THE SOCIETY OF THORACIC SURGEONS

633 N. SAINT CLAIR STREET, SUITE 2320
CHICAGO, ILLINOIS 60611-3658
Phone: 312/202-5800
Fax: 312/202-5801
E-mail: sts@sts.org
Web: http://www.sts.org

September 7, 2010

Pediatric Cardiac Surgery Steering Committee
National Quality Forum
601 Thirteenth Street NW
Suite 500 North
Washington, DC 20005

Re: National Voluntary Consensus Standards for Pediatric Cardiac Surgery: A Consensus Report-Draft Report for Commenting

Dear Steering Committee Members:

On behalf of The Society of Thoracic Surgeons (STS), a member of the National Quality Forum (NQF), we would like to take this opportunity to provide comments regarding the aforementioned document.

STS is a not-for-profit organization representing more than 6,000 cardiothoracic surgeons, researchers, and allied health professionals worldwide who are dedicated to ensuring the best surgical care for patients with diseases of the chest. Founded in 1964, the mission of STS is to enhance the ability of cardiothoracic surgeons to provide the highest quality care through education, research and advocacy. STS supports data-driven approaches to the improvement of the quality of cardiothoracic surgical care.

STS would like to comment specifically about two measures recommended for NQF endorsement:

1. **PCS-018-09**: Operative Mortality Stratified by the Five STS-EACTS Mortality Levels [Society of Thoracic Surgeons (STS)]

2. **PCS-021-09**: Standardized Mortality Ratio for Congenital Heart Surgery, Risk Adjustment for Congenital Heart Surgery (RACHS-1 Method) [Children's Hospital Boston (CHB)]

This document contains three parts:

1. General discussion about the use of administrative data versus clinical data for the evaluation of quality of care for patients undergoing pediatric cardiac surgery
2. Comparison between **PCS-018-09** and **PCS-021-09**
3. Critique of **PCS-021-09** organized according to the four criteria proposed by NQF for evaluation of measures

I. Administrative Data versus Clinical Data for the Evaluation of Quality of Care for Patients Undergoing Pediatric Cardiac Surgery

STS advocates the use of clinical databases rather than administrative databases for the evaluation of the quality of care for patients undergoing treatment for pediatric cardiac disease. Evidence from three recent investigations suggests that the validity of coding of lesions seen in the congenitally malformed heart via the International Classification of Diseases (ICD) by administrative databases is likely to be poor [1, 2, 3]:

Among 373 infants with congenital cardiac defects at the Children's Hospital of Wisconsin, investigators reported that only 52% of the cardiac diagnoses in the medical records had a corresponding ICD code in the hospital discharge database [1].

The Hennepin County Medical Center discharge database in Minnesota identified all infants born during 2001 with an ICD-9 code for congenital cardiac disease. Physician review of these 66 medical records confirmed the accuracy of only 41% of the codes contained in the administrative database from the ICD [2].

The Metropolitan Atlanta Congenital Defect Program of the Centers for Disease Control and Prevention's Birth Defect Branch carried out surveillance of infants and fetuses with cardiac defects delivered to mothers residing in Atlanta during the years 1988 through 2003 [3]. These records were reviewed and classified using both administrative coding from the ICD and the clinical nomenclature used in the STS Congenital Heart Surgery Database. It was concluded that analyses based on the codes available in the ICD are likely to "have substantial misclassification" of congenital cardiac disease.

The following are potential reasons for the poor diagnostic accuracy of administrative databases and codes from the ICD:

- accidental miscoding
- coding performed by medical records clerks who have never seen the actual patient
- contradictory or poorly described information in the medical record
- lack of diagnostic specificity for congenital cardiac disease in ICD codes
- inadequately trained medical coders

Based on the above discussion, STS supports the endorsement of NQF National Voluntary Consensus Standards for Pediatric and Congenital Cardiac Surgery based on data from clinical databases rather than administrative databases.

