Invited editorial for JOCS-2021-RA-875

Title: To understand a meta-analysis, best read the fine print.

Author: Kevin L. Greason MD

Institutional affiliation: Department of Cardiovascular Surgery, Mayo Clinic, Rochester, Minnesota, USA

Short running title: Meta-analysis commentary

Address for correspondence: Department of Cardiovascular Surgery, Mayo Clinic, 200 First Street, Southwest, Rochester, Minnesota, USA; Tel (507) 255-7067; Telefax (507) 255-7378

Sources of funding: None

Abstract:

The results of a meta-analysis are more than just the reported odds ratio, 95% confidence interval, and P value. Of equal importance is the fine print of the study which should include assessment of risk of bias, certainty in evidence, and heterogeneity in the individual point estimates and confidence intervals. These areas all have influence on the quality of the data in the analysis. Reading and understanding the fine print is important.

Readers of the Journal will find interest in this systematic review and meta-analysis on the outcomes of sutureless aortic valve replacement (SuAVR) for aortic valve stenosis in comparison to standard surgical valve replacement (SAVR) or transcatheter valve replacement (TAVR) (1). The article provides an in-depth examination of SuAVR in comparison to the other techniques; but more importantly to me, it is the fine print of this meta-analysis that demonstrates the importance of factors other than summation odds ratios (OR), 95% confidence intervals (95% CI), and P values. The purpose of this commentary is to define the importance and context of the work and to go further in speculating where the field might be heading in the future. In full disclosure, I am a surgeon. I think like a surgeon, not a statistician.

The authors did a methodical review of the available literature and data comparing SuAVR to SAVR and SuAVR to TAVR (1). They reported that 51 observational studies were reviewed which identified 8,688 patients with aortic valve stenosis: SuAVR versus SAVR group (n=7,378) and SuAVR versus TAVR group (n=1,310). The authors studied the important endpoints of short-term (30-day) and mid-term mortality (up to 2 years), stroke, permanent pacemaker implantation, and paravalvular regurgitation. Additional endpoints included acute kidney injury, postoperative aortic valve mean gradient, and where relevant, cardiopulmonary bypass and aortic cross-clamp time.

In the SuAVR versus SAVR group analysis, the authors reported SuAVR was associated with similar mortality risk at 30 days (OR 1.02; 95% CI 0.72, 1.43) and 2 years (OR 0.99; 95% CI 0.43, 2.30) (1). Risk of stroke was also similar (SuAVR group OR 1.20; 95% CI 0.77, 1.87). The SuAVR group had similar risk of mild paravalvular regurgitation (OR 2.70; 95% CI 0.91, 8.01) and moderate paravalvular regurgitation (OR 1.71; 95% CI 0.42, 6.95), but increased risk of permanent pacemaker implantation (OR 2.45; 95% CI 1.93, 3.10; P<0.001). There was similar risk of acute kidney injury and postoperative aortic valve mean gradient. And as expected, SuAVR was associated with shorter aortic cross-clamp time (MD -23.02 minutes; 95% CI -33.17, -12.87, P<0.001) and cardiopulmonary bypass time (MD -19.8 minute; 95% CI -27.5, -12.2, P<0.001).

In the SuAVR versus TAVR group analysis, the authors reported SuAVR was associated with lower mortality risk at 30 days (OR 0.36; 95% CI 0.17, 0.73; P=0.005) and 2 years (OR 0.39; 95% CI 0.17, 0.88; P=0.02), but there was no difference in effects when stratified a priori by intermediate or high-risk surgical risk (1). Stroke risk was similar in the two groups (SuAVR OR 0.77; 95% CI 0.37, 1.61). The SuAVR group had reduced risk of mild paravalvular regurgitation (OR 0.09; 95% CI 0.03, 0.26; P<0.001), moderate paravalvular regurgitation (OR 0.11; 95% CI 0.02-0.61; P=0.01), and acute kidney injury (OR 0.50; 95% CI 0.27, 0.91; P=0.02), but similar risk of permanent pacemaker implantation and similar postoperative aortic valve mean gradient.

