To the Editor:

We read with interest the recent study by Muneretto and colleagues¹ that compared outcomes of patients with severe aortic stenosis at intermediate surgical risk undergoing either sutureless aortic valve replacement (SuAVR) or transcatheter aortic valve replacement (TAVR). In their propensity score–matched analysis of 291 patient pairs treated between 2008 and 2015, Muneretto and colleagues¹ found that SuAVR reduced rates of all-cause mortality at both 30 days and 5 years relative to TAVR. Notwithstanding these results, however, we believe that their findings must be interpreted in the context of several important limitations, and as such, the conclusions of Muneretto and colleagues’ regarding the superiority of SuAVR relative to TAVR and the resulting need for a randomized trial comparing these 2 approaches directly should be tempered.

Any observational study must be carefully evaluated for confounding. This is particularly true for open-label surgical studies, in which surgeons' experience and acumen play a strong role in treatment selection for a variety of reasons that are often not measured or adequately captured, even with the use of advanced statistical methods such as propensity matching.² In the study of Muneretto and colleagues,¹ there are several indicators that propensity score methods may have inadequately adjusted for confounding. First, patients were treated between 2008 and 2015, a period during which TAVR use was relatively uncommon in the intermediate-risk population. As such, it is possible that many of the patients who underwent TAVR may not have been at intermediate risk despite their Society of Thoracic Surgeons Predicted Risk of Mortality score. Careful review of the study indicates that in addition to having a Society of Thoracic Surgeons Predicted Risk of Mortality score greater than 4%, patients were included if they were at least 75 years old. Indeed, more than a quarter of all patients in the propensity-matched sample had a Society of Thoracic Surgeons Predicted Risk of Mortality score less than 4%. Additional information regarding the documented reasons as to why patients underwent TAVR instead of conventional aortic valve replacement during this period would be helpful in this regard.

Beyond our understanding of practice patterns and the limitations of surgical risk scores, there are additional hints of residual confounding in the outcomes data. The most striking example of this issue can be seen in the increase in mortality from 30 days to 5 years, which should be a reliable indicator of the underlying complexity of the study population. For example, in the randomized PARTNER 2A trial, the increases in mortality between 30 days and 5 years were 38% with SAVR and 42% with TAVR (a relative risk of approximately 1.1).³ In contrast, in the study of Muneretto and colleagues,¹ these values were 15% and 23%, respectively, suggesting a relative risk of 1.5. The much higher relative mortality in the study of Muneretto and colleagues¹ suggests that despite propensity matching, the TAVR group was substantially “sicker” than the SuAVR group. Although it is possible that some of this late mortality difference reflects differences in valve performance, these findings are inconsistent with a meta-analysis of intermediate-risk patients that found no difference between TAVR and conventional aortic valve replacement for both early and midterm mortality.⁴ Finally, it is unlikely that these improved outcomes are a result of the use of sutureless valve technology, because of studies to date,⁵ including a recent publication in the Journal, that have not shown improved clinical outcomes with rapid-deployment valves relative conventional surgical aortic valve replacement.⁶

Comparative effectiveness studies that rely on observational study methods are challenging to conduct and must be interpreted in the context of various types of bias and confounding.⁷ Propensity score methods are often used to adjust for observed confounders, but they are by no means a panacea for this difficult problem. As such, we must consider other means to verify our assumptions or consider other methods to adjust for confounding. One way to evaluate the validity of risk adjustment includes the use of one or more “falsification end points” as a negative control.⁸ In essence, outcomes that are not related to treatment (eg, hip fracture) are compared between groups to see if their incidences differ during the study period. A difference in these falsification end points would suggest that the groups are not comparable despite the use of statistical adjustment. Other methods that may help better adjust for confounders include instrumental variables⁹ and overlap weighting.¹⁰

Ultimately, despite extensive adjustment and evidence of balance on measured covariates, propensity-matched studies are not randomized trials. As such, in observational studies (such as this one) we should avoid causal language in the interpretation of our findings. Although the confounders may be unmeasured, unknown, and hidden, we as investigators must find ways to account carefully for these factors and to interpret observational studies with appropriate caution.

The Editor welcomes submissions for possible publication in the Letters to the Editor section that consist of commentary on an article published in the Journal or other relevant issues. Authors should: Include no more than 500 words of text, three authors, and five references. Type with double-spacing. See http://jtcvs.ctsnetjournals.org/misc/fora.shtml for detailed submission instructions. Submit the letter electronically via jtcvs.editorialmanager.com. Letters commenting on an article published in the JTCVS will be considered if they are received within 6 weeks of the time the article was published. Authors of the article being commented on will be given an opportunity of offer a timely response (2 weeks) to the letter. Authors of letters will be notified that the letter has been received. Unpublished letters cannot be returned.
Dr Cohen receives research grant support and consulting income from Edwards LifeSciences, Medtronic, Abbott, and Boston Scientific. All other authors reported no conflicts of interest.

The Journal policy requires editors and reviewers to disclose conflicts of interest and to decline handling or reviewing manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

References

https://doi.org/10.1016/j.jtcvs.2020.07.095

Reply to the Editor:
I read with great interest the letter by Tam and colleagues1 regarding the application of the statistical methodology in propensity score matching studies such as in the recent study by Muneretto and colleagues.2 Tam and colleagues correctly raised an important statistical issue inherent in many retrospective clinical trials. They pointed out that in several clinical studies, the confounding variables could lead to improper results owing to inadequate selection of the confounding variables. In particular, Tam and colleagues deeply criticized some aspects of the study of Muneretto and colleagues, stating that the use of propensity score matching was inadequate, for several reasons. According to their criticism, the patient selection period (2008-2015) was improper, because transcatheter aortic valve replacement (TAVR) was not commonly used in patients at intermediate risk, all the patients included in the study were age >75 years, and more than one-quarter of the matched patients had a Society of Thoracic Surgeons Predicted Risk of Mortality <4%. In addition, the 5-year relative mortality rate was 15% in the sutureless aortic valve replacement (SuA VR) patients, compared with 23% in the TAVR patients (relative risk, 1.5). Tam and colleagues doubted that the TAVR patients were sicker than the SuA VR patients, however.

Randomized controlled trials (RCTs) represent the cornerstone when comparing 2 treatments; however, RCTs are somewhat affected by inclusion and exclusion criteria, which are sometimes very rigid, so that such patients recruited in RCTs do not always reflect daily clinical practice. In contrast, retrospective studies allow analysis of the clinical results coming from “real-life” data, but such studies are affected by an excessive number of confounding variables and bias. That said, Tam and colleagues’ statements and criticisms about Muneretto and colleagues’ study are very shareable, as are the suggestions regarding the application of even more effective statistical methods than propensity score matching when adjusting for confounding.

In a recent Commentary published in the Journal,1 I highlighted that the patients included in this study had a lower incidence of preoperative risk factors compared with the PARTNER-2 and SUR-TAVI trials, and I stated that it can be assumed the heart team followed a careful patient selection process to address the patients either to SuA VR or TAVR.

However, we cannot absolutely deny that the clinical results reported by Muneretto and colleagues demonstrate better efficacy and safety for conventional surgery...