgical debridement: a randomized clinical study. *Dermatology* 2011;222:244–9. <https://doi.org/10.1159/000326116>.
- [42] Holmes JHT, Molnar JA, Shupp JW, Hickerson WL, King BT, Foster KN, et al. Demonstration of the safety and effectiveness of the RECELL System combined with split-thickness meshed autografts for the reduction of donor skin to treat mixed-depth burn injuries. *Burns* 2019;45:772–82. <https://doi.org/10.1016/j.burns.2018.11.002>.
- [43] Iglesia CB, Sokol AI, Sokol ER, Kudish BI, Gutman RE, Peterson JL, et al. Vaginal mesh for prolapse: a randomized controlled trial. *Obstet Gynecol* 2010;116:293–303. <https://doi.org/10.1097/AOG.0b013e3181e7d7f8>.
- [44] Jones AL, Bucholz RW, Bosse MJ, Mirza SK, Lyon TR, Webb LX, et al. Recombinant human BMP-2 and allograft compared with autogenous bone graft for reconstruction of diaphyseal tibial fractures with cortical defects. A randomized, controlled trial. *J Bone Joint Surg Am* 2006;88:1431–41. <https://doi.org/10.2106/jbjs.E.00381>.
- [45] Jull A, Wadham A, Bullen C, Parag V, Weller C, Waters J. Wool-derived keratin dressings versus usual care dressings for treatment of slow healing venous leg ulceration: a randomised controlled trial (Keratin4VLU). *BMJ Open* 2020;10:e036476. <https://doi.org/10.1136/bmjopen-2019-036476>.
- [46] Jull A, Walker N, Parag V, Molan P, Rodgers A. Randomized clinical trial of honey-impregnated dressings for venous leg ulcers. *Br J Surg* 2008;95:175–82. <https://doi.org/10.1002/bjs.6059>.
- [47] Kadish A, Nademanee K, Volosin K, Krueger S, Neelagaru S, Raval N, et al. A randomized controlled trial evaluating the safety and efficacy of cardiac contractility modulation in advanced heart failure. *Am Heart J* 2011;161:329–337.e1-2. <https://doi.org/10.1016/j.ahj.2010.10.025>.
- [48] Kern P, Ramelet AA, Wütschert R, Hayoz D. Compression after sclerotherapy for telangiectasias and reticular leg veins: a randomized controlled study. *J Vasc Surg* 2007;45:1212–6. <https://doi.org/10.1016/j.jvs.2007.02.039>.
- [49] Landsman AS, Robbins AH, Angelini PF, Wu CC, Cook J, Oster M, et al. Treatment of mild, moderate, and severe onychomycosis using 870- and 930-nm light exposure. *J Am Podiatr Med Assoc* 2010;100:166–77. <https://doi.org/10.7547/1000166>.
- [50] Lewin AB, Peris TS, De Nadai AS, McCracken JT, Piacentini J. Agreement between therapists, parents, patients, and independent evaluators on clinical improvement in pediatric obsessive-compulsive disorder. *J Consult Clin Psychol* 2012;80:1103–7. <https://doi.org/10.1037/a0029991>.
- [51] Li Q, Xiao M, Liu X, Zhao Y, Zhang H, Yin Y, et al. Enhancing open clinical trials through blinded evaluations: an exploration with diabetic foot infections. *Trials* 2023;24:716. <https://doi.org/10.1186/s13063-023-07652-y>.
- [52] Lorentzen J, Nielsen D, Holm K, Baagøe S, Grey MJ, Nielsen JB. Neural tension technique is no different from random passive movements in reducing spasticity in patients with traumatic brain injury. *Disabil Rehabil* 2012;34:1978–85. <https://doi.org/10.3109/09638288.2012.665132>.
- [53] Martin DF, Sierra-Madero J, Walmsley S, Wolitz RA, Macey K, Georgiou P, et al. A controlled trial of valganciclovir as induction therapy for cytomegalovirus retinitis. *N Engl J Med* 2002;346:1119–26. <https://doi.org/10.1056/NEJMoa011759>.
- [54] Meltzer HY, Alphs L, Green AI, Altamura AC, Anand R, Bertoldi A, et al. Clozapine treatment for suicidality in schizophrenia: international suicide prevention trial (InterSePT). *Arch Gen Psychiatry* 2003;60:82–91. <https://doi.org/10.1001/archpsyc.60.1.82>.
