



Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey

Lars G Hemkens,^{1,2} Despina G Contopoulos-Ioannidis,^{3,4} John P A Ioannidis^{1,4-6}

¹Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

²Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Basel, Switzerland

³Department of Pediatrics, Division of Infectious Diseases, Stanford University School of Medicine, Stanford, California, USA

⁴Meta-Research Innovation Center at Stanford (METRICS)

⁵Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California, USA

⁶Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, California, USA

Correspondence to: J P A Ioannidis
jioannid@stanford.edu

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;352:i493
<http://dx.doi.org/10.1136/bmj.i493>

Accepted: 08 January 2016

ABSTRACT

OBJECTIVE

To assess differences in estimated treatment effects for mortality between observational studies with routinely collected health data (RCD; that are published before trials are available) and subsequent evidence from randomized controlled trials on the same clinical question.

DESIGN

Meta-epidemiological survey.

DATA SOURCES

PubMed searched up to November 2014.

METHODS

Eligible RCD studies were published up to 2010 that used propensity scores to address confounding bias and reported comparative effects of interventions for mortality. The analysis included only RCD studies conducted before any trial was published on the same topic. The direction of treatment effects, confidence intervals, and effect sizes (odds ratios) were compared between RCD studies and randomized controlled trials. The relative odds ratio (that is, the summary odds ratio of trial(s) divided by the RCD study estimate) and the summary relative odds ratio were calculated across all pairs of RCD studies and trials. A summary relative odds ratio greater than one indicates that RCD studies gave more favorable mortality results.

RESULTS

The evaluation included 16 eligible RCD studies, and 36 subsequent published randomized controlled trials

investigating the same clinical questions (with 17 275 patients and 835 deaths). Trials were published a median of three years after the corresponding RCD study. For five (31%) of the 16 clinical questions, the direction of treatment effects differed between RCD studies and trials. Confidence intervals in nine (56%) RCD studies did not include the RCT effect estimate. Overall, RCD studies showed significantly more favorable mortality estimates by 31% than subsequent trials (summary relative odds ratio 1.31 (95% confidence interval 1.03 to 1.65; $I^2=0\%$)).

CONCLUSIONS

Studies of routinely collected health data could give different answers from subsequent randomized controlled trials on the same clinical questions, and may substantially overestimate treatment effects. Caution is needed to prevent misguided clinical decision making.

Introduction

Routinely collected health data (RCD), such as electronic health records or patient registries, are proposed to assess comparative treatment effects of medical interventions. In theory, observational studies collecting this type of data could complement randomized controlled trials.¹ The most important limitation of RCD studies is their inherent risk of bias due to confounding by indication. While only proper randomization can pre-emptively eliminate such bias, approaches such as propensity scores are frequently used to deal with bias in observational research. The propensity score reflects the probability that a patient will be selected for a treatment and is estimated by use of information on known factors affecting the treatment choice, for example, disease severity.^{2,3} Many other methods are increasingly used, but propensity scores are probably the most popular method used to inform healthcare decisions.^{3,4,5} Studies using data not collected for the purpose of a specific research project face many challenges and are prone to various specific biases related to the very nature of this data.¹ A major challenge is the accuracy and reliability of the collected data, which is typically lower than many clinical trials with standardized and predefined outcome assessments. This might be less problematic for mortality, because it is an unambiguous outcome and less prone to data accuracy problems.

Although their limitations should not be underestimated,¹ RCD studies could provide the best available evidence to inform healthcare decisions when randomized controlled trials are not available. However, it is unknown whether such studies offer highly reliable answers on vital clinical questions, for example,

WHAT IS ALREADY KNOWN ON THIS TOPIC

Observational studies using routinely collected data (RCD studies) are increasingly used to inform healthcare decisions when RCTs are not available
However, observational studies have an inherent risk of bias due to confounding by indication
Another difficulty is the accuracy and reliability of routinely collected data

WHAT THIS STUDY ADDS

RCD studies systematically and substantially overestimate mortality benefits of medical treatments compared with subsequent trials investigating the same question
Observational RCD studies might not necessarily provide very reliable answers on how to best treat patients; caution is needed to prevent misguided clinical decision making
If no randomized trials exist, clinicians and funders of care should consider that treatment effects are probably more uncertain and substantially smaller than RCD studies suggest; decisions for widespread adoption and reimbursement of expensive interventions might be best withheld until trial evidence becomes available

whether the estimated treatment effects from RCD studies agree with effects demonstrated in subsequent randomized controlled trials. Most RCD studies are published on questions where there is already available evidence from trials. For example, a 2010 survey showed that almost 70% of 337 RCD studies based on propensity scores already had randomized controlled trials published on the same question.⁶ It is likely that the authors of these RCD studies may be consciously or unconsciously influenced by the already available results of the respective trials. To directly assess whether RCD studies can predict the results of subsequent randomized controlled trials, one needs to focus on topics where no prior trial evidence is available to influence what might be considered as reasonable effects to report by the RCD studies.

We therefore aimed to obtain insights on the concordance between RCD studies and randomized controlled trials with a comprehensive meta-epidemiological study. The present study used RCD studies that analyzed a critical healthcare question, used propensity scores to deal with bias, and evaluated effects on mortality. We systematically compared the findings from such studies on various clinical questions (which have never been addressed in trials before), with the findings from subsequent randomized controlled trials.

Methods

Eligibility criteria and identification of routine data studies

Eligible RCD studies compared one treatment with another or no intervention, usual care, or standard treatment; were performed before any randomized controlled trial on the same clinical question; assessed mortality effects; and used propensity scores based analyses for mortality. We considered studies that used only data that were routinely collected. Any type of such data was considered eligible,^{7,8} including those from health insurance claims, electronic health or medical records, and registries (even if registries also comprised some actively collected data for the purpose of the registry rather than only passive, routine data collection).⁹ We considered studies evaluating drugs, biologics, dietary supplements, devices, diagnostic procedures, surgeries, or radiotherapies in any patient population with any condition, and mortality outcome (all cause or cause specific) that were published in English. We included studies published up to 2010 to ensure sufficient time for randomized evidence, if any, to appear.

We searched PubMed (last search November 2014) combining terms for RCD (such as “routine*”, “database*”, “claim*”, “health record*”, “registr*”, and covering all terms used in the National Library of Medicine search strategy for electronic health or medical records¹⁰), with terms for mortality and propensity scores. For further details on inclusion criteria, definitions, and search strategies, see reference 6. One reviewer (LGH) screened titles and abstracts and obtained full texts of potentially relevant articles and determined eligibility.

Data extraction from RCD studies

For each eligible study, we extracted all clinical questions reported in the abstract following the PICO structure (patient, intervention, comparison, outcome).¹¹ We formulated separate clinical questions for each combination of patients and compared interventions (experimental and comparator) for which any result was reported in the abstract. We considered clinically relevant variations of treatment characteristics (such as timing or dose) or patient conditions (eg, comorbidities) as separate PICO clinical questions. We also considered specific subquestions separately—such as when the main comparison looked at coronary stenting versus no stenting, and subanalyses compared drug eluting stents with bare metal stents separately. We did not consider separately specific age subgroups within adult populations and demographic subpopulations (sex, race, or ethnicity).

For each clinical question, we searched the complete article for a comparative effect between the compared interventions on mortality outcomes based on analyses that used propensity scores in any way (adjustment, selection of compared populations, both, or other). If we identified such an effect estimate, we screened the full text and references for randomized evidence on the same clinical question (not necessarily evaluating mortality outcomes). We excluded any clinical questions with existing prior trial evidence. We then extracted data on RCD study characteristics and the mortality effect estimate with 95% confidence intervals. If a study reported multiple estimates, we used the analysis with results first mentioned in the abstract (as a prespecified rule to avoid subjectivity in the selection of effects). One reviewer (LGH) extracted the data and screened the articles.