What can the reader discern from the reported results of this analysis? Based on the odds ratios, 95% confidence intervals, and P values, conclusions can be made that SAVR is better than SuAVR (i.e., less permanent pacemaker implantation) and that SuAVR is better than TAVR (i.e., reduced mortality, paravalvular regurgitation, and acute kidney injury) (1). Are those fair conclusions? The answer to that question can be found in in-depth examination of the methods, results, and appendix sections of the meta-analysis. In my opinion, this is the most important and informative part of the study (i.e., the so called “fine print”). The data and analyzes must be reviewed in the 58 page appendix.

I believe the authors did due diligence in the makings of this meta-analysis. They included a description of the search criteria that went beyond the basics of sutureless, surgical, and transcatheter aortic valve replacement. For example, the search included 37 other import specific terms such as minimally invasive, rapid deployment, replacement, implantation, procedure, and specific valve types (1). The authors performed the study in accordance to PRISMA guidelines (2). The PRISMA flow diagram (Figure 1, Appendix) reports the flow of the review with an initial screening of 1299 references with a final 59 studies in the SuAVR versus SAVR group (including 1 randomized trial) and 21 studies in the SuAVR versus TAVR group. But, these numbers don’t match what is reported in the results section of the manuscript (i.e., “… 1489 references screened…” and “…no randomized trial but 51 observational studies…”). Hmm?

The analysis gets more interesting when the reader digs into the meta-analysis specifics of the assessment of risk of bias, certainty in evidence, and heterogeneity in the individual point estimates and confidence intervals (1). These are common terms for the statistician, but what about for the surgeon? When it comes to meta-analysis, they should be as it is imperative to understand these critically important factors. And, for me, this relates to the question often asked by the journal editor of the reviewers, “do you trust the data?” How does a reviewer, or reader, make that determination about the data? Let us explore the tools used in this meta-analysis by Makhdoum and colleagues who did a fine job in that regard.

Risk of bias assessment was done using the CLARITY tool (i.e., Clinical Advances through Research and Information Technology) and funnel plots for symmetry (1). The author’s reference was incomplete about the CLARITY tool (3), but I found another readable and understandable on-line reference that proved useful to me (4). Tables 1A-C , 2A, and 2B provide crucial insight about the bias in the data (i.e., the studies reviewed). My review of the Tables shows all of the studies were rated by the authors as unclear (majority) or high risk of bias. Of note, there were 57 studies reviewed in the SuAVR versus SAVR group and 22 in the SuAVR versus TAVR group (Hmm?). The presence of bias is an important distinction upon which the foundation of the present analysis rests. Bias in clinical research can influence all aspects of study including trial planning, implementation, data analysis, and publication (5).

Certainty of evidence was evaluated using the GRADE framework (Grading of Recommendations, Assessment, Development, and Evaluations) (6). First and foremost, GRADE is not an objective assessment (7). It can’t be implemented mechanically, and I believe is therefore relatively subjective in nature. GRADE certainty ratings go from very low (i.e., the true effect is probably markedly different from the estimated effect) to high (i.e., the authors have a lot of confidence that the true effect is similar to the estimated effect) (7). Using the GRADE framework, the authors assigned the certainty rating of “very low” for all the reported clinical outcomes (i.e., mortality, stroke, etc; see Appendix 4). The authors dutifully reported their negative assessment of the certainty of evidence in the results section of the manuscript.

The authors assessed heterogeneity in the individual point estimates and confidence intervals in a forest plot, X2 test for homogeneity, and the I2 index (1). They reported that large variance in point estimates, minimal overlap in confidence intervals, a significant value in the X2 test, and higher I2 values all indicated heterogeneity. That sentence alone was quite helpful to my understanding of the heterogeneity of the studies. Appendix 3 reports the Forest plots which provide a visual reference as to the heterogeneity of the point estimates; for instance, Figure 6A of Appendix 3 shows wide heterogeneity in postoperative aortic mean gradient in the SuAVR versus SAVR group (X2=358.32 and I2=96%). How do you report the true result when the mean difference ranges from -5.4 to +3.30 mm Hg? The authors reported heterogeneity in the outcomes of mortality, stroke, permanent pacemaker implantation, paravalvular regurgitation, and postoperative aortic valve mean gradient, but not in aortic cross-clamp or cardiopulmonary bypass time (the last one not surprising to me).

