The Society of Thoracic Surgeons National Adult Cardiac Database: A continuing commitment to excellence

David M. Shahian, MD,a Fred Edwards, MD,b Fred L. Grover, MD,c Jeffrey P. Jacobs, MD,d Cameron D. Wright, MD,a Richard L. Prager, MD,e Jeffrey B. Rich, MD,f Michael J. Mack, MD,g and Douglas J. Mathisen, MDh

In the August issue of the Journal, 2 articles1,2 raised questions regarding the Society of Thoracic Surgeons (STS) National Adult Cardiac Surgery Database (ACD). Over the past 2 decades, the STS has devoted enormous time and energy to develop and progressively refine the largest clinical cardiac surgery data registry in the world. The issues discussed in these 2 articles have always been of paramount importance to database leadership.

The ACD has been the foundation for a set of sophisticated risk adjustment models3-6 for isolated coronary artery bypass grafting (CABG) surgery, isolated valve procedures, and combined CABG plus valve procedures. It has also been the basis for a comprehensive quality measurement system for CABG.7,8 The ACD has provided data for more than a hundred clinical research studies that have advanced patient care, and its educational content has enhanced the knowledge of both cardiothoracic trainees and experienced surgeons. The Centers for Medicare and Medicaid Services (CMS) RUC process for the valuation of cardiothoracic procedures has been greatly facilitated by having “hard” data from the ACD.9 Overall, patient care has been improved in a myriad of ways by virtue of this rich repository of information. All these efforts depend on the underlying accuracy of our data.

The Society is now about to embark on a new and very significant database initiative, the voluntary public release of STS Composite CABG Scores for participants. New York State initiated mandatory public reporting of CABG results 2 decades ago, and a few other states have subsequently implemented similar programs. However, the STS voluntary public reporting initiative, initially conducted in collaboration with Consumers Union and subsequently planned for the STS Web site, is the first of its kind by any medical specialty organization. It again establishes the leadership of cardiothoracic surgery in quality measurement and transparency.

As the Society begins this new initiative, it is appropriate to ask whether anything in the two recent articles1,2 poses a concern. For example, can we trust the STS ACD as the basis for assessing the performance of our participants? Because we now enroll more than 90% of the cardiac programs in the United States, these are questions of great relevance to our entire specialty and not just the STS. We will examine each of the recent Journal articles in detail and respond to their specific criticisms and concerns.

ELEPHANT IN THE ROOM?

Brown and colleagues1 imply that the STS has just recently awakened to the need to verify data accuracy: “The STS has begun to conduct national audits on the accuracy of the data in the NCD. Their stated goal is to audit 20 centers per year for 62 data elements in 20 cases of isolated CABG.” The commentary by Grunkemeier and Furnary2 reinforces this notion, beginning with the title “Data variability and validity: the elephant in the room.” This unfortunately chosen metaphor evokes the image of a huge and obvious problem that has been largely ignored and curiously neglected.

Such characterizations are both inaccurate and completely at odds with the philosophy of ACD staff and leadership, who continually strive to improve and ensure the accuracy of all data elements. No single issue has been more important to the ACD, and it is a major agenda item on every call and meeting of database leadership. We agree with Brown and colleagues1 and Grunkemeier and Furnary2 that the ultimate limiting factor in all ACD initiatives is the accuracy and completeness of the underlying data. The Society and its database leadership are committed to ensuring that the validity of our data is second to none.

For the fifth consecutive year, the Iowa Foundation for Medical Care is conducting a comprehensive external audit of randomly selected STS sites to determine the accuracy and completeness of their data coding. This initiative was voluntarily instituted by the Society, which completely funds its cost. At each audited site (40 this year and not 20 as stated by Brown and colleagues1), 75 variables are audited for 15 CABG and 5 valve procedures, using the paper or electronic medical record as the gold standard. Hospital operative logs are also examined to determine whether the data are complete and accurate.

