

1 **Title**

2 **Are there mortality improvements with newer interventions in adult cardiac surgery? Evidence**
3 **from 73 meta-analyses**

4

5 Short Title

6 Mortality improvements in adult cardiac surgery

7

8 Austin Parish, MD, George Tolis, Jr., MD, John P.A. Ioannidis, MD, DSc

9

10 Department of Emergency Medicine, Brookdale Hospital Medical Center, Brooklyn, New York,
11 11212, USA (AP)

12

13 Division of Cardiac Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA
14 02115 (GT)

15

16 Meta-Research Innovation Center at Stanford (METRICS), Stanford University, and Departments of
17 Medicine and of Epidemiology and Population Health, Stanford University School of Medicine,
18 Stanford, CA 94305, USA (JPAI)

19

20 Address for Correspondence

21 George Tolis, Jr. MD, Division of Cardiac Surgery, Brigham and Women's Hospital, Harvard
22 Medical School, 75 Francis Street, Boston MA 02115 Email : gtolis@bwh.harvard.edu

23

24 Funding: none

- 25 Disclosure statement: The authors have no conflicts of interest.
- 26 Data sharing statement: All data are available from the authors upon request.
- 27 Independent data access and analysis: The corresponding author as well as the remaining two authors
- 28 had full access to all the data in the study. The corresponding author takes responsibility for its
- 29 integrity and the data analysis.
- 30

31 **ABSTRACT**

32 **Background** In the last two decades, many new interventions have been introduced with the ultimate
33 goal of improving overall postoperative outcomes after cardiac operations in adults. We aimed to
34 assess how often randomized controlled trials (RCTs) in adult cardiac surgery found significant
35 mortality benefits for newer interventions versus older ones, whether observed treatment effect
36 estimates changed over time and whether RCTs and non-randomized observational studies gave
37 similar results.

38 **Methods** We searched journals likely to publish systematic reviews on adult cardiac surgery for
39 meta-analyses of mortality outcomes and that included at least one RCT, with or without
40 observational studies. Relative treatment effect sizes were evaluated overall, over time, and per study
41 design.

42 **Results** 73 meta-analysis comparisons (824 study outcomes on mortality, 519 from RCTs, 305 from
43 observational studies) were eligible. The median mortality effect size was 1.00, IQR 0.54-1.30 (1.00
44 among RCTs, 0.91 among observational studies, $p=0.039$). 4 RCTs and 6 observational studies
45 reached $p<0.005$ favoring newer interventions. 2/73 meta-analyses reached $p<0.005$ favoring the
46 newer interventions. Effect size for experimental interventions relative to controls did not change
47 over time overall ($p=0.64$) or for RCTs ($p=0.30$), and there was a trend for increase in observational
48 studies ($p=0.027$). In 34 meta-analyses with both RCTs ($n=95$) and observational studies ($n=305$), the
49 median relative summary effect (summary effect in observational studies divided by summary effect
50 in RCTs) was 0.87 (IQR, 0.55-1.29); meta-analysis of the relative summary effects yielded a
51 summary of 0.93 (95% CI, 0.74-1.18).

52 **Conclusions** The vast majority of newer interventions had no mortality differences over older ones
53 both overall and in RCTs in particular, while benefits for newer interventions were reported more
54 frequently in observational studies.

55 **Keywords**

56 cardiac surgery, mortality, death, meta-analysis, bias

57

58 **Abbreviations**

59 HR: hazard ratio

60 OR: odds ratio

61 RCT: randomized controlled trial

62 RR: risk ratio

63 SRMA: systematic review and meta-analysis

64

65 **INTRODUCTION**

66 Although adult cardiac surgery was officially introduced as a subspecialty of thoracic surgery
67 100 years ago with the first closed mitral commissurotomy (1), it did not become widely clinically
68 applicable until more than three decades later, with the development of cardiopulmonary bypass (2).
69 The expanded treatment options afforded by cardiopulmonary bypass transformed the initial scattered
70 heroic interventions by a handful of surgeons into a surgical discipline with a strong clinical and
71 research record. Adult cardiac surgery has been shown to offer very effective interventions for
72 common conditions such as coronary artery disease, various valvulopathies and diseases of the great
73 vessels (3-6), and for some less common ones, e.g. intracardiac and thoracic tumors or abdominal
74 tumors with vascular extension into cardiac chambers (7,8). Many well-established effective
75 interventions are reproducible by a variety of academic and community-based surgeons.
76 Nevertheless, there is also a large and rapidly growing literature in the field attempting to assess
77 newer interventions or modifications of established interventions. Benefits and advantages could
78 come in different forms. A newer intervention may represent an invasive procedure treating a

79 previously medically observed, non-intervenable condition, or may involve a more
80 invasive/aggressive approach than the older procedure which has historically represented the gold
81 standard. Typically, the goal would be to show that the new intervention affords a mortality benefit in
82 order to justify it. Alternatively, a new intervention may not aim to improve mortality, but may be
83 considered to have less invasiveness (and hence perhaps fewer complications and/or lesser cost),
84 therefore rendering it preferable to the established standard-of-care if non-inferiority regarding
85 mortality could be demonstrated.

86 The gold standard for testing and evaluating treatment interventions is afforded by
87 randomized controlled trials (RCTs), especially when expected effects are likely to be modest and
88 risk-benefit ratios need careful examination. Moreover, it is widely established that when many
89 RCTs exist, careful systematic review and meta-analysis may help examine the overall evidence to
90 have guidance for eventual clinical decision-making. Many non-randomized studies are also
91 conducted and published. This is even more true for surgical specialties where RCTs may be difficult
92 to conduct (9-11). These non-randomized (henceforth called observational) studies can also be
93 examined in meta-analyses and may also influence eventual guidance and decision-making. The
94 relative merits and comparative outcomes of observational versus randomized evidence has been
95 heavily contested and debated across diverse clinical areas across medicine. Some meta-research
96 evaluations have combined data from multiple meta-analyses where both RCTs and observational
97 studies are available on the same topic (12-14). Some evaluations suggest that, on average, RCTs and
98 observational studies yield similar estimates of the treatment effect (12). Others have found that
99 observational studies tend to have exaggerated estimates of benefit compared with RCTs (13,14). To
100 our knowledge, no such systematic evaluation has been performed including diverse topics in the
101 field of adult cardiac surgery, a field where many observational studies publish estimates of
102 intervention effectiveness.

103 Here, we undertook a meta-research evaluation that systematically identified and analyzed
104 meta-analyses of adult cardiac surgery interventions where the outcome was mortality and where at
105 least one RCT was available. We aimed to assess how often RCTs in the field found significant
106 benefits for newer interventions versus older ones or doing nothing; whether treatment effect
107 estimates changed over time in the last several decades; and whether RCTs and observational studies
108 on the same intervention comparison give similar results.

109 **METHODS**

110 *Search strategy and eligibility criteria*

111 This is a meta-research project (15) reported following the PRISMA 2020 checklist for
112 systematic reviews and meta-analyses (16), adapted to our specific meta-research design.

113 We considered as eligible all systematic reviews and meta-analyses (SRMAs) related to adult
114 cardiac surgery that contained randomized controlled trials (RCTs) (regardless of whether
115 observational non-randomized studies were also included or not) and summarized all-cause mortality
116 outcomes. We searched in publication venues likely to publish most of these adult cardiac surgery
117 SRMAs. Specifically, the search strategy captured for further screening all SRMAs published in the
118 Journal of Thoracic and Cardiovascular Surgery and Annals of Thoracic Surgery (the two primary
119 journals specializing in the field of interest) and all SRMAs that were likely to address cardiac
120 surgery in any of the 4 highest impact factor surgical journals, the 4 highest impact factor general
121 medical journals, and the comprehensive Cochrane Database of Systematic Reviews. The complete
122 search string was ((Annals of Surgery [SO] OR Surgery [SO] OR JAMA Surgery [SO] OR British
123 Journal of Surgery [SO] OR New England Journal of Medicine [SO] OR JAMA [SO] OR Lancet
124 [SO] OR BMJ [SO] OR Cochrane Database of Systematic Reviews [SO]) AND ('cardiac surgery' OR
125 'heart surgery' OR 'coronary bypass' OR CABG OR 'valve surgery')) OR (ann thorac surg[Journal]

126 OR j thorac cardiovasc surg[Journal]) AND (systematic[sb] OR meta-analysis[sb]). The search was
127 last updated on August 1st, 2023.