II. Comparison of Pediatric Cardiac Surgery Outcome Measures Submitted to NQF:

A. PCS-021-09: Standardized Mortality Ratio for Congenital Heart Surgery, Risk Adjustment for Congenital Heart Surgery (RACHS-1 Method) [Children's Hospital Boston (CHB)]

Published in 2002, RACHS-1 is a consensus-based method for risk adjustment for congenital heart surgery. At its inception, a panel of 11 experts ordered procedures by likelihood of short-term mortality, evaluated the results, and made adjustments based on data observed from two large databases: Pediatric Cardiac Care Consortium (1996, 32 institutions) and hospital discharge data purchased from three states (Illinois 1994, Massachusetts 1995, California 1996). After establishing initial assignments to categories by a consensus process, the panel decided that for some operations, age at surgery or specific cardiac diagnoses were potentially important additional risk factors. The procedures were then assigned to risk categories. Then the panel reviewed the information from the "reference data sets," and revised the categorizations of some procedures because the actual mortality rate differed considerably from the initial subjective judgment about risk for death. Case selection was largely dependent upon ICD-9 and CPT-4 codes.

In 2002, in the Journal of Thoracic and Cardiovascular Surgery (JTCVS), Jenkins et al. published an evaluation of center-specific differences in mortality using the RACHS-1 method. By using 1996 hospital discharge data from six states, centers performing at least 100 operations for congenital heart disease, in patients age <18 years, were identified. Using the RACHS-1 method, procedures were grouped into six risk categories, and institutions were ranked in order of increasing mortality rate. Among 109 centers performing 7,177 operations for congenital heart disease, 22 performed at least 100 cases (72.3% of total operations). Unadjusted mortality rates ranged from 2.5% to 11.4%. A total of 4,318 of the 7,177 cases could be placed into one of the six risk categories. Few deaths occurred in risk category 1, and few institutions performed procedures in risk categories 5 and 6, making institutional comparisons in these categories uninformative. Considering mortality rates in categories 2 through 4, institutions displayed either relatively consistent ranks, a threshold increase in mortality as higher-risk procedures were performed, or a threshold decrease in mortality. Used in many settings as an expression of performance, Standardized Mortality Ratios (SMRs) were calculated with the intent of describing which institutions performed better or worse than expected on the basis of their case mix.

RACHS-1 levels have been widely used as a tool to express relative risk of in-hospital mortality for various procedures. In addition to procedure information, the complete RACHS-1 model incorporates certain patient factors (i.e. age, prematurity, major non-cardiac structural anomaly) into the risk adjustment process.

The proposed measure utilizes assignment of cases to RACHS-1 levels and other patient and procedural variables (i.e. age, prematurity, presence of non-cardiac anomalies, combination procedures) to determine an institution's SMR, which is defined as its actual or observed in-hospital mortality rate divided by its expected in-hospital mortality rate. The expected rate is calculated based on the patient case mix at the institution relative to the case mix in the reference data set as a whole.

Advantages:

1. Having been used in many published evaluations of outcomes, the RACHS-1 system of procedure categorization is widely recognized.
2. The development of RACHS-1 was based upon expert panel opinion.

3. When SMR is calculated using logistic regression based on coefficients that appropriately pertain to the correct reference data set and if all centers are participants in the same reference data set, then institutions may be "ranked" according to SMR (lowest to highest).

Disadvantages:

1. RACHS-1 is based largely upon expert opinion, rather than objective evidence.
2. The measure steward has made reference to the possible release of "RACHS-2 levels" in 2010, which presumably will supplant and thus render obsolete the current system.
3. The development of RACHS-1 was based on administrative data, now more than 10 years old. The measure description describes the use of at least four independent, disparate reference data sets, the majority of which depend entirely upon administrative data. Many important and frequently performed procedures do not have ICD-9 codes (e.g., Norwood Stage 1), and thus, they must be inferred from aggregate codes in administrative data sets. Coding of patient factors such as non-cardiac anomalies may be quite variable between data sets, and these factors lack rigorous, consistent definitions.
4. Many pediatric cardiac surgical procedures cannot be classified by RACHS-1. Specifically, only 85.8% of procedures in the STS Congenital Heart Surgery Database can be classified by RACHS-1 [4]. In the previously mentioned study by Jenkins et al. on center-specific differences in mortality, only 4,318 of 7,177 congenital cardiac surgical operations (60%) could be placed in one of the six RACHS-1 categories.
5. The mainstay of the proposed measure is an SMR, which is determined by the calculation of observed to expected (O-E) mortality ratio for a given center. This process is dependent upon the availability, accuracy, and applicability of a reference data set, and the solution to a multivariate equation with specific coefficients for each of the "risk variables." The reference data set used to develop RACHS-1 was based mainly on administrative data, now more than 10 years old. It is unknown whether there is a new "contemporary universal reference data set." **No such unique reference data set is described or identified in the measure description. As such, the mechanism by which expected mortality will be calculated for a given center is unknown.** This problem leads to many unanswered questions:
    a. Will index institutions need to be among the centers from which the reference data are derived?
    b. Can an institution base their measurement and reporting on a reference data set which does not include their own data?
    c. Will one of several sets of coefficients for the logistic regression need to be provided to each institution?
    d. Will this set of coefficients depend on the participation of that particular center in one of several consortia or databases from which a reference data set is derived?
    e. Will each institution need to enlist the services of a biostatistician to analyze their own outcome data, apply the institution-specific logistic regression model, and calculate their observed and expected mortalities and SMR?
    f. Can the SMR of an institution that utilizes one reference data set (which determines the coefficients used to calculate their expected mortality) be compared to the SMR of another institution that utilizes a different reference data set?
6. SMR incompletely illustrates outcome and does not appropriately address questions regarding a center's performance within the individual strata of complexity or risk. From SMR, it is not possible to determine whether an institution is dealing at all with cases in a specific stratum of risk or complexity. For example, two centers may have the same

calculated SMR but one of those centers may be dealing with many complex "high risk" cases, while the other center deals only with cases of lesser complexity and thus lower risk.

B. PCS-018-09: Operative Mortality Stratified by the Five STS-EACTS Mortality Levels [The Society of Thoracic Surgeons (STS)]

Historically, STS has recognized the limitations of using raw, unadjusted mortality rates as a measure of outcome, quality, and performance. Given the enormous diversity of congenital heart operations, it is understandable that the initial approach to the characterization of case mix (i.e., description of relative complexity and relative risk of mortality) was largely based on expert opinion. For nearly a decade, the STS Congenital Heart Surgery Database reported outcomes of cases stratified by RACHS-1 and the Aristotle Complexity Score, two popular and widely used tools adapted to this purpose. The importance of strict definitions (e.g., procedural terms, patient factors, mortality, and time intervals) was emphasized in the methodology. Ultimately, the goal was to transition from subjectively derived estimates of risk or complexity to an empirically-based tool for analyzing mortality associated with congenital heart surgery.

In 2009, the STS-EACTS Congenital Heart Surgery Mortality Categories (STS-EACTS Categories) were published [5]. Mortality risk was estimated for 148 types of pediatric and congenital cardiac surgical operations using data from 77,294 operations entered into the STS Congenital Heart Surgery Database (43,934 patients) and the European Association for Cardiothoracic Surgery (EACTS) Congenital Heart Surgery Database (33,360 operations) between 2002 and 2007. Procedure-specific mortality rate estimates were calculated using a Bayesian model that adjusted for small denominators. Procedures were then sorted by increasing mortality risk and grouped into five categories (STS–EACTS Congenital Heart Surgery Mortality Categories [2009]) that were chosen to be optimal with respect to minimizing within-category variation and maximizing between-category variation. Model performance was subsequently assessed using an independent validation sample (n = 27,700; 2007-2008 data) and compared with two existing methods: RACHS-1 Categories and the Aristotle Basic Complexity Score.

Estimated mortality rates ranged across procedure types from 0.3% (atrial septal defect repair with patch) to 29.8% (truncus + interrupted aortic arch repair). The STS–EACTS Categories demonstrated good discrimination for predicting mortality in the validation sample (C-index= 0.773).

In the subset of procedures for which STS–EACTS Category, RACHS-1 Category, and Aristotle Basic Complexity Score are defined, discrimination was highest for the STS–EACTS Categories (C-index = 0.778), followed by RACHS-1 Categories (C-index = 0.745), and Aristotle Basic Complexity scores (C-index = 0.687).

Table 1 shows the results of comparing the STS–EACTS Categories (2009) to the RACHS-1 Categories and the Aristotle Basic Complexity Score using an independent validation sample of 27,700 operations performed in 2007 and 2008.

