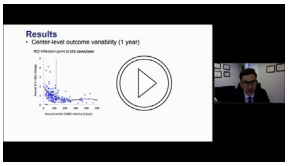
CONCLUSIONS

Center-level, risk-adjusted CABG mortality varies significantly from one year to the next, especially in centers performing less than approximately 110 cases per year. Aggregating the outcome data by volume, instead of time period, may improve the reliability of performance measures.

Webcast

You can watch a Webcast of this AATS meeting presentation by going to: https://aats.blob.core.windows.net/media/20AM/Presentations/Association%20Between%20CABG%20Center%20Volu.mp4.



Conflict of Interest Statement

The 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

1. Weintraub WS, Garratt KN. Challenges in risk adjustment for hospital and provider outcomes assessment. *Circulation*. 2017;135:317-9.
2. Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. *Ann Intern Med*. 1997;127:764-8.
3. Shahian D, Normand S-L. Statistical and clinical aspects of hospital outcomes profiling. *Stat Sci*. 2007;22:206-26.
4. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg*. 2012;255:703-7.
5. New York State Department of Health. Cardiovascular Disease Data and Statistics; 2019. Available at: https://www.health.ny.gov/statistics/diseases/cardiovascular/. Accessed July 1, 2019.
6. California Hospital Performance Ratings for Coronary Artery Bypass Graft (CABG) Surgery; 2019. Available at: https://oshpd.ca.gov/data-and-reports/healthcare-quality/cabg-reports/. Accessed July 1, 2019.
7. State of New Jersey Department of Health. Hospital Performance Report; 2019. Available at: https://web.doh.state.nj.us/apps2/hpr/usingcabg.aspx. Accessed July 1, 2019.
8. Glance LG, Dick AW, Mukamel DB, Li Y, Osler TM. How well do hospital mortality rates reported in the New York State CABG report card predict subsequent hospital performance? *Med Care*. 2010;48:466-71.
9. Krumholz HM, Wang Y, Mattera JA, Han LF, Ingber MJ, Roman S, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation*. 2006;113:1683-92.
10. Krumholz HM, Wang Y, Mattera JA, Han LF, Ingber MJ, Roman S, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. *Circulation*. 2006;113:1693-701.
11. Chikwe J, Toyoda N, Anyanwu AC, Itagaki S, Egorova NN, Boateng P, et al. Relation of mitral valve surgery volume to repair rate, durability, and survival. *J Am Coll Cardiol*. 2017;69:2397-406.
12. Mori M, Bin Mahmood SU, Zhuo H, Yousef S, Green J, Mangi AA, et al. Persistence of risk of death after hospital discharge to locations other than home after cardiac surgery. *J Thorac Cardiovasc Surg*. March 6, 2019 [Epub ahead of print].
13. Thomas N, Longford NT, Rolph JE. Empirical Bayes methods for estimating hospital-specific mortality rates. *Stat Med*. 1994;13:889-903.
14. Krell RW, Hozain A, Kao LS, Dimick JB. Reliability of risk-adjusted outcomes for profiling hospital surgical quality. *JAMA Surg*. 2014;149:467-74.
15. Silber JH, Rosenbaum PR, Brachet TJ, Ross RN, Bressler LJ, Even-Shoshan O, et al. The Hospital compare mortality model and the volume-outcome relationship. *Health Serv Res*. 2010;45(5 Pt 1):1148-67.
16. Badhwar V, Rankin JS, Thourani VH, D'Agostino RS, Habib RH, Shahian DM, et al. The Society of Thoracic Surgeons adult cardiac surgery database: 2018 update on research: outcomes analysis, quality improvement, and patient safety. *Ann Thorac Surg*. 2018;106:8-13.
17. Mori M, Shahian DM, Suter LG, Geirsson A, Lin Z, Krumholz HM. Relevance of cardiac surgery outcome reporting 3 years later in a New York and California statewide analysis. *JAMA Surg*. 2020;155:442-4.
18. Scientific Registry of Transplant Recipients. Available at: https://www.srtr.org/. Accessed April 5, 2020.
19. Hannan EL, Samadashvili Z, Cozzens K, Chikwe J, Adams DH, Sundt TM, et al. Out-of-hospital 30-day deaths following cardiac surgery are often under-reported. *Ann Thorac Surg*. 2020;110:183-8.
20. McPadden J, Durant TJ, Bunch DR, Coppi A, Price N, Rodgerson K, et al. Health care and precision medicine research: analysis of a scalable data science platform. *J Med Internet Res*. 2019;21:e13043.
21. Casalino LP, Gans D, Weber R, Cea M, Tuchovsky A, Bishop TF, et al. US physician practices spend more than $15.4 billion annually to report quality measures. *Health Aff (Millwood)*. 2016;35:401-6.
22. Omoigui NA, Miller DP, Brown KJ, Annan K, Cosgrove D, Lytle B, et al. Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes. *Circulation*. 1996;93:27-33.
23. Romano PS, Marcin JP, Dai JJ, Yang XD, Kravitz RL, Rocke DM, et al. Impact of public reporting of coronary artery bypass graft surgery performance data on market share, mortality, and patient selection. *Med Care*. 2011;49:1118-25.
24. CMS.gov. Hospital-level 30-day All-Cause Unplanned Readmission Following Coronary Artery Bypass Graft Surgery; 2014. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Coronary-Artery-Bypass-Graft-CABG-Readmission.zip. Accessed February 10, 2020.