The Journal of Thoracic and Cardiovascular Surgery • Volume 140, Number 5 955

From the Department of Surgery, a Massachusetts General Hospital, Boston, Mass; Division of Cardiothoracic Surgery, e University of Michigan Health Center, Ann Arbor, Mich; Mid-Atlantic Cardiothoracic Surgeons, f Norfolk, Va; and Cardiothoracic Surgery Associates of North Texas, g Dallas, Tex.

Received for publication Aug 13, 2010; revisions received Sept 10, 2010; accepted for publication Sept 20, 2010.

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

Copyright © 2010 by The American Association for Thoracic Surgery doi:10.1016/j.jtcvs.2010.09.004
all cases were reported in the STS ACD. In 2009, aggregate agreement rates across all audited variables ranged from 92.3\% to 98.6\% (mean, 96.1\%) among the 29 sites. Several individual variables had substantially lower agreement rates at some hospitals, including many of the valve disease variables studied by Brown and colleagues.\textsuperscript{1} Similar audit initiatives, also conducted by the Iowa Foundation for Medical Care, are used in the STS Congenital Heart Surgery Database\textsuperscript{10} and, starting this year, the STS General Thoracic Surgery Database.

Several initiatives were introduced this year to further enhance the audit process. First, the most time-consuming and costly aspect of the audit has been the need for reviewers to travel to distant hospitals and to stay there for an average of 2 days. This year, the STS Audit Task Force conducted a pilot study of an initiative to remotely audit electronic hospital records, thus eliminating the onsite requirement. Results thus far are encouraging, and this approach might make it possible to audit many more sites each year. Subsequently in this commentary, we address additional new initiatives to enhance the accuracy of 30-day mortality status.

One essential step in auditing the accuracy of registry data is to ensure that all procedures performed are actually reported. To that end, we recently performed analyses based on the successful linkage of the STS ACD to CMS data from 2000 to 2007, and these results are also reassuring. Completeness of case inclusion at STS sites (defined as the number of CMS CABG cases at STS sites linked to STS records divided by the total number of CMS CABG cases at STS sites) increased progressively from 88\% to 97\% during these 8 years of analysis.\textsuperscript{11} In 2007, 88,857 (97\%) of 91,363 CMS CABG hospitalizations at STS sites were linked to an STS record.

WHY STUDY THE CODING PERFORMANCE OF UNTRAINED ABSTRACTORS?

The study by Brown and colleagues\textsuperscript{1} is an attempt to examine the variability of information contained in the STS ACD, but the design and methodology of their study are problematic for this purpose. The authors used untrained cardiothoracic residents and fellows, only 30 patients (a 10-patient “convenience sample” for each of 3 different procedures), 5-year old data, and a version of the database no longer in use since late 2007.

We concur with Grunkemeier and Furnary\textsuperscript{2} that studying the accuracy of coding by untrained physicians seems inappropriate. The study design by Brown and colleagues\textsuperscript{1} created a straw man that was predestined to produce unfavorable results. It bears little resemblance to the sophisticated data collection infrastructure that is actually used at most centers, typified by the example provided by Grunkemeier and Furnary.\textsuperscript{2} The clinicians in the study by Brown and colleagues\textsuperscript{1} had no significant prior experience with the ACD, and their brief introduction to STS coding practices in anticipation of their reviews could hardly be expected to produce high-quality results. Furthermore, these individuals undoubtedly lacked the many support structures that are available to full-time data managers.

Data managers are the backbone of our data collection efforts nationally, and they have access to multiple levels of assistance regarding difficult coding issues. These include the training manual that accompanies each database release; regularly updated FAQs; an E-mail listserv for distributing important coding updates; a regularly published newsletter; regional consortia of data managers who share expertise and offer coding assistance; and a 3-day annual Advances in Quality & Outcomes national conference attended by hundreds of data managers and devoted almost completely to coding issues.

