Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Bias Estimation In Study Design: A Meta-Epidemiological Analysis of Transcatheter Versus Surgical Aortic Valve Replacement

Saerom Youn (syoun2022@meds.uwo.ca)
Schulich School of Medicine and Dentistry, University of Western Ontario https://orcid.org/0000-0001-5745-954X

Shannon Avery Wong
James Cook University

Caitlin Chrystoja
Institute of Health Policy Management and Evaluation, University of Toronto

George Tomlinson
Institute of Health Policy Management, and Evaluation, University of Toronto

Harindra C Wijeysundera
Institute of Health Policy Management and Evaluation, University of Toronto

Chaim M Bell
Institute of Health Policy Management and Evaluation, University of Toronto

Anna R Gagliardi
Institute of Health Policy Management and Evaluation, University of Toronto

Nancy N Baxter
Institute of Health Policy Management and Evaluation, University of Toronto

Lakhir Sandhu
Department of Surgery, University of Toronto

Julie Takata
Women's College Hospital Research Institute

David Robert Urbach
Institute of Health Policy Management and Evaluation

Research article

Keywords: Medline, Medline In-Process/ePubs, postoperative mortality, Bias

DOI: https://doi.org/10.21203/rs.3.rs-71534/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Objective: To estimate the bias associated with specific nonrandomized study attributes among studies comparing transcatheter aortic valve implantation with surgical aortic valve replacement for the treatment of severe aortic stenosis.

Data sources and study selection: We searched 7 databases from inception to June 2017: Medline, Medline In-Process/ePubs, Embase, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Scopus, and Web of Science. We included all RCTs and nonrandomized studies that reported outcomes of interest.

Data extraction and synthesis: We categorized studies according to study design, and evaluated 41 nonrandomized study attributes as potential sources of bias. We calculated odds ratios (OR) and other effect measures with 95% confidence intervals (CI) using random effects models.

Main outcomes: One month postoperative mortality, and length of stay. Bias was defined as the difference in estimates of treatment effects between nonrandomized studies and high quality (low risk of bias) RCTs, which were considered to provide “gold standard” estimates.

Results: We included 6 RCTs and 87 nonrandomized studies. Surgical risk scores were similar for comparison groups in RCTs, but were higher for patients having transcatheter aortic valve implantation in nonrandomized studies. Nonrandomized studies underestimated the benefit of transcatheter aortic valve implantation compared with RCTs. For example, nonrandomized studies without adjustment estimated a higher risk of postoperative mortality for transcatheter aortic valve implantation compared with surgical aortic valve replacement (OR 1.43 [95% CI, 1.26 to 1.62]) than high quality RCTs (OR 0.78 [95% CI, 0.54 to 1.11). Nonrandomized studies using propensity score matching (OR 1.13 [95% CI, 0.85 to 1.52]) and regression modelling (OR 0.68 [95% CI, 0.57 to 0.81]) to adjust results estimated treatment effects closer to high quality RCTs. Nonrandomized studies describing losses to follow-up estimated treatment effects that were significantly closer to high quality RCT than nonrandomized studies that did not.

Conclusion: Studies with different attributes produce different estimates of treatment effects. Study design attributes related to the completeness of follow-up may explain biased treatment estimates in nonrandomized studies, as in the case of aortic valve replacement where high-risk patients were preferentially selected for the newer (transcatheter) procedure.

Introduction

Frameworks of study designs often specify hierarchies based on the likelihood of estimating biased treatment effects, with well-designed randomized controlled trials (RCT) and their meta-analyses considered to provide the least biased estimates.[1–3] However, there are limited RCTs of non-drug technologies such as medical devices and surgical techniques,[4,5] leading to widespread dependence on non-randomized studies for the evaluation of non-drug health technologies.

Not surprisingly, there is variation in the treatment effects estimated by different study designs, with non-randomized studies frequently reporting larger benefits for the experimental treatment than RCTs.[6–12] Differences in the conclusions of non-randomized studies and RCTs vary according to the clinical context.[13–16] Among RCTs, study quality is associated with estimates of treatment effects; lower quality RCTs often overestimate the benefit of an experimental procedure as compared to high quality RCTs.[17–21] The relationship between study attributes and biased treatment effects is less clear for nonrandomized studies—a better understanding of this relationship would help inform readers, providers, patients, and policy makers, especially when data from high-quality RCTs are not available .

There are many nonrandomized studies and RCTs comparing transcatheter and surgical aortic valve replacement for the treatment of aortic stenosis, providing an ideal opportunity to study the influence of study designs and characteristics on estimates of treatment effectiveness. We sought to empirically explore the direction and magnitude of bias associated with different study attributes using a meta-epidemiological analysis of published studies.

Methods

The study was performed in accordance with the PRISMA guidelines for meta-epidemiological studies.[22]

Clinical context

We studied transcatheter and surgical aortic valve replacement for aortic stenosis because there were both high quality RCTs and a large number of non-randomized studies. Transcatheter aortic valve implantation is a relatively new technique, and its safety and efficacy is of current clinical interest.

Study selection

We included all RCTs that randomly assigned patients to transcatheter or surgical aortic valve replacement and followed patients over time. We also included all comparative cohort studies that reported primary data on outcomes of interest after transcatheter or surgical aortic valve replacement.

We excluded non-randomized studies that were not comparative cohort studies, defined the population by excluding the outcome of interest, combined patients from RCTs and non-randomized studies, conference abstracts, poster presentations, non-peer reviewed publications, unpublished literature, systematic reviews that lacked primary data, and studies that used other surgical aortic valve replacement methods (e.g., minimally invasive, sutureless).

For multiple publications using the identical cohort we included the publication with the most representative sample, determined by sample size or duration of follow up.