Eligibility criteria and identification of randomized controlled trials

For each eligible clinical question, we systematically searched PubMed (to November 2014) for randomized controlled trials or systematic reviews or meta-analyses of trials that also addressed this question and reported any mortality outcome. We created standardized search strategies for each topic by combining search terms for the intervention, comparator, and condition. We used the PubMed standard filters for study design, limited results to the English language, and added terms for mortality to increase specificity when we searched for trials and diagnostic topics (web appendix 1 and reference 6). For RCD studies published up to 2007, we also searched all relevant modules of the Cochrane Library, but found no pertinent randomized controlled trial that was not also identified via PubMed; thus for newer RCDs, we only searched PubMed.

We screened titles and abstracts, obtained full texts of potentially relevant articles and determined eligibility. The resulting randomized controlled trials derived from these searches were considered for further analyses. We tested the completeness of our search by using the related articles function in PubMed for each eligible trial (screening the first 20 related articles), and in no

case we found an additional trial. These processes were all done by one reviewer (LGH) who marked studies if he was uncertain about eligibility. These studies were discussed with a second reviewer (DCI), who also confirmed the eligibility of all identified pertinent trials and spot checked all excluded full texts for verification. Discrepancies were discussed to reach consensus. We excluded from further analyses any clinical questions for which preceding trials (that is, published up until the year before the RCD study was published) were identified with the above searches.

Data extraction from randomized controlled trials

For each eligible trial, we extracted the number of randomized patients and deaths per treatment group (we preferred intention to treat data wherever possible). If a trial had multiple mortality endpoints, we preferred the same type of outcome definition as in the RCD study (all cause or cause specific mortality) and the most similar follow-up period (eg, in-hospital and 30 day mortality). We extracted the proportions of patients not initiating the randomized treatment and patients switching to the non-allocated treatment during the study (treatment crossover). Data extraction was performed by one reviewer (LGH).

Risk of bias assessment

We assessed the risk of bias for RCD studies (DCI, JPAI) and randomized controlled trials (LGH, and an external researcher experienced in systematic reviewing), using the Cochrane risk of bias tools.^{12,13} Discrepancies were discussed to reach consensus.

Statistical analysis

For consistency, we inverted the RCD effect estimates where necessary so that each RCD study indicated an odds ratio less than 1 (that is, swapping the study groups so that the first study group has lower mortality risk than the second). We assumed that reported relative risks or hazard ratios were approximations to the odds ratio, a reasonable assumption because death was a relatively uncommon event (median across treatment comparisons 3% (interquartile range 2-9%)). For each clinical question, we also calculated the odds ratio for mortality using data from randomized controlled trials for the same clinical question. Multiple trials were meta-analytically combined with random effects models to obtain a summary odds ratio.¹⁴ We used Peto's approach for event rates less than 1%.¹⁵

We recorded how frequently the treatment effect estimates from RCD studies and randomized controlled trials were in the opposite direction, how often the confidence intervals did not overlap, and how often the RCD study's confidence interval did not include the effect estimate demonstrated by later available trials.

We also calculated for each clinical question the relative odds ratio (ratio of odds ratios) by dividing the summary odds ratio of all subsequent randomized controlled trials by the estimated odds ratio in the RCD study. Confidence intervals of relative odds ratios were calculated by use of the sum of the variances of the trial

summary odds ratio and of the RCD study odds ratio estimate. We then combined the individual relative odds ratios across all questions to calculate the summary value. A summary relative odds ratio greater than 1 indicates that the RCD study found more favorable mortality outcomes than subsequent trials. Calculations were done after log-transformation.

We conducted several sensitivity analyses:

- Used fixed effect models instead of random effects models to combine effect sizes from randomized controlled trials¹⁴
- Excluded trials with a high risk of bias
- Excluded trials reporting high treatment crossover rates (>20% in any group) or asymmetric crossover (between group difference >10%)
- Included only trials clearly reporting low treatment crossover rates (<10% in all groups)
- Excluded trials with frequent non-initiation of randomized treatment (>10% in any group)
- Excluded trials in which the median age differed by more than two standard deviations from the median age in the RCD study
- Used the effect estimates from two mutually exclusive patient subgroups instead of the main effect from one RCD study¹⁶ and compared them with the summary odds ratio for the trials representing effects specifically for these subgroups
- Excluded one clinical question where all pertinent trials were already used for another treatment comparison^{17,18}
- Used only trials identified by search strategies of existing systematic reviews
- Included only RCD studies with low risk of bias for all assessed domains (with the exception of "bias due to confounding," which was deemed moderate for all RCD studies).

We used Stata 13.1 (Stata Corp) for all analyses and reported 95% confidence intervals. All P values were two tailed.

Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results

In the search for RCD studies, we identified 929 records and evaluated 420 in full text (fig 1). We found preceding randomized evidence on all evaluated clinical questions in 231 studies, did not find any subsequent randomized controlled trials in 90 studies, and excluded 83 studies for different reasons (fig 1). We eventually analyzed 16 RCD studies on clinical questions that did not have preceding trials and for which subsequent pertinent trials were identified (table 1). One study reported on three clinical questions with one

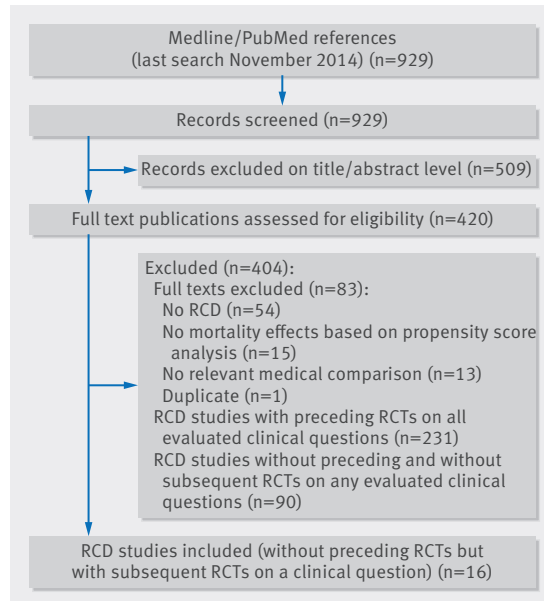


Fig 1 | Study flow diagram. RCT=randomized controlled trial

primary result (which we included in our main analysis) and two subgroup effects (included alternatively in sensitivity analyses).¹⁶

RCD studies were published between 2000 and 2010 and used diverse types of routine data including registries, hospital databases, and administrative data. Most studies were relevant to cardiology (12 (75%) of 16), and 11 (69%) compared two active interventions. All RCD studies assessed all cause mortality, and comparative effect estimates were based on a median of 2086 patients per analysis (interquartile range 734-8658; table 1). While we deemed the risk of bias due to confounding moderate for all studies, most had a low risk of bias for other types of bias. The overall risk of bias was therefore low to moderate for all RCD studies (web appendix 2).

We identified 36 subsequent randomized controlled trials³²⁻³³⁻⁶⁷ with 17 275 patients and 835 deaths overall, addressing the same clinical question as the RCD studies. All trials reported all cause mortality, and were published between 2003 and 2014, a median of three years after the RCD study. For each clinical question, we

Table 1 | Description of analyzed treatment comparisons in routinely collected data studies

