The Society of Thoracic Surgeons Voluntary Public Reporting Initiative

The First 4 Years

David M. Shahian, MD, Frederick L. Grover, MD,† Richard L. Prager, MD,‡ Fred H. Edwards, MD,§ Giovanni Filardo, PhD,¶ Sean M. O'Brien, PhD,|| Xia He, MS,|| Anthony P. Furnary, MD,** J. Scott Rankin, MD,†† Vinay Badhwar, MD,‡‡ Joseph C. Cleveland, Jr, MD,† Frank L. Fazzalari, MD, MBA,‡ Mitchell J. Magee, MD,§§ Jane Han, MS,¶¶ and Jeffrey P. Jacobs, MD||||*

Objectives: To evaluate participant characteristics and outcomes during the first 4 years of the Society of Thoracic Surgeons (STS) public reporting program.

Background: This is the first detailed analysis of a national, voluntary, cardiac surgery public reporting program using STS clinical registry data and National Quality Forum–endorsed performance measures.

Methods: The distributions of risk-adjusted mortality rates, multidimensional composite performance scores, star ratings, and volumes for public reporting versus nonreporting sites were studied during 9 consecutive semiannual reporting periods (2010–2014).

Results: Among 8929 unique observations (~1000 STS participant centers, 9 reporting periods), 916 sites (10.3%) were classified low performing, 6801 (76.2%) were average, and 1212 (13.6%) were high performing. STS public reporting participation varied from 22.2% to 46.3% over the 9 reporting periods. Risk-adjusted, patient-level mortality rates for isolated coronary artery bypass grafting were consistently lower in public reporting versus nonreporting sites (P value range: <0.001–0.0077). Reporting centers had higher composite performance scores and star ratings (23.2% high performing and 4.5% low performing vs 7.6% high performing and 13.8% low performing for nonreporting sites). STS public reporting sites had higher mean annualized coronary artery bypass grafting volumes than nonreporting sites (169 vs 145, $P < 0.0001$); high-performing programs had higher mean coronary artery bypass grafting volumes (n = 241) than average (n = 139) or low-performing (n = 153) sites. Risk factor prevalence (except reoperation) and expected mortality rates were generally stable during the study period.

Conclusions: STS programs that voluntarily participate in public reporting have significantly higher volumes and performance. No evidence of risk aversion was found.

Keywords: outcomes, public reporting, quality measurement, ratings, report card

(*Ann Surg* 2015;262:526–535)

The Society of Thoracic Surgeons (STS) National Database was initiated in 1989 in response to the release by the federal government of inadequately adjusted mortality data, including coronary artery bypass grafting (CABG) surgery results.[1–3] Based on robust, audited, clinical registry data, STS developed and has periodically updated risk models for isolated CABG, valve, and combined procedures.[4–6] These were subsequently incorporated into National Quality Forum (NQF)–endorsed performance metrics, including multidimensional composite measures (risk-adjusted mortality and morbidity, plus 2 additional process measures for CABG).[7–10] Participants received domain-specific and overall composite scores (higher scores indicate better performance) and star-rating reports (1-star, 2-star, and 3-star, corresponding to low performance, average performance, and high performance, respectively), which they used to benchmark and improve their performance.

In 2010, STS leaders embarked upon an evolutionary new initiative in accountability and transparency—voluntary public reporting. Hospital or practice-level CABG results were first published in *Consumer Reports* in September 2010[11] and on the STS Web site in January 2011[12]; subsequent reports have included isolated aortic valve replacement (AVR) and AVR + CABG. This study describes the first 4 years of STS voluntary public reporting.

METHODS

Study Period

STS Adult Cardiac Surgery Database (ACSD) data harvests occur quarterly, and composite scores and public reporting are available with the first and third of these each year. Beginning in 2010, STS participants (usually hospital departments or cardiac surgery groups) were given the option of participating in semiannual public reporting with *Consumer Reports*,[11] the STS Web site,[12] or both.

In this report, data harvest or public reporting periods are designated by the calendar year (eg, 2010) followed by the numbers 1 or 3 to indicate which specific semiannual period. For this study, detailed analyses began with the first harvest of 2010 (2010–1), as these data (January 1 to December 31, 2009) were used for the first public reports. Analyses are provided for each subsequent STS semiannual public reporting period, ending with the first harvest of 2014 (2014–1). Some analyses in the present study include data that antedate STS

From the *Department of Surgery and Center for Quality and Safety, Massachusetts General Hospital, Boston, MA; †Division of Cardiothoracic Surgery, University of Colorado, Aurora, CO; ‡Department of Cardiac Surgery, University of Michigan Health System, Ann Arbor, MI; §Division of Cardiothoracic Surgery, University of Florida, Gainesville, FL; ¶Institute for Health Care Research and Improvement, Baylor Health Care System, Dallas, TX; ||Duke Clinical Research Institute, Durham, NC; **Starr-Wood Cardiac Group, Portland, OR; ††Vanderbilt University, Nashville, TN; ‡‡Department of Cardiothoracic Surgery, University of Pittsburgh Medical Center, Pittsburgh, PA; §§HCA North Texas Division, Dallas, TX; ¶¶Society of Thoracic Surgeons, Chicago, IL; and ||||Johns Hopkins All Children's Heart Institute, All Children's Hospital, St. Petersburg, FL.

Presented at the 135th Annual Meeting of the American Surgical Association, April 23–25, 2015, San Diego, CA.

Disclosure: Supported by Society of Thoracic Surgeons. The authors declare no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Reprints: David M. Shahian, MD, Department of Surgery and Center for Quality and Safety, Massachusetts General Hospital, 55 Fruit St., Bulfinch 284, Boston, MA 02114. E-mail: dshahian@partners.org.

ISSN: 0003-4932/15/26203-0526
DOI: 10.1097/SLA.0000000000001422

public reporting in order to demonstrate preexisting trends in risk factors and outcomes (2004–2010) or to study composite measure results since their introduction in 2007.

Study Population

The initial CABG public reporting population consisted of 9118 unique observations (~1000 STS participant sites, 9 semiannual data harvest and reporting periods). Twenty-two observations were excluded because they involved non-North American participants or those having no CABG cases; an additional 167 were excluded because participants were ineligible to receive a star rating (no data submitted to STS, <10 eligible cases submitted during the measurement period, or high missing data rate), although they may have consented to public reporting.

The remaining 8929 observations (97.9% of the initial 9118) constitute the final study population for CABG. Similar analyses of AVR and AVR + CABG public reporting have been performed, but experience with these is limited to recent harvests and is not the primary focus of this study.

Although nearly contemporaneous STS public reports were available on *Consumer Reports* and STS Web sites, slight differences existed between the number of consents for each, and several reporting periods had no *Consumer Reports* publication. To simplify data presentation, this report focuses primarily on public reporting of CABG results on the STS Web site.