- [55] Miller RS, Steward DL, Tami TA, Sillars MJ, Seiden AM, Shete M, et al. The clinical effects of hyaluronic acid ester nasal dressing (Merogel) on intranasal wound healing after functional endoscopic sinus surgery. *Otolaryngol Head Neck Surg* 2003;128:862–9. <https://doi.org/10.1016/s0194-59980300460-1>.
- [56] Murtha AP, Kaplan AL, Paglia MJ, Mills BB, Feldstein ML, Ruff GL. Evaluation of a novel technique for wound closure using a barbed suture. *Plast Reconstr Surg* 2006;117:1769–80. <https://doi.org/10.1097/01.prs.0000209971.08264.b0>.
- [57] Narins RS, Coleman W, Donofrio L, Jones DH, Maas C, Monheit G, et al. Nonanimal sourced hyaluronic acid-based dermal filler using a cohesive polydensified matrix technology is superior to bovine collagen in the correction of moderate to severe nasolabial folds: results from a 6-month, randomized, blinded, controlled, multicenter study. *Dermatol Surg* 2010;36:730–40. <https://doi.org/10.1111/j.1524-4725.2010.01553.x>.
- [58] Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 1994;44:16–20. <https://doi.org/10.1212/wnl.44.1.16>.
- [59] Oesterle SN, Sanborn TA, Ali N, Resar J, Ramee SR, Heuser R, et al. Percutaneous transmyocardial laser revascularisation for severe angina: the PACIFIC randomised trial. *Potential Class Improvement from Intramyocardial Channels*. *Lancet* 2000;356:1705–10. [https://doi.org/10.1016/s0140-6736\(00\)03203-7](https://doi.org/10.1016/s0140-6736(00)03203-7).
- [60] Paithankar DY, Sakamoto FH, Farinelli WA, Kositratna G, Blomgren RD, Meyer TJ, et al. Acne treatment based on selective photothermolysis of sebaceous follicles with topically delivered light-absorbing gold microparticles. *J Invest Dermatol* 2015;135:1727–34. <https://doi.org/10.1038/jid.2015.89>.
- [61] Patsko E, Godolphin PJ, Thomas KS, Hepburn T, Mitchell EJ, Craig FE, et al. Investigating the effect of independent, blinded digital image assessment on the STOP GAP trial. *Trials* 2017;18:53. <https://doi.org/10.1186/s13063-017-1779-9>.
- [62] Powell NB, Zonato AI, Weaver EM, Li K, Troell R, Riley RW, et al. Radiofrequency treatment of turbinate hypertrophy in subjects using continuous positive airway pressure: a randomized, double-blind,

- placebo-controlled clinical pilot trial. *Laryngoscope* 2001;111:1783–90. <https://doi.org/10.1097/00005537-200110000-00023>.
- [63] Purdue GF, Hunt JL, Still JM Jr, Law EJ, Herndon DN, Goldfarb IW, et al. A multicenter clinical trial of a biosynthetic skin replacement, Dermagraft-TC, compared with cryopreserved human cadaver skin for temporary coverage of excised burn wounds. *J Burn Care Rehabil* 1997;18:52–7. <https://doi.org/10.1097/00004630-199701000-00009>.
- [64] Realmuto GM, Erickson WD, Yellin AM, Hopwood JH, Greenberg LM. Clinical comparison of thiothixene and thioridazine in schizophrenic adolescents. *Am J Psychiatry* 1984;141:440–2. <https://doi.org/10.1176/ajp.141.3.440>.
- [65] Reddiough DS, King JA, Coleman GJ, Fosang A, McCoy AT, Thomason P, et al. Functional outcome of botulinum toxin A injections to the lower limbs in cerebral palsy. *Dev Med Child Neurol* 2002;44:820–7. <https://doi.org/10.1017/s0012162201002997>.
- [66] Reynolds T, Russell L. Evaluation of a wound dressing using different research methods. *Br J Nurs* 2004;13:S21–4. <https://doi.org/10.12968/bjon.2004.13.Sup2.13236>.
- [67] Reynolds T, Russell L, Deeth M, Jones H, Birchall L. A randomised controlled trial comparing Drawtex with standard dressings for exuding wounds. *J Wound Care* 2004;13:71–4. <https://doi.org/10.12968/jowc.2004.13.2.26577>.
