

See Article page 1035.

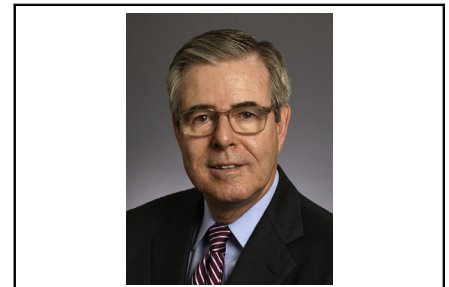


## Commentary: So many and yet so few—How many is not enough?

Gaetano Paone, MD, MHSA

Reliable and actionable evaluation of program performance is important for understanding quality of care, informing patient decision making, and providing a foundation for quality improvement. In this issue of the *Journal*, Mori and colleagues<sup>1</sup> use data from 155 centers in New York (n = 36) and California (n = 119) to analyze annual variability in center-level risk-adjusted observed-to-expected (O/E) mortality ratios for isolated coronary artery bypass grafting (CABG) surgery between 2012 and 2016. The median annual case volume per center was 89 (range, 55-160). The authors demonstrate significant year-to-year variation, identifying an inflection point at which variation was considered to have stabilized at a volume of 111 cases. The median year-to-year O/E ratio change was 0.83 (range, 0.26-1.59) for those centers with <111 cases and 0.49 (range, 0.22-0.87) for centers with  $\geq 111$  cases ( $P < .001$ ).

Impacting variation in the analysis of small-volume centers ultimately resides in somehow increasing the number of cases in the denominator. By extending the analysis across a 2-year period, the number of centers reaching the volume threshold increased from 93 to 118. However, compared over consecutive 2-year periods (2012-2013 vs 2014-2015), changes in the 2-year O/E ratio differences (median, 0.54; range, 0.23-1.02) were not significantly different than the 1-year comparisons (median, 0.70; range, 0.26-1.22) ( $P = .095$ ). Nonetheless, based on this information, the authors hypothesize that volume-based analysis may provide “a more reliable quality metric” of risk-adjusted mortality. Despite extending the study period to 2 years, approximately 25% of centers remained below the threshold of 111 cases. This then leads to the further recommendation



Gaetano Paone, MD, MHSA

### CENTRAL MESSAGE

Small-volume programs exhibit greater year to year variability in outcomes, which may limit the capacity to evaluate comparative performance and confidently assess the quality of care for coronary artery bypass grafting surgery.

to analyze what I assume is a time-independent “rolling average of the last  $\sim 110$  cases,” a number undoubtedly dependent on the specifics of the procedure and the associated dataset.

For “high”-volume programs, however defined, determining which cases to select from which time frame would be necessary. Recognizing this potential limitation, the authors alternatively suggest adding this approach “to supplement but not replace the conventional time-based aggregation of the data” only for the evaluation of small-volume programs. In 2017, 33 of 126 programs in California performed fewer than 50 procedures; 9 of these performed 25 or fewer.<sup>2</sup> Thus, for many programs, this would necessarily require the inclusion of cases performed over a period of 3 or more years. Given the already 3-year delay between the procedure, data acquisition and analysis, and finally public reporting, many programmatic changes, for better or worse, may have occurred over this extended time frame, limiting the contemporary relevance of the information.<sup>3</sup>

Although the value of this proposed approach is unclear, the authors have done a service by bringing the issue forward for additional discussion. As they suggest, advances in direct data acquisition with mapping and machine learning techniques stand to impact at least the time delay limitations. On the other hand, and although it is certainly possible to be a small-volume program of high quality,<sup>4-7</sup>

From the Division of Cardiothoracic Surgery, Structural Heart and Valve Center, Emory University School of Medicine, Atlanta, Ga.

Disclosure: The author 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.

Received for publication July 31, 2020; revisions received July 31, 2020; accepted for publication Aug 3, 2020; available ahead of print Aug 7, 2020.

Address for reprints: Gaetano Paone, MD, MHSA, Division of Cardiothoracic Surgery, Structural Heart and Valve Center, Emory University Midtown Hospital, 550 Peachtree St NE, Davis-Fischer Bldg, 4th floor, Atlanta, GA 30308 (E-mail: [gaetano.paone@emory.edu](mailto:gaetano.paone@emory.edu)).

J Thorac Cardiovasc Surg 2021;161:1042-3  
0022-5223/\$36.00

Copyright © 2020 Published by Elsevier Inc. on behalf of The American Association for Thoracic Surgery

<https://doi.org/10.1016/j.jtcvs.2020.08.005>

the number of low- and especially very-low-volume programs performing CABG seems unwarranted. A solution to this problem is somewhat less obvious.