Table 1: Method of Modeling Procedures	Model without patient covariates	Model with patient covariates	Percent of operations that can be classified
STS-EACTS Congenital Heart Surgery Mortality Categories (2009)	**C = 0.778**	**C = 0.812**	**99%**
RACHS-1 Categories	**C = 0.745**	**C = 0.802**	**86%**
Aristotle Basic Complexity Score	**C = 0.687**	**C = 0.795**	**94%**

The STS–EACTS Categories are now incorporated into the outcomes reports of the STS Congenital Heart Surgery Database. In addition, the five STS-EACTS Categories are the basis for stratification of mortality outcomes in the measure submitted to NQF.

Advantages:

1. The STS–EACTS Categories are based on objective data from a clinical data set that is internally validated and checked by site verification on an ongoing basis.
2. The accuracy and discrimination of the STS–EACTS Categories were validated using 2007-2008 data.
3. The majority (99%) of pediatric and congenital cardiac surgical operations that are coded by centers participating in the STS Congenital Heart Surgery Database can be assigned to one of the five STS-EACTS Categories. In comparison, 94% can be categorized using the Aristotle Basic Complexity Score, and 86% can be assigned to a RACHS-1 Category.
4. Comparison (by C-statistic) to Aristotle Basic Complexity Score and RACHS-1 Categories, both with and without patient factors, reveals best performance by STS–EACTS Categories.
5. Using the STS–EACTS Categories eliminates the uncertainties of coding that are associated with the use of administrative data, both for center reporting and for development of a reference data set.
6. The proposed measure is maximally informative, allowing the stakeholder to draw inferences concerning mortality outcomes associated with procedures of low, intermediate, and high levels of complexity.
7. Calculation and reporting of mortality outcomes by STS-EACTS Categories does not require each institution to solve any multivariable equations and does not require biostatistician-level skill. For centers participating in the STS Congenital Heart Surgery Database, the calculation and reporting is already completed at six month intervals.
8. Measurement and reporting is not dependent upon a reference data set and is therefore not susceptible to error or misinterpretation that could result from use of incorrect or inappropriate reference data.

Disadvantages: As a consequence of having been developed over the past three years, this empirically derived system of mortality levels may be less familiar to some stakeholders than Aristotle or RACHS-1.

STS recommends that ONLY the STS–EACTS Congenital Heart Surgery Mortality Categories (2009) are used for complexity stratification of mortality. Main rationale for this recommendation is two-fold:

1. The C-statistic for the STS-EACTS Congenital Heart Surgery Mortality Categories (2009) is better than for the RACHS-1 Categories and the Aristotle Basic Complexity Score.
2. 86% of pediatric and congenital cardiac operations can be assessed by the RACHS-1 Categories, 94% by the Aristotle Basic Complexity Score, and 99% by the STS-EACTS Congenital Heart Surgery Mortality Categories (2009).

III. Critique of PCS-021-09, Organized According to the Four Criteria Proposed by NQF for Measure Evaluation

This critique is organized according to the four NQF evaluation criteria:
(1) I=Importance to Measure and Report;
(2) S=Scientific Acceptability of Measure Properties;
(3) U=Usability;
(4) F=Feasibility

(1) Importance to Measure and Report

Measuring mortality in a way that includes a demonstration of case complexity is generally accepted as being of great importance. For quality improvement, this is essential and fundamental. For public reporting, the issues are different and even more complex. Despite explanations and disclaimers, the concept of what does and does not reach a level of statistical significance is difficult to convey in the public reporting of surgical outcomes. This becomes even more difficult and complex when reporting mortality outcomes by assigning a specific institution with a calculated numerical value such as the SMR. Use of a derived value such as an SMR leads to misunderstandings that result from the erroneous assumption that the process of "standardizing" enables the interested party to draw inferences concerning the significance of an institution's SMR as being higher or lower than that of another institution (see below). An impression of "good performance" is generally inferred from a low SMR (i.e., SMR<1), while at the same time, it conveys no information about whether or not an institution is managing cases in the higher strata of risk or complexity, and to what extent these high complexity cases account for the observed mortality rate. While the individual factors that contribute to the complexity of this problem are numerous, one that is of overriding importance is the accuracy, reliability, and applicability of the "reference data set" from which the "expected mortality" (the denominator in SMR) is derived. ***Unless there is a single, universally applicable reference data set, then reporting of individual institution SMRs based upon a variety of disparate, non-harmonious data sets is likely to misinform those to whom it is reported.***

(2) Scientific Acceptability of Measure Properties

The elements of an SMR are the observed mortality (numerator), the expected mortality (denominator), and the calculated ratio. In the case of this measure, both the numerator and denominator are derived, or calculated values.