Data sources

We searched Medline, Medline In-Process/ePubs, Embase, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Scopus, and Web of Science from inception to June 2017 (eTable 1). We used DistillerSR (Evidence Partners, Ottawa, Canada) to check for duplicate citations, and to screen titles, abstracts, and full text.

Data extraction

A single reviewer collected study characteristics, patient characteristics, and outcomes of interest; questions were resolved by consensus among the study team. Agreement of re-abstracted outcomes for a sample of 15 nonrandomized studies (17%) by a second reviewer demonstrated excellent inter-rater reliability (ICC 0.99 [95% CI, 0.98 to 0.99]).[23]

Study characteristics

We collected study sample size, publication year and country, surgical approach, and the study time period. We collected surgical risk scores (e.g., EuroSCORE II) as a measure of potential selection bias among comparison groups.

Outcomes

We defined postoperative mortality as death due to any cause within 1-month or in hospital after the procedure regardless of location. We defined length of stay as the number of days the patient stayed in the hospital after the procedure. We extracted the necessary components of each outcome to calculate the pooled estimates of treatment effects. We calculated missing data points using given information where possible.

Explanatory variables: Study designs

We categorized studies into 8 groups according to study design: (1) All (all RCT and nonrandomized studies), (2) All RCT, (3) High quality RCT, (4) Low quality RCT, (5) All non-randomized studies, (6) Nonrandomized studies without adjustment, (7) Nonrandomized studies adjusted using propensity score matching (PSM), and (8) Nonrandomized studies adjusted using regression.

RCTs were divided into high or low quality RCTs based on the Cochrane Risk Of Bias (ROB) tool[24] based on the content of the published articles; authors were not contacted for additional information (eTable 3). No RCT blinded study participants; hence RCTs that satisfied all other criteria were categorized as high quality. Non-randomized studies reported unadjusted estimates, adjusted estimates, or both. Non-randomized studies estimates were pooled into 3 groups: without adjustment, adjusted using PSM, and adjusted using regression.

Finally, we previously developed a set of 41 non-randomized studies attributes that could bias studies (Appendix B). These attributes were based on existing frameworks of bias and quality assessment tools

for nonrandomized studies, and were extensively pilot tested and iteratively developed for clarity and reliability.

Data synthesis

Study characteristics

We compared overall study characteristics between RCTs and non-randomized studies using descriptive statistics. To combine continuous variables across studies, the weighted mean of estimates was calculated, and the pooled standard deviation (SD) was either calculated directly (where reported) or imputed from the pooled variance of included studies in the relevant group if missing [25].

Pooled estimates of treatment effects

The effect of treatment on postoperative mortality was estimated using odds ratio (OR). OR < 1 indicated lower risk of death for transcatheter aortic valve implantation. For Bayesian RCTs, we assumed the median estimate represented the percentage with events.[26,27] The treatment effect on length of stay was estimated using mean difference (MD, with values < 0 indicating shorter length of stay for transcatheter aortic valve implantation).

All effect sizes were pooled using a random effects model to account for potential between-study heterogeneity. For postoperative mortality, we used the DerSimonian-Laird method,[28] with the exception of estimates that incorporated adjusted ORs from nonrandomized studies adjusted using regression, which were calculated using the generic inverse variance method [25]. For length of stay, we used the inverse variance method.[25] All pooled estimates were presented visually using forest plots with point estimates and 95% CI. Estimates from high-quality RCTs were considered to represent the “gold standard” treatment effects.

We evaluated the impact of the 41 nonrandomized study attributes on estimates of treatment effect by calculating the ratio of odds ratios (ROR) for postoperative mortality and difference of mean differences (DMD) for length of stay with 95% CI using random effects meta regression. The ROR is the ratio of the OR in one group of studies and the OR in another group of studies[18]; the DMD is the difference between MD reported in one group of studies and the MD in another group of studies.[29] We compared the pooled estimates between study categories, and also between nonrandomized studies with attributes hypothesized to be associated with bias. ROR < 1 and DMD < 0 indicated that studies with ‘better’ study characteristics favored transcatheter aortic valve implantation.

All statistical analyses were conducted using R studio version 1.0.136 (2016).[30] The analysis of whether the attributes of nonrandomized studies were associated with statistical differences in pooled effect sizes was an exploratory analysis; a less restrictive 2-sided P value of 0.10 was used to determine potentially important attributes. In all other analyses a P value of 0.05 or less was considered statistically

significant. P values for comparisons of estimates between types of study were those of the ROR or DMD for the comparison.

Results

Study selection

Of 2,061 RCTs identified in our search, six (described in 15 publications) met the inclusion criteria, and of 10,409 nonrandomized studies, 87 (described in 88 publications) met the inclusion criteria (Fig. 1 and eTable 2). We included four additional nonrandomized studies from publications that were not identified in the initial search.

Study Characteristics

The six RCTs included 5,352 patients, and the 87 non-randomized studies included 239,433 patients (Table 1). RCTs and nonrandomized studies had similar years of publication, were conducted mostly in Europe and North America, and often used multiple surgical approaches.