Analyses

Public Reporting Rates and Participant Characteristics

STS composite CABG scores were first available in 2007, 3 years before public reporting began.[7,8] For each data harvest beginning that year, the number and proportion of STS participant programs that received 1, 2, or 3-star CABG composite scores was determined; for more recent years, similar data were compiled for AVR and AVR + CABG.

For each reporting period between 2010 and 2014, the number and proportion of all eligible STS programs participating in CABG public reporting were determined, and the volumes of reporting and nonreporting sites were compared. For each star-rating category, the proportion of programs that consented to report was calculated.

CABG Performance at Reporting and Nonreporting Programs

For each reporting period beginning in 2010, the mean, median, and interquartile range of risk-adjusted CABG mortality rates for reporting and nonreporting programs were compared using nonparametric Mann-Whitney *U* tests; analyses were performed at both center and patient levels.

Composite CABG scores were compared at reporting and nonreporting sites (nonparametric Mann-Whitney *U* tests) and data displayed using box and whisker plots. The distributions of star ratings in reporting and nonreporting groups were also compared for each reporting period (Pearson χ^2 test; null hypothesis that star-rating distributions were equal).

For all STS ACSD participants (reporting and nonreporting), the CABG volumes of 3-star versus 1- and 2-star sites were compared (nonparametric Mann-Whitney *U* tests) to assess potential volume-outcome associations.

Risk Aversion

To assess the potential impact of STS public reporting on risk aversion, quarterly expected mortality rates and the prevalence of selected high-risk predictors were calculated, beginning in 2004 (6 years before public reporting began) and extending through early 2014 (4 years after the start of public reporting). A second similar analysis, limited to the 4 years (9 reporting periods) of public reporting, was performed. This analysis also stratified results by reporting and nonreporting sites, and it used rolling 12-month data collection periods, the same approach employed in computing composite scores.

RESULTS

Public Reporting Rates and Participant Characteristics

Overall star-rating distributions among STS ACSD participants for the 3 index procedures are summarized in Supplemental Digital Content Table 1, available at http://links.lww.com/SLA/A835. Between 2007 and 2014, the percentages for each star-rating category have been relatively stable for CABG (about 10%–13% 1-star, 12%–15% 3-star), although a small but statistically significant ($P < 0.001$) decrement in 1-star programs has occurred over time.

The number and percentage of all STS participant programs that voluntarily consented to public reporting between 2010 and early 2014 are presented in Supplemental Digital Content Table 2, available at http://links.lww.com/SLA/A835. Participation rates increased steadily from 22.2% in 2010 to a peak of 46.3% in late 2013, then declined to 37.1% in early 2014 because of a transient contractual issue. The most recent STS public reporting rates (late 2014 and early 2015), not included in this study, were 41.7% and 44.2%.

Figure 1 displays CABG volumes for reporting and nonreporting programs, using box and whisker plots. The upper whisker extends from the hinge (top of box) to the highest value within 1.5 * interquartile range of the hinge. The lower whisker extends from the hinge (bottom of box) to the lowest value within 1.5 * interquartile range of the hinge. Outliers beyond the end of the whiskers are plotted as dots. Reporting sites had significantly higher volumes at every reporting period (*P* value range: <0.0001–0.0007), although some of the highest volume programs (black dots) were nonreporting.

Supplemental Digital Content Table 3, available at http://links.lww.com/SLA/A835, shows the percentage of all 1-star (range: 6.9%–25.8%, mean 16.8%), 2-star (range: 20.7%–44.5%, mean 36.3%), and 3-star (range: 46.4%–73.8%, mean 65.2%) programs that agreed to participate in public reporting for each period and overall.

CABG Performance at Reporting and Nonreporting Programs

Estimated at the participant level, the risk-adjusted isolated CABG mortality rates for reporting and nonreporting programs are shown in Supplemental Table 4, available at http://links.lww.com/SLA/A835. A 20% to 35% lower risk-adjusted mortality rate was observed at reporting sites compared with nonreporting sites, and the former had statistically significantly lower mortality in every reporting period (*P* value range: <0.0001–0.0503). A similar analysis at the individual patient level (ie, all patients operated upon at reporting vs nonreporting sites) is provided in Supplemental Table 5, available at http://links.lww.com/SLA/A835, with similar findings (*P* value range: <0.0001–0.0077). Data from Supplemental Table 5, available at http://links.lww.com/SLA/A835, are shown graphically in Figure 2.

Estimated CABG composite scores for reporting and nonreporting sites are shown in Figure 3. Scores were significantly higher ($P < 0.0001$) for reporting programs during each public reporting period; furthermore, the lowest scoring programs in each period were consistently in the nonreporting group.

Distributions of star ratings at STS reporting versus nonreporting sites are shown in Table 1, and these differed significantly

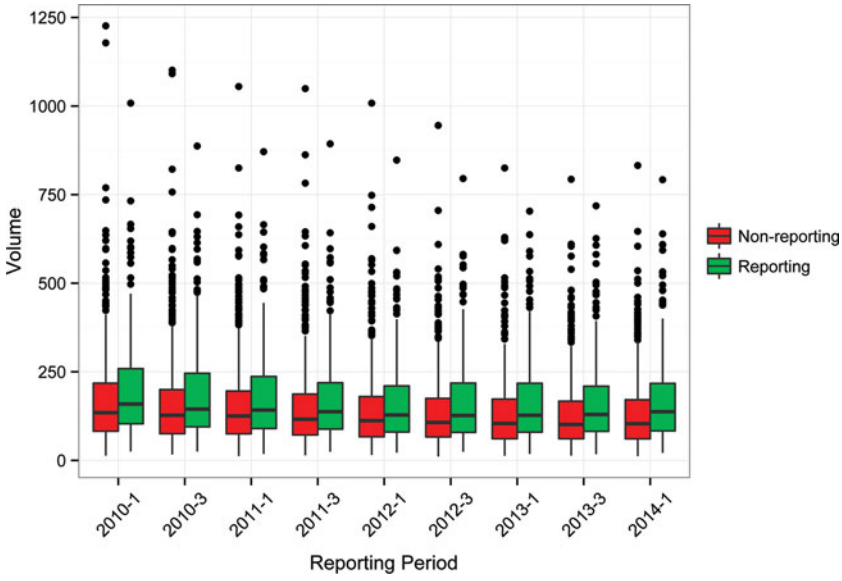

FIGURE 1. Volumes for reporting versus nonreporting sites. Reporting sites had higher volumes at every reporting period (*P* values <0.0001–0.0007).