The present study has several strengths and limitations. On the strength side, the authors used a medical librarian, formalized broad search strategy of the available literature, and independent reviewers. They used all the appropriate and best available meta-analysis tools (i.e., PRISMA guidelines, CLARITY tools, GRADE framework, and heterogeneity assessment). It is interesting to me that the strengths of the study are what highlighted the main weakness of the study, which is the “low quality of evidence” as rightly identified and concluded by the authors. This is not a newly identified issue with the meta-analysis method of study. Bailar referenced this back in 1997 (8). It can be problematic to present a single statistical stratagem estimate as sum of all that is known, especially when that strategem is based on data and analyzes with varying degrees of bias and heterogeneity.

The authors identify the usual limitations associated with the meta-analysis method of study: bias, lack of differential subgroups (e.g., device type), and low event rates. I would add to the list that the review ended in 2018 which neglects that last 3 years of data and analyzes. Is that an “updated” systemic review? A recent publication from the STS Adult Cardiac Surgery database may have addressed many of the present study limitations with reporting a large number of patients (over 17,000) with reduced bias by using the advanced statistical technique of propensity matching (9). And common to all of the reported studies, no long-term follow-up data are available (1, 9). This neglects the important negative influence of permanent pacemaker and paravalvular regurgitation on long-term quality of life (e.g., repeat hospitalization) and mortality (10, 11).

I applaud the authors on conducting a thorough meta-analysis study. They identified the low quality of the data and put that conclusion right out front. This sets the stage for future study, and I agree with Makhdoum and colleagues, comparative randomized data with long-term follow-up are required to clarify the role of SuAVR (1). Surgeons need to understand the limitations of the meta-analysis method of study. They should use that understanding to dig deep into the data and analysis and read the fine print. The truth does not always lie in the reported odds ratio, 95% confidence interval, or P value.

References:

1. Makhdoum A, Koziarz A, Kim K, Guta S, Alsagheir, A, Pandey A, et al. Outcomes of sutureless aortic valve replacement versus conventional aortic valve replacement and transcatheter aortic valve replacement, updated systematic review and meta-analysis. J Card Surg; in press.
2. PRISMA Moher D, Liberati A, Tetzlaff J, Altman DG, and GroupP. Perferred reporting items for systemic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009 Jul 21;6(7):e1000097. doi: 10.1371/journal.pmed.1000097. Epub 2009 Jul 21.PMID: 19621072.
3. CLARITY group. Tool to assess risk of bias in cohort studies .; 2017.
4. CLARITY <https://www.evidencepartners.com/wp-content/uploads/2021/03/Tool-to-Assess-Risk-of-Bias-in-Case-Control-Studies-DistillerSR.pdf>.
5. Pannucci CJ and Wilkins EG. Identifying and avoid bias in research. Plast Reconstr Surg 2010;126(2):619-625.
6. GRADE Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008 Apr 26;336(7650):924-6. doi: 10.1136/bmj.39489.470347.AD.
7. BMJ best practice. <https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/#:~:text=GRADE%20(Grading%20of%20Recommendations%2C%20Assessment,for%20making%20clinical%20practice%20recommendations>. Accessed September 26, 2021.
8. Bailer JC 3rd. The promise and problems of meta-analysis. N Engl J Med 1997;337:559-81.
9. Erfe JM, Malaisrie SC, Andrei AC,, Pham DT, Churyla A, Kruse J, et al. Outcomes of sutureless/rapid deployment valves comparted to traditional bioprosthetic aortic valves. Ann Thorac Surg. 2021;111:1884-92.
10. Greason KL, Lahr BD, Stulak JM, Cha YM, Rea RF, Schaff HV, et al. Long-term mortality effect of early pacemaker implantation after surgical aortic valve replacement. Ann Thorac Surg 2017;104:1259-66.
11. Pibarot P, Hahn RT, Weissman NJ, Arsenault M, Beaudoin J, Bernier M, et al. Association of paravalvular regurgitation with 1-year outcomes after transcatheter aortic valve replacement with the SAPIEN 3 valve. JAMA Cardiol 2017;2(11):1208-16. doi:10.1001/jamacardio.2017.3425