Table 1
Descriptive characteristics of included studies

	RCTs		NRSs	
Number of studies	6		87	
Total number of patients	5352		239433	
Year published*	2014 (2012, 2016)		2014 (2012, 2016)	
Region				
Europe	2 (33.3%)		47 (54.0%)	
North America	2 (33.3%)		16 (18.4%)	
Asia	0		7 (8.0%)	
Other	0		3 (3.5%)	
Multiple	2 (33.3%)		2 (2.3%)	
Unclear	0		12 (13.8%)	
TAVI Approach				
Any	5 (83.3%)		57 (65.5%)	
Transfemoral	0		10 (11.5%)	
Transapical	1 (16.7%)		13 (14.9%)	
Other	0		7 (8.0%)	
	TAVI	SAVR	TAVI	SAVR
Number of patients	2771	2581	78254	161179
Year enrolment began	2010 (2008, 2011) (n = 6)	2010 (2008, 2011) (n = 6)	2009 (2006–2011) (n = 79)	2007 (2005–2009) (n = 76)
Year enrolment ended	2012 (2011, 2013) (n = 6)	2012 (2011, 2013) (n = 6)	2012 (2010–2013) (n = 75)	2012 (2010, 2013) (n = 72)
Baseline surgical risk†				
STS	6.13 ± 2.25 (n = 5)	6.20 ± 2.32 (n = 5)	9.83 ± 5.03 (n = 34)	6.76 ± 3.68 (n = 33)
EuroSCORE I	NA	NA	18.25 ± 8.61 (n = 8)	11.16 ± 5.26 (n = 8)

	RCTs		NRSs	
LogEuroSCORE	16.21 ± 8.77 (n = 5)	16.30 ± 8.64 (n = 5)	22.32 ± 11.29 (n = 44)	14.19 ± 8.73 (n = 44)
EuroSCORE II	NA	NA	8.52 ± 6.58 (n = 5)	8.09 ± 5.74 (n = 5)
NYHA	2.75 (n = 3)	2.74 (n = 3)	3.40 (n = 12)	2.62 (n = 12)
Abbreviations: RCT, Randomized Controlled Trial; NRS, Nonrandomized Study; TAVI, Transcatheter Aortic Valve Implantation; SAVR, Surgical Aortic Valve Replacement; STS, Society of Thoracic Surgeons; NYHA, New York Heart Association; NA, Not Applicable.				
All continuous variables are reported as either median (25th, 75th percentile) or mean ± SD. All discrete variables are reported as n (%).				
Values describing the characteristics of patients in each arm of the studies are followed by the number of studies each category that reported the value (n).				
* For studies with multiple publications, the year of the first publication was used.				
† STS, EuroSCORE I, LogEuroSCORE and EuroSCORE II are measures of predicted operative mortality. NYHA classifies the extent of heart failure into 4 classes I to IV, with I being least severe and IV being most severe. The numbers indicate the weighted average NYHA class of each cohort. 'Other' TAVI approaches included non-iliofemoral, transthoracic, or transvascular approaches.				
a) Postoperative mortality				

The proportion of studies including patients of all surgical risk categories was higher in non-randomized studies than RCTs (67.8% vs 33.3%,). In general, transcatheter aortic valve implantation subjects in nonrandomized studies had higher surgical risk compared to transcatheter aortic valve implantation subjects in RCTs or surgical aortic valve replacement subjects in nonrandomized studies and RCTs.

Three RCTs satisfied modified ROB assessment criteria for "high quality" and three were "low quality" (Appendix C).

Comparison Of Treatment Effects Between Rcts And Non-randomized Studies

For postoperative mortality, nonrandomized studies adjusted using regression significantly favored transcatheter aortic valve implantation (OR 0.68 [95% CI, 0.57 to 0.81], P for comparison with high quality RCT 0.61). High quality RCTs (OR 0.78 [95% CI, 0.54 to 1.11]), low quality RCTs (OR, 0.8 [95% CI, 0.58 to 1.65], P for comparison with high quality RCT 0.48) and nonrandomized studies adjusted using PSM (OR, 1.13 [95% CI, 0.85 to 1.52], P for comparison with high quality RCT 0.18) found no statistical difference, while nonrandomized studies without adjustment significantly favored surgical aortic valve replacement (OR, 1.43 [95% CI, 1.26 to 1.62], P for comparison with high quality RCT 0.01).

For length of stay, all categories of study design except for PSM-adjusted nonrandomized studies significantly favored transcatheter aortic valve implantation. However, there were differences in the magnitudes of the pooled point estimates. High quality RCTs reported a point estimate for the length of stay in the transcatheter group (MD -4.50 [95% CI, -5.05 to -3.96]) that was about 1.5 days shorter than low quality RCTs (MD -2.87 [95% CI, -5.13 to -0.61], P for comparison 0.26), nonrandomized studies adjusted using PSM (MD -3.01 [95% CI, -6.01 to 0], P for comparison 0.62), and nonrandomized studies without adjustment (MD -3.06 [95% CI, -3.89 to -2.24], P for comparison 0.33). No nonrandomized studies adjusted length of stay using regression.

Influence of non-randomized study characteristics on estimates of treatment effect

For each outcome, some attributes of nonrandomized studies were significantly ($P < 0.10$) associated with pooled estimates of treatment effect closer to those from high quality RCTs (Fig. 3). For postoperative mortality, these attributes were: losses to follow up described ($P = 0.05$), follow up equal in duration ($P = 0.10$), and conflict of interest disclosure for non-first/last authors ($P = 0.10$). For length of stay, these attributes were: losses to follow up described ($P = 0.08$), missing data addressed ($P = 0.09$), and outcome measured from interviews ($P = 0.06$).

Discussion

When comparing estimates of treatment effects in randomized and nonrandomized studies comparing transcatheter aortic valve implantation with surgical aortic valve replacement, we found that point estimates of the effect sizes of study designs with lower risk of bias tended to show larger benefit for transcatheter aortic valve implantation than study designs with higher risk of bias. Statistical adjustment using regression, but not propensity score matching, brought estimated effect sizes closer to high quality RCTs for postoperative mortality. Among nonrandomized studies, accounting for loss to follow up was associated with estimates of treatment effect closer to those from RCTs.