Additionally, difficult coding questions submitted by data managers to the STS are reviewed and answered routinely within a few days, either by STS staff, senior experienced data managers who serve as consultants to STS, or by STS ACD surgeon leaders. Similar processes are in place for data managers involved with the STS Congenital Heart Surgery Database\textsuperscript{10} and the STS General Thoracic Surgery Database. National conference calls are held every other week to address the most vexing coding issues, and a senior database leader always participates. All questions raised by database managers are collated and added to the FAQs. They are also used as the basis for considering changes in data element specifications with each major upgrade.

Although we agree with many of the suggestions of Grunkemeier and Furnary,\textsuperscript{2} we are concerned about the use of an anecdote regarding the “dismal” response time to one data manager’s request, which they assert remained unanswered for 6 months. This is completely at odds with existing STS ACD practice, which is to respond to data manager inquiries within 24 to 48 hours. We would encourage the data manager in question, or any data manager having similar difficulties, to contact us directly to resolve such issues.

WHY USE OLD DATA AND AN OUTDATED DATABASE VERSION?

Cases for review in the study of Brown and colleagues\textsuperscript{1} were selected from 2005 to remove the “impact of recall bias” by residents and fellows serving as reviewers because they would not have been present at the Mayo Clinic during that earlier time period. However, the use of 5-year-old data and an outdated ACD version are highly problematic. ACD version 2.52.1 was introduced in early 2004 and phased out in December 2007. Its replacement, version 2.61, has itself recently undergone extensive revision in anticipation of the release of version 2.70 in January 2011. With each iteration of database versions, definitions are refined, questionable data elements are clarified, and the training manual is
updated. It is difficult to justify or respond to criticisms of a database version that was replaced years ago, and this makes the implications of the study by Brown and colleagues\(^1\) less relevant to current practice. Nonetheless, we will respond to 2 specific comments by Grunkemeier and Furnary.\(^2\) Regarding the changed definition of stroke, the 24-hour cutoff is consistent with current national recommendations.\(^12\text{-}14\) In fact, some have considered much shorter time windows to differentiate transient ischemic attacks from strokes because many presumed transient ischemic attacks are found on magnetic resonance imaging studies to have sustained infarcts.\(^13\) The question regarding how best to code coma in version 2.61 is legitimate and does point out an ambiguous definitional issue that has been corrected in the upcoming version 2.70.

**SELECTION OF STUDY VARIABLES**

Table 1 of the article by Brown and colleagues\(^1\) enumerates the variables in their study. With the exception of surgeon and assistant name, reason for reoperation, and urgency, the remaining 24 study variables are all related to the etiology and classification of valve disease and the operative details of valve, CABG, or aortic procedures. The valve variables that dominate this study are among the most difficult to specify in the entire database, and they often generate discussion, even among attending surgeons and cardiologists. The authors acknowledge that they did not assess “...preoperative risk factors, outcomes measures, or the completeness of the number of cases abstracted...” This is an unfortunate omission because previous studies have shown that these most important variables for risk modeling and provider performance are much more accurately collected in the STS Database than in alternative sources.\(^10\text{-}15\text{-}17\)

Despite this adverse selection of study variables, the findings of Brown and colleagues\(^1\) do present some reassuring findings. Although there was some lack of agreement among untrained physician abstractors, the degree of difference was often minimal and completely understandable to any clinician who deals with such ambiguities on a daily basis (e.g., a difference of one category [mild to moderate, moderate to severe] in the assessment of valvular regurgitation, often resulting from multiple different data sources and patient assessments). Importantly, among the increasingly common cohort of mitral repair procedures, the degree of mitral regurgitation and number of diseased vessels were collected with median percentage agreement of greater than 95%. In aortic valve or aortic procedure groups, aortic stenosis and arch replacement both had median agreement percentages of 93%.

The Fleiss $\kappa$ statistic for coronary-specific variables, which apply to the most commonly performed cardiac surgery procedure, ranged from 0.69 to 1.0, but Brown and colleagues\(^1\) fail to provide readers with commonly accepted interpretations of this statistic. For example, one classic reference on strength of agreement classifies $\kappa$ values of 0.41 to 0.60 as moderate agreement, 0.61 to 0.80 as substantial agreement, and 0.81 to 1.0 as almost perfect agreement.\(^15\) Thus despite untrained abstractors, agreement on the coding of these variables was very high.