128 Two authors (GT/AP) screened the abstracts of all retrieved items for eligibility. Items were
129 excluded if they were not SRMAs, not studying the effect of interventions on all-cause mortality,
130 contained no RCTs, or were studying topics unrelated to current adult cardiac surgery. When
131 SRMAs on the same comparison and outcomes (duplicates) were identified, the SRMA with most
132 RCTs was included; with ties, we included the one that had more observational studies.

133 *Data extraction from eligible meta-analyses*

134 For all included eligible SRMAs, we extracted journal, year of publication, and data for all-
135 cause mortality outcomes, including the specific outcome, effect metric (odds ratio [OR], risk ratio
136 [RR], hazard ratio [HR], or incident rate ratio [IRR]), specific comparison (experimental and control
137 groups), the overall reported meta-analysis summary effect and effects from individual studies
138 (summarized in tables or forest plots). For each individual study, we extracted the effect size, 95%
139 confidence interval (CI), and the 2x2 table whenever available. When SRMAs presented all-cause
140 mortality outcomes for different timepoints, all timepoints were extracted.

141 The summary estimate presented by the authors in the SRMA was extracted. When both
142 adjusted and unadjusted study results were synthesized, we preferred the latter. For 22 SRMAs and a
143 total of 221 individual studies, a summary estimate was not available, but we could calculate OR
144 from 2x2 tables of the individual studies and obtain the summary OR thereof.

145 For each comparison, we recorded whether the comparison was between an active
146 experimental group and usual care, placebo (sham) or an active control group. We also assessed and
147 recorded the direction of each eligible comparison. Whenever an older intervention was the
148 experimental group and a newer intervention was the control group, we inverted the effect sizes (e.g.
149 3 became $1/3=0.33$), so that eventually all comparisons in our analysis reflect the comparison of a

150 newer intervention versus an older one (or nothing/placebo/sham). Accordingly, effect sizes (relative
151 risks) <1.0 represent a reduction in mortality favoring the newer intervention. Whether such coining
152 (inversion) was needed was determined by consensus between all authors (AP/GT/JPAI) regarding
153 which intervention is newer.

154 *Validation of data from primary studies*

155 We randomly selected 30 individual studies from the 34 SRMAs that contained both RCTs
156 and observational studies (from a total of 400 individual studies, n=95 RCTs and n=305
157 observational studies) and re-calculated the study effect to confirm the numerical effect reported by
158 the SRMA authors.

159 *Statistical techniques*

160 Random effects meta-analysis was carried out using the Sidik-Jonkman estimator (17). All
161 effects were log-transformed before meta-analysis. Heterogeneity was estimated with the Higgins and
162 Thompson's I^2 statistic and Cochran's Q test (18,19). Meta-analysis calculations were performed
163 using the meta package in R (20). Continuous data was also summarized using medians and
164 interquartile ranges (IQR); the Spearman correlation coefficient was calculated between continuous
165 variables (21).

166 All comparisons of interventions were classified into two types, type A (comparing a newer
167 less invasive method versus an older more invasive method looking at non-inferiority) and type B
168 (comparing a newer more invasive method versus an older less invasive method looking at
169 superiority). Type B includes also all comparisons against doing nothing (or giving placebo/sham).
170 Sensitivity analyses were performed where the summary effect was inverted for all type A
171 comparisons.

172 For those SRMAs that included also observational studies, we compared the results of RCTs
173 versus observational studies in each topic by obtaining the relative summary effect, i.e. the fixed

174 effects summary effect across observational studies divided by fixed effects summary effect across
175 RCTs) (22). We then synthesized the relative summary effects across all topics that included the
176 RCTs and observational studies. A relative summary effect <1.0 suggests that the observational
177 studies show better results for the newer intervention regarding mortality than RCTs.

178 Statistical significance was assessed at two levels, $\alpha = 0.05$ (suggestive significance) and $\alpha =$
179 0.005 (formal statistical significance), as proposed by Benjamin et al. (23).

180 R version 4.1.0 was used for all calculations (R Core Team 2021).

181 **RESULTS**

182 *Eligible SRMAs and single studies*

183 Our literature search identified 927 items. Of these, 531 (57%) were unrelated to cardiac
184 surgery, 170 (18%) included no RCTs (18%), 50 (5%) did not include mortality outcomes, 26 (3%)
185 were not SRMAs presenting extractable data, 23 (2%) were not studies of interventions, 31 (3.3%)
186 were repeat topics, and 1 was withdrawn. Of the 95 remaining publications of SRMAs, 61 (64%)
187 were found to have extractable data on mortality from forest plots and were included (**Figure 1**). For
188 the other 34 that had no clearly extractable data per study in forest plots, the corresponding authors
189 were contacted, but none offered usable data. Of the 61 publications, 25 (41%) were published in the
190 Journal of Thoracic and Cardiovascular Surgery, 22 (36%) in Annals of Thoracic Surgery, 12 (20%)
191 in the Cochrane Database of Systematic Reviews, and 2 (3%) in Annals of Surgery.

192 10 of the 61 papers contained two or more different eligible comparisons ($n=2$ papers) or
193 timepoints for the comparison of the same intervention ($n=8$ papers). Overall, 73 different
194 comparisons were eligible for our analyses; of them, 6 (8%) were inverted to consistently have the
195 newer (experimental) intervention compared against the older (control) intervention. The 73
196 comparisons are shown in **Supplementary Table 1**. They included a total of 824 mortality study
197 outcomes, 519 RCTs and 305 observational ones. Of the 824, 490 were presented as OR, 242 RR, 55

198 HR, and 37 IRR metrics. Across the 73 meta-analyses, each contained a median of 10 studies (IQR 6
199 to 17). Of the 73 meta-analyses, 29 (416 study outcomes) were type A comparisons and 44 (408
200 study outcomes) were type B comparisons.

201 *Effect sizes and statistical significance in single studies*

202 The median mortality effect size across 824 individual study outcomes was 1.00 (IQR, 0.54-
203 1.30), suggesting no difference in mortality, on average, between experimental and control
204 interventions. Across 519 RCTs, the median effect size was also 1.00 (IQR, 0.58-1.32), whereas
205 across 305 observational studies it was 0.91 (IQR, 0.50-1.26). The difference between RCTs and
206 observational studies suggested statistical significance (Wilcoxon rank sum $p=0.039$).

207 Across the 519 RCTs, 36 (6%) had suggestive significance ($p<0.05$), but only 11 (2%) were
208 formally statistically significant with $p=0.005$. Suggestive and formal significance was seen in 18
209 and 4 RCTs in favor of the newer (experimental) intervention and in 18 and 7 RCTs in favor of the
210 control intervention. When limited to the 277 RCTs of type B (presumably superiority) comparisons,
211 16 (6%) and 5 (2%) had suggestive and formal significance and almost always this favored the newer
212 intervention (**Table 1**).

213 Across the 305 observational studies, there were 38 (13%) and 10 (3%) with suggestive and
214 formal statistical significance. Most of those favored the newer (experimental) intervention (27 and 6,
215 respectively) and fewer favored the control intervention (11 and 4, respectively). When limited to 131
216 type B (presumably superiority) comparisons, 25 (19%) and 4 (3%) showed suggestive and formal
217 significance, almost always favoring the newer intervention (**Table 1**).

218 *Treatment effects over time*

219 A total of 122 RCTs and 125 observational studies had been published in the last decade
220 (2012 or later). The median effect size of these 247 studies was 1.00 (IQR, 0.50-1.26) overall (1.00,
221 IQR, 0.60-1.36 in RCTs and 0.86, IQR, 0.46-1.05 in observational studies). The majority of

222 statistically significant RCTs with $p < 0.005$ (8/11, 73%) were published in the last decade and of
223 these, 6/8 (75%) were in favor of the control intervention. Conversely, among observational studies
224 with $p < 0.005$, 6/10 (60%) had been published in the last decade and 5/6 (83%) were in favor of the
225 newer (experimental) intervention (**Table 1**).