Our findings are consistent with meta-analyses that found RCTs favored transcatheter aortic valve implantation more than nonrandomized studies with respect to postoperative mortality.[31] Interestingly, while meta-epidemiological studies of other clinical topics found that lower quality studies tend to *overestimate* the benefit of newer treatments,[19–21,32–36] higher risk of bias studies of transcatheter aortic valve implantation *underestimated* treatment benefit. There are several possible reasons for the discrepancy. Our analysis included recent studies, which are generally of higher quality and follow better reporting guidelines than older studies.[37] The difference may also be specific to the clinical context we studied. Allocation of patients to treatment groups is highly influenced by differences in case-mix.[38] The surgical risk of postoperative mortality was higher in patients who had transcatheter aortic valve implantation in nonrandomized studies, presumably because the transcatheter procedure was largely restricted to patients who were too high risk for surgical aortic valve replacement in the early years of its clinical use. This situation is different than other clinical situations, where newer or innovative procedures are preferentially used in lower-risk patients.[39]

In our study, propensity score matching did not consistently shift estimates from nonrandomized studies closer to RCT estimates. Others found that propensity score-matched effect sizes from nonrandomized studies were closer to RCTs than regression modeling [38].

Of the various design attributes of nonrandomized studies we analyzed, studies that described loss to follow up yielded estimates of treatment effects that were closer to high quality RCTs. Study attributes related to baseline characteristics did not substantially affect effect estimates. Loss to follow up is a major source of selection bias in cohort studies; it is associated with socioeconomic status,[40–47] substance abuse,[41] smoking,[45,48–51] alcohol abuse,[43,52] physical inactivity,[49,52,53] and poor diet.[52] Quantifying the extent of bias due to loss to follow up may be helpful in understanding biased estimation of treatment effects in nonrandomized studies. Our study had important strengths. We focused on a single clinical question, allowing us to focus on the influence of study characteristics on estimated treatment effects without introducing other sources of variation from studying a heterogeneous group of interventions. We stratified RCTs by risk of bias, instead of pooling all RCTs together. Studies comparing transcatheter and surgical aortic valve replacement included several high quality RCTs, and many recent large and well-reported nonrandomized studies, enabling us to disentangle the influence of study quality and study characteristics on estimated treatment effects. Thirteen nonrandomized studies reported both adjusted and unadjusted treatment effects. Surgical risk scores allowed us to examine confounding by indication.

Our study has limitations. Our literature review may not have included every potentially eligible study. However, this would not affect the internal consistency and generalizability of our findings, which focused on differential estimates between RCTs and nonrandomized studies, rather than estimating the independent treatment effect of aortic valve replacement techniques. Although we limited our analysis to a single clinical question, there is nevertheless substantial heterogeneity among the articles we analyzed that must be taken into account. Although we categorized studies by design, different studies included subjects from very different clinical populations (e.g., a high-risk population is completely different from an intermediate risk population). However, this situation is typical of the medical literature, and pooled measures of effect are commonly reported in meta-analyses even when clinical heterogeneity exists among included studies. Although some of the high-quality RCTs were designed as non-inferiority studies, they would still be expected to provide unbiased estimates of the relative effectiveness of transcatheter aortic valve implantation with respect to the outcomes we evaluated.

We specified a priori a liberal P value threshold of 0.10 and performed multiple univariate comparisons to identify nonrandomized study attributes potentially associated with biased effect estimates. The intent of these exploratory analyses was to generate hypotheses about these study attributes for future analyses rather than test specific hypotheses. Many of these attributes are correlated, and further research could test specific hypothesis regarding the effect of a limited number of pre-specified attributes on bias. Further studies on the reliability of measured attributes of non-randomized studies and how they influence effect estimates compared with RCTs will help improve the interpretation of the results of nonrandomized studies. Finally, while a single reviewer collected the data for this study, analyses of inter-

rater reliability demonstrated excellent correlation among a sample of key variables that were re-abstracted by a second reviewer.

We found that higher quality studies reported a larger benefit than lower quality studies for transcatheter aortic valve replacement compared with surgical valve replacement, although differences were not statistically significant. While adjusted estimates of treatment effects in nonrandomized studies were generally closer to high quality RCT estimates, propensity score snatching and regression modelling varied in the extent to which they were able to adjust effect estimates closer to RCT estimates. Risk adjustment methods may not reliably account for biases in nonrandomized studies. Consideration of loss to follow up appears to be an important attribute of nonrandomized studies with respect to estimating treatment effects that are closer to RCT estimates.

Declarations

Ethics approval and consent to participate: The study was exempted from ethics review by the Research Ethics Board at the University of Toronto.

Consent for publication: Not applicable.

Funding: The study was funded by the Canadian Institutes of Health Research (CIHR) operating grant MOP-136787.

Competing interests: The authors declare that they have no competing interests.

Availability of data and materials: The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

Author's contributions: SY conducted data screening, collection, analysis, and interpretation. SY drafted and revised the manuscript. SAW collected data. CC assisted with data collection and analysis. GT assisted with data analysis and interpretation. HCW, CMB, ARG, NNB and LS assisted with data interpretation. JT provided administrative assistance. DRU is the corresponding author. All authors approved of the submitted version.

Acknowledgement: We thank Eugene Kim, BEE JD (University of Ottawa) for help with coding and editing; Dr. Douglas Lee, MD PhD (Institute of Health Policy Management and Evaluation, University of Toronto) and Dr. Asim Cheema, MD PhD (Institute of Medical Sciences, University of Toronto) for helpful comments; the Human Brain Project (HBP) for supporting the presentation of project at the 4th HBP School – Future computing: Brain Science and Artificial Intelligence. None received compensation for their roles in the study.