RCD study	Treatment comparison	Condition or disease	Total no of patients (groups 1/2)	Follow-up	Data source
Holman 2000* ¹⁹	Preincision prophylactic intra-aortic balloon pump v no pump	CABG; stable, high risk patients	7581 (592/6989)	Unclear	Administrative data: "Alabama CABG Cooperative Project database," USA
Shavelle 2002 ²⁰	Very early angiography (within 6h) v early conservative treatment (no or later angiography)	NSTEMI	2810 (1405/1405)	In hospital	National Registry of Myocardial Infarction 2 database, USA
Winkelmayer 2002 ²¹	Hemodialysis v peritoneal dialysis	Renal replacement therapy	2503 (1966/537)	3 months	Medicare/Medicaid, USA
Karthik 2003 ²²	Off pump v on pump CABG	CABG, non-elective	828 (417/411)	In hospital	Hospital database: "cardiac surgery registry," UK
Guru 2006 ²³	2 or 3 arterial grafts v 1 arterial graft	CABG	10982 (5491/5491)	4-6 years (average)	Linked clinical and administrative data: "Cardiac Care Network Database," Canada
Wu 2008 ²⁴	CABG v drug eluting stent	Unprotected, left main coronary artery stenosis	112 (56/56)	8 months	Registries: "New York State Cardiac Surgery Reporting System" and "Percutaneous Coronary Intervention Reporting System," USA
Ascione 2003 ²⁵	On pump v off pump CABG	CABG, severe left ventricular dysfunction	250 (176/74)	In hospital	Hospital database: "Patient Analysis and Tracking System," UK
Polkinghorne 2004 ²⁶	Native arteriovenous fistula v arteriovenous grafts	Hemodialysis	2632 (2261/371)	1.07 years	Australian and New Zealand Dialysis and Transplant Association Registry, Australia and New Zealand
Gnerlich 2007 ²⁷	Surgical removal of the primary tumor v no surgical removal	Metastatic breast cancer (stage IV)	9734 (4578/5156)	≤15 years	Registry: "Surveillance, Epidemiology, and End Results," USA
Lindenauer 2004 ²⁸	Early perioperative lipid lowering treatment (during first 2 hospital days) v no or later treatment	Major non-cardiac surgery	204885 (73050/131835)	In hospital	Hospital discharge and pharmacy records of 329 hospitals, USA
Butler 2009 ¹⁸	Clopidogrel treatment duration ≥12 months v ≤6 months	Drug eluting stenting	1669 (1022/647)	1 year	Melbourne Interventional Group Registry, Australia
Cabell 2005† ²⁹	Surgery v non surgical treatment	Infective endocarditis, native valve	299 (94/205)	In hospital	Registry: "International Collaboration on Endocarditis Merged Database," Europe and USA
Kim 2009 ¹⁶	Clopidogrel initiation after CABG v no clopidogrel	CABG, preoperative treatment with aspirin	15067 (3268/11799)	In hospital	Clinical, administrative and financial data: "University Health System Database," USA
Moss 2003 ³⁰	Mitral valve repair v mitral valve replacement	Mitral valve surgery	644 (322/322)	3.4 years	British Columbia Cardiac Registry, Canada
Fonarow 2008 ³¹	Beta blocker continuation v beta blocker withdrawal	Heart failure, decompensated	1429 (79/1350)	3 month post-discharge	Registry: OPTIMIZE/HF program, USA
Hahn 2010‡ ¹⁷	Clopidogrel treatment duration 3 month v >3 months	Drug eluting stenting	823 (661/151)	1 year	Duration of Dual Antiplatelet Therapy After Implantation of Endeavour Stent Registry, Republic of Korea

CABG=coronary artery bypass graft; HF=heart failure; NSTEMI=non-ST elevation myocardial infarction; RCD=routinely collected health data.

*We used the only available mortality effect reported as relative effects measure with confidence intervals because no mortality results were reported in the abstract.

†We used mortality rates reported for the median propensity stratum and calculated odds ratios because no mortality results were reported in the abstract.

‡For this treatment comparison, both identified subsequent randomized controlled trials are also relevant for another treatment comparison (Butler 2009).

included a median of 985 randomized patients (interquartile range 287-1696; fig 2 and fig 3). We deemed the risk of bias high for 10 trials, mainly due to lack of blinding (web appendix 3).

Agreement of treatment effects

Across 16 clinical questions, eight RCD studies found significant treatment effects (fig 4). Confidence intervals were wide and overlapped between RCD studies and randomized controlled trials in all 16 treatment comparisons. However, in more than half of cases (nine of 16; 56%), the confidence intervals of the RCD based estimate did not include the mortality effect found in subsequent randomized trial evidence. For five (31%) of 16 clinical questions, treatment effects from randomized evidence were in the opposite direction to the RCD study estimate. None of these five trial estimates was significant, and one RCD study estimate was significant.

When data were synthesized, RCD studies showed significantly inflated results compared with randomized

controlled trials, with an average overestimation of mortality benefits by 31% (summary relative odds ratio 1.31 (95% confidence interval 1.03 to 1.65); table 2, fig 4). There was no heterogeneity between topics ($I^2=0%$ (0% to 45%)). The results were quite similar in all sensitivity analyses (table 2), with estimates of summary relative odds ratios ranging between 1.20 and 1.34 and their 95% confidence intervals excluding the null in six of the 10 sensitivity analyses. We found the smallest estimate of a difference between RCD studies and trials (summary relative odds ratio 1.20) when we considered only RCD studies with a low risk of bias on all dimensions (except for confounding bias, where a moderate risk is probably the best one can expect for this type of study design).

Discussion

Principal findings

In our comprehensive analysis of various clinical questions on topics never evaluated in randomized controlled trials before, we found that studies using

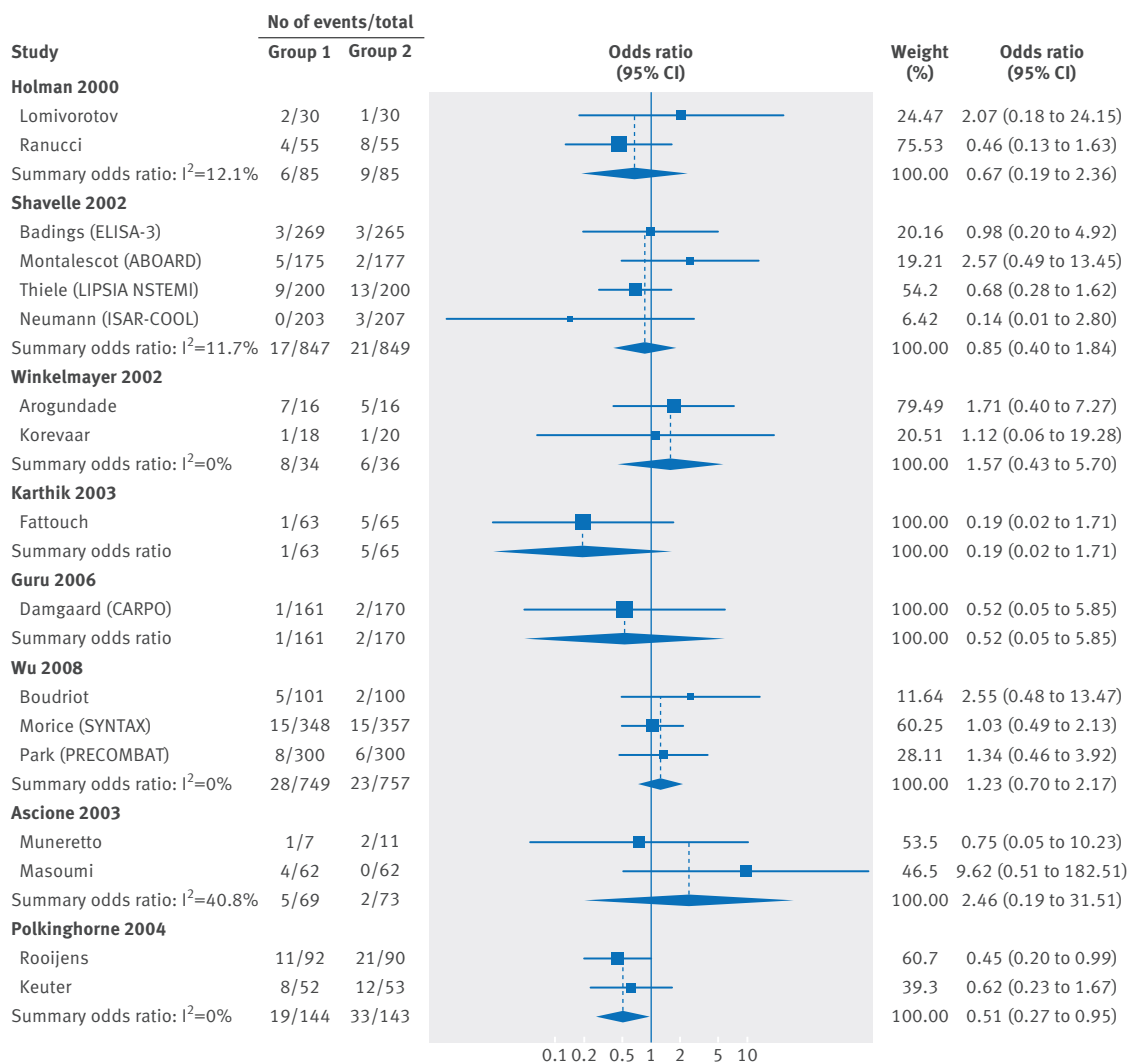


Fig 2 | Meta-analyses of comparative effects of medical interventions on mortality reported in randomized controlled trials published after the same clinical question was investigated in RCD studies (part one). For each clinical question investigated in a RCD study, the trials published subsequently are shown. Diamonds=result of meta-analyses combining these subsequent trials as summary odds ratios (using random effects models)