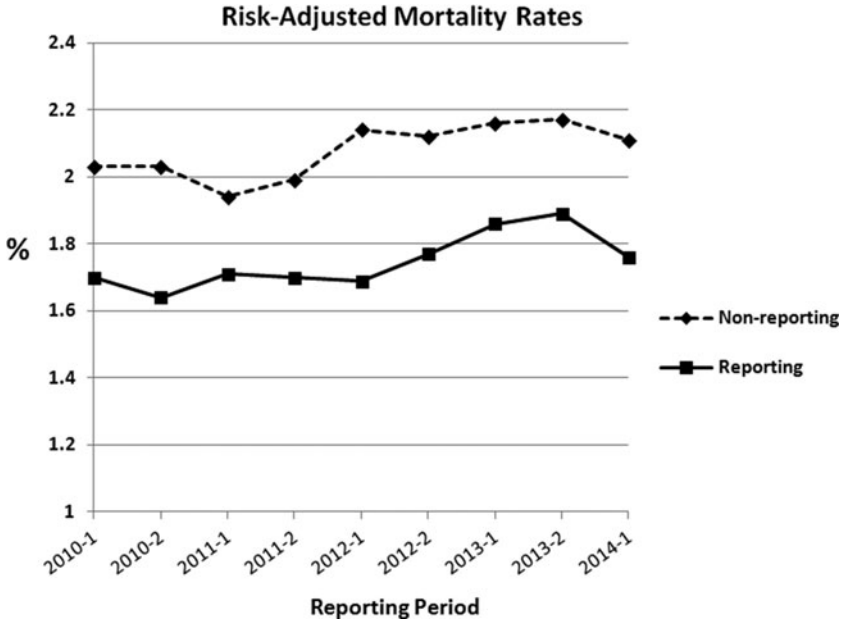

FIGURE 2. Patient-level, risk-adjusted mortality rates for isolated CABG: public reporting versus nonreporting sites (*P* values <0.0001–0.0077).

(all *P* values <0.0001). For every reporting period, the public reporting group had approximately 3 times the frequency (mean 23.2%) of above average (3-star) performing sites as the nonreporting group (mean 7.6%). Conversely, the proportion of 1-star (low performing) programs in the nonreporting group (mean 13.8%) was about 3 times that of the reporting group (4.5%). Within the public reporting group, the percentage of 3-star programs (mean 23.2%) was approximately 5 times the percentage of 1-star programs (mean 4.5%). This pattern was reversed in the nonreporting group, in which the percentage of 1-star programs (mean 13.8%) was twice that of 3-star programs (mean 7.6%).

Figure 4 shows that 3-star programs had consistently and significantly ($P < 0.0001$) higher CABG volume than the remaining low and average performing programs. Equivalent tabular data are provided in Supplemental Digital Content Table 6, available at http://links.lww.com/SLA/A835.

Risk Aversion

Quarterly observed and expected mortality rates, O/E (observed to expected) ratios, and the prevalence of 6 major preoperative risk factors are illustrated in Figure 5, beginning with data from 2004, 6 years before public reporting began. Expected mortality was flat across this decade; observed mortality and O/E ratios steadily declined ($P < 0.0001$ for trend) until 2010 and then stabilized, with consistently lower than predicted observed mortality (eg, O/E ratios <1). Although large sample sizes result in most of the trend tests being statistically significant for individual predictors, from a practical perspective emergent status, shock and age were relatively stable in frequency, whereas severe chronic lung disease and dialysis both increased markedly. Conversely, the incidence of previous cardiovascular surgery decreased by about 50% over the decade; this trend antedated public reporting and did not change when reporting was introduced.

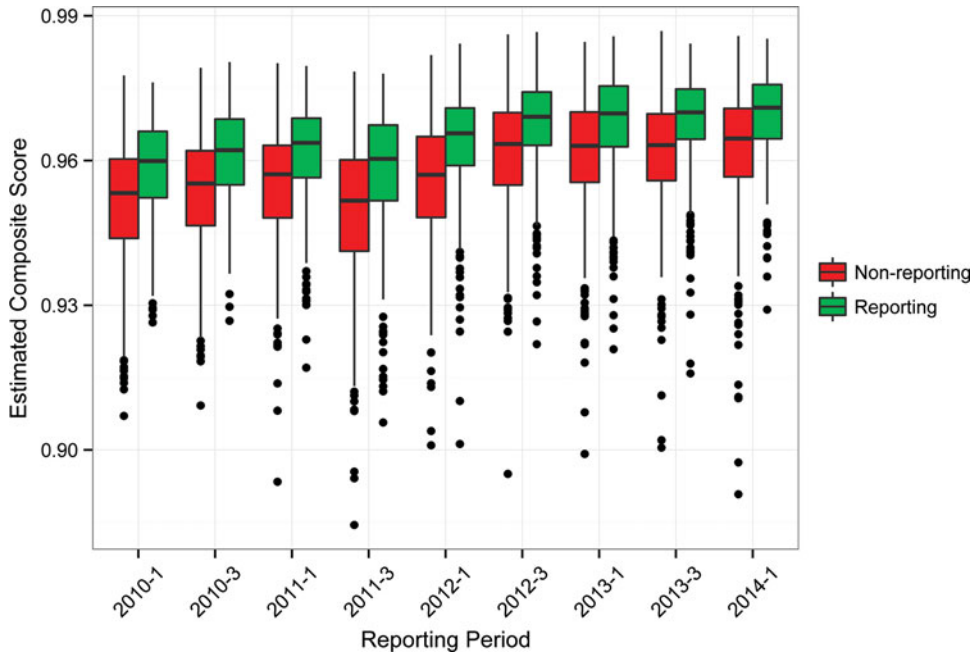

FIGURE 3. STS CABG composite scores for reporting and nonreporting sites, by reporting period. Scores were significantly higher ($P < 0.0001$) for reporting programs at each harvest, and lowest scoring programs were consistently in the nonreporting group.

TABLE 1. Society of Thoracic Surgeons Isolated Coronary Artery Bypass Grafting Star Ratings by Reporting Status

	Nonreporting			Reporting			
Year—Report Period	1-Star N (Row %)	2-Star N (Row %)	3-Star N (Row %)	1-Star N (Row %)	2-Star N (Row %)	3-Star N (Row %)	Overall
2010–1	100 (13.6)	575 (78.3)	59 (8.0)	9 (4.3)	150 (71.4)	51 (24.3)	944
2010–3	95 (15.0)	493 (77.9)	45 (7.1)	7 (2.1)	246 (72.8)	85 (25.1)	971
2011–1	92 (14.1)	508 (78.0)	51 (7.8)	17 (5.0)	239 (70.5)	83 (24.5)	990
2011–3	84 (13.8)	477 (78.5)	47 (7.7)	24 (6.2)	266 (68.9)	96 (24.9)	994
2012–1	77 (13.2)	461 (79.1)	45 (7.7)	19 (4.6)	307 (74.0)	89 (21.4)	998
2012–3	80 (14.0)	447 (78.0)	46 (8.0)	22 (5.2)	313 (73.8)	89 (21.0)	997
2013–1	74 (13.4)	440 (79.7)	38 (6.9)	21 (4.6)	327 (71.9)	107 (23.5)	1007
2013–3	72 (13.3)	434 (79.9)	37 (6.8)	25 (5.3)	348 (74.4)	95 (20.3)	1011
2014–1	88 (13.8)	498 (77.8)	54 (8.4)	10 (2.7)	272 (72.1)	95 (25.2)	1017
Overall	762 (13.8)	4333 (78.5)	422 (7.6)	154 (4.5)	2468 (72.3)	790 (23.2)	8929

Figure 6 contains the same variables as in Figure 5 but focuses on the 9 public reporting periods (2010–2014) and compares reporting and nonreporting programs. Overall results are similar to those in Figure 5, and the patterns for reporting and nonreporting groups mirror each other. Baseline prevalence of several risk factors was somewhat lower in the reporting group but overall expected mortality rates for the 2 groups remain close throughout the study period. The O/E ratio for the public reporting group is consistently lower than that for the nonreporting group (P value range: <0.0001–0.0077).