## References

1. Mori M, Weininger GA, Shang M, Brooks C II, Mullan CW, Najem M, et al. Association between coronary artery bypass graft center volume and year-to-year outcome variability: New York and California statewide analysis. *J Thorac Cardiovasc Surg.* 2021;161:1035-41.e1.
2. California Office of Statewide Health Planning and Development. Hospital performance ratings for CABG surgery. Available at: <https://og-production-open-data-chelseama-892364687672.s3.amazonaws.com/resources/51e78126-5217-49ab-bf4a-9d96767ce073/2017-hospital-results.pdf?Signature=e4sPcVpDjZoUZAjKumMkCScLhmQ%3D&Expires=1598389960&AWSAccessKeyId=AKIAJJIENTAPKHZMIPXQ>. Accessed July 28, 2020.
3. 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.
4. Lancey RA. How valid is the quantity and quality relationship in CABG surgery? A review of the literature. *J Card Surg.* 2010;25:713-8.
5. Shahian DM, Normand S-LT. Low-volume coronary artery bypass surgery: measuring and optimizing performance. *J Thorac Cardiovasc Surg.* 2008;135:1202-9.
6. Kurlansky PA, Argenziano M, Dunton R, Lancey R, Nast E, Stewart A, et al. Quality, not volume determines outcome of coronary artery bypass surgery in a university-based community hospital network. *J Thorac Cardiovasc Surg.* 2012;143:287-93.
7. Shahian DM, O'Brien SM, Normans S-LT, Peterson ED, Edwards FH. Association of hospital coronary artery bypass volume with processes of care, mortality, morbidity, and the Society of Thoracic Surgeons composite quality score. *J Thorac Cardiovasc Surg.* 2010;139:273-82.

See Article page 1035.



## Commentary: Safety in numbers

David M. Shahian, MD

Using publicly reported data from New York and California, Mori and colleagues<sup>1</sup> found substantial year-to-year variation in publicly reported, hospital-level ratios of observed to expected coronary artery bypass grafting (CABG) mortality, which they interpret as measure instability related to small sample sizes. Based on inflection point analyses, they recommend adding mortality metrics derived from a hospital's most recent 111 CABG cases (ie, a standardized denominator sample size) as a complement to traditional annual or biennial reports.

### THE CURSE OF SMALL NUMBERS

Notwithstanding its methodological issues (eg, admixture of 2 states with markedly different cardiac surgery structures and oversight; attribution of all year-to-year

From the Division of Cardiac Surgery, Department of Surgery, and Center for Quality and Safety, Massachusetts General Hospital, Boston, Mass.

Disclosures: The author 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.

Received for publication July 17, 2020; revisions received July 17, 2020; accepted for publication July 17, 2020; available ahead of print July 22, 2020.

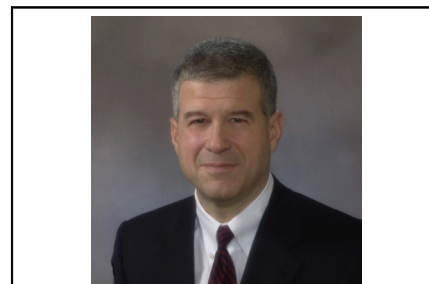
Address for reprints: David M. Shahian, MD, Division of Cardiac Surgery, Department of Surgery, and Center for Quality and Safety, Massachusetts General Hospital, 55 Fruit St, Boston, MA 02114 (E-mail: [dshahian@partners.org](mailto:dshahian@partners.org)).

*J Thorac Cardiovasc Surg* 2021;161:1043-5

0022-5223/\$36.00

Copyright © 2020 by The American Association for Thoracic Surgery

<https://doi.org/10.1016/j.jtcvs.2020.07.058>



David M. Shahian, MD, Massachusetts General Hospital and Harvard Medical School.

### CENTRAL MESSAGE

Quality measures based on small sample sizes have low statistical power and reliability. Mitigation may include multiyear samples, standardized denominator sample sizes, composite metrics, shrinkage estimators, or graphical plots.

variation in observed to expected mortality as random “noise”) and obvious implementation challenges (eg, confusion related to simultaneous time- and sample size–specified measures), this study illustrates a pervasive challenge in health care quality measurement—small sample size.<sup>2-7</sup> Annual hospital discharge volumes for individual conditions and procedures are often relatively low (eg, <100 discharges), which limits accurate performance measurement. For binary outcomes