226 There was no significant change in the effect size for newer versus control interventions over
227 time ($p = 0.56$, $p = 0.55$ after controlling for specific comparisons). For RCTs, there was no significant
228 time effect ($p = 0.42$; after adjusting for specific comparisons, $p = 0.40$). For observational studies, there
229 was a suggestion of a time association with a trend for increasing effects over time ($p = 0.049$; after
230 adjusting for specific comparisons, $p = 0.065$) (**Figure 2**).

231 *Treatment effects in meta-analyses*

232 As reported by the authors, the median meta-analysis summary effect across 73 comparisons
233 was 0.90 (IQR, 0.69-1.12). Of the 73 meta-analyses, as reported by the authors, 15 (21%) had
234 summary effects with suggestive significance with $p < 0.05$ and 4 (5%) had summary effects that were
235 statistically significant with $p < 0.005$ level in favor of the experimental arm, while 6 (8%) and 3 (4%)
236 meta-analyses had effects at these significance levels favoring the control arm. When we re-analyzed
237 the data with random effects and the SJ method, the respective numbers were 5, 2, 3 and 1.

238 When limited to data from the 519 RCTs outcomes, the median meta-analysis summary effect
239 was 0.94 (IQR, 0.74 to 1.19); of the 73 meta-analyses based on RCT data, only 2 were statistically
240 significant ($p < 0.005$) in favor of the experimental arm and 2 in favor of the control arm.

241 When limited to type B (presumably superiority) comparisons (44 meta-analyses total), 14
242 meta-analyses had summary effects with suggestive significance with $p < 0.05$ and 4 had summary
243 effects that were statistically significant with $p < 0.005$ in favor of the experimental arm; conversely,
244 only 1 and 0 meta-analyses had effects at these significance levels favoring the control arm. When
245 limited to data from the RCTs, the respective numbers were 7, 2, 0, and 0.

246 *Clinical topics with statistically significant differences*

247 **Table 2** shows the 7 clinical topics where there were statistically significant differences in
248 mortality at the $p < 0.005$ level in the meta-analyses, as originally reported by the systematic review
249 authors, 3 of which retained $p < 0.005$ in our re-calculated SJ random effects estimates.

250 *Relative effects in RCTs and observational studies*

251 Of 73 meta-analyses, 34 (47%) contained at least one RCT and at least one observational
252 study; across these 34 meta-analyses there were 400 individual studies (95 were RCTs (24%) and 305
253 observational studies). The median observational summary effect was 0.81 (IQR, 0.68-1.05) vs 0.93
254 (IQR, 0.77-1.22) in the respective RCTs (paired Wilcoxon $p = 0.17$). The Spearman correlation
255 between RCT and observational summary effect sizes was -0.023 ($p = 0.90$).

256 The relative summary effect (summary effect in observational studies divided by summary
257 effect in RCTs) had a median of 0.87 (IQR, 0.55-1.29); meta-analysis of the relative summary effects
258 yielded a summary of 0.93 (95% CI 0.74-1.18, $p = 0.57$, $I^2 = 0\%$, 0-39%) (**Figure 3**). Respective results
259 were 0.86 (IQR, 0.41-1.04) and 0.85 (95% CI, 0.55-1.34, $p = 0.85$, $I^2 = 0\%$, 0%-57%) for type A
260 comparisons and 1.08 (IQR, 0.78-1.44) and 0.98 (95% CI, 0.75-1.27, $p = 0.85$, $I^2 = 0\%$, 0%-47%) for
261 type B comparisons.

262 Sensitivity analyses yielded similar results (see **Supplementary Results**).

263 *Validation of primary data*

264 In the 30 randomly selected studies (from 22 SRMAs), our re-calculated study effect was
265 always the same or very similar (maximum deviation 0.04 in relative risk scale) to the effect reported
266 by the SRMA authors.

267 **DISCUSSION**

268 Our analysis of 73 comparisons of newer interventions versus control management in adult
269 cardiac surgery shows that mortality benefits of new interventions versus older approaches were

270 uncommon. The median effect size of 1.00 suggests no difference in mortality between the compared
271 groups. The same picture was seen when examining strictly the data from RCTs. Conversely, on
272 average, observational data suggested some benefit for the experimental intervention groups.
273 Evidence from RCTs rarely suggested statistically significant benefits of the experimental
274 intervention at the level of single studies or meta-analyses thereof. Conversely, a modest fraction of
275 non-randomized observational data suggested significant benefits for the experimental arm. Effect
276 sizes in RCTs have been steady over time, while effect sizes for non-randomized observational data
277 may tend to become larger in more recent years.

278 Only about half of the 73 topics that we examined included both RCTs and observational data
279 in their meta-analyses. Moreover, very few RCTs were included in these topics where a direct
280 comparison of effect sizes RCTs versus observational studies would be feasible within the same
281 topic. Therefore, the results from these direct comparisons have large CIs and are underpowered to
282 detect clear differences between the two designs. Nevertheless, the picture in this subset is
283 compatible with the overall picture across all 73 comparisons.

284 There is debate about the merits and uses of non-randomized observational studies to
285 evaluate comparative effectiveness (24,25). Our findings suggest caution in making definitive
286 inferences and formulating guidelines about comparative effectiveness based on observational data in
287 adult cardiac surgery. Observational data have more degrees of freedom and this may result in higher
288 chances of selective reporting that fits some expected narrative. RCTs certainly have their own
289 biases and they are not infallible gold standards. Nevertheless, the effort to promote the use of RCTs
290 in surgery is welcome (26).

291 The dearth of well-documented benefits in mortality with newer, experimental interventions
292 in adult cardiac surgery does not mean that the field does not make progress. Half of the data that we
293 analyzed pertained to type A comparisons, where the newer intervention was a less invasive one and

294 thus the goal would probably not be to show necessarily better survival, but other comparative
295 benefits, e.g. fewer complications, better convenience or lesser cost. Moreover, among the
296 comparisons that had a superiority outlook with the newer intervention being more aggressive than
297 the control, we could identify two where the newer intervention did significantly better in mortality
298 outcomes than the control at $p < 0.005$. Both of them reflected classic, widely-used interventions
299 where the magnitude of the survival benefit was very large, i.e. use of intra-aortic balloon pump and
300 surgery for asymptomatic severe aortic stenosis. For new and future interventions, one may need to
301 be prepared to detect much smaller but still clinically meaningful survival benefits. This would
302 require larger studies to have sufficient power to detect such benefits with certitude.

303 We evaluated results both at the level of effect sizes and at the level of statistical significance.
304 Effect sizes are preferable to convey the magnitude of the potential benefit (27). Nevertheless, we
305 should caution that the data are typically presented in relative effect metrics. Clinical decision-
306 making should also examine carefully the absolute magnitude of benefits. Also for statistical
307 significance, some authors argue that dichotomizing p-values may be spurious and misleading and
308 have even urged to abandon the notion of statistical significance (28). Nevertheless, for clinical trials
309 and their SRMAs (as analyzed here), the notion of statistical significance still has relevance: its
310 abandonment may open the door to even more spurious practices, where investigators may move the
311 goalposts without any statistical rules (29). The use of $p < 0.005$ instead of $p < 0.05$ is recommended to
312 avoid false-positive results (23). This suggestion has also empirical grounding in RCTs and meta-
313 analyses thereof (30).