DISCUSSION

Building the Foundation for a Successful Public Reporting Program

The foundation of the STS voluntary public reporting program is the STS National Database. External audit of these clinical registry data, encompassing 10% of sites and approximately 100,000 data elements annually, demonstrates 96% to 97% accuracy, which should be reassuring to STS ACSD participants contemplating public reporting.

Using these granular clinical data, the STS Quality Measurement Task Force developed successively more sophisticated risk models[4–6] and performance metrics for most major cardiac procedures and outcomes. Recognizing that mortality alone was an inadequate measure of quality, particularly as short-term mortality rates declined, the Quality Measurement Task Force published the first (isolated CABG)[7,8] of its multidimensional composite metrics in 2007 and began providing these results to STS ACSD participants. Similar composite measures have now been developed for isolated AVR and AVR + CABG,[9,10] and models for mitral valve and mitral valve + CABG procedures will be available in 2015.

An important feature of all these risk models and performance metrics is their publication in the peer-reviewed literature, including highly detailed descriptions of methodology.[4–6] STS also routinely submits its performance measures to the rigorous external endorsement process of the NQF. As of 2015, STS has 33 NQF-endorsed

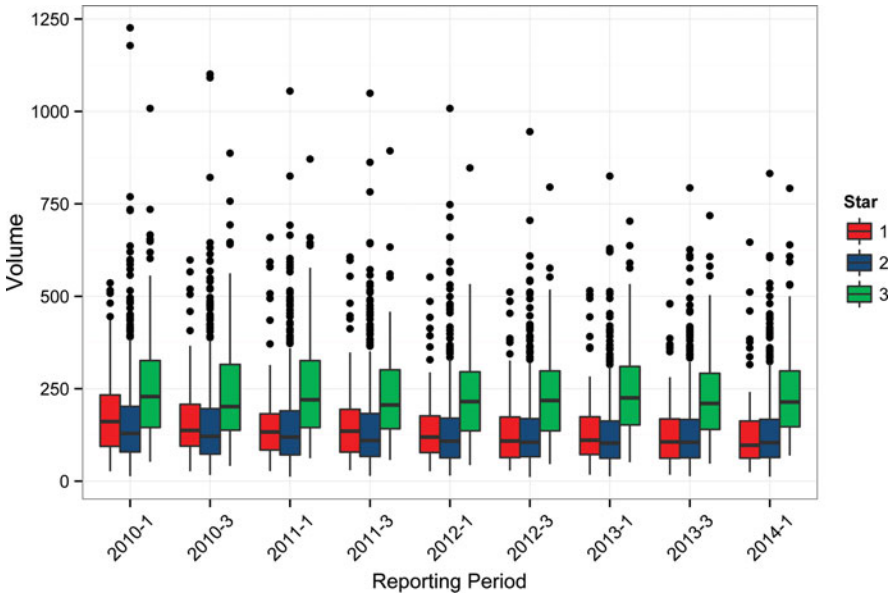

FIGURE 4. STS database participant (reporting and nonreporting combined) volumes for each star-rating category between 2010 and 2014. Three-star programs had statistically significantly ($P < 0.0001$) higher volume than the remaining low and average performing programs.

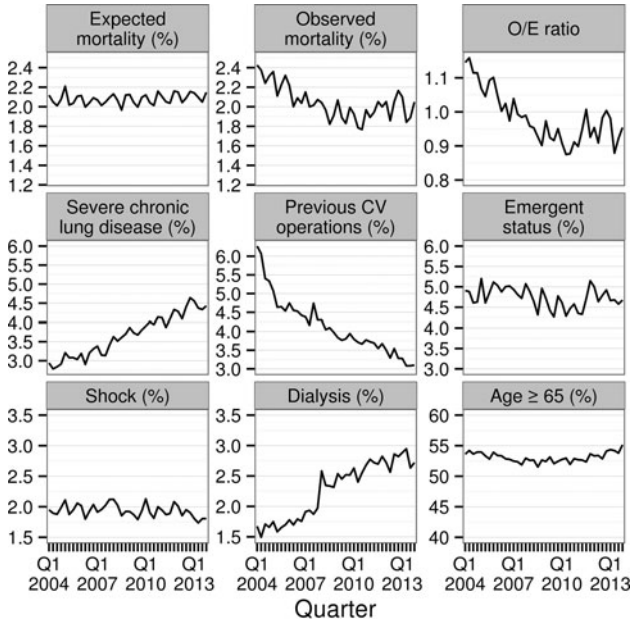

FIGURE 5. Observed and expected mortality rates, O/E ratios, and prevalence of 6 major risk factors, by quarter, 2004–2014.

measures, the largest number of any health care professional organization. All these efforts are aimed at engendering trust among STS members and external stakeholders in the accuracy of these performance measures.

Change Management—Leadership and Education

Notwithstanding years of confidential feedback using credible performance metrics, public reporting represented a major cultural change for the cardiothoracic surgery community. Robust STS leadership support and a vigorous educational campaign were critical elements in this change management process.

STS leaders consistently communicated to their members the principle that public reporting is a professional ethical responsibility.[13–15] Transparency promotes the ethical right of patient autonomy by providing patients and referring physicians with data that facilitate informed decision making. Public reporting also may advance the ethical principle of beneficence by improving surgical outcomes. Most evidence supporting this concept comes from states that have implemented mandatory reporting of cardiac surgical outcomes, such as New York.[16] Tempering this finding, however, are data showing that regions with robust but confidential data collection and feedback initiatives had similar improvements in outcomes to those in public reporting states,[17,18] and that public reporting added to existing confidential feedback reporting resulted in little incremental improvement in outcomes.[19] Overall, the evidence suggests that public reporting is one, but not the only, effective way to improve outcomes. Furthermore, when public reporting is implemented, there should be mechanisms to monitor and prevent risk aversion, the inappropriate denial of care to high-risk patients, which would be contrary to the ethical principle of justice.