The numerator used in the calculation of SMR is the observed mortality, which is calculated using data from the participant. Observed mortality is defined as the number of pediatric cases of congenital heart surgery resulting in in-hospital death that can be placed into a RACHS-1 risk category divided by the total number of pediatric cases of congenital heart surgery that can be placed into a RACHS-1 risk category. ***The use of the RACHS-1 risk categories limits the cases in both the numerator and denominator to those that can be classified according to RACHS-1. It has been determined that currently, 86% of the cardiac procedure types coded by participants in the STS Congenital Heart Surgery Database can be assigned to a RACHS-1 risk category. Thus, one in seven cases would be excluded from the SMR calculation*** [4, 5].

The measure specifies consideration of "total cases of congenital heart surgery among patients <18 years of age" rather than limiting the analysis to index cases, as is done in other proposed

measures. By using this criterion for the numerator and the denominator, the methodology introduces the potential for significant underestimation of mortality in both the institutional calculation and the reference data set. Mortality calculation based on the number of operative cases as opposed to the number of surgical admissions (i.e., the number of index cases) can be very misleading. This phenomenon is explained and exemplified in the following example: Consider a center that performs Norwood Stage 1 operations on ten patients in a year, with one of those patients returning to the operating room during the same admission for additional atrial septectomy, one returning for conversion from a modified Blalock-Taussig shunt to a Right-Ventricle to Pulmonary Artery Conduit, two of the ten patients being re-explored for bleeding, and seven of the ten patients being treated with delayed sternal closure operations. One patient dies prior to discharge. Twenty-one operations (total cases) were performed. But ten index operations were performed on ten patients. If mortality is calculated based upon total cases, the mortality rate would be $^1/_{21}$ or 4.8%. Based upon index cases, the mortality rate is $^1/_{10}$ or 10%, which of course reflects the true outcome for the group of ten patients. ***The description of PCS-021-09 (Standardized Mortality Ratio for Congenital Heart Surgery, Risk Adjustment for Congenital Heart Surgery [RACHS-1 Method]) does not specify calculation based on index cases in either the institutional calculation of observed mortality or in the mortality determination for the reference data set. By using all cases (total cases) rather than index cases, the methodology of PCS-021-09 introduces the potential for significant underestimation of mortality in both the institutional calculation (numerator) and the reference data set (denominator).***

The denominator used in the calculation of SMR is the expected mortality rate, which is calculated specifically for each group or participant. In this calculation, a multivariable logistic regression model, with the outcome "in-hospital death," is fitted. Five clinical characteristics are incorporated as covariates:

1. RACHS-1 risk categories 2, 3, 4, 5, and 6 as binary covariates, with category 1 as the reference group;
2. Age ≤30 days and age 31 days to 1 year, with age ≥ 1 year as the reference group;
3. Prematurity;
4. Presence of a major non-cardiac structural anomaly; and
5. Presence of combinations of cardiac surgical procedures

This logistic model is used to calculate the predicted probability of death for each individual case in the data set. The average predicted probability of death for all cases is computed by summing the predicted probabilities for each case and dividing by the total number of cases that can be placed into a RACHS-1 risk category; this average predicted probability of death for all cases represents the expected mortality rate for the group, adjusting for case mix.

The SMR is then calculated as the observed mortality rate divided by the expected mortality rate. ***A very important problem is that the measure proponents include in their measure description three distinctly different sets of coefficients for the risk adjustment model. Each pertains to a given "reference data set."*** Each set of coefficients was used by the measure stewards in validation exercises in which they considered outcome data from institutions that participated in different consortia or databases.