Key Words: CABG, outcome reporting, center volume, volume-outcome relationship

Discussion

Presenter: Dr Arnar Geirsson

Dr Clifford W. Barlow *(Southampton, United Kingdom)*. My first question relates to combining data from California and New York. There are important differences between the 2 states including the number of centers, but also their methodology, analysis, and reporting of data. Could you comment on that? Second, do the results still apply when the states are analyzed separately?

Dr Arnar Geirsson *(New Haven, Conn)*. The 2 states have several key differences in factors pertinent to the studies. First, the number of centers is lower in New York—36 versus 119 in California. We didn't really focus on the difference in the article. The mean annual case volume is also different; there were 60 cases per year in California versus 220 in New York.

Perhaps this relates to the differences in geographical concentration of the 2 states. Both states use clinical registry, but the variables included differ slightly between the 2 states, as well as the inclusion criteria in the model. For example, New York excludes patients with preoperative shock, although it's a small portion of the patients. California also excludes what they define as salvage status undergoing active cardiopulmonary resuscitation.

In considering these differences, we elected to use O/E ratio rather than metrics of performance such as risk-standardized mortality rates, which is common, meaning in both states the O/E ratio of 1 indicates as expected performance giving patients risk, whereas risk-standardized mortality rate of, for example, 2%, may be above the state average in one or below in the other. In other words, the risk-standardized mortality rate would be at risk to the overall mortality rates within the state, and using this value would not allow us to compare the outcomes of centers between 2 different states. In short, we certainly acknowledged the differences and accounted for them, we believe, by using the ratio measures.

Dr Barlow. That's a very thorough answer. This is such an important and topical subject because public reporting of performance profiles has such major implications to hospitals and surgeons. One of the positive implications, for example, is that suboptimally performing centers are sometimes inspired and learn from the presentation of public data. On the other hand, of course, low-volume centers are protected when time-based reporting takes place by the presence of wide confidence intervals if the data is potentially wrong.

So, if you perceive using your method with volume reporting, will the protection provided by wide confidence intervals potentially disappear? Could hospitals and surgeons then be wrongly labeled as being poorly performing for many of the other reasons that outcome data are frequently wrong?

Dr Geirsson. I'd say that a relevant analogy may be that you would consider it unethical to conduct an underpowered experiment by sacrificing 50 animals 3 times and not achieving meaningful interrogation of the result. Whereas if you do a single 150-animal experiment that would be adequately powered to provide adequate power of analytic statistics and get meaningful results.

In that sense, we propose that the protection of the low-volume centers by truncating the data annually and not powering them sufficiently to achieve a statistically significant difference is perhaps depriving the centers from knowing whatever is actually going on that could be improved once the problem is identified analytically as a problem. At the same time, your point is well taken: Let's say it takes a long time—let's say it takes 5 years for a low-volume center to achieve sufficient denominator case volume. Let's say 100 cases. Then it's possible that poor outcomes concentrating the first 2 years may lead hospitals to make changes to improve outcomes yet still be penalized for the initial poor outcomes.

So we agree that a volume-based approach is limited in that sense. We would propose that reporting both time-based and volume-based analysis may paint the better pictures. You would think that resource-wise, it would not be prohibited in addition to what's already being done, using various outcomes measures reporting this.

Dr Barlow. That makes sense but leads to my last question, which relates to the practical implications of the reporting of outcomes by volume. Reporting by time periods is understandable to the public—a 1-year period for example. How would you practically propose reporting every 110 cases? Some centers could be reporting several times a year but others only every 3 years potentially. By the same token, what would the delay be in the data release upgrade? One of the reasons that the current delay of up to 1 year, and sometimes 2, takes place before the upgrade is because we're checking the data. How would you confirm that the data are valid?

Dr Geirsson. That's another important question—to address the limitation related to the issues we propose. The practical output of the data is important. Our stance is to advocate for reporting of both the current time-based aggregation in addition to volume-based aggregation of data.

As for the mechanics, we may propose, for example, using a rolling average of the last X number of cases as an option. The question is how often one should update these data. Ideally, it is obviously a real-time update. It's probably not practical at this stage because it only takes actually 3 years for these 2 state reports to report the outcomes. The 2016 report for both of those states actually published in 2019.

There are some other studies. The Scientific Registry of Transplant Recipients updates results twice a year, and the Society of Thoracic Surgeons provides quarterly reports. We think that reporting the last X number of cases at the frequencies around that range should be practical and reasonable. Once we start using artificial intelligence–based data mapping or natural language processing, there may be a time that real-time reporting becomes a possibility. I don't think we're there quite yet, though. Ensuring the fidelity of the data is certainly important, but the variations in delay (several months vs 3 years across various organizations) make us think it's really more of a logistic and resource issue than something inherent to the process itself. So we should continue to acknowledge it as an important problem that requires attention and improvement.

Dr Barlow. Thank you, Dr Geirsson, for your presentation and thorough answers.