STS also educated its members regarding the practical realities of health care reform. The federal government, other payers, and consumer groups are increasingly focused on quality measurement, transparency, and value-based purchasing, and commercially developed health care "report cards" have proliferated. Many of these report card methodologies are proprietary and nontransparent; their algorithms are not published in the peer-reviewed literature; they are not endorsed by organizations such as the NQF; and they frequently produce divergent results for the same institutions.[20–23] By voluntarily reporting results based on the best available data and analytics, STS protects its members against these problematic measurement and reporting practices.

Implementation

STS began confidential feedback to ACSD participants of performance data 3 years before it initiated voluntary public reporting. Although the rapidly increasing demand for transparency may no longer allow the luxury of such a long "burn-in," some period of confidential feedback and pilot testing is critical to ensure process validity and to gain the trust of participants before public reporting.

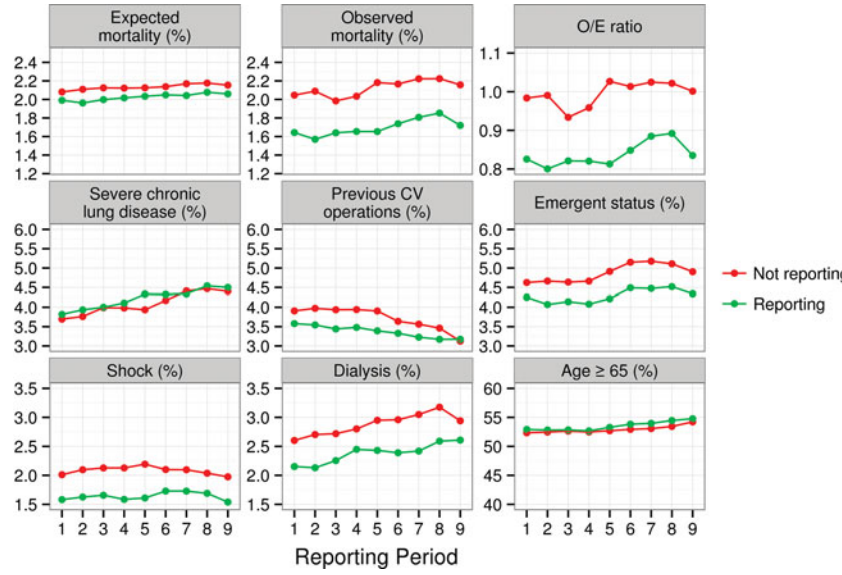

FIGURE 6. Observed and expected mortality rates, O/E ratios, and prevalence of 6 major risk factors for STS public reporting and nonreporting programs across 9 public reporting periods (2010–2014), rolling 12-month data. O/E ratios were consistently and significantly lower for public reporting programs (*P* value <0.0001–0.0077).

To ensure smooth implementation, STS and *Consumers Reports* devoted considerable resources to various operational aspects of public reporting. First, the consent to publish hospital or group-level results requires the written agreement of at least 1 surgeon member of the STS participant group in addition to the relevant hospital administrator. Consequently, STS has had no complaints from ACSD participants regarding unintended or unwanted reporting. Programs were provided with their own results before public reporting, but they generally made their reporting decisions before seeing these results. Consents for reporting do not have an expiration date; programs wishing to discontinue participation must notify STS and *Consumer Reports* in writing.

STS and *Consumer Reports* also recognized that through consolidation or ownership changes, the names of hospitals or cardiac surgical programs may change. Because of this concern, public reporting programs are asked to confirm the correct designation of their hospital and practice group, and links are provided to these Web sites if available.

Finally, STS and *Consumer Reports* formats are specifically designed to be consumer-friendly to avoid information overload and incorrect interpretation, problems that are commonly associated with health care report cards.[24–28] Extensive educational materials are available to facilitate consumer comprehension. Finally, STS provides varying levels of granularity, ranging from point estimates with credible intervals for statistically sophisticated users, to star ratings for typical consumers. When a composite measure encompasses multiple procedures or performance domains, STS provides "drill-down" functionality.

What Have We Learned From the Current Study?

Mortality Rates and Star Ratings Before and After Public Reporting

When observed mortality rates for CABG are examined over the past 15 years, well before the introduction of STS public reporting, a steadily declining trend is apparent.[29] In the current study, observed rates generally have been below expected since 2007, and they remained relatively unchanged after public reporting began. These findings are consistent with previous reports from Guru and colleagues.[19] When confidential feedback of CABG results to hospitals was initiated, mortality rates fell in Ontario compared with the rest of Canada. When public reporting was added, no further impact on these declining rates was observed. In a study by Peterson et al[17] of Medicare CABG mortality rates between 1989 and 1992, the lowest absolute rates and greatest reductions were achieved by 2 regions with completely different approaches. New York State was the leader in using public reporting to drive performance improvement, but The Northern New England Cardiovascular Disease Study Group achieved virtually identical results with confidential feedback to participants and a regional collaborative quality improvement initiative.

STS had been collecting, analyzing, and providing database participants with feedback for 2 decades before the institution of public reporting. Because of this long history of preexisting measurement and improvement activities, and resulting marked reductions in CABG mortality, minimal if any incremental effect of public reporting on subsequent outcomes was noted. By the time STS public reporting began, CABG mortality rates were so low that further dramatic improvement was unlikely to occur. In specialties without a long history of confidential feedback, the impact of public reporting would likely be much more substantial.

No Evidence of Risk Aversion

The most worrisome unintended consequence of public reporting is risk aversion. Fearing loss of reputation and referrals due to low performance scores, programs might decline to accept high-risk patients for surgery, despite the fact that it is precisely such patients who might benefit most.[30,31] Risk aversion was observed after the introduction of CABG public reporting in New York and Pennsylvania,[32–35] and numerous studies of high-risk percutaneous coronary intervention (PCI) patients strongly suggest that risk aversion has occurred.[36–40] In Massachusetts, declining access to high-risk PCI was reversed when compassionate use criteria were added to risk adjustment.[41–43] The best solution to risk aversion is to convince providers that they are adequately protected by existing risk models, as evidenced by recent studies using the STS and NCDR clinical registries.[44,45]

Data in the current study reveal no evidence of risk aversion. Throughout the 9 public reporting opportunities analyzed (Fig. 6), expected mortality remained stable or slightly increased for both reporting and nonreporting groups, as did the prevalence of several important individual risk factors such as dialysis-dependent renal

failure, severe chronic lung disease, age, and emergent status. In fact, expected mortality rates by calendar quarter beginning in 2004 (6 years before STS public reporting began) remained essentially unchanged through 2014 (Fig. 5); existing trends in risk factor frequency did not change appreciably when public reporting was introduced in 2010. Importantly, numerous external audits have revealed no evidence that programs have "gamed" the reporting system by inflating the prevalence of their patients' risk factors.

The only major risk factor that decreased significantly over time is previous cardiovascular surgery (eg, reoperation). This finding probably relates to improvements in care including greater use of internal mammary artery grafts and statins, which reduce the likelihood of conduit failure and progression of native coronary disease; and the increasing availability of interventional cardiology techniques to address bypass graft stenoses or new native coronary lesions, thus reducing the need for repeat surgery.