- [68] Schultheis M, Staubach P, Nikolakis G, Grabbe S, Ruckes C, von Stebut E, et al. LAight therapy significantly enhances treatment efficacy of 16 Weeks of topical clindamycin solution in hurley I and II hidradenitis suppurativa: results from period A of relieve, a multicenter randomized, controlled trial. *Dermatology* 2022;238:476–86. <https://doi.org/10.1159/000518540>.
- [69] Shanshanwal SJ, Dhurat RS. Superiority of dutasteride over finasteride in hair regrowth and reversal of miniaturization in men with androgenetic alopecia: a randomized controlled open-label, evaluator-blinded study. *Indian J Dermatol Venereol Leprol* 2017;83:47–54. <https://doi.org/10.4103/0378-6323.188652>.
- [70] Smith S, Busso M, McClaren M, Bass LS. A randomized, bilateral, prospective comparison of calcium hydroxylapatite microspheres versus human-based collagen for the correction of nasolabial folds. *Dermatol Surg* 2007;33(Suppl 2):S112–21. <https://doi.org/10.1111/j.1524-4725.2007.33350.x>. discussion S21.
- [71] Spinelli MG, Endicott J, Goetz RR. Disagreement between therapist raters and independent evaluators in a controlled clinical trial of interpersonal psychotherapy for depressed pregnant women. *J Psychiatr Pract* 2015;21:114–23. <https://doi.org/10.1097/01.pra.0000462604.79606.4e>.
- [72] Still J, Glat P, Silverstein P, Griswold J, Mozingo D. The use of a collagen sponge/living cell composite material to treat donor sites in burn patients. *Burns* 2003;29:837–41. [https://doi.org/10.1016/s0305-4179\(03\)00164-5](https://doi.org/10.1016/s0305-4179(03)00164-5).
- [73] Swiontkowski MF, Aro HT, Donell S, Esterhai JL, Goulet J, Jones A, et al. Recombinant human bone morphogenetic protein-2 in open tibial fractures. A subgroup analysis of data combined from two prospective randomized studies. *J Bone Joint Surg Am* 2006;88:1258–65. <https://doi.org/10.2106/jbjs.E.00499>.
- [74] Taber LH, Knight V, Gilbert BE, McClung HW, Wilson SZ, Norton HJ, et al. Ribavirin aerosol treatment of bronchiolitis associated with respiratory syncytial virus infection in infants. *Pediatrics* 1983;72:613–8.
- [75] Ulm G, Schüler P. Cabergolin versus pergolid: a video-blinded, randomised multicenter cross-over study. *Akt Neurol* 1999;26:360–5.
- [76] Wahidi MM, Reddy C, Yarmus L, Feller-Kopman D, Musani A, Shepherd RW, et al. Randomized trial of pleural fluid drainage frequency in patients with malignant pleural effusions. The ASAP trial. *Am J Respir Crit Care Med* 2017;195:1050–7. <https://doi.org/10.1164/rccm.201607-1404OC>.
- [77] Waibel KH, Golding H, Manischewitz J, King LR, Tuchscherer M, Topolski RL, et al. Clinical and immunological comparison of smallpox vaccine administered to the outer versus the inner upper arms of vaccinia-naïve adults. *Clin Infect Dis* 2006;42:e16–20. <https://doi.org/10.1086/499957>.
- [78] Wang X, Wu Y, Li B, Mu X, Li L. Lifting the midface using a hyaluronic acid filler with lidocaine: a randomized multi-center study in a Chinese population. *J Cosmet Dermatol* 2022;21:6710–6. <https://doi.org/10.1111/jocd.15286>.
- [79] Watson JM, Kang'ombe AR, Soares MO, Chuang LH, Worthy G, Bland JM, et al. VenUS III: a randomised controlled trial of therapeutic ultrasound in the management of venous leg ulcers. *Health Technol Assess* 2011;15:1–192. <https://doi.org/10.3310/hta15130>.
- [80] Weaver FM, Follett K, Stern M, Hur K, Harris C, Marks WJ Jr, et al. Bilateral deep brain stimulation vs best medical therapy for patients with advanced Parkinson disease: a randomized controlled trial. *JAMA* 2009;301:63–73. <https://doi.org/10.1001/jama.2008.929>.
- [81] Wedekind D, Broocks A, Weiss N, Engel K, Neubert K, Bandelow B. A randomized, controlled trial of aerobic exercise in combination with paroxetine in the treatment of panic disorder. *World J Biol Psychiatry* 2010;11:904–13. <https://doi.org/10.3109/15622975.2010.489620>.