With regard to the classification of valve etiology, anatomy, pathophysiology, and procedure types, we acknowledge that these areas have been incomplete in the past, mainly in an attempt to keep the database to a more manageable number of data elements. In preparation for version 2.70, a completely new valve section has been developed over the past 3 years by a group of surgeons and cardiologists with a special interest in this area. The ability to precisely specify valve parameters for an individual patient has been greatly expanded. Similarly, a revised and expanded CABG section will provide detailed information unavailable in recent versions, such as the specific conduits used to graft particular target vessels.

Repeated comments and questions from data managers have led ACD leadership to the conclusion that we are often asking too much of nonsurgeon coders and that surgeons are sometimes too detached from more complex coding questions. This does a disservice to them, and it potentially compromises database accuracy. Accordingly, we are designing a series of concise, focused data forms that would be filled out by the operating surgeon at the time of the procedure and transmitted to the data managers for entry into the database (or entered directly through an operating room computer, if that capability exists). For the most difficult and technical coding issues, this places the responsibility for accurate coding on the most knowledgeable team member: the surgeon. Recognizing the demands on the surgeon’s time, these forms will be constructed in such a way that they will require only a few seconds to check off the appropriate responses.

**THE VALUE OF CLINICAL REGISTRIES VERSUS ADMINISTRATIVE DATABASES**

Brown and associates\(^1\) cite studies comparing the STS ACD with administrative sources. Then, because their current study of a limited number of the most difficult STS ACD variables revealed some inconsistencies in STS ACD coding, they conclude that “Variability among databases draws into question the reliability of the content of each database being compared.” This statement essentially places the audited, clinical STS ACD into the same category as administrative databases collected for billing purposes and lacking important clinical variables. In reality, every major study comparing clinical and administrative data sources for CABG, some dating back nearly 2 decades, has shown the inferiority of the latter, often with substantial negative implications for accurate performance measurement.\(^15\text{-}17\text{-}19\text{-}22\) Similar studies have clearly documented
the inferiority of administrative data in comparison with clinical data for the analysis of outcomes after congenital cardiac surgery.23-25

PREVIOUS ANALYSES

Brown and colleagues1 refer to a study of coding accuracy at a single institution, presumably as further evidence supporting their current study:

Another study of isolated CABG procedures examined the rates of error within a single institution’s STS database, including risk factors, operative mortality, major complications, and other outcomes. Herbert and coworkers found complete correlation in only 2% of cases, 1 to 10 discrepancies in 82% of cases, and more than 10 fields of disagreement in 17% of cases. This study did not examine other types of operations or etiology of disease.1

One could logically conclude from reading this statement that the cited study of Herbert and associates16 found substantial concerns regarding STS ACD data accuracy. To the contrary, these are actual excerpts from the abstract of that study:

Discrepancies were noted in 5% (16) or fewer of the audited fields for 98.8% of the records. Of the 32 variables in the mortality risk algorithm, discrepancies were present in fewer than 10% of the audits on 30 of the 32 variables. More than 95% of the audited charts had zero or one discrepancy in the seven most important variables in the mortality risk models. Operative mortality was determined to be completely accurate with no discrepancies between the database and the audited clinical record. Among major complications, the error rate was less than 1% for all complications except prolonged ventilation (4.0%)... A detailed audit of a clinical outcomes registry database demonstrated that the major fields within this specific database including operative mortality, major complications, and the significant factors in the risk algorithm were highly accurate." [italics added]16

Thus in contrast to the negative impression conveyed by Brown and associates,1 this careful internal audit study by Herbert and associates16 actually revealed high accuracy in the most critical variables necessary for provider profiling. This is reassuring, particularly because the data were from 2001 and predated the current STS external audit process. Because one of the authors on this response (MJM) was the senior author on the study by Herbert and associates,16 we take particular exception to this misleading characterization of its findings.