Public Reporting Programs Have Generally Higher Volume and Better Performance

Higher volume programs are generally more likely to publicly report, although some do not (Fig. 1). Similarly, high-performing programs in this study were 4 times more likely to report than lower performing programs (see Supplemental Digital Content Table 3, available at http://links.lww.com/SLA/A835). Three-star programs had 3 times the frequency in the reporting group compared with the nonreporting groups; low-performing programs had 3 times the frequency in nonreporting groups compared with the reporting group (Table 1). During every reporting period since 2010, the STS CABG composite scores for reporting programs have been statistically significantly higher than those at nonreporting sites (Fig. 3). Finally, at both the participant level and the patient level, risk-adjusted CABG mortality rates for reporting programs are consistently and statistically significantly lower than those for nonreporting programs (see Supplemental Digital Content Tables 4 and 5, available at http://links.lww.com/SLA/A835; Fig. 2).

These findings do not address causation. Programs may choose to report because they know their results are excellent. Alternatively, they may make a decision to report on the basis of principle, and this decision may be an incentive to institute quality improvement activities.

Participation Levels

Despite continuing education and encouragement from STS leadership, the rate of increase in public reporting participation has slowed after initially rapid growth driven by early adopters. Overall participation rates currently are somewhat less than half of all ACSD participants. In informal surveys conducted to better understand the reasons for nonreporting, no ethical or philosophical objections were voiced, but a variety of practical rationales (eg, risk adjustment, consumer misinterpretation, "waiting for more programs to join") were cited. Some centers perceive no advantage, and possibly downside risk, to reporting. This concern applies not only to low-performing programs but also to some regarded by the public as high performing, whose reputations might suffer if their superiority were not confirmed by objective data.

The Future

STS will continue to aggressively encourage voluntary public reporting. With continuing education and promotion by STS, and increasing demands for outcomes data by many stakeholders, participation levels will likely increase. Certain incentives would be helpful, such as inclusion of credit for public reporting in hospital ratings programs. However, only mandates by federal or state governments, regulators, or commercial payers will ensure universal participation.

In addition to its current public reporting of adult cardiac and congenital outcomes, STS will expand the range of composite performance metrics to include additional procedures (eg, mitral valve and mitral valve + CABG; lobectomy; esophagectomy) and outcomes (eg, patient satisfaction, patient reported outcomes, costs).

CONCLUSIONS

About 44% of STS ACSD participants are currently engaged in voluntary national public reporting. Participating programs are higher volume, have lower risk-adjusted mortality rates, and achieve higher STS CABG composite scores and star ratings. There is no evidence of risk aversion related to public reporting.

REFERENCES

1. Blumberg MS. Comments on HCFA hospital death rate statistical outliers. Health Care Financing Administration. *Health Serv Res.* 1987;21: 715–739.
2. Blumberg MS. Risk adjusting health care outcomes: a methodologic review. *Med Care Rev*. 1986;43:351–393.
3. Kouchoukos NT, Ebert PA, Grover FL, et al. Report of the Ad Hoc Committee on risk factors for coronary artery bypass surgery. *Ann Thorac Surg.* 1988;45:348–349.
4. Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1—coronary artery bypass grafting surgery. *Ann Thorac Surg.* 2009;88:S2–S22.
5. O'Brien SM, Shahian DM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. *Ann Thorac Surg.* 2009;88:S23–S42.
6. Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 3—valve plus coronary artery bypass grafting surgery. *Ann Thorac Surg.* 2009;88:S43–S62.
7. Shahian DM, Edwards FH, Ferraris VA, et al. Quality measurement in adult cardiac surgery: part 1—conceptual framework and measure selection. *Ann Thorac Surg.* 2007;83:S3–S12.
8. O'Brien SM, Shahian DM, Delong ER, et al. Quality measurement in adult cardiac surgery: part 2—statistical considerations in composite measure scoring and provider rating. *Ann Thorac Surg.* 2007;83:S13–S26.
9. Shahian DM, He X, Jacobs JP, et al. The Society of Thoracic Surgeons Isolated Aortic Valve Replacement (AVR) Composite Score: a report of the STS Quality Measurement Task Force. *Ann Thorac Surg.* 2012;94:2166–2171.
10. Shahian DM, He X, Jacobs JP, et al. The STS AVR + CABG composite score: a report of the STS quality measurement task force. *Ann Thorac Surg.* 2014;97:1604–1609.
11. Ferris TG, Torchiana DF. Public release of clinical outcomes data-online CABG report cards. *N Engl J Med.* 2010;363:1593–1595.
12. Available at: http://www.sts.org/quality-research-patient-safety/sts-public-reporting-online. Accessed March 3, 2015.
13. Shahian DM, Edwards FH, Jacobs JP, et al. Public reporting of cardiac surgery performance: part 1—history, rationale, consequences. *Ann Thorac Surg.* 2011;92:S2–S11.
14. Shahian DM, Edwards FH, Jacobs JP, et al. Public reporting of cardiac surgery performance: part 2—implementation. *Ann Thorac Surg.* 2011;92:S12–S23.
15. STS rationale for public reporting. Available at: http://www.sts.org/quality-research-patient-safety/sts-public-reporting-online/rationale-public-reporting. Accessed March 12, 2015.
16. Hannan EL, Cozzens K, King SB III, et al. The New York State cardiac registries: history, contributions, limitations, and lessons for future efforts to assess and publicly report healthcare outcomes. *J Am Coll Cardiol*. 2012;59:2309–2316.
17. Peterson ED, Delong ER, Jollis JG, et al. The effects of New York's bypass surgery provider profiling on access to care and patient outcomes in the elderly. *J Am Coll Cardiol*. 1998;32:993–999.
18. Chassin MR. Achieving and sustaining improved quality: lessons from New York State and cardiac surgery. *Health Aff (Millwood)*. 2002;21:40–51.
19. Guru V, Fremes SE, Naylor CD, et al. Public versus private institutional performance reporting: what is mandatory for quality improvement? *Am Heart J*. 2006;152:573–578.