314 Our study has some limitations. First, we only focused on SRMAs published in the key
315 journals that are likely to publish SRMAs in this specialty, as it would have been inconvenient and
316 low-yield to screen hundreds of thousands of SRMAs published across the entire medical literature.
317 Second, SRMAs may not have been performed yet on some interesting topics and this may affect

318 more prominently recent developments in the field with very recent data. Third, while most of the
319 included data used OR metrics, some used other relative effect metrics and these are not identical.
320 Nevertheless, for the mostly small effects analyzed here, differences between OR and HR, RR, or
321 IRR are probably small. Fourth, while we took meticulous care with data, we depended in existing
322 published SRMAs and these evidence syntheses may include their own biases and errors. There are
323 many efforts to improve the methodological rigor of conduct and the quality of reporting of both
324 primary studies and of SRMAs (31). The importance of enhanced attention to these matters cannot be
325 overstated. Fifth, we could not evaluate whether improved outcomes with new interventions may
326 exist specifically for subgroups of patients, e.g. in recent years there has been an expansion of the
327 pool of surgical candidates. Patients who were considered inoperable in the past due to advanced age
328 or underlying conditions are routinely offered surgery in the current era. Finally, no single study and
329 no single SRMA can be taken as perfect evidence, no matter how well it is done. Therefore,
330 uncertainty about the comparative effects of different interventions may be larger than actually
331 observed.

332 Allowing for these caveats, our meta-research analysis offers a bird's eye view of over 800
333 study effect size estimates from comparisons involving adult cardiac surgery. These data can be used
334 as a benchmark for what might be reasonable expectations for future RCT and comparative
335 effectiveness research and SRMAs in the field.

336

337 **References**

- 338 (1) Cutler EC, Levine SA. Cardiomy and valvulotomy for mitral stenosis. *Boston Med Surg J*
339 1923;188:1023-7.
- 340 (2) Gibbon JH Jr. Application of a mechanical heart and lung apparatus to cardiac surgery. *Minn*
341 *Med* 1954;37(3):171-185.
- 342 (3) Coronary Artery Surgery Study (CASS): a randomized trial of coronary artery bypass
343 surgery. Survival data. *Circulation* 1983;68(5):939-950.
- 344 (4) Carabello BA. Clinical practice. Aortic stenosis. *N Engl J Med* 2002;346-677.
- 345 (5) David TE. Aortic valve repair and aortic valve-sparing operations *J Thorac Cardiovasc Surg*
346 2015;149:9.
- 347 (6) Carpentier A. Cardiac valve surgery: the “grench correction.” *J Thorac Cardiovasc Surg*
348 1983;86:323.
- 349 (7) Dein JR, Frist WH, Stinson EF, et al: Primary cardiac neoplastms: early and late results of
350 surgical treatment in 42 patients. *J Thorac Cardiovasc Surg* 1987;93:502.
- 351 (8) Bossert Torsten B, Gummert JF, Battellini, et al. Surgical experience with 77 primary cardiac
352 tumors. *Interact Cardiovasc Thorac Surg* 2005;4:311-315.
- 353 (9) Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Balliol Collaboration, et al.
354 Evaluation and stages of surgical innovations. *Lancet*. 2009;374:1089–96.
- 355 (10) Gelijns AC, Ascheim DD, Parides MK, Kent KC, Moskowitz AJ. Randomized trials
356 in surgery. *Surgery*. 2009;145:581–7.
- 357 (11) Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien P-A, Reeves BC, et al.
358 Challenges in evaluating surgical innovation. *Lancet*. 2009;374:1097–104.

- 359 (12) Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational
360 study designs compared with those assessed in randomized trials. *Cochrane Database Syst*
361 *Rev.* 2014 Apr 29;2014(4):MR000034.
- 362 (13) Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Agreement of treatment
363 effects for mortality from routinely collected data and subsequent randomized trials: meta-
364 epidemiological survey. *BMJ.* 2016 Feb 8;352:i493.
- 365 (14) Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG,
366 Contopoulos-Ioannidis DG, Lau J. Comparison of evidence of treatment effects in
367 randomized and nonrandomized studies. *JAMA.* 2001 Aug 15;286(7):821-30.
- 368 (15) Lozada-Martinez ID, Ealo-Cardona CI, Marrugo-Ortiz AC, Picón-Jaimes YA,
369 Cabrera-Vargas LF, Narvaez-Rojas AR. Meta-research studies in surgery: a field that should
370 be encouraged to assess and improve the quality of surgical evidence. *Int J Surg.* 2023 Jun
371 1;109(6):1823-1824.
- 372 (16) IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for
373 random effects meta-analysis is straightforward and considerably outperforms the standard
374 DerSimonian-Laird method. *BMC Med Res Methodol.* 2014 Feb 18;14:25.
- 375 (17) Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD,
376 Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM,
377 Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA,
378 Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020
379 statement: An updated guideline for reporting systematic reviews. *Int J Surg.* 2021
380 Apr;88:105906.
- 381 (18) Sedgwick P. Meta-analyses: what is heterogeneity? *BMJ.* 2015 Mar 16;350:h1435.

- 382 (19) von Hippel PT. The heterogeneity statistic I^2 can be biased in small meta-analyses.
383 BMC Med Res Methodol. 2015 Apr 14;15:35.
- 384 (20) Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a
385 practical tutorial. Evid Based Ment Health. 2019 Nov;22(4):153-160.
- 386 (21) de Winter JCF, Gosling SD, Potter J. Comparing the Pearson and Spearman
387 correlation coefficients across distributions and sample sizes: A tutorial using simulations and
388 empirical data. Psychological Methods 2016;21(3):273–290.
- 389 (22) Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and
390 discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med.
391 2001 Dec 4;135(11):982-9.
- 392 (23) Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al.
393 Redefine statistical significance. Nat Hum Behav. 2018 Jan;2(1):6-10.
- 394 (24) Gerstein HC, McMurray J, Holman RR. Real-world studies no substitute for RCTs in
395 establishing efficacy. Lancet. 2019 Jan 19;393(10168):210-211.
- 396 (25) Djulbegovic B, Glasziou P, Chalmers I. The importance of randomised vs non
397 randomised trials. Lancet. 2019 Aug 24;394(10199):634-635.
- 398 (26) McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al.
399 No surgical innovation without evaluation: the IDEAL recommendations. Lancet.
400 2009;374:1105–12.
- 401 (27)
402 Ioannidis JPA. Options for publishing research without any P-values. Eur Heart J. 2019 Aug
403 14;40(31):2555-2556.
- 404 (28) Amrhein V, Greenland S, McShane B. Scientists rise up against statistical
405 significance. Nature. 2019 Mar;567(7748):305-307.

- 406 (29) Ioannidis JPA. Retiring statistical significance would give bias a free pass. *Nature*.
407 2019 Mar;567(7749):461.
- 408 (30) Koletsi D, Solmi M, Pandis N, Fleming PS, Correll CU, Ioannidis JPA. Most
409 recommended medical interventions reach $P < 0.005$ for their primary outcomes in meta-
410 analyses. *Int J Epidemiol*. 2020 Jun 1;49(3):885-893.
- 411 (31) Kolaski K, Logan LR, Ioannidis JPA. Guidance to best tools and practices for
412 systematic reviews. *JBIEvid Synth*. 2023;21(9):1699-1731.
- 413

414 **Table 1.** RCTs and non-randomized observational studies with suggestive ($p < 0.05$) and formally
 415 statistically significant ($p < 0.005$) results according to type of comparison and year of publication.

| | p < 0.05, favoring experimental intervention | p < 0.005, favoring experimental intervention | p < 0.05, favoring control intervention | p < 0.005, favoring control intervention |
|------------------------------------|--|---|---|--|
| PER TYPE OF COMPARISON | | | | |
| RCT, type A (n=242) | 3 (1%) | 0 (0%) | 17 (7%) | 6 (2%) |
| RCT, type B (n=277) | 15 (5%) | 4 (1%) | 1 (0.4%) | 1 (0.4%) |
| Observational, type A (n=174) | 7 (4%) | 3 (2%) | 6 (3%) | 3 (2%) |
| Observational, type B (n=131) | 20 (15%) | 3 (2%) | 5 (4%) | 1 (1%) |
| PER YEAR OF PUBLICATION | | | | |
| RCT 2012-21 (n = 122) | 7 (5.7%) | 2 (1.6%) | 14 (11.5%) | 6 (4.9%) |
| RCT pre-2012 (n = 397) | 11 (2.8%) | 2 (0.5%) | 4 (1.0%) | 1 (0.3%) |
| Observational 2012-21 (n = 125) | 14 (11.2%) | 5 (4.0%) | 2 (1.6%) | 1 (0.8%) |

| | | | | |
|-------------------------------------|-----------|----------|----------|----------|
| Observational pre-2012 (n = 180) | 13 (7.2%) | 1 (0.6%) | 9 (5.0%) | 3 (1.7%) |
|-------------------------------------|-----------|----------|----------|----------|

416

417 Type A comparisons involve comparing a newer less invasive method versus an older more invasive
418 method, thus presumably demonstration of non-inferiority might be sufficient to make the newer
419 approach attractive. Type B comparisons involve comparing a newer more invasive method versus
420 an older less invasive method, thus presumably demonstration of superiority would be needed to
421 show that the added invasiveness is worthwhile. Type B includes also all comparisons against doing
422 nothing (or giving placebo/sham).