- [82] Weiss RA, Moradi A, Bank D, Few J, Joseph J, Dover J, et al. Effectiveness and safety of large gel particle hyaluronic acid with lidocaine for correction of midface volume deficit or contour deficiency. *Dermatol Surg* 2016;42:699–709. <https://doi.org/10.1097/dss.00000000000000771>.
- [83] Wenkui Y, Ning L, Jianfeng G, Weiqin L, Shaoqiu T, Zhihui T, et al. Restricted peri-operative fluid administration adjusted by serum lactate level improved outcome after major elective surgery for gastrointestinal malignancy. *Surgery* 2010;147:542–52. <https://doi.org/10.1016/j.surg.2009.10.036>.
- [84] Wunderink RG, Giamarellos-Bourboulis EJ, Rahav G, Mathers AJ, Bassetti M, Vazquez J, et al. Effect and safety of meropenem-vaborbactam versus best-available therapy in patients with carbapenem-resistant enterobacteriaceae infections: the TANGO II randomized clinical trial. *Infect Dis Ther* 2018;7:439–55. <https://doi.org/10.1007/s40121-018-0214-1>.
- [85] US Food and Drug Administration. FDA summary of safety and effectiveness data: Hylaform. Silver Spring (MD): The Administration; 2004. Available at: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P030032>. Accessed August 5, 2011.
- [86] US Food and Drug Administration. FDA executive summary Medicis Aesthetics PMA P40024/s051. The Administration; 2011. Available at: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P040024S051>. Accessed August 5, 2011.
- [87] Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010;340:c723. <https://doi.org/10.1136/bmj.c723>.
- [88] Wang Y, Parpia S, Couban R, Wang Q, Armijo-Olivo S, Bassler D, et al. Compelling evidence from meta-epidemiological studies demonstrates overestimation of effects in randomized trials that fail to optimize randomization and blind patients and outcome assessors. *J Clin Epidemiol* 2024;165:111211. <https://doi.org/10.1016/j.jclinepi.2023.11.001>.
- [89] Dello Russo C, Cappoli N, Navarra P. A comparison between the assessments of progression-free survival by local investigators versus blinded independent central reviews in phase III oncology trials. *Eur J Clin Pharmacol* 2020;76:1083–92. <https://doi.org/10.1007/s00228-020-02895-z>.
- [90] Ndounga Diakou LA, Trinquart L, Hróbjartsson A, Barnes C, Yavchitz A, Ravaud P, et al. Comparison of central adjudication of outcomes and onsite outcome assessment on treatment effect estimates. *Cochrane Database Syst Rev* 2016;3:Mr000043. <https://doi.org/10.1002/14651858.MR000043.pub2>.
- [91] Dodd LE, Korn EL, Freidlin B, Jaffe CC, Rubinstein LV, Dancy J, et al. Blinded independent central review of progression-free survival

- in phase III clinical trials: important design element or unnecessary expense? *J Clin Oncol* 2008;26:3791–6. <https://doi.org/10.1200/jco.2008.16.1711>.
- [92] Godolphin PJ, Bath PM, Algra A, Berge E, Brown MM, Chalmers J, et al. Outcome assessment by central adjudicators versus site investigators in stroke trials: a systematic review and meta-analysis. *Stroke* 2019;50:2187–96. <https://doi.org/10.1161/strokeaha.119.025019>.
- [93] Moustgaard H, Bello S, Miller FG, Hróbjartsson A. Subjective and objective outcomes in randomized clinical trials: definitions differed in methods publications and were often absent from trial reports. *J Clin Epidemiol* 2014;67:1327–34. <https://doi.org/10.1016/j.jclinepi.2014.06.020>.
- [94] Hansson L, Hedner T, Dahlöf B. Prospective Randomized Open Blinded End-Point (PROBE) study. A novel design for intervention trials. prospective randomized open blinded end-point. *Blood Press* 1992;1:113–9. <https://doi.org/10.3109/08037059209077502>.
- [95] Moustgaard H, Jones HE, Savović J, Clayton GL, Sterne JA, Higgins JP, et al. Ten questions to consider when interpreting results of a meta-epidemiological study-the MetaBLIND study as a case. *Res Synth Methods* 2020;11:260–74. <https://doi.org/10.1002/jrsm.1392>.
- [96] Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:14898. <https://doi.org/10.1136/bmj.14898>.