References

1. Sackett DL. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Little, Brown; 1991.

2. Brighton B, Bhandari M, Tornetta P, Felson DT. Hierarchy of evidence: from case reports to randomized controlled trials. Clin Orthop Relat Res. 2003;Aug(413):19–24. doi:10.1097/01.blo.0000079323.41006.12.
3. Burns P, Rohrich R, Chong K. The Levels of Evidence and their role in Evidence-Based Medicine. Plast Reconstr Surg. 2011;128(1):305–10. doi:10.1097/PRS.0b013e318219c171.The.
4. Wente MN, Seiler CM, Uhl W, Büchler MW. Perspectives of evidence-based surgery. Dig Surg. 2003;20(4):263–9. doi:10.1159/000071183.
5. Califf RM, Zarin DA, Kramer JM, Sherman RE, Aberle LH, Tasneem A. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007–2010. Jama. 2012;307(17):1838–47. doi:10.1001/jama.2012.3424.
6. Huynh T, Perron S, O'Loughlin J, et al. Comparison of primary percutaneous coronary intervention and fibrinolytic therapy in ST-segment-elevation myocardial infarction: bayesian hierarchical meta-analyses of randomized controlled trials and observational studies. Circulation. 2009;119(24):3101–9. doi:10.1161/CIRCULATIONAHA.108.793745.
7. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA. 2001;286(7):821–30. doi:10.1001/jama.286.7.821.
8. Nicolaides K, Brizot MDL, Patel F, Snijders R. Comparison of chorionic villus sampling and amniocentesis for fetal karyotyping at 10–13 weeks' gestation. Obstet Gynecol Surv. 1995;50(2):96–7. doi:10.1097/00006254-199502000-00008.
9. Jha P, Flather M, Lonn E, Farkouh M, Yusuf S. The antioxidant vitamins and cardiovascular disease. A critical review of epidemiologic and clinical trial data. Ann Intern Med. 1995;123(11):860–72. http://www.ncbi.nlm.nih.gov/pubmed/7486470.
10. Pyorala S, Huttunen NP, Uhari M. A review and meta-analysis of hormonal treatment of cryptorchidism. J Clin Endocrinol Metab. 1995;80(9):2795–9. doi:10.1210/jcem.80.9.7673426.
11. Kunz R, Vist GE, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev. 2007;(2). doi:10.1002/14651858.MR000012.pub2.
12. Kirtane AJ, Gupta A, Iyengar S, et al. Safety and efficacy of drug-eluting and bare metal stents: Comprehensive meta-analysis of randomized trials and observational studies. Circulation. 2009;119(25):3198–206. doi:10.1161/CIRCULATIONAHA.108.826479.
13. Odgaard-Jensen J, Vist G, Timmer A, et al. Randomisation to protect against selection bias in healthcare trials (Review). Cochrane database Syst Rev. 2011;4:MR000012. doi:10.1002/14651858.MR000012.pub3.www.cochranelibrary.com.
14. Shikata S, Nakayama T, Noguchi Y, Taji Y, Yamagishi H. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. Ann Surg. 2006;244(5):668–76. doi:10.1097/01.sla.0000225356.04304.bc.
15. Antman K, Amato D, Wood W, et al. Selection bias in clinical trials. J Clin Oncol. 1985;3(8):1142–7. .
16. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies:a systematic review. Heal Technol Assess. 1998;2(1998):13.

doi:10.1136/bmj.317.7167.1258a. 1–136.

17. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273(5):408–12. doi:10.1001/jama.273.5.408.
18. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. BMJ. 2008;336(7644):601–5. doi:10.1136/bmj.39465.451748.AD.
19. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352(9128):609–13. doi:10.1016/S0140-6736(98)01085-X.
20. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med. 2001;135(11):982–9. doi:10.7326/0003-4819-149-3-200808050-00023.
21. Pildal J, Hróbjartsson A, Jörgensen KJ, Hilden J, Altman DG, Gøtzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol. 2007;36(4):847–57. doi:10.1093/ije/dym087.
22. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med. 2017;22(4):139–42. doi:10.1136/ebmed-2017-110713.
23. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016. doi:10.1016/j.jcm.2016.02.012.
24. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J. 2011;343:889–93. doi: 10.1136/bmj.d5928.
25. Higgins J, S G, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [Updated March 2011]. The Cochrane Collaboration. . Published 2011.
26. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol. 2014. doi:10.1186/1471-2288-14-135.
27. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol. 2005. doi:10.1186/1471-2288-5-13.
28. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986. doi:10.1016/0197-2456(86)90046-2.
29. Page MJ, Higgins JPT, Clayton G, Sterne JAC, Hróbjartsson A, Savović J. Empirical evidence of study design biases in randomized trials: Systematic review of meta-epidemiological studies. PLoS One. 2016. doi:10.1371/journal.pone.0159267.
30. RStudio Team. RStudio: Integrated Development for R. 2016. http://www.rstudio.com/.
31. Gargiulo G, Sannino A, Capodanno D, et al. Transcatheter aortic valve implantation versus surgical aortic valve replacement: A Systematic review and meta-analysis. Ann Intern Med. 2016;165(5):334–

44. doi:10.7326/M16-0060.