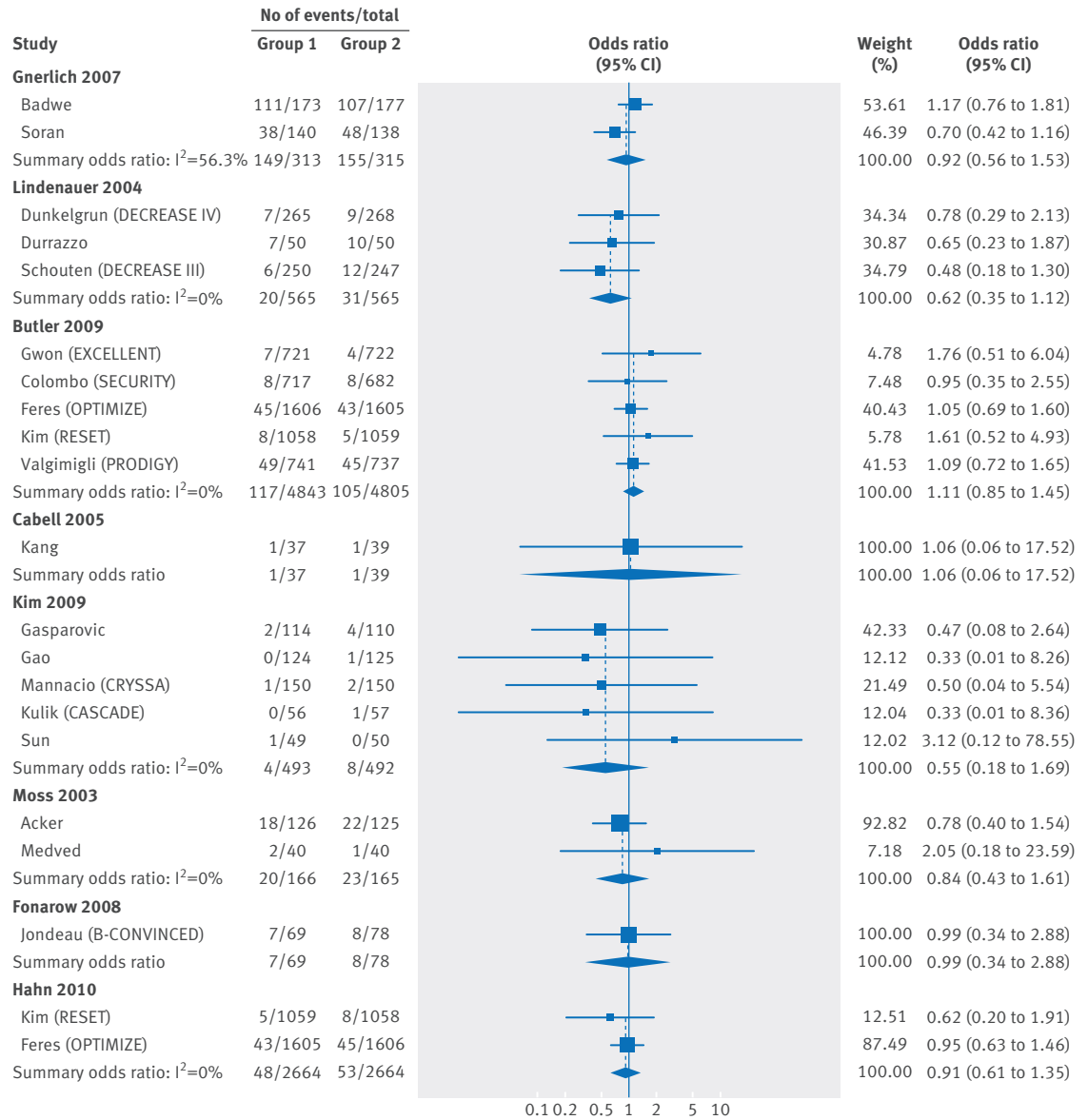


Fig 3 | Meta-analyses of comparative effects of medical interventions on mortality reported in randomized controlled trials published after the same clinical question was investigated in RCD studies (part two). For each clinical question investigated in a RCD study, the trials published subsequently are shown. Diamonds=result of meta-analyses combining these subsequent trials as summary odds ratios (using random effects models)

routinely collected health data frequently do not agree with subsequent randomized trials. We analyzed 16 clinical questions with 36 corresponding subsequent trials published a median of three years later. Although our results need to be interpreted cautiously given the relatively small numbers of studies, the emerging pattern was that RCD studies systematically and substantially overestimated the mortality benefits of medical treatments compared with subsequent trials investigating the same question.

The overall findings suggest that results from RCD studies in the absence of randomized controlled trials need to be seen with substantial caution. RCD studies might not necessarily provide reliable answers on how to best treat patients. As an example, the clinical consequences might be illustrated by the clinical question in

our analysis with the largest body of randomized evidence—that is, on the duration of clopidogrel treatment after use of drug eluting stents.¹⁸ Here, the RCD based estimate suggested substantial and significant reductions in mortality (odds ratio 0.59 (95% confidence interval 0.35 to 0.99)), leaving the study authors to conclude that “longer (≥12 months) planned duration of clopidogrel results in reduced 12-month mortality . . . Randomized studies are urgently needed to address this issue.”¹⁸ However, later trial evidence showed no benefit of longer clopidogrel treatment and rather indicated harm, and the confidence intervals were not compatible with the early findings in the RCD study (odds ratio 1.11 (95% confidence interval 0.85 to 1.45)). This shows that RCD studies have a substantial risk of misleading patient care.¹