In the "original model validation set" the coefficients for RACHS-1 levels 2 and 6 are 1.7477 and 4.0022, with odds ratios of 5.74 and 54.7, respectively. For a second reference data set, the Kids' Inpatient Database 2006, the corresponding coefficients are 0.0202 and 1.8726, with odds ratios

of 1.02 and 6.51. For the Pediatric Health Information System 2002-2006, a third reference data set, the corresponding coefficients are 0.4081 and 2.2412, with odds ratios of 1.50 and 9.40. Thus, for one reference data set the odds ratio for mortality in RACHS-1 level 6 is 9.53 times the odds ratio for mortality in RACHS-1 level 2. For another reference data set, it is 6.38 times greater. For the third reference data set it is 6.27 times higher. The odds ratio for mortality associated with the presence of a Major Non-cardiac Structural Anomaly is 2.20, 1.29, or 1.70, depending upon the reference data set. The odds ratio for this variable in the second reference data set is outside of the 95% confidence intervals for that odds ratio in the first reference data set, and vice versa.

On page 25 of the Measure Submission Form, in the table entitled: Mortality Rates by Risk Category (RACHS-1) Single Procedures, the measure steward presents calculated mortality rates with 95% confidence intervals in each of six RACHS-1 risk categories in five different reference data sets (four American, one European). For Risk Category 3, the mortality in the KID 2003 data set is well outside the 95% confidence intervals of the PHIS 2002-2006 data set for the same risk category. The same is true in Risk Categories 5 and 6. In fact in Risk Category 6, the mortality in the KID 2003 database is outside the 95% confidence intervals of all of the other American data sets. In Risk Category 2, the mortality in the PCCC 2002-2004 data set is well outside the 95% confidence intervals of the KID 2003 data set in the same risk category. ***Since the detailed description of the measure does not actually specify what will constitute the reference data set (which will be the basis for calculation of expected mortality, the denominator in SMR), it is assumed that it may be chosen by the reporting institution, presumably from the universe of consortia, registries, or administrative data sets in which it is included. Obviously, the calculated value of SMR can vary tremendously, based upon the choice of reference data set.***

Another important measure property that affects scientific acceptability is ascertainment (i.e., to find out or learn with certainty). In the words of the measure stewards, page 3 section 11, "Data elements may be obtained from an administrative database (e.g., Healthcare Cost and Utilization Project (HCUP) Kids' Inpatient Database (KID), Pediatric Health Information System (PHIS)); from a clinical database (e.g., Pediatric Cardiac Care Consortium (PCCC), Society of Thoracic Surgeons (STS) Congenital Heart Surgery Database); from hospital-specific electronic medical records; or from paper medical records." ***Of all of the potential sources of data, the STS Congenital Heart Surgery Database is the only data source for which there is a formalized process of data verification. Data from other sources are not verified.***

As discussed above, ***the diagnostic and procedural information in the various administrative data sets has been shown to be less reliable, accurate, and predictable than that in the clinical registry database of STS***. Strickland and associates at the National Center on Birth Defects and Developmental Disabilities of the Centers for Disease Control and Prevention described their findings in comparing coding of congenital cardiac anomalies and procedures by ICD-9-CM and by the clinical nomenclature used in the STS Congenital Heart Surgery Database [3]. Their investigation revealed that the sensitivity of ICD diagnosis codes was 83% for tetralogy of Fallot, 100% for transposition of the great arteries, and 95% for hypoplastic left heart syndrome. The false positive fraction was 2% for tetralogy of Fallot, 49% for transposition, and 11% for HLHS. They concluded that "analyses based on International Classification of Diseases diagnosis codes may have substantial misclassification of congenital heart disease. Isolating the major defect is difficult, and certain codes do not differentiate between variants that are clinically and developmentally different."

As a corollary, the use of ICD-9 (and ICD-10) codes to assign cases to RACHS-1 procedural risk levels is fraught with uncertainty because of the incomplete nature and vagaries of codes.

For example, there is no ICD-9 code for the Norwood stage 1 operation. Accordingly, determination from hospital charge records (the "front sheet") or from codes in administrative databases is reliant upon the probability that a patient coded in the following fashion actually underwent a Norwood procedure:

Stage 1 Repair Risk Category 6
Require:
Proc 35.41 Enlargement of existing ASD
or 35.42 Creation of septal defect in heart
Proc 39.0 Systemic to pulmonary artery shunt
or 35.92 Creation of conduit between RV and PA
Proc 38.35 *or* 38.45 Resection of thoracic vessel
or 38.34 *or* 38.44 Resection of abdominal aorta
or 38.64 *or* 38.65 Other excision of vessel/aorta
or 38.84 *or* 38.85 Other surg occlusion of vessel/aorta
or 39.56, 39.57, 39.58 Repair of blood vessel
or 39.59
or 36.99 Other operation on vessel of heart
Cannot have:
Proc 35.94 Creation of conduit between atrium-PA
Proc 35.95 Revision corrective procedure on heart
Proc 39.21 Caval-pulmonary artery anastomosis
Allow:
Dx 745.5 Ostium secundum atrial septal defect
Dx 746.3 Congenital stenosis of aortic valve
Dx 746.5 Congenital mitral stenosis
Dx 747.0 Patent ductus arteriosus
Dx 747.10 Coarctation of aorta
Dx 747.22 Atresia and stenosis of aorta
Dx 747.89 Other anomalies of great veins
Proc 37.33 Excision of other lesion/tissue of heart
Cannot have: Any other cardiac dx

The Norwood procedure is one of the most frequently performed congenital heart operations, and is one for which operative mortality is substantial. To rely upon probabilistic matching to ascertain whether a given patient actually underwent a Norwood procedure introduces a substantial degree of uncertainty that undermines the usefulness of the measure for both quality improvement and public reporting purposes.

Another similar example is the Ross operation. This procedure consists of replacement of the aortic valve with the patient's own pulmonary valve (autograft) and replacement of the pulmonary valve with a prosthesis or homograft. Under the heading, "Define individual cardiac procedures assigned to a risk category," the specifications for the Ross procedure in Measure PCS-021-09 appear as listed below:

Ross Procedure Risk Category 3
Require:
Proc 35.21 *or* 35.22 Aortic valve replacement
Proc 35.25 *or* 35.26 Pulmonary valve replacement
Allow:
Proc 35.01 Aortic valvotomy

Proc 35.11 Aortic valvuloplasty
Proc 35.03 Pulmonary valvotomy
Proc 35.13 Pulmonary valvuloplasty
Proc 35.33 Annuloplasty
Proc 35.99 Other operation on valves of heart
Proc 39.56, 39.57, 39.58 Repair of blood vessel
or 39.59
Proc 36.99 Other operation on vessel of heart
Proc 37.33 Excision of other lesion/tissue of heart
PDA surgery
ASD2 repair

This complex method of coding the Ross procedure introduces considerable potential for error, since any combination of aortic and pulmonary valve replacement would meet these criteria, whether or not it involved pulmonary autograft replacement of the aortic valve, which is the sine qua non for the Ross procedure.

These examples of the Norwood and Ross procedures are merely two of many examples of the imprecision that can result from coding of congenital cardiac surgical procedures based upon administrative data.

(3) Usability

Given the requirement that each institution must apply a logistic model to calculate the predicted probability of death for each individual case in the data set, it seems clear that each reporting institution will be obligated to devote to this measure the efforts of a biostatistician or comparable member of the work force.

In addition, ***SMR can only be calculated once the coefficients specific to a given reference data set are determined, verified, and provided to the institution. The measure description for PCS-021-09 does not specify how it identifies what reference data set is to be used for a given institution, or how the appropriate coefficients for the logistic model will be determined and made available to each institution. Thus, the method by which this measure would be usable at the level of the measuring and reporting institution is unclear.*** The examples of validation exercises described by the measure steward involve at least four reference data sets. For some institutions, one data set pertains; for a few institutions, more than one data set pertain. If the measure can ultimately be used by all centers and institutions, then there will inevitably be some institutions for which none of the aforementioned reference data sets are applicable and some for which it would be necessary to choose among reference data sets or to calculate more than one SMR.

In section 23 of the measure description, the measure steward states that, "quality improvement efforts can be enhanced and stimulated by a clear understanding of how an entity (e.g., an institution) is performing in comparison to other entities." ***In terms of usability by stakeholders to make such determinations of relative performance, the measure is severely hampered by:***

1. ***Its failure to include reporting of any directly measured outcomes data (e.g., observed mortality, stratified by level of complexity); and***
2. ***The fact that comparison of SMR between institutions "affiliated" with different reference data sets and thus applying different risk variable coefficients in the logistic***

model, can be expected to mislead and misinform, which is certainly not the desired outcome.