20. Rothberg MB, Morsi E, Benjamin EM, et al. Choosing the best hospital: the limitations of public quality reporting. *Health Aff (Millwood)*. 2008;27:1680–1687.
21. HANYS' Report on Report Cards; Understanding publicly reported hospital quality measures. Available at: http://www.hanys.org/quality/data/report_cards/2013/docs/2013_hanys_report_card_book.pdf. Published 2013. Accessed November 14, 2013.
22. Leonardi MJ, McGory ML, Ko CY. Publicly available hospital comparison web sites: determination of useful, valid, and appropriate information for comparing surgical quality. *Arch Surg*. 2007;142:863–868.
23. Austin JM, Jha AK, Romano PS, et al. National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Aff (Millwood)*. 2015;34:423–430.
24. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. *Annu Rev Public Health*. 2003;24:413–433.
25. Hibbard JH, Peters E, Slovic P, et al. Making health care quality reports easier to use. *Jt Comm J Qual Improv*. 2001;27:591–604.
26. Hibbard JH, Greene J, Sofaer S, et al. An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care. *Health Aff (Millwood)*. 2012;31:560–568.
27. Peters E, Dieckmann N, Dixon A, et al. Less is more in presenting quality information to consumers. *Med Care Res Rev*. 2007;64:169–190.
28. Donelan K, Rogers RS, Eisenhauer A, et al. Consumer comprehension of surgeon performance data for coronary bypass procedures. *Ann Thorac Surg.* 2011;91:1400–1405.
29. ElBardissi AW, Aranki SF, Sheng S, et al. Trends in isolated coronary artery bypass grafting: an analysis of the Society of Thoracic Surgeons adult cardiac surgery database. *J Thorac Cardiovasc Surg*. 2012;143:273–281.
30. Jones RH. In search of the optimal surgical mortality. *Circulation*. 1989;79:I132–I136.
31. Lee TH, Torchiana DF, Lock JE. Is zero the ideal death rate? *N Engl J Med*. 2007;357:111–113.
32. Omoigui NA, Miller DP, Brown KJ, et al. Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes. *Circulation*. 1996;93:27–33.
33. Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care—a survey of cardiovascular specialists. *N Engl J Med*. 1996;335:251–256.
34. Dranove D, Kessler D, McClellan M, et al. Is more information better? The effects of "report cards" on health care providers. *J Political Economy.* 2003;111:555–588.
35. Burack JH, Impellizzeri P, Homel P, et al. Public reporting of surgical mortality: a survey of New York State cardiothoracic surgeons. *Ann Thorac Surg.* 1999;68:1195–1200.
36. Moscucci M, Eagle KA, Share D, et al. Public reporting and case selection for percutaneous coronary interventions: an analysis from two large multicenter percutaneous coronary intervention databases. *J Am Coll Cardiol*. 2005;45:1759–1765.
37. Apolito RA, Greenberg MA, Menegus MA, et al. Impact of the New York State Cardiac Surgery and Percutaneous Coronary Intervention Reporting System on the management of patients with acute myocardial infarction complicated by cardiogenic shock. *Am Heart J*. 2008;155:267–273.
38. Joynt KE, Blumenthal DM, Orav EJ, et al. Association of public reporting for percutaneous coronary intervention with utilization and outcomes among Medicare beneficiaries with acute myocardial infarction. *JAMA*. 2012;308:1460–1468.
39. Waldo SW, McCabe JM, O'Brien C, et al. Association between public reporting of outcomes with procedural management and mortality for patients with acute myocardial infarction. *J Am Coll Cardiol*. 2015;65:1119–1126.
40. Moscucci M. Public reporting of percutaneous coronary intervention outcomes: harm or benefit? *J Am Coll Cardiol*. 2015;65:1127–1129.
41. Resnic FS, Welt FG. The public health hazards of risk avoidance associated with public reporting of risk-adjusted outcomes in coronary intervention. *J Am Coll Cardiol*. 2009;53:825–830.
42. Resnic FS, Normand SL, Piemonte TC, et al. Improvement in mortality risk prediction after percutaneous coronary intervention through the addition of a "compassionate use" variable to the National Cardiovascular Data Registry CathPCI Dataset: a study from the Massachusetts Angioplasty Registry. *J Am Coll Cardiol*. 2011;57:904–911.
43. Peterson ED. The need for "compassionate provider profiling" refining risk assessment for percutaneous coronary intervention. *J Am Coll Cardiol*. 2011;57:912–913.
44. Englum BR, Saha-Chaudhuri P, Shahian DM, et al. The impact of high-risk cases on hospitals' risk-adjusted coronary artery bypass grafting mortality rankings. *Ann Thorac Surg.* 2015;99:856–862.
45. Sherwood MW, Brennan JM, Ho KK, et al. The impact of extreme-risk cases on hospitals' risk-adjusted percutaneous coronary intervention mortality ratings. *JACC Cardiovasc Interv*. 2015;8:10–16.

DISCUSSANTS

D. Hoyt (Chicago, IL):

No conflict of interest.

The Affordable Care Act was created on the basis of 3 policy principles: (1) to increase access through insurance reform; (2) to initiate cost reduction by exploring different payment models; and (3) to redevelop health care delivery through quality measurement and transparency creating market-driven choices.

In the area of insurance reform, progress has been made, and 15 million more people are currently covered. Cost reduction has been evident. For the first time, the cost curve seems to be bending. The measurement of quality has developed tremendously through measurement science. What has been slow to develop is the transparency of measurement that will allow consumers to judge value and drive choice and health-delivery reform.

Michael Porter has talked about the drivers of transformation and links transparency in payment and transparency in results as the critical forces that evolve the marketplace. These principles are embedded in the health policy upon which the ACA was developed.

The paper we have heard today has provided leadership in medicine and in surgery. The cardiac surgeons have been among the first doing this through the well-known vehicle of *Consumer Reports*. Data from other areas have suggested that this is important to consumers. Many have argued that we have a "moral" responsibility to be transparent about outcomes. Data about cardiac surgery outcomes are not new, and Mark Chassin, the current executive director for The Joint Commission, led this effort in New York State when he was Health Care Commissioner.

The American College of Surgeons has shared National Surgical Quality Improvement Program data for the last 4 years on Hospital Compare, and currently 25% of hospitals participate. The Commission on Cancer has also shared cancer data from the National Cancer Database on a national and regional basis.

The current study demonstrates that since its introduction, participation has increased. There seems to be a persistent group that has been slow to participate. Of those reporting, hospitals with higher volumes and, in general, higher performance are participating. There seems to be no evidence of risk aversion per se. Cardiothoracic surgeons are to be congratulated for their forward thinking and their leadership in bringing these data forward. This will push us all to bring this forward to patients.

The experience does not seem to suggest significant changes over time within hospitals. It does not seem to have affected referral patterns over time either. It is not clear from the data, but one might question whether this actually affects or guides patients' selection of a provider.

This leads me to several questions:

1. Do you have evidence that higher-quality hospitals seem to attract patients or drive patients to select care?
2. Is availability in a publication such as *Consumer Reports* better than just society or CMS Web sites?
3. Why has there been no greater penetration over time? Is this due to hospital leadership, cultural change, convincing people that this is the right thing to do? What are your thoughts? What is the carrot or stick to be used in the future to increase participation?

4. How should these programs improve to meet or reach their goals? How will hospitals be ranked if everyone improves and hits high targets, will there still be underperformers reported or is there an ideal state?

Response From D. Shahian:

The market impact of public reporting has been debated and studied for years. The best data come from heart surgery, specifically in New York State. There have been 2 particularly good studies, 1 authored by Mark Chassin, looking at the 4-year results of public reporting, and then a 15-year follow-up by Ashish Jha and Arnie Epstein from the Harvard School of Public Health.