423

424

425 **Table 2.** Clinical topics with statistically significant survival differences at $p < 0.005$ in meta-analyses

426

| Intervention (type A or B) | Condition/ Setting | Metric | Author reported meta-analysis effect | SJ random effects meta-analysis effect | RCTs only random effects meta-analysis effect | Mortality follow-up |
|---|---------------------------------|---------------|---|---|--|----------------------------|
| Loading dose vs regular dose of statins (B) | Prophylactic after CABG | OR | 0.74 (0.60-0.91) | 0.78 (0.57-1.06) | 0.78 (0.57-1.06) | 1 to 10 years |
| Off-pump vs on-pump (A) | CABG | HR | 1.07 (1.03-1.11) | 1.04 (0.95-1.14) | 1.20 (0.81-1.76) | 5 year |
| IABP vs No IABP (B) | Pre-op prophylactic before CABG | OR | 0.35 (0.23-0.53) | 0.36 (0.20-0.65) | 0.25 (0.11-0.56) | In-hospital or 30-day |
| Radial artery versus saphenous vein as the second conduit (B) | CABG | IRR | 0.74 (0.63-0.87) | 0.72 (0.54-0.95) | 0.94 (0.61-1.44) | 0.7 to 11.9 years |
| Early surgery vs conservative management (B) | Asymptomatic severe AS | HR | 0.49 (0.36-0.68) | 0.48 (0.32-0.72) | 0.33 (0.12-0.91) | 1.8 to 6.2 years |
| TAVR PCI vs SAVR CABG (A) | CABG | HR | 1.35 (1.11-1.65) | 1.34 (1.05-1.71) | 1.34 (1.05-1.71) | 1 to 3 years |
| PCI vs CABG (A) | Diabetic patients | OR | 1.64 (1.30-2.08) | 1.63 (1.23-2.17) | 1.63 (1.23-2.17) | 5 year |

427

428 AS: aortic stenosis, CABG: Coronary artery bypass graft, CI: confidence interval, IABP: Intra-aortic

429 balloon pump, PCI: percutaneous coronary intervention

430 Follow-up shows the time of mortality assessment.

431

432

433

434

435

436

437 **Figure legends**

438 **Figure 1.** Flow chart for eligible meta-analyses

439 **Figure 2.** Study effect size vs year of publication, with associated line of best fit, for observational
440 studies (n = 305) and RCTs (n = 519)

441 **Figure 3.** Forest plot of relative summary effect (summary effect in observational studies divided by
442 summary effect in RCTs).