32. Hrobjartsson A, Thomsen A, Emanuelsson F, 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. doi:10.1136/bmj.e1119.
33. Hróbjartsson A, Thomsen ASS, Emanuelsson F, 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(3):937–48. doi:10.1093/ije/dyt270.
34. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. BMJ. 2009;339(oct19 1):b4012–2. doi:10.1136/bmj.b4012.
35. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. Int J Epidemiol. 2005;34(1):79–87. doi:10.1093/ije/dyh300.
36. Nüesch E, Trelle S, Reichenbach S, et al. The effects of excluding patients from the analysis in randomised controlled trials: Meta-epidemiological study. BMJ. 2009;339(7722):679–83. doi:10.1136/bmj.b3244.
37. Mcculloch P, Feinberg J, Philippou Y, et al. Progress in clinical research in surgery and IDEAL. Lancet. 2018;0(0). doi:10.1016/S0140-6736(18)30102-8.
38. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess (Rockv). 2003;7(27). doi:10.3310/hta7270.
39. Amer MA, Herbison GP, Smith MD, Grainger SH, Khoo CH, McCall J. Bias in surgical randomised trials: a meta-epidemiological study using laparoscopic versus open surgery as an example. In: *Abstracts of the 25th Cochrane Colloquium, Edinburgh, UK.* Cochrane Database of Systematic Reviews; 2018:9 Suppl 1.
40. Tin ST, Woodward A, Ameratunga S. Estimating bias from loss to follow-up in a prospective cohort study of bicycle crash injuries. Inj Prev. 2014;20(5):322–9. doi:10.1136/injuryprev-2013-040997.
41. Corrigan JD, Harrison-Felix C, Bogner J, Dijkers M, Terrill MS, Whiteneck G. Systematic bias in traumatic brain injury outcome studies because of loss to follow-up. Arch Phys Med Rehabil. 2003;84(2):153–60. doi:10.1053/apmr.2003.50093.
42. Osler M, Kriegbaum M, Christensen U, Lund R, Nybo Andersen AM. Loss to follow up did not bias associations between early life factors and adult depression. J Clin Epidemiol. 2008;61(9):958–63. doi:10.1016/j.jclinepi.2007.11.005.
43. Osler M, Kriegbaum M, Christensen U, Holstein B, Nybo Andersen AM. Rapid Report on Methodology: Does Loss to Follow-up in a Cohort Study Bias Associations Between Early Life Factors and Lifestyle-Related Health Outcomes? Ann Epidemiol. 2008;18(5):422–4. doi:10.1016/j.annepidem.2007.12.008.
44. Wolke D, Waylen A, Samara M, et al. Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. Br J Psychiatry. 2009;195(3):249–56. doi:10.1192/bjp.bp.108.053751.

45. Greene N, Greenland S, Olsen J, Nohr EA. Estimating bias from loss to follow-up in the danish national birth cohort. Epidemiology. 2011;22(6):815–22. doi:10.1097/EDE.0b013e31822939fd.
46. Carter KN, Imlach-Gunasekara F, McKenzie SK, Blakely T. Differential loss of participants does not necessarily cause selection bias. Aust N Z J Public Health. 2012;36(3):218–22. doi:10.1111/j.1753-6405.2012.00867.x.
47. Howe LD, Tilling K, Galobardes B, Lawlor DA. Loss to follow-up in cohort studies: Bias in estimates of socioeconomic inequalities. Epidemiology. 2013;24(1):1–9. doi:10.1097/EDE.0b013e31827623b1.
48. Nilsen RM, Vollset SE, Gjessing HK, et al. Self-selection and bias in a large prospective pregnancy cohort in Norway. Paediatr Perinat Epidemiol. 2009;23(6):597–608. doi:10.1111/j.1365-3016.2009.01062.x.
49. Bjertness E, Sagatun Å, Green K, Lien L, Søgaard AJ, Selmer R. Response rates and selection problems, with emphasis on mental health variables and DNA sampling, in large population-based, cross-sectional and longitudinal studies of adolescents in Norway. BMC Public Health. 2010;10. doi:10.1186/1471-2458-10-602.
50. Barchielli A, Balzi D. Nine-year follow-up of a survey on smoking habits in Florence (Italy): higher mortality among non-responders. Int J Epidemiol. 2002;31(5):1038–42. doi:10.1093/ije/31.5.1038.
51. Verlato G, Melotti R, Olivieri M, et al. Asthmatics and ex-smokers respond early, heavy smokers respond late to mailed surveys in Italy. Respir Med. 2010;104(2):172–9. doi:10.1016/j.rmed.2009.09.022.
52. Kypri K, Samaranayaka A, Connor J, Langley JD, Maclennan B. Non-response bias in a web-based health behaviour survey of New Zealand tertiary students. Prev Med (Baltim). 2011;53(4–5):274–7. doi:10.1016/j.ypmed.2011.07.017.
53. Van Loon AJM, Tijhuis M, Picavet HSJ, Surtees PG, Ormel J. Survey non-response in the Netherlands: Effects on prevalence estimates and associations. Ann Epidemiol. 2003;13(2):105–10. doi:10.1016/S1047-2797(02)00257-0.