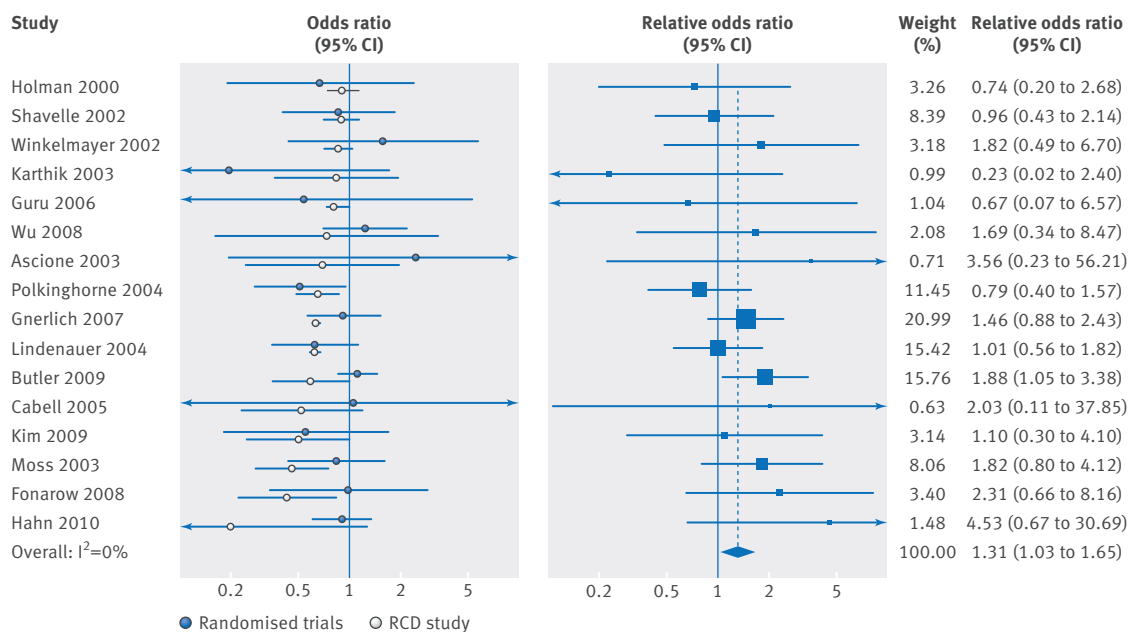


Fig 4 Treatment effects on mortality in RCD studies and randomized controlled trials. Left panel shows comparative effects of medical interventions on mortality reported in RCD studies and results of subsequently published trials on the same treatment comparisons. White circles=effect estimates reported in RCD studies; blue circles=pooled summary effects from subsequent trials (corresponding meta-analyses are shown in fig 2 and fig 3); lines=95% confidence intervals. Right panel shows for each clinical question the ratio of mortality effects reported in trial evidence versus RCD study effects (as relative odds ratios). Blue squares (lines)=relative odds ratio (95% confidence intervals); blue diamond=pooled summary relative odds ratio (meta-analysis of relative odds ratio) across all clinical questions. A relative odds ratio greater than 1 indicates more favorable mortality outcomes in RCD studies than in subsequent trials

Comparison with other studies

A recent Cochrane review identified 14 previous meta-epidemiological studies comparing randomized and observational study results.⁶⁸ Most focused on traditional observational epidemiology rather than on RCD studies, and only two meta-epidemiological analyses compared propensity score analyses with randomized controlled trials.^{69,70} A further empirical evaluation was excluded from the Cochrane review.⁷¹

In their analysis of mortality effects across 22 clinical questions in the field of surgery, Lonjon and colleagues found a point estimate of a summary relative

odds ratio that was similar to our analysis (1.20, 95% confidence interval 0.96 to 1.54; original results inverted to allow comparison with this study).⁶⁹ For subjective outcomes, they found a summary relative odds ratio close to 1 (0.93, 95% confidence interval 0.75 to 1.15). The authors interpreted the lack of statistical difference between study designs as evidence for equivalent effects. However, 20-30% relative changes in the odds of mortality are substantial, because most differences in mortality with treatments across medicine are of this magnitude or even smaller.⁷² Kuss and colleagues analyzed only one treatment comparison

Table 2 | Agreement of treatment effects reported in RCD studies and subsequent randomized trial evidence

Analysis	No of treatment comparisons	Summary ROR (95% CI)	I ² (%; 95% CI)
Main analysis			
Random effect models to combine RCTs	16	1.31 (1.03 to 1.65)	0 (0 to 45)
Sensitivity analyses			
Fixed effect models to combine RCTs	16	1.34 (1.09 to 1.63)	0 (0 to 45)
Exclusion of RCTs with high risk of bias	14	1.21 (0.92 to 1.59)	0 (0 to 47)
Exclusion of RCTs with high treatment crossover rates or with asymmetric crossover	15	1.34 (1.05 to 1.70)	0 (0 to 46)
RCTs that had low treatment crossover rates	12	1.27 (0.92 to 1.76)	0 (0 to 51)
Exclusion of RCTs with frequent non-initiation of randomized treatment	16	1.31 (1.04 to 1.65)	0 (0 to 45)
Exclusion of RCTs with age differences >2 SD	16	1.28 (1.01 to 1.62)	0 (0 to 45)
Two subquestions instead of main question	15	1.33 (1.05 to 1.68)	0 (0 to 46)
Exclusion of treatment comparison with RCTs that are also pertinent for another treatment comparison	15	1.28 (1.01 to 1.62)	0 (0 to 46)
RCTs that were identified in available systematic reviews	10	1.21 (0.88 to 1.65)	0 (0 to 53)
RCD studies with low risk of bias except for "bias due to confounding"*	11	1.20 (0.89 to 1.62)	0 (0 to 51)

RCD=routinely collected health data; RCT=randomized controlled trial; ROR=relative odds ratio; SD=standard deviation.

*All RCD studies were deemed to have moderate risk of bias due to confounding.

(off pump on pump cardiac bypass surgery) and similarly interpreted lack of statistical difference as signaling equivalence.⁷⁰ Dahabreh and colleagues analyzed mortality effects of treatments in the setting of acute coronary syndrome.⁷¹ They also found that propensity score analyses gave significantly larger effect sizes than RCTs.⁷⁰

Strengths and limitations of study

All these previous empirical evaluations were restricted to specific topics and none evaluated clinical questions where all the data from randomized controlled trials were published subsequently to the RCD studies. However, many RCD studies are specifically undertaken to explore whether trials results can be replicated in the real world.⁶ In such cases, the trial evidence provides some prior knowledge that could inhibit the publication of findings that deviate greatly from the trial experience. Thus, our approach provides a more clean assessment of the ability of RCD results to predict the results of trials.

Some caveats should be considered in our study. Although we screened many RCD studies using propensity scores, only a fraction of the entire RCD literature was eligible for our analyses. This was largely due to the high number of clinical questions that were already addressed by some randomized trials, as we have previously discussed.¹⁶ However, we followed a systematic approach to derive a reproducible sample of RCD studies that covers a wide range of diverse healthcare questions. Although many relate to cardiovascular conditions, they represent various types of interventions, including surgery, devices, drug treatment, or treatment concepts. The generalizability to other conditions and diseases might also need to be assessed in the future.

The RCD studies included in our sample encompass a wide spectrum of data sources, from administrative hospital databases to committed registries. These data sources might differ with regard to their granularity, validation processes, and completeness. The sample was too small to allow a meaningful evaluation of differences across different subgroups of routine data sources. We have no detailed information on the accuracy of the key information of interest for our analyses (mortality and treatment allocation). Although we assume high data accuracy given the type of outcome (death is difficult to err on) and the clinical prominence of the assessed interventions, we cannot rule out that accuracy problems further reduce the reliability of such research.

Our PubMed search strategy for subsequent randomized controlled trials was relatively specific. It would be difficult to conduct thorough systematic reviews from scratch with highly sensitive search strategies for all the 106 RCD studies without preceding trials that we evaluated. Instead, we used a standardized search approach, systematically integrated existing systematic reviews and validated the search results with alternative identification algorithms—that is, the related article function in PubMed. Although the number of included clinical

questions with pairs of RCD study and trials could have been higher with a more sensitive strategy for subsequent trials, we had similar results in sensitivity analyses restricted to trial results obtained from search strategies of existing systematic reviews.

We assessed only mortality effects. Other more subjective clinical outcomes would probably be collected less accurately in the routinely collected datasets. This might further reduce the validity of treatment effect estimates and further limit the reliability of RCD studies to guide clinical decision making. Conversely, some other types of outcomes might have much larger treatment effects than mortality, and thus it might be easier to separate from noise due to bias in RCD studies. However, treatment benefits for other outcomes (eg, hospital admission) might not necessarily translate to benefits for mortality or other hard benefits.⁷³

We compared the RCD effects with early evidence from subsequent randomized trials that sometimes overestimates treatment effects.⁷⁴ Thus, our results even might be conservative and we may have underestimated the inflated and optimistic effects from RCD studies.

Randomized controlled trials are not necessarily a perfect gold standard. When their results differ against those of observational studies on the same question,^{75 76} it may not be certain that the trials are correct and the observational data are wrong. We explored the potential effect of risk of bias in the randomized and non-randomized studies. None of the RCD studies and only a few trials were deemed to have high risk of bias. When we compared only the effects from studies without high bias potential, we found similar effects as in the main analysis.

We used intention to treat effects for our comparison with RCD studies, because this is the most robust approach against bias. Such effects could be conservative in trials without active controls, low adherence to the allocated treatment, or high dropout rates. However, most trials compared active treatments, most had only very few patients not starting the allocated treatment or switching to the other treatment during the study, and none had a high risk of bias due to missing outcome information (dropouts). In various sensitivity analyses, we found no indication that use of intention to treat effects affected our main findings.