In both those studies, and in virtually every other investigation of which I am aware, public reporting has seemed to have little impact on patients choosing better hospitals or avoiding lower-rated hospitals. So, although public reporting addresses the ethical principle of patient autonomy by providing data upon which to base informed choices, we cannot force patients to use these data the way we think they should.

Why have patients not optimally utilized these data? In many cases, they are simply unaware of the existence of report cards. Furthermore, in cardiac surgery, the urgency of the patient's condition may make it difficult to consider alternative providers. Patients may not understand how to correctly interpret report cards, or they may not appreciate the significance of different ratings. Finally, a particular patient's insurance may not cover all providers, thus limiting their range of choices.

This is all about to change. We are now in an era of unprecedented quality and cost transparency, and these metrics are also increasingly linked to provider reimbursement. Together with differential co-pays, preferred provider networks, reference pricing, and similar consumer-focused initiatives, there will be much more pressure on patients to make data-driven decisions.

With regard to our *Consumer Reports* initiative, we would like our STS data and metrics to be used preferentially by any entity, commercial or governmental, wanting to assess cardiothoracic surgery quality. In addition to our clinical data and thoroughly vetted analytical methodologies, we also have about 95% national penetration among cardiothoracic providers, which ensures broad representativeness.

I think there are a number of reasons why we seem to have plateaued at just under 50% participation in voluntary public reporting. We have done some informal surveys and webinars, and the most encouraging finding is that I have never heard a surgeon say, "We don't think this is the right thing to do." Usually, any objections are practical in nature and typically fall into several categories. First, there is concern about the methodology. Is there adequate sample size? Is there adequate risk adjustment? Will reporting lead to risk aversion? Will patients understand the results?

There are cogent responses to all those objections, most of which rely on our highly vetted, state-of-the-art statistical methodologies. Adequate sample sizes are achieved by using composite measures with multiple endpoints, multiple years of data, and sometimes multiple procedures. With regard to risk aversion, my colleagues and I demonstrated in a paper just last year (*Ann Thorac Surg*. 2015;99:856–862.) that STS risk adjustment may slightly overpredict mortality in the very highest risk patients, thereby more than adequately protecting providers who care for such patients. Finally, it is critical that we present our results in a variety of ways that are understandable to all patients, many of whom have little experience interpreting complex data.

How do we continue to drive improvement? Feedback of STS data to our surgeons and hospitals is necessary but not sufficient. Quality measurement and quality improvement are related but somewhat different skillsets. You cannot rely solely on measurement and reporting, as important as they are. I think we need to look at the structured, continuous improvement programs exemplified by organizations such as the Northern New England Cardiovascular Disease Study Group, the Michigan Society of Thoracic and Cardiovascular Surgeons, the Virginia Cardiac Surgery Quality Initiative, and the STS Task Force on Quality Initiatives, led by my colleague Dr Rich Prager.

What is the ideal future state? I would love to see all cardiothoracic programs improve to the point that a patient could use features such as geographic proximity and convenience as the major selection criteria. Performance differences between programs will have become very small, because all are functioning at high levels. That would be my goal.

DISCUSSANTS

R.S. Jones (Charlottesville, VA):

First of all, I want to express deep gratitude to you and your colleagues and to the STS for your outstanding leadership, which dates back to 1989 when the program started. It has clearly been a model for data-driven quality improvement.

There's little that I can add to Dave Hoyt's comments, but I do want to bring up an issue and would like to hear your comments on it.

What we were talking about are professional reporting and *Consumer Reports*, but today the CMS value-based purchasing program for hospitals is clearly in effect, and the CMS physician quality reporting system is in effect, and the latter is certainly complex and confusing.

My question is, what strategies do you have or can we develop to incorporate the high-quality programs such as the STS, like you've just demonstrated? How do we get CMS to acknowledge that effort and allow that to be the quality metrics for their purposes?

We talk about CMS, but we have had little discussion about the private insurance industry. Each one of those companies has its own approach to evaluating doctors' performance. It seems to me that's yet another challenge that we need to think how can we get STS and similar kinds of provider-developed quality measures to be acknowledged by both public and private insurers?

Response From D. Shahian:

You raise a number of important points. We recently developed, under contract from CMS, a readmission measure for CABG based on STS clinical data linked to Medicare claims. We hoped that CMS would preferentially use this readmission measure for their various accountability programs, but it is going to use a completely claims-based measure instead. Although the STS Adult Cardiac Surgery Database has 95% national penetration, CMS feels that it cannot mandate a metric that requires participation in the STS database. My suggestion for this and similar scenarios is to simply designate STS metrics as an available option, along with claims-based metrics for those who are not STS participants.

You also mentioned the CMS PQRS program for physicians. The traditional approach to measuring surgeon's performance in cardiac surgery was to estimate risk-adjusted mortality for their coronary bypass surgery cases. Given the declining numbers of such cases and falling mortality rates, this is statistically no longer an ideal metric. Furthermore, measuring surgeon performance using the results from just 1 procedure is not a fair reflection of their overall performance.

We have now developed a surgeon-level composite measure on the basis of the 5 most common procedures for adult cardiac surgeons, encompassing about 80% of a typical practice. We use rolling 3-year data periods. Our composite includes both risk-adjusted mortality and the occurrence of any 1 of the 5 most common serious complications

of cardiac surgery. Because of the multiple procedures and endpoints, the reliability or signal-to-noise ratio of this composite surgeon metric is 0.8, which is quite high.

We are planning to confidentially provide surgeons with their composite results over the next year or so to prepare them for what I think might be coming, a national requirement for physician-level performance reporting.

DISCUSSANTS

R.C. Thirlby (Seattle, WA):

Cardiac surgeons and neurosurgeons are in a unique position. Dr Birkmeyer has shown that there are the only 2 elective procedures, craniotomy and CABG, where mortality is a relevant outcome measure. There was a very important paper from the Michigan Collaborative in this month's *Annals of Surgery*, which looked at colectomy complications. There's 1 surgeon in the entire state of Michigan who has an adequate volume to have statistically significant assessment of his or her colectomy complications.

My question to you is "as general surgeons, since mortality is not a relevant outcome, how do we design a composite score that's relevant and statistically valid for scoring individual surgeon performance?"

Response From D. Shahian:

Focusing on single procedures is a problem. I think you are going to have to take the same approach that we did. There are probably a handful of procedures that are commonly performed by most general surgeons. Mortality rates may be low for many of those procedures, but I'm sure there are various complications that occur more frequently. I would suggest developing multiprocedural, multiple endpoint composite measures on the basis of multiple years of data.

DISCUSSANTS

H. Polk (Louisville, KY):

Do you have any evidence that the more lowly performing programs have used any of these data to improve their performance and their status?

Response From D. Shahian:

Only anecdotal. We do have programs that have contacted us requesting help in improving their results. STS hopes to increase its capability to render that kind of assistance.