Figures

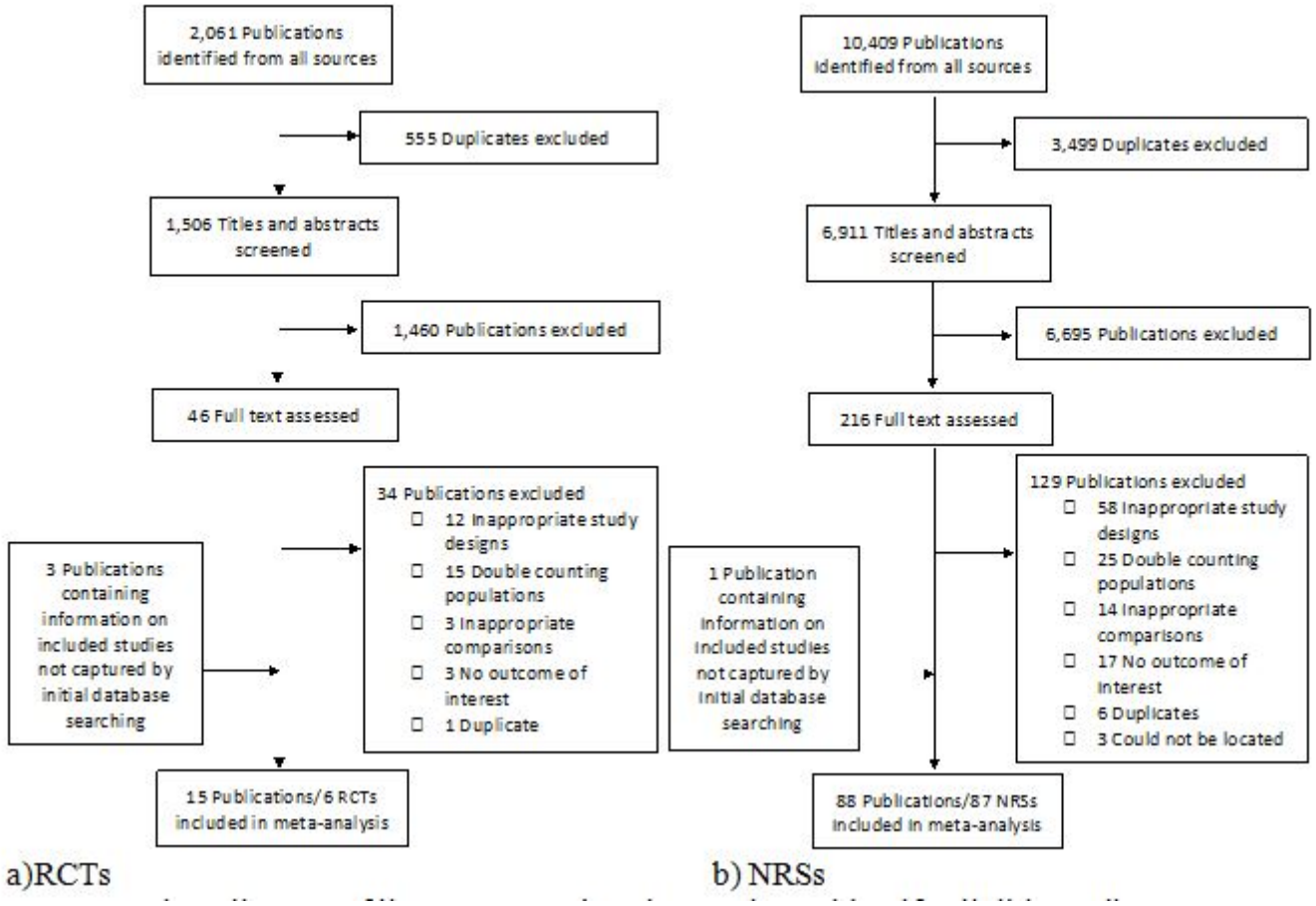

Figure 1

Flow diagram of literature search and screening to identify eligible studies RCT, randomized controlled trials; NRS, nonrandomized studies