443

444

445

446

447

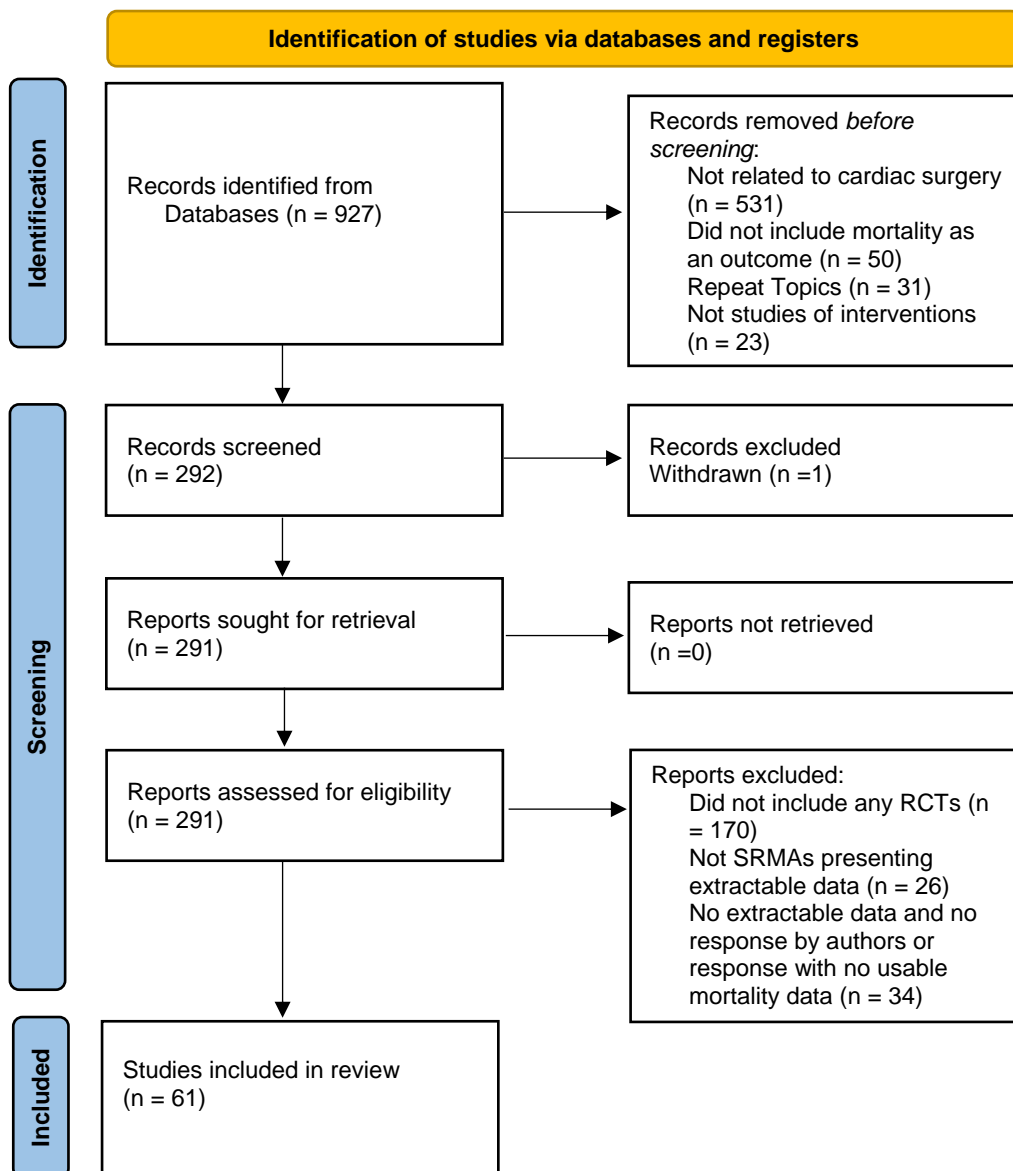
448

449

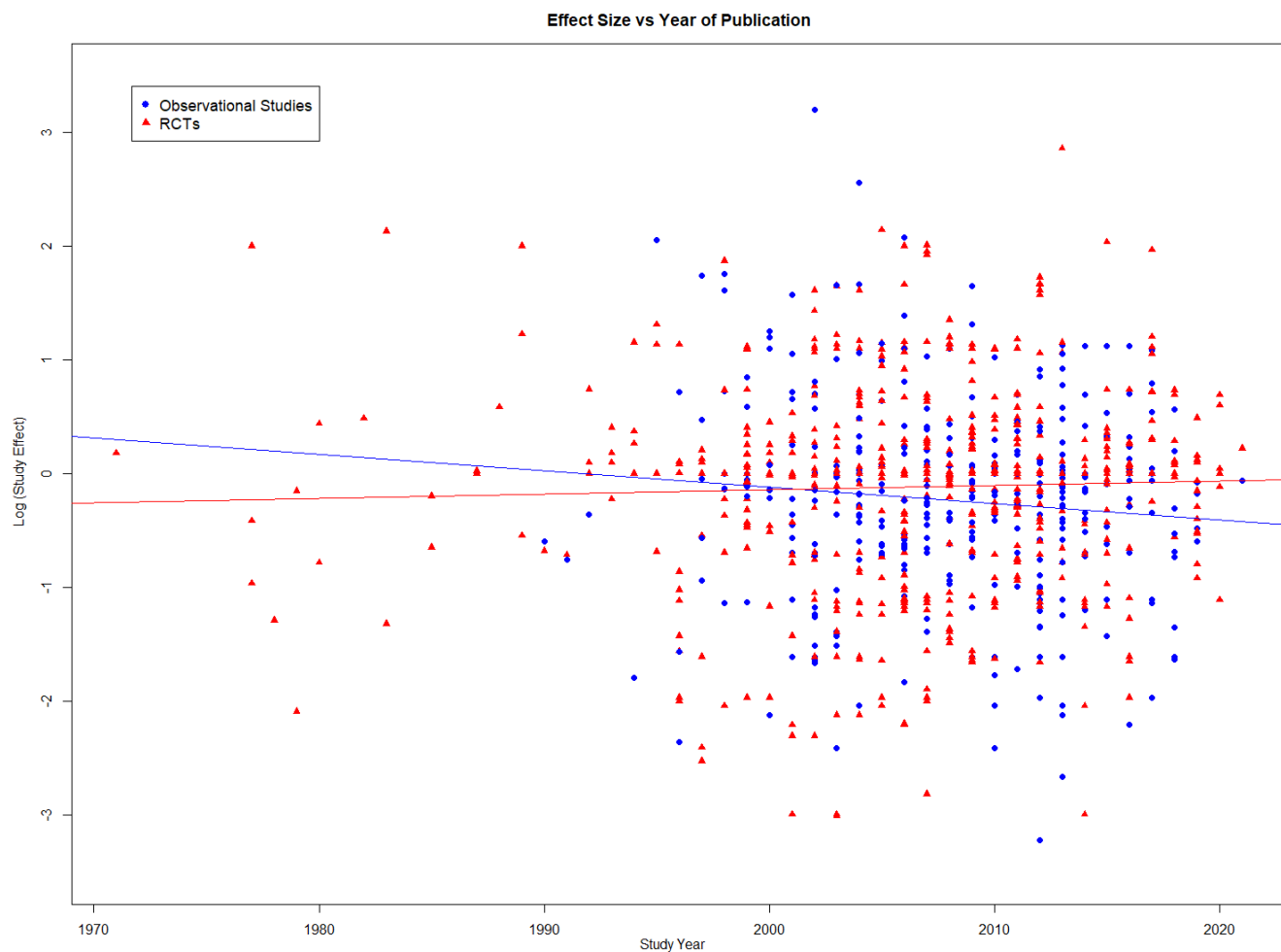
450

451 **FIGURE 1**

452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486



487 **FIGURE 2**

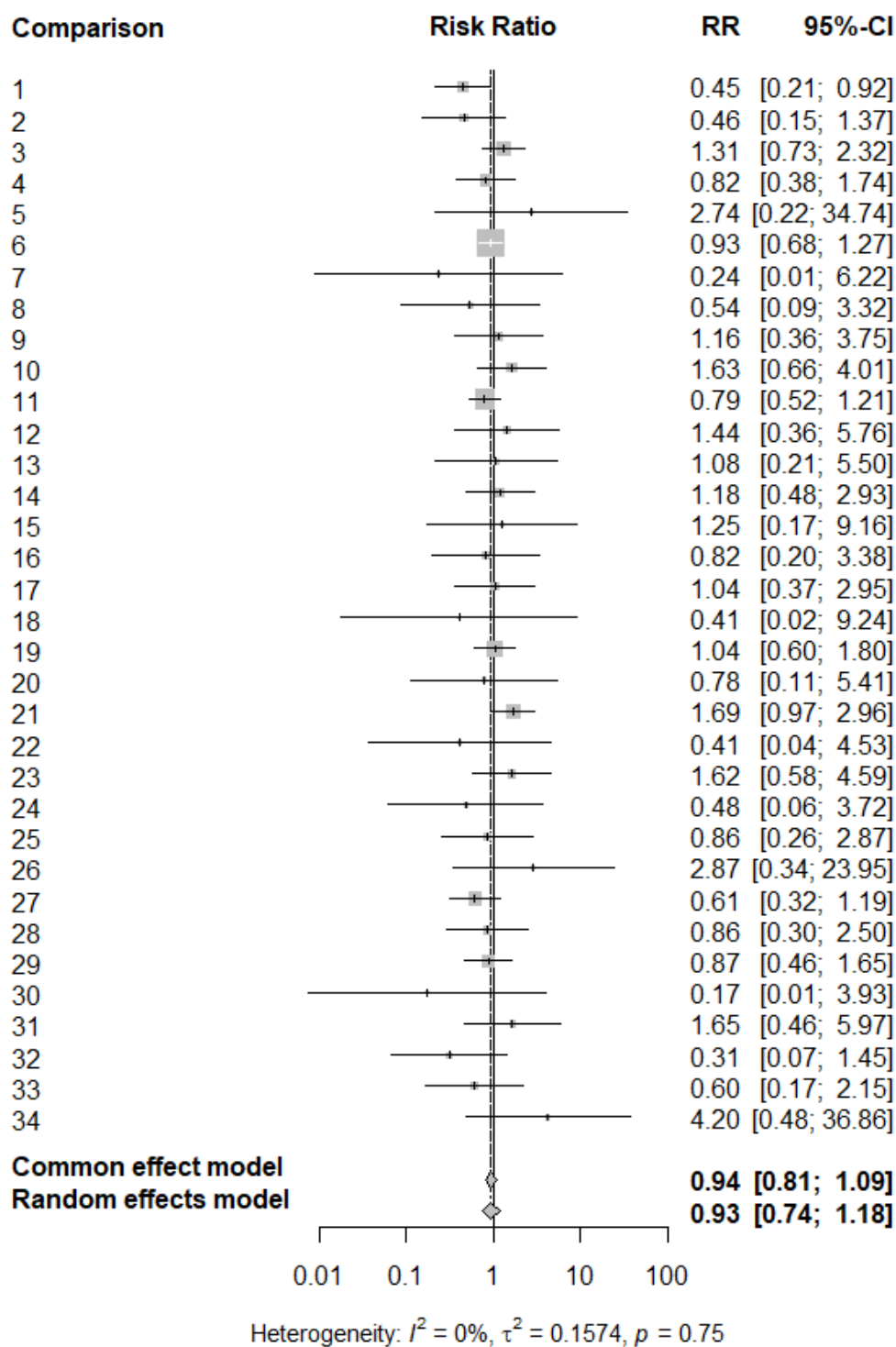


488

489

490

491 **FIGURE 3**



492

493

494

495 **Supplementary Table 1.** 73 comparisons (comparison-outcome pairs) across 61 SRMAs.
 496 *Abbreviations:* ECMO: Extra Corporeal Membrane Oxygenation. PCI: Percutaneous Coronary
 497 Intervention. CABG: Coronary Artery Bypass Grafting. ARDS: Acute Respiratory Distress
 498 Syndrome. DAPT: Dual Anti-Platelet Therapy. SAPT: Single Anti-Platelet Therapy. TAVR:
 499 Transcatheter Aortic Valve Replacement. AVR: Aortic Valve Replacement. IABP: Intra-Aortic
 500 Balloon Pump. TVA: Tricuspid Valve Annuloplasty. OPCAB: Off Pump Coronary Artery Bypass.
 501 SAVR: Surgical Aortic Valve Replacement (same as AVR). MIDCAB: Minimally Invasive Direct
 502 Coronary Artery Bypass.
 503