For RCD studies, the assessment of the risk of bias is not straightforward. Use of propensity score methods helps to reduce confounding, but it is unlikely that confounding can be eliminated. It is difficult even to judge to what exact extent confounding has been reduced with different propensity adjustments or other approaches. For other dimensions of potential bias beyond confounding, our selected studies might have been at lower risk for bias than many other RCD studies that look at outcomes other than mortality. For non-mortality outcomes, missing information, measurement errors, and availability of diverse definitions and analyses could be more prominent than for death. Our results remained largely similar in different sensitivity analyses, although we did see the lowest estimate for a summary relative odds ratio (indicating closest

convergence of results from randomized controlled trials and RCD studies) when we considered RCD studies with low risk of bias in all dimensions (other than confounding). We cannot exclude the possibility that RCD studies become better in predicting trial results when bias is minimized, although much more data are needed to make a conclusive statement about this.

Genuine differences in estimated effect sizes could still exist between the two methods. Nevertheless, we tried to make the PICO structure highly comparable in the juxtaposed RCD studies and randomized controlled trials that we evaluated. It is also unclear whether those questions where subsequent trials were performed are qualitatively different from those where subsequent trials are never performed once an effect has been described in the observational literature. When strong, conclusive effects are seen in RCD studies, there may be less likelihood to perform a subsequent trial.⁷² However, it is unlikely that such strong, conclusive effects are commonly seen.

Conclusions and policy implications

Despite the wide and increasing use of routinely collected health data in comparative effectiveness research, the reliability of this approach needs to be questioned, especially when effectiveness outcomes are concerned and randomized controlled trials might be feasible to conduct. Of course, for some outcomes (especially on safety or harms), it may be difficult to obtain definitive evidence from large trials, and RCD data could then offer the best possible guidance.

If no randomized trials exist, clinicians and funders of care can still act on the results from observational RCD and other evidence, but they should consider that treatment effects could be more uncertain and substantially smaller than what RCD studies suggest. Therefore, decisions for widespread adoption and reimbursement of expensive interventions with evidence based entirely on RCD may be best withheld until trial evidence becomes available. Large randomized trials might still be needed to address critically important clinical questions for patient relevant outcomes.^{177 78}

We thank Hannah Ewald, University of Basel, for support in the risk of bias assessment.

Contributors: LGH and JPAI conceived the study. All authors extracted and analyzed the data and interpreted the results. LGH wrote the first draft and all authors made revisions on the manuscript. All authors read and approved the final version of the paper. JPAI is the guarantor.

Funding: This study was supported by the Commonwealth Fund, a private independent foundation based in New York City. The views presented here are those of the authors and not necessarily those of the Commonwealth Fund, its directors, officers, or staff. The Basel Institute for Clinical Epidemiology and Biostatistics received support from Santésuisse, the umbrella association of Swiss social health insurers. The Meta-Research Innovation Center at Stanford is funded by a grant by the Laura and John Arnold Foundation. The work of JPAI is supported by an unrestricted gift from Sue and Bob O'Donnell. The funders had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript or its submission for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare: DCI and JPAI had no financial support for this project; LGH had support from the Commonwealth Fund for the submitted work; all authors declare no financial relationships with any organization that might

have an interest in the submitted work in the previous three years and no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required for this study.

Data sharing: No additional data available.

The corresponding author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>.

- Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *CMAJ* [forthcoming].
- Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984;79:516-24. doi:10.1080/01621459.1984.
- Sox HC, Goodman SN. The methods of comparative effectiveness research. *Annu Rev Public Health* 2012;33:425-45. doi:10.1146/annurev-publhealth-031811-124610.
- Hlatky MA, Winkelmayr WC, Setoguchi S. Epidemiologic and statistical methods for comparative effectiveness research. *Heart Fail Clin* 2013;9:29-36. doi:10.1016/j.hfc.2012.09.007.
- Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR good research practices for retrospective database analysis task force report—Part III. *Value Health* 2009;12:1062-73. doi:10.1111/j.1524-4733.2009.00602.x.
- Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Do routinely collected health data complement randomized evidence? A survey. *CMAJ Open* [forthcoming].
- Safran C. Using routinely collected data for clinical research. *Stat Med* 1991;10:559-64. doi:10.1002/sim.4780100407.
- Spasoff RA. *Epidemiologic methods for health policy*. Oxford University Press, Inc, 1999.
- Gliklich R, Dreyer N, Leavy M, eds. Registries for evaluating patient outcomes: a user's guide. 3rd ed. (Prepared by the Outcome DECIDE Center [Outcome Sciences, a Quintiles company] under contract no 290 2005 00351 TO7) AHRQ publication no 13(14)-EHC111. Agency for Healthcare Research and Quality. April 2014. www.effectivehealthcare.ahrq.gov/registries-guide-3.cfm.
- National Library of Medicine. MEDLINE / PubMed Search Strategy & Electronic Health Record Information Resources. Secondary MEDLINE / PubMed Search Strategy & Electronic Health Record Information Resources 2015. https://www.nlm.nih.gov/services/queries/ehr_details.html.
- Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995;123:A12-3.
- The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Secondary Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] 2011. www.cochrane-handbook.org.
- Sterne JAC, Higgins JPT, Reeves BC, et al. A Cochrane risk of bias assessment tool: for non-randomized studies of interventions (ACROBAT-NRSI), version 1.0.0, 24 September 2014, 2014. www.riskofbias.info.
- Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997;127:820-6. doi:10.7326/0003-4819-127-9-199711010-00008.
- Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med* 2007;26:53-77. doi:10.1002/sim.2528.
- Kim DH, Daskalakis C, Silvestry SC, et al. Aspirin and clopidogrel use in the early postoperative period following on-pump and off-pump coronary artery bypass grafting. *J Thorac Cardiovasc Surg* 2009;138:1377-84. doi:10.1016/j.jtcvs.2009.07.027.
- Hahn J-Y, Song YB, Choi J-H, et al. DATE Registry Investigators. Three-month dual antiplatelet therapy after implantation of zotarolimus-eluting stents: the DATE (Duration of Dual Antiplatelet Therapy After Implantation of Endeavor Stent) registry. *Circ J* 2010;74:2314-21. doi:10.1253/circj.CJ-10-0347.
- Butler MJ, Eccleston D, Clark DJ, et al. Melbourne Interventional Group. The effect of intended duration of clopidogrel use on early and late mortality and major adverse cardiac events in patients with drug-eluting stents. *Am Heart J* 2009;157:899-907. doi:10.1016/j.ahj.2009.02.018.