ARE 30-DAY MORTALITY DATA ACCURATE IN THE STS ACD?

Brown and colleagues1 state that “Unfortunately, the patient status at 30 days (dead or alive) was only in agreement in 83% of cases in 2008. This is of great concern, because ‘30-day mortality’ or ‘operative mortality’ is the most common outcome used for both quality indicators and research."

We acknowledge that ascertainment of 30-day status is a problem for all clinical data registries and not just the STS ACD. It is particularly difficult for major referral centers whose patients often return to the care of distant primary providers. Through a combination of reviewing office and laboratory visit records and directly calling patients and their referring doctors, accurate information can usually be obtained, but it is a costly and time-consuming process.

In our most recent 2009 audit, in-hospital mortality was captured with essentially 100% accuracy, and 30-day mortality status agreement was 89%. Although the accuracy of 30-day status improved substantially over the prior year’s audit, it is still a concern of database leadership. Our presumption is that most programs with poor agreement on 30-day status have used in-hospital mortality as a convenient (but inappropriate) surrogate.

At the request of STS ACD leadership, the Duke Clinical Research Institute (DCRI) has conducted a number of studies to assess the potential effect of inaccurate 30-day mortality coding on our CABG risk models, composite scores, and star ratings. The results of these studies are reassuring. Linkage of STS data with the Social Security Death Master File (SSDMF) shows that the majority of 30-day deaths occur in the hospital, and these deaths are recorded with high fidelity. Most errors occur in the coding of patients who are discharged alive but subsequently die outside the hospital within the 30-day window and whose status is not correctly ascertained. Differences between in-hospital and operative mortality (which includes 30-day mortality) have only a mild effect on risk model regression coefficients, odds ratios, and star ratings. The Pearson correlation coefficient between STS CABG composite scores estimated by using STS ACD operative mortality versus operative mortality based on the SSDMF was 0.99. This close correlation presumably results from the fact that a majority of 30-day deaths occur in the hospital and are accurately collected.

Notwithstanding these reassuring results, STS ACD leadership is committed to achieving 100% accuracy of the 30-day mortality status. In 2009, we required that hospitals document the method by which they verified 30-day mortality status, and in 2010, they are required to present this documentation for each case (eg, the record of a 6-week check-up). Furthermore, through the leadership of the STS Database Task Force on Longitudinal Follow-up and External Linkages and DCRI, we are creating linkages of our ACD with the SSDMF for those programs that provide Social Security numbers.26,27
Many institutions currently audit their reported 30-day mortality against the SSDMF. Unfortunately, there is a lag period of up to 6 months on Social Security data, making “real-time” correlation problematic. Ultimately, however, this will enable the ACD to independently confirm death status not only at 30 days but also at the longer follow-up intervals that some have recommended.

**CONCLUSION**

We welcome the interest of Brown and colleagues in the validity of the STS ACD, and we invite their participation in any of our ongoing initiatives. Obtaining complete and accurate data has been a major focus of STS ACD leadership for more than 2 decades, and the Society is absolutely committed to this goal.

Determining optimum clinical care and developing evidence-based guidelines requires the highest quality study data. As we enter an era of greater transparency and accountability, data accuracy has even more widespread implications. Patients might be misled in their selection of providers if ratings are based on inaccurate data, performance improvement initiatives might be inappropriately prioritized, and the reputations and reimbursement of providers are at stake. Hospitals and STS ACD participant groups must recognize the need for highly trained data abstractors, robust internal data quality checks to complement those implemented by STS and DCRI, and more direct attending surgeon involvement in the coding of those technical data elements for which they are best qualified.

The STS ACD is not perfect, but we strive every day to continually improve it. The ACD is a process in evolution and not a static product. Even with its imperfections, it is widely and correctly regarded as one of the best registries in all of health care, providing information that is far more accurate and useful than that available from administrative sources. It provides value to patients, our profession, and to other stakeholders who want to better understand the complex world of cardiac surgery. We will continue to refine it and to improve its accuracy, and we welcome all who would contribute to these efforts.

**References**