(4) Feasibility

The measure description template includes the following instruction: "Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure." The measure steward states that "because this measure can be applied in administrative databases, it can be subject to the coding inaccuracies sometimes associated with these databases."

Indeed, a growing prevalence of data supports the use of clinical databases rather than administrative databases for the evaluation of quality of care for patients undergoing treatment for congenital cardiac disease. As described previously, evidence from three recent studies suggested that the validity of ICD coding of lesions seen in the congenitally malformed heart as used in administrative databases is likely to be poor [1, 2, 3].

Unintended consequences

Notwithstanding the best of intentions, public reporting of a methodologically flawed numerical index (i.e., SMR based upon a variety of disparate reference data sets) has the potential to misinform rather than educate and enlighten stakeholders and consumers. In addition, it would be worse to do so utilizing a measure that is based partially or entirely on data from administrative sources which utilize coding nomenclature that fails to address the diversity and granularity of congenital cardiac anomalies and the surgical procedures used to treat them. STS's position should not be taken as an across-the-board condemnation of the use of administrative data; we acknowledge that administrative data are a rich and essential tool that are best suited for numerous applications and across many domains. Rather, our position reflects the current state of nomenclature and coding of congenital cardiac diseases and their treatments.

Thank you for this opportunity to submit comments. Please do not hesitate to contact Jane Han, STS Manager of Quality Initiatives, at jhan@sts.org or (312) 202-5856, with any questions you may have. We appreciate your time.

Sincerely,

Jeffrey P. Jacobs, MD
Chair, STS Congenital Heart Surgery Database Task Force
Chair, STS Public Reporting Task Force

Marshall L. Jacobs, MD
Member, STS Congenital Heart Surgery Database Task Force
Member, STS Public Reporting Task Force

Fred H. Edwards, MD
Chair, STS Council on Quality, Research, and Patient Safety Operating Board

David M. Shahian, MD
Chair, STS Workforce on National Databases
Chair, STS Quality Measurement Task Force

References

1. Cronk CE, Malloy ME, Pelech AN, et al. **Completeness of state administrative databases for surveillance of congenital heart disease.** Birth Defects Res A Clin Mol Teratol 2003;67:597-603.

2. Frohnert BK, Lussky RC, Alms MA, Mendelsohn NJ, Symonik DM, Falken MC. **Validity of hospital discharge data for identifying infants with cardiac defects.** J Perinatol 2005;25:737-42.

3. Strickland MJ, Riehle-Colarusso TJ, Jacobs JP, Reller MD, Mahle WT, Botto LD, Tolbert PE, Jacobs ML, Lacour-Gayet FG, Tchervenkov CI, Mavroudis C, Correa A. **The importance of nomenclature for congenital cardiac disease: implications for research and evaluation.** In: 2008 Supplement to Cardiology in the Young: Databases and The Assessment of Complications associated with The Treatment of Patients with Congenital Cardiac Disease, Prepared by: The Multi-Societal Database Committee for Pediatric and Congenital Heart Disease, Jeffrey P. Jacobs, MD (editor). Cardiology in the Young, Volume 18, Issue S2 (Suppl. 2), pp 92–100, December 9, 2008.

4. Jacobs JP, Jacobs ML, Lacour-Gayet FG, Jenkins KJ, Gauvreau K, Bacha E, Maruszewski B, Clarke DR, Tchervenkov CI, Gaynor JW, Spray TL, Stellin G, O'Brien SM, Elliott MJ, Mavroudis C. **Stratification of complexity improves the utility and accuracy of outcomes analysis in a Multi-Institutional Congenital Heart Surgery Database: Application of the Risk Adjustment in Congenital Heart Surgery (RACHS-1) and Aristotle Systems in the Society of Thoracic Surgeons (STS) Congenital Heart Surgery Database.** Pediatric Cardiology, 2009, DOI 10.1007/s00246-009-9496-0.

5. O'Brien SM, Clarke DR, Jacobs JP, Jacobs ML, Lacour-Gayet FG, Pizarro C, Welke KF, Maruszewski B, Tobota Z, Miller WJ, Hamilton L, Peterson ED, Mavroudis C, Edwards FH. **An empirically based tool for analyzing mortality associated with congenital heart surgery.** The Journal of Thoracic and Cardiovascular Surgery, 2009 Nov;138(5):1139-53.PMID: 19837218, November 2009