- 19 Holman WL, Li Q, Kiefe CI, et al. Prophylactic value of preincision intra-aortic balloon pump: analysis of a statewide experience. *J Thorac Cardiovasc Surg* 2000;120:1112-9. doi:10.1067/mtc.2000.110459.
- 20 Shavelle DM, Parsons L, Sada MJ, French WJ, Every NR. National Registry of Myocardial Infarction 2. Is there a benefit to early angiography in patients with ST-segment depression myocardial infarction? An observational study. *Am Heart J* 2002;143:488-96. doi:10.1067/mhj.2002.120970.
- 21 Winkelmayer WC, Glynn RJ, Mittleman MA, Levin R, Pliskin JS, Avorn J. Journal of the American Society of Nephrology. Comparing mortality of elderly patients on hemodialysis versus peritoneal dialysis: a propensity score approach. *J Am Soc Nephrol* 2002;13:2353-62. doi:10.1097/01.ASN.0000025785.41314.76.
- 22 Karthik S, Musleh G, Grayson AD, et al. Effect of avoiding cardiopulmonary bypass in non-elective coronary artery bypass surgery: a propensity score analysis. *Eur J Cardiothorac Surg* 2003;24:66-71. doi:10.1016/S1010-7940(03)00255-0.
- 23 Guru V, Fremes SE, Tu JV. How many arterial grafts are enough? A population-based study of midterm outcomes. *J Thorac Cardiovasc Surg* 2006;131:1021-8. doi:10.1016/j.jtcvs.2005.09.036.
- 24 Wu C, Hannan EL, Walford G, Faxon DP. Utilization and outcomes of unprotected left main coronary artery stenting and coronary artery bypass graft surgery. *Ann Thorac Surg* 2008;86:1153-9. doi:10.1016/j.athoracsur.2008.05.059.
- 25 Ascione R, Narayan P, Rogers CA, Lim KH, Capoun R, Angelini GD. Early and midterm clinical outcome in patients with severe left ventricular dysfunction undergoing coronary artery surgery. *Ann Thorac Surg* 2003;76:793-9. doi:10.1016/S0003-4975(03)00664-7.
- 26 Polkinghorne KR, McDonald SP, Atkins RC, Kerr PG. Vascular access and all-cause mortality: a propensity score analysis. *J Am Soc Nephrol* 2004;15:477-86. doi:10.1097/01.ASN.0000109668.05157.05.
- 27 Gnerlich J, Jeffe DB, Deshpande AD, Beers C, Zander C, Margenthaler JA. Surgical removal of the primary tumor increases overall survival in patients with metastatic breast cancer: analysis of the 1988-2003 SEER data. *Ann Surg Oncol* 2007;14:2187-94. doi:10.1245/s10434-007-9438-0.
- 28 Lindenaer PK, Pekow P, Wang K, Gutierrez B, Benjamin EM. Lipid-lowering therapy and in-hospital mortality following major noncardiac surgery. *JAMA* 2004;291:2092-9. doi:10.1001/jama.291.17.2092.
- 29 Cabell CH, Abrutyn E, Fowler VG Jr, et al. International Collaboration on Endocarditis Merged Database (ICE-MD) Study Group Investigators. Use of surgery in patients with native valve infective endocarditis: results from the International Collaboration on Endocarditis Merged Database. *Am Heart J* 2005;150:1092-8. doi:10.1016/j.ahj.2005.03.057.
- 30 Moss RR, Humphries KH, Gao M, et al. Outcome of mitral valve repair or replacement: a comparison by propensity score analysis. *Circulation* 2003;108(Suppl 1):II90-7. doi:10.1161/01.cir.0000089182.44963.bb.
- 31 Fonarow GC, Abraham WT, Albert NM, et al. OPTIMIZE-HF Investigators and Coordinators. Influence of beta-blocker continuation or withdrawal on outcomes in patients hospitalized with heart failure: findings from the OPTIMIZE-HF program. *J Am Coll Cardiol* 2008;52:190-9. doi:10.1016/j.jacc.2008.03.048.
- 32 Muneretto C, Bisleri G, Negri A, et al. Off-pump coronary artery bypass surgery technique for total arterial myocardial revascularization: a prospective randomized study. *Ann Thorac Surg* 2003;76:778-82, discussion 783. doi:10.1016/S0003-4975(03)00564-2.
- 33 Masoumi M, Saidi MR, Rostami F, Sepahi H, Roushani D. Off-pump coronary artery bypass grafting in left ventricular dysfunction. *Asian Cardiovasc Thorac Ann* 2008;16:16-20. doi:10.1177/021849230801600105.
- 34 Arogundade FA, Ishola DA Jr, Sanusi AA, Akinsola A. An analysis of the effectiveness and benefits of peritoneal dialysis and haemodialysis using Nigerian made PD fluids. *Afr J Med Med Sci* 2005;34:227-33.
- 35 Korevaar JC, Feith GW, Dekker FW, et al. NECOSAD Study Group. Effect of starting with hemodialysis compared with peritoneal dialysis in patients new on dialysis treatment: a randomized controlled trial. *Kidney Int* 2003;64:2222-8. doi:10.1046/j.1523-1755.2003.00321.x.
- 36 Thiele H, Rach J, Klein N, et al. LIPSIA-NSTEMI Trial Group. Optimal timing of invasive angiography in stable non-ST-elevation myocardial infarction: the Leipzig Immediate versus early and late Percutaneous coronary Intervention trial in NSTEMI (LIPSIA-NSTEMI Trial). *Eur Heart J* 2012;33:2035-43. doi:10.1093/eurheartj/ehr418.
- 37 Montalescot G, Cayla G, Collet JP, et al. ABOARD Investigators. Immediate vs delayed intervention for acute coronary syndromes: a randomized clinical trial. *JAMA* 2009;302:947-54. doi:10.1001/jama.2009.1267.
- 38 Neumann FJ, Kastrati A, Pogatsa-Murray G, et al. Evaluation of prolonged antithrombotic pretreatment ("cooling-off" strategy) before intervention in patients with unstable coronary syndromes: a randomized controlled trial. *JAMA* 2003;290:1593-9. doi:10.1001/jama.290.12.1593.
- 39 Badings EA, The SH, Dambrink JH, et al. Early or late intervention in high-risk non-ST-elevation acute coronary syndromes: results of the ELISA-3 trial. *EuroIntervention* 2013;9:54-61. doi:10.4244/EIJV9I1A9.
- 40 Medved I, Anić D, Ostrić M, Zrnić B, Ivancić A, Tomulić V. Is mitral valve repair safe procedure in elderly patients? *Coll Antropol* 2010;34(Suppl 2):213-5.
- 41 Acker MA, Parides MK, Perrault LP, et al. CTSN. Mitral-valve repair versus replacement for severe ischemic mitral regurgitation. *N Engl J Med* 2014;370:23-32. doi:10.1056/NEJMoa1312808.
- 42 Fattouch K, Guccione F, Dioguardi P, et al. Off-pump versus on-pump myocardial revascularization in patients with ST-segment elevation myocardial infarction: a randomized trial. *J Thorac Cardiovasc Surg* 2009;137:650-6, discussion 656-7. doi:10.1016/j.jtcvs.2008.11.033.
- 43 Dunkelgrun M, Boersma E, Schouten O, et al. Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echocardiography Study Group. Bisoprolol and fluvastatin for the reduction of perioperative cardiac mortality and myocardial infarction in intermediate-risk patients undergoing noncardiovascular surgery: a randomized controlled trial (DECREASE-IV). *Ann Surg* 2009;249:921-6. doi:10.1097/SLA.0b013e3181a77d00.
- 44 Durazzo AE, Machado FS, Ilikeka DT, et al. Reduction in cardiovascular events after vascular surgery with atorvastatin: a randomized trial. *J Vasc Surg* 2004;39:967-75, discussion 975-6. doi:10.1016/j.jvs.2004.01.004.
- 45 Schouten O, Boersma E, Hoeks SE, et al. Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echocardiography Study Group. Fluvastatin and perioperative events in patients undergoing vascular surgery. *N Engl J Med* 2009;361:980-9. doi:10.1056/NEJMoa0808207.
- 46 Rooijens PP, Burgmans JP, Yo TI, et al. Autogenous radial-cephalic or prosthetic brachial-antecubital forearm loop AVF in patients with compromised vessels? A randomized, multicenter study of the patency of primary hemodialysis access. *J Vasc Surg* 2005;42:481-6, 487. doi:10.1016/j.jvs.2005.05.025.
- 47 Keuter XH, De Smet AA, Kessels AG, van der Sande FM, Welten RJ, Tordoir JH. A randomized multicenter study of the outcome of brachial-basilic arteriovenous fistula and prosthetic brachial-antecubital forearm loop as vascular access for hemodialysis. *J Vasc Surg* 2008;47:395-401. doi:10.1016/j.jvs.2007.09.063.
- 48 Kang DH, Kim YJ, Kim SH, et al. Early surgery versus conventional treatment for infective endocarditis. *N Engl J Med* 2012;366:2466-73. doi:10.1056/NEJMoa1112843.
- 49 Soran A, Ozmen V, Ozbas S, et al. Abstract S2-03: Early follow up of a randomized trial evaluating resection of the primary breast tumor in women presenting with de novo stage IV breast cancer; Turkish study (protocol MF07-01). *Cancer Res* 2013;73(24 suppl):S2-03-S2-03.
- 50 Badwe R, Parmar V, Hawaldar R, et al. Abstract S2-02: Surgical removal of primary tumor and axillary lymph nodes in women with metastatic breast cancer at first presentation: A randomized controlled trial. *Cancer Res* 2013;73(24 suppl):S2-02-S2-02.
- 51 Damgaard S, Wetterstevig J, Lund JT, et al. One-year results of total arterial revascularization vs. conventional coronary surgery: CARRPO trial. *Eur Heart J* 2009;30:1005-11. doi:10.1093/eurheartj/ehp048.
- 52 Morice MC, Serruys PW, Kappetein AP, et al. Outcomes in patients with de novo left main disease treated with either percutaneous coronary intervention using paclitaxel-eluting stents or coronary artery bypass graft treatment in the Synergy Between Percutaneous Coronary Intervention with TAXUS and Cardiac Surgery (SYNTAX) trial. *Circulation* 2010;121:2645-53. doi:10.1161/CIRCULATIONAHA.109.899211.
- 53 Boudriot E, Thiele H, Walther T, et al. Randomized comparison of percutaneous coronary intervention with sirolimus-eluting stents versus coronary artery bypass grafting in unprotected left main stem stenosis. *J Am Coll Cardiol* 2011;57:538-45. doi:10.1016/j.jacc.2010.09.038.21272743.
- 54 Park SJ, Kim YH, Park DW, et al. Randomized trial of stents versus bypass surgery for left main coronary artery disease. *N Engl J Med* 2011;364:1718-27. doi:10.1056/NEJMoa1100452.21463149.
- 55 Ranucci M, Castelvécchio S, Biondi A, et al. Surgical and Clinical Outcome Research (SCORE) Group. A randomized controlled trial of preoperative intra-aortic balloon pump in coronary patients with poor left ventricular function undergoing coronary artery bypass surgery. *Crit Care Med* 2013;41:2476-83. doi:10.1097/CCM.0b013e3182978dfc.
- 56 Lomivorotov VV, Boboshko VA, Efreimov SM, et al. Levosimendan versus an intra-aortic balloon pump in high-risk cardiac patients. *J Cardiothorac Vasc Anesth* 2012;26:596-603. doi:10.1053/j.jvca.2011.09.006.
- 57 Sun JC, Teoh KH, Lamy A, et al. Randomized trial of aspirin and clopidogrel versus aspirin alone for the prevention of coronary artery bypass graft occlusion: the Preoperative Aspirin and Postoperative Antiplatelets in Coronary Artery Bypass Grafting study. *Am Heart J* 2010;160:1178-84. doi:10.1016/j.ahj.2010.07.035.
- 58 Mannacio VA, Di Tommaso L, Antignan A, De Amicis V, Vosa C. Aspirin plus clopidogrel for optimal platelet inhibition following off-pump coronary artery bypass surgery: results from the CRYSSA (prevention of Coronary artery bypass occlusion After off-pump procedures) randomised study. *Heart* 2012;98:1710-5. doi:10.1136/heartjnl-2012-302449.