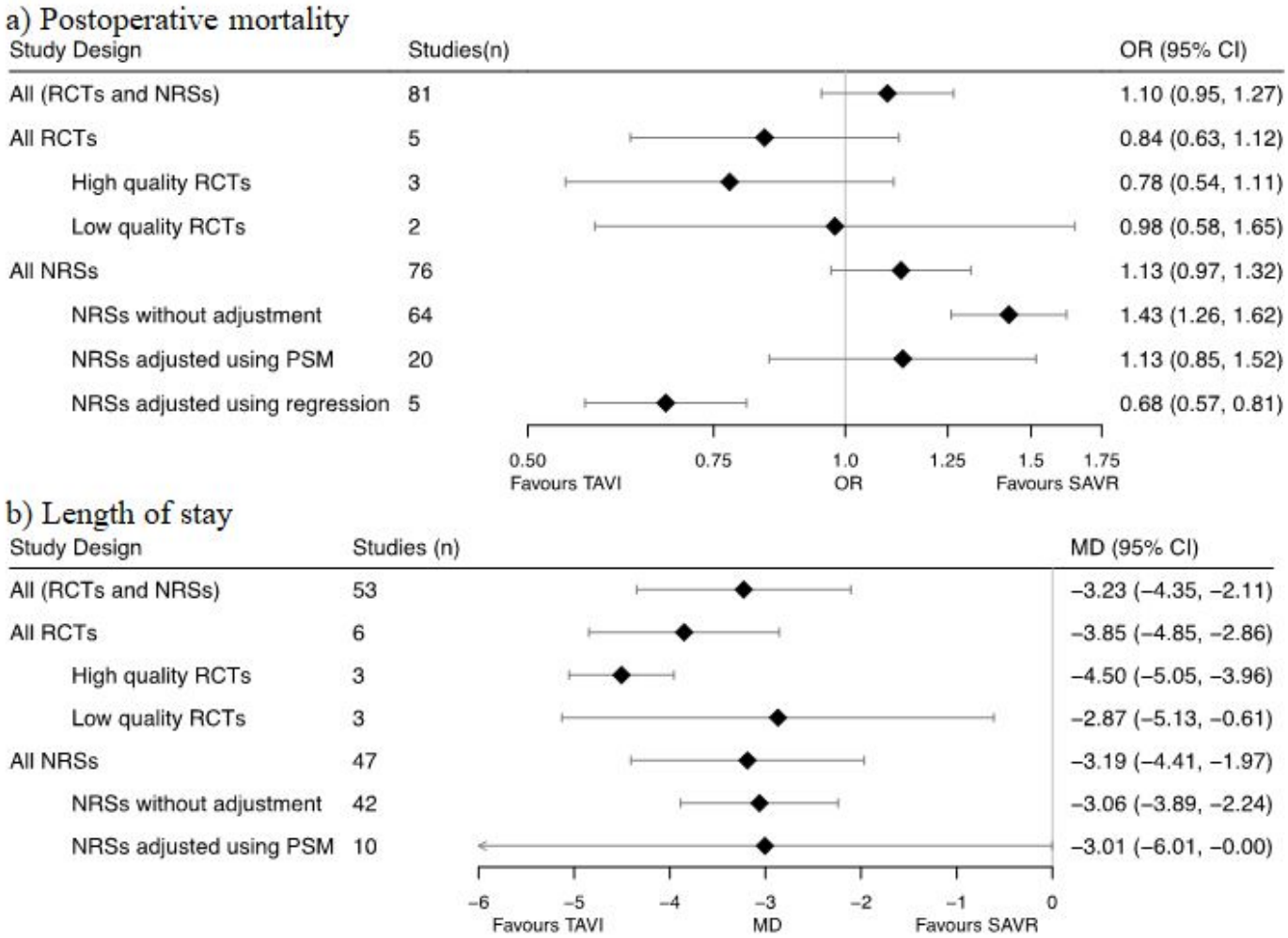

Figure 2

Pooled estimates of treatment effects in different study designs Abbreviations: RCT, Randomized Controlled Trial; NRS, Non-Randomized Studies; PSM, Propensity Score Matching; OR, Odds Ratio; CI, Confidence Interval; MD, Mean Difference. Early means ≤30 days post-op. Diamond is the point estimate of treatment effect. Horizontal lines are 95% CI. Odds ratios were plotted in log scale. Treatment effects were plotted to exact values, but were reported rounded to 2 decimal places.

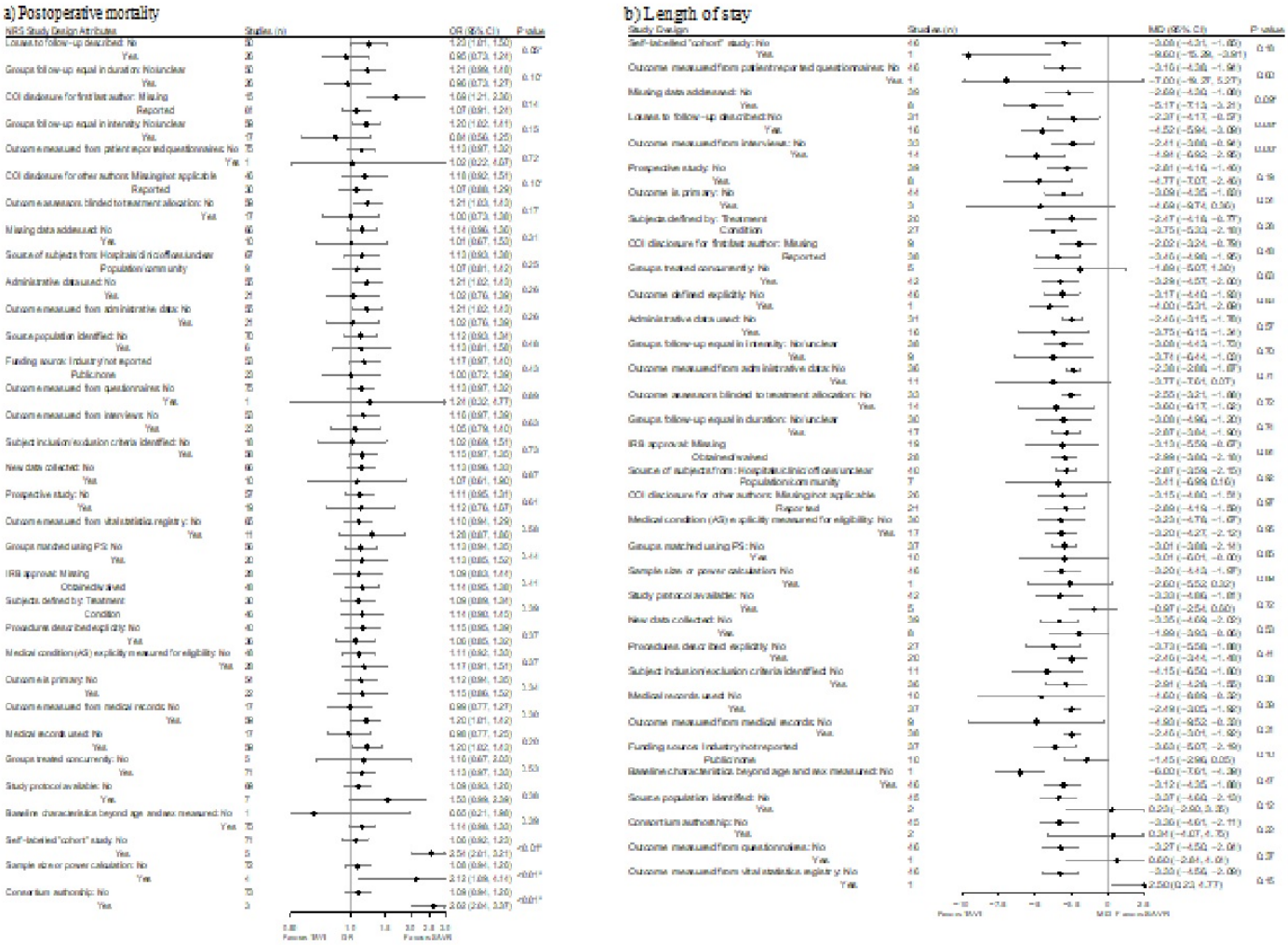

Figure 3

Comparison of pooled estimates of treatment effect in NRSs stratified by specific NRS attributes Abbreviations: NRS, Non-Randomized Studies; OR, Odds Ratio; MD, Mean Difference; CI, Confidence Interval; COI, Conflict of Interest; IRB, Institutional Review Board; AS, aortic stenosis. Attributes were ordered by increasing Ratio of Odds Ratios (RORs) and Difference in Mean Differences (DMD) between the pooled estimates in each stratification. Diamond is the point estimate. Horizontal lines are 95% CI. Odds ratios were plotted in log scale. Treatment effects are plotted to exact values, but are reported rounded to 2 decimal places. Studies column shows how many studies out of total NRS were pooled to produce each estimate. Grey and white bars group rows that belong to each attribute for easier visibility.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 200902PRISMA.doc
- 200902OnlineSupplement.docx