| Comparison Number From Fig 3 | Title of SRMA | Comparison | Outcome Name |
|-------------------------------------|--|---|--|
| 1 | Expert consensus guidelines: Examining surgical ablation for atrial fibrillation | Surgical ablation vs No ablation, just index operation | 30 day mortality |
| | Percutaneous coronary intervention versus coronary artery bypass grafting: a meta-analysis | PCI vs CABG | Death at latest available follow-up time point |
| | Contemporary extracorporeal membrane oxygenation therapy in adults: Fundamental principles and systematic review of the evidence | ECMO vs "Endotracheal ventilation +/- mechanical assist | Death at primary endpoint |

| | | | |
|---|---|---|--------------------------------|
| | | devices" for ARDS patients | |
| | N-acetylcysteine in cardiovascular-surgery-associated renal failure: a meta-analysis | NAC vs no adjunct medication | Mortality |
| | Preoperative chlorhexidine mouthwash to reduce pneumonia after cardiac surgery: A systematic review and meta-analysis | Chlorhexidine mouthwash vs No preop skin prep (outside of operating room) | In-hospital mortality |
| 2 | Comparing Single- and Dual-Antiplatelet Therapies After Transcatheter Aortic Valve Implantation | DAPT vs SAPT after TAVR | Short-term all-cause mortality |
| 3 | Comparing Single- and Dual-Antiplatelet Therapies After Transcatheter Aortic Valve Implantation | DAPT vs SAPT after TAVR | Long-term all-cause mortality |
| 4 | Ring or suture annuloplasty for tricuspid regurgitation? A meta-analysis review | RING vs NO RING | Early mortality after surgery |
| 5 | A Meta-Analysis of Miniaturized Versus Conventional Extracorporeal Circulation in Valve Surgery | MECC vs CECC | Mortality |
| | Closure of the sternum with anchoring of the steel wires: Systematic review and meta-analysis | Anchoring vs no anchoring | Death |

| | | | |
|---|---|--|-----------------------------------|
| | Meta-Analysis of Medium and Long-Term Efficacy of Loading Statins After Coronary Artery Bypass Grafting | Loading dose of statins vs regular dose of statins | Mortality |
| 6 | Worse long-term survival after off-pump than on-pump coronary artery bypass grafting | Off-pump vs on-pump CABG | 5 year all-cause mortality |
| 7 | Beating-Heart Versus Conventional On-Pump Coronary Artery Bypass Grafting: A Meta-Analysis of Clinical Outcomes | BH-ONCAB vs C-ONCAB | Early mortality |
| 8 | Mitral valve surgery: right lateral minithoracotomy or sternotomy? A systematic review and meta-analysis | MIV vs Conv | All-cause mortality up to 30 days |
| | Outcomes of Coronary Artery Bypass and Stents for Unprotected Left Main Coronary Stenosis | Stenting vs CABG | Medium-term death |
| | A systematic review of biocompatible cardiopulmonary bypass circuits and clinical outcome | Coated vs Standard cardiopulmonary bypass tubing | Death |
| 9 | Gentamicin-collagen sponge reduces the risk of sternal wound infections after heart surgery: Meta-analysis | Gentamycin sponge vs No prophylaxis | Mortality |
| | Volatile anesthetics in preventing acute kidney injury after cardiac surgery: a | Volatile vs TIVA | All-cause mortality |

| | | | |
|----|--|--|--|
| | systematic review and meta-analysis | | |
| | Preoperative Prophylactic Intraaortic Balloon Pump Reduces the Incidence of Postoperative Acute Kidney Injury and Short-Term Death of High-Risk Patients Undergoing Coronary Artery Bypass Grafting: A Meta-Analysis of 17 Studies | IABP vs No IABP | Short-term death in high-risk patients undergoing CABG |
| 11 | Radial artery versus saphenous vein as the second conduit for coronary artery bypass surgery: A meta-analysis | Radial artery versus saphenous vein as the second conduit for coronary artery bypass surgery | Long-term mortality |
| 12 | Treatment options for ischemic mitral regurgitation: A meta-analysis | MitraCLIP + OMT vs OMT | Mortality |
| | Treatment options for ischemic mitral regurgitation: A meta-analysis | CABG + RMA vs CABG | Mortality |
| 13 | Treatment options for ischemic mitral regurgitation: A meta-analysis | RMA + subvalvular vs subvalvular | Mortality |
| | Antithrombotic Strategies After Bioprosthetic Aortic Valve Replacement: A Systematic Review | Dual antiplatelet therapy vs AVA after TAVR | 30-day mortality |
| | Antithrombotic Strategies After | Warfarin vs ASA | 90-day |

| | | | |
|----|---|---|--|
| | Bioprosthetic Aortic Valve Replacement: A Systematic Review | after surgical bioprosthetic AVR | mortality |
| 14 | Biatrial Versus Bicaval Orthotopic Heart Transplantation: A Systematic Review and Meta-Analysis | Bicaval vs biatrial | Early mortality |
| 15 | Supplemental Cardioplegia During Donor Heart Implantation: A Systematic Review and Meta-Analysis | Supplemental cardioplegia vs not | Perioperative mortality |
| 16 | Skeletonized internal thoracic artery harvest improves prognosis in high-risk population after coronary artery bypass surgery for good quality grafts | Skeletonized vs pedicled harvesting | Mortality |
| 17 | Single- versus multidose cardioplegia in adult cardiac surgery patients: A meta- analysis | Single vs multi- dose cardioplegia | Operative mortality |
| 18 | Automated Fastener vs Hand-tied Knots in Heart Valve Surgery: A Systematic Review and Meta-analysis | COR-KNOT vs Hand-tie | Mortality rates |
| 19 | Drug-eluting stents versus coronary artery bypass grafting for the treatment of coronary artery disease: a meta-analysis of randomized and nonrandomized studies | Drug eluting stent vs CABG | All-cause mortality at 12 months |
| | A meta-analysis of transcatheter aortic | TAVI vs surgical | Early all-cause |

| | | | |
|----|---|---|---------------------------|
| | valve implantation versus surgical aortic valve replacement | AVR for aortic stenosis | mortality |
| 20 | Repair of Less Than Severe Tricuspid Regurgitation During Left-Sided Valve Surgery: A Meta-Analysis | Concomitant tricuspid valve repair vs non-repair during left sided valve surgery for mild or moderate tricuspid valve regurgitation | Early mortality |
| 21 | Repair of Less Than Severe Tricuspid Regurgitation During Left-Sided Valve Surgery: A Meta-Analysis | Concomitant tricuspid valve repair vs non-repair during left sided valve surgery for mild or moderate tricuspid valve regurgitation | Late, all-cause mortality |
| | Difference in spontaneous myocardial infarction and mortality in percutaneous versus surgical revascularization trials: A systematic review and meta-analysis | PCI vs CABG | All-cause mortality |

| | | | |
|----|--|---|--|
| 22 | Rapid deployment or sutureless versus conventional bioprosthetic aortic valve replacement: A meta-analysis | RDAVR vs CAVR | Early death (this study also reports HR of all-cause mortality during follow-up period) |
| 23 | Early surgery versus conservative management of asymptomatic severe aortic stenosis: A meta-analysis | Early surgery vs conservative management for asymptomatic, severe aortic stenosis | All-cause mortality |
| 24 | Rigid Plate Fixation Versus Wire Cerclage for Sternotomy After Cardiac Surgery: A Meta-Analysis | Rigid plate fixation vs wire cerclage in cardiac surgery pts | 30-day mortality |
| 25 | A meta-analysis of minimally invasive versus conventional sternotomy for aortic valve replacement | MIAVR vs CAVR | Perioperative mortality in patients undergoing minimally invasive aortic |