- 59 Gao G, Zheng Z, Pi Y, Lu B, Lu J, Hu S. Aspirin plus clopidogrel therapy increases early venous graft patency after coronary artery bypass surgery: a single-center, randomized, controlled trial. *J Am Coll Cardiol* 2010;56:1639-43. doi:10.1016/j.jacc.2010.03.104.
- 60 Kulik A, Le May MR, Voisine P, et al. Aspirin plus clopidogrel versus aspirin alone after coronary artery bypass grafting: the clopidogrel after surgery for coronary artery disease (CASCADE) Trial. *Circulation* 2010;122:2680-7. doi:10.1161/CIRCULATIONAHA.110.978007.
- 61 Gasparovic H, Petricevic M, Kopjar T, Djuric Z, Svetina L, Biocina B. Impact of dual antiplatelet therapy on outcomes among aspirin-resistant patients following coronary artery bypass grafting. *Am J Cardiol* 2014;113:1660-7. doi:10.1016/j.amjcard.2014.02.024.
- 62 Valgimigli M, Borghesi M, Tebaldi M, Vranckx P, Parrinello G, Ferrari R. PROlonging Dual antiplatelet treatment after Grading stent-induced Intimal hyperplasia studyY Investigators. Should duration of dual antiplatelet therapy depend on the type and/or potency of implanted stent? A pre-specified analysis from the PROlonging Dual antiplatelet treatment after Grading stent-induced Intimal hyperplasia studyY (PRODIGY). *Eur Heart J* 2013;34:909-19. doi:10.1093/eurheartj/ehs460.
- 63 Gwon HC, Hahn JY, Park KW, et al. Six-month versus 12-month dual antiplatelet therapy after implantation of drug-eluting stents: the Efficacy of Xience/Promus Versus Cypher to Reduce Late Loss After Stenting (EXCELLENT) randomized, multicenter study. *Circulation* 2012;125:505-13. doi:10.1161/CIRCULATIONAHA.111.059022.
- 64 Kim BK, Hong MK, Shin DH, et al. RESET Investigators. A new strategy for discontinuation of dual antiplatelet therapy: the RESET Trial (REal Safety and Efficacy of 3-month dual antiplatelet Therapy following Endeavor zotarolimus-eluting stent implantation). *J Am Coll Cardiol* 2012;60:1340-8. doi:10.1016/j.jacc.2012.06.043.
- 65 Feres F, Costa RA, Abizaid A, et al. OPTIMIZE Trial Investigators. Three vs twelve months of dual antiplatelet therapy after zotarolimus-eluting stents: the OPTIMIZE randomized trial. *JAMA* 2013;310:2510-22. doi:10.1001/jama.2013.282183.
- 66 Colombo A, Chieffo A, Frasieri A, et al. Second-generation drug-eluting stent implantation followed by 6- versus 12-month dual antiplatelet therapy: the SECURITY randomized clinical trial. *J Am Coll Cardiol* 2014;64:2086-97. doi:10.1016/j.jacc.2014.09.008.
- 67 Jondeau G, Neuder Y, Eicher JC, et al. B-CONVINCED Investigators. B-CONVINCED: Beta-blocker CONTinuation Vs. INterruption in patients with Congestive heart failure hospitalizED for a decompensation episode. *Eur Heart J* 2009;30:2186-92. doi:10.1093/eurheartj/ehp323.
- 68 Anglemeyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* 2014;4:MR000034. doi:10.1002/14651858.MR000034.pub2.
- 69 Lonjon G, Boutron I, Trinquart L, et al. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg* 2014;259:18-25. doi:10.1097/SLA.0000000000000256.
- 70 Kuss O, Legler T, Börgermann J. Treatments effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. *J Clin Epidemiol* 2011;64:1076-84. doi:10.1016/j.jclinepi.2011.01.005.
- 71 Dahabreh IJ, Sheldrick RC, Paulus JK, et al. Do observational studies using propensity score methods agree with randomized trials? A systematic comparison of studies on acute coronary syndromes. *Eur Heart J* 2012;33:1893-901. doi:10.1093/eurheartj/ehs114.
- 72 Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. *JAMA* 2012;308:1676-84. doi:10.1001/jama.2012.13444.
- 73 Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Concordance of effects of medical interventions on hospital admission and readmission rates with effects on mortality. *CMAJ* 2013;85:E827-37. doi:10.1503/cmaj.130430.
- 74 Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *J Clin Epidemiol* 2011;64:1060-9. doi:10.1016/j.jclinepi.2010.12.012.
- 75 Ioannidis JP, Haidich AB, Lau J. Any casualties in the clash of randomised and observational evidence? *BMJ* 2001;322:879-80. doi:10.1136/bmj.322.7291.879.
- 76 Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;286:821-30. doi:10.1001/jama.286.7.821.
- 77 Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984;3:409-22. doi:10.1002/sim.4780030421.
- 78 Ioannidis JP. Mega-trials for blockbuster. *JAMA* 2013;309:239-40. doi:10.1001/jama.2012.168095.

Web appendix: Supplementary material