| | | | |
|----|--|---|---------------------------------|
| | | | valve replacement |
| | Meta-Analysis of Trials on Prophylactic Use of Levosimendan in Patients Undergoing Cardiac Surgery | Levosimendan vs No additional medication administration | Mortality at 30 days |
| 26 | Adjunct retrograde cerebral perfusion provides superior outcomes compared with hypothermic circulatory arrest alone: A meta-analysis | Aortic arch surgery with HCA alone versus HCA with RCP | 30 day or in-hospital mortality |
| 27 | Off-pump versus on-pump coronary revascularization: meta-analysis of mid- and long-term outcomes | Off-pump versus on-pump coronary revascularization | Long-term mortality (N=3 RCTs) |
| 28 | Complete transcatheter versus surgical approach to aortic stenosis with coronary artery disease: A systematic review and meta-analysis | TAVR PCI vs SAVR CABG | 30-day mortality |
| | Complete transcatheter versus surgical approach to aortic stenosis with coronary artery disease: A systematic review and meta-analysis | TAVR PCI vs SAVR CABG | All-cause mortality follow-up |
| 29 | A systematic review and meta-analysis of in situ versus composite bilateral internal thoracic artery grafting | In situ vs composite BITA | All-cause mortality |

| | | | |
|----|--|---|------------------------------|
| 30 | Systematic review and meta-analysis of chordal replacement versus leaflet resection for posterior mitral leaflet prolapse | Chordal replacement vs leaflet resection techniques | Perioperative mortality |
| 31 | Mitral valve surgery and coronary artery bypass grafting for moderate-to-severe ischemic mitral regurgitation: Meta-analysis of clinical and echocardiographic outcomes | CABG + mitral valve surgery vs CABG alone | Perioperative mortality |
| | Systematic review and meta-analysis of randomized controlled trials assessing safety and efficacy of posterior pericardial drainage in patients undergoing heart surgery | Intervention vs No additional posterior pericardial drainage tube placement | Death after heart surgery |
| | Meta-analysis of valve hemodynamics and left ventricular mass regression for stentless versus stented aortic valves | Stentless vs stented aortic valves | Mortality at 1 year followup |
| | Effects of dipyridamole in combination with anticoagulant therapy on survival and thromboembolic events in patients with prosthetic heart valves. A meta-analysis of the randomized trials | Warfarin + dipyrimadole vs warfarin alone | Total mortality |
| | Minimally invasive direct coronary artery | MIDCAB vs PCI | Early mortality |

| | | | |
|----|---|---|---------------------------------|
| | bypass graft surgery or percutaneous coronary intervention for proximal left anterior descending artery stenosis: a meta-analysis | for proximal left anterior descending artery stenosis | |
| | Minimally invasive direct coronary artery bypass graft surgery or percutaneous coronary intervention for proximal left anterior descending artery stenosis: a meta-analysis | MIDCAB vs PCI for proximal left anterior descending artery stenosis | Late mortality |
| 34 | Stenting versus coronary artery bypass grafting for unprotected left main coronary artery disease: a meta-analysis of comparative studies | PCI-S vs CABG for unprotected left main coronary artery disease | Death |
| | Pexelizumab in ischemic heart disease: a systematic review and meta-analysis on 15,196 patients | Pexelizumab vs Placebo for CABG | Death |
| | Preoperative physical therapy for elective cardiac surgery patients | Preoperative physical therapy versus no preoperative physical therapy | Post-operative death all causes |
| | Remote ischaemic preconditioning for coronary artery bypass grafting (with or without valve surgery) | RIPC versus no RIPC (sham interventino) in | All-cause mortality at 30 days |

| | | | |
|--|---|--|---------------------------------------|
| | | people undergoing CABG (with or without valve surgery) | |
| | Prophylactic corticosteroids for cardiopulmonary bypass in adults | Corticosteroids vs control | Mortality |
| | Epidural analgesia for adults undergoing cardiac surgery with or without cardiopulmonary bypass | Epidural analgesia vs systemic analgesia | 30 day mortality |
| | Fast-track cardiac care for adult cardiac surgical patients | Low-dose opioids vs high-dose opioids in opioid based cardiac anesthesia | Death at any time after surgery |
| | Fast-track cardiac care for adult cardiac surgical patients | Early extubation after surgery vs usual care | Death at any time after surgery |
| | HMG CoA reductase inhibitors (statins) for preventing acute kidney injury after surgical procedures requiring cardiac bypass | Statin vs placebo/no treatment after cardiac bypass procedures | Mortality |
| | Transcatheter aortic valve implantation | Transcatheter | Short-term all |

| | | | |
|--|--|--|-------------------------------|
| | versus surgical aortic valve replacement for severe aortic stenosis in people with low surgical risk | aortic valve implantation versus surgical aortic valve replacement for severe aortic stenosis in people with low surgical risk | cause mortality |
| | Transcatheter aortic valve implantation versus surgical aortic valve replacement for severe aortic stenosis in people with low surgical risk | Transcatheter aortic valve implantation versus surgical aortic valve replacement for severe aortic stenosis in people with low surgical risk | Long-term all cause mortality |
| | Coronary Revascularization for Patients with Diabetes Mellitus: A Contemporary Systematic Review and Meta-Analysis | PCI vs CABG | 30 day mortality |
| | Coronary Revascularization for Patients with Diabetes Mellitus: A Contemporary | PCI vs CABG | 1 year mortality |

| | | | |
|--|--|--|------------------------------------|
| | Systematic Review and Meta-Analysis | | |
| | Coronary Revascularization for Patients with Diabetes Mellitus: A Contemporary Systematic Review and Meta-Analysis | PCI vs CABG | 5 year mortality |
| | Off-pump versus on-pump coronary artery bypass grafting for ischaemic heart disease | Off-pump CABG vs on-pump CABG | All cause mortality, any follow up |
| | Transmyocardial laser revascularization versus medical therapy for refractory angina | Transmyocardial laser revascularization versus medical treatment for refractory angina | Overall mortality |
| | Psychological interventions for coronary heart disease | Psychological intervention vs usual care | Total mortality |
| | Fresh frozen plasma for cardiovascular surgery | Fresh frozen plasma vs no plasma | Short term mortality (30 days) |
| | Long-term and Temporal Outcomes of Transcatheter Versus Surgical Aortic-valve Replacement in Severe Aortic Stenosis: A Meta-analysis | TAVR vs SAVR | 5 year all cause mortality |
| | Perioperative beta-blockers for preventing | Perioperative | Early all-cause |

| | | | |
|--|---|--|-------------------------------|
| | surgery-related mortality and morbidity in adults undergoing cardiac surgery | beta-blockers vs control | mortality |
| | Perioperative beta-blockers for preventing surgery-related mortality and morbidity in adults undergoing cardiac surgery | Perioperative beta-blockers vs control | Long term all-cause mortality |

504 **Supplementary Results: Sensitivity analyses**

505 In the 34 meta-analyses containing both RCTs and observational studies, the earliest
506 published study was an RCT in 8 (24%). In the 26 meta-analyses where observational studies were
507 published first, the median years until an RCT was published was 3 (IQR 1 to 6). Across these meta-
508 analyses, comparing only the observational studies published before the first RCT appeared, the
509 summary of the relative summary effects was 0.93 (95% CI, 0.67-1.28), $p=0.64$, $I^2=4%$ (0 to 34%).

510 Of 6 meta-analyses where the observational effect was statistically significant ($p<0.005$), only
511 1 also showed the RCT effect to be statistically significant. Of 2 meta-analyses where the RCT effect
512 was statistically significant ($p<0.005$), 1 also showed a significant observational effect.

513 Of the 824 studies, 416 (50%) were type A comparisons (comparing a newer less invasive
514 method versus an older more invasive method looking at non-inferiority). In sensitivity analysis
515 flipping the coining of type A comparisons, the median observational summary effect across the 34
516 comparisons with both RCTs and observational studies was 0.89 (IQR, 0.1 to 1.16) versus an RCT
517 effect of 0.86 (IQR, 0.59 to 1.03) ($p=0.47$). The median relative summary effect was 1.16 (IQR, 0.80
518 to 1.63) and the summary of the relative summary effects 1.03 (0.82 to 1.31, $p=0.79$).

519