



Published in final edited form as:

Ann Thorac Surg. 2016 November ; 102(5): 1580–1587. doi:10.1016/j.athoracsur.2016.04.034.

Congenital Heart Surgery Case-mix across North American Centers and Impact on Performance Assessment

Sara K. Pasquali, MD MHS¹, Amelia S. Wallace, MS², J. William Gaynor, MD³, Marshall L. Jacobs, MD^{4,5}, Sean M. O'Brien, PhD², Kevin D. Hill, MD MS², Michael G. Gaies, MD MPH¹, Jennifer C. Romano, MD MS⁶, David M. Shahian, MD⁷, John E. Mayer, MD⁸, and Jeffrey P. Jacobs, MD^{4,5}

¹Department of Pediatrics and Communicable Diseases, C.S. Mott Children's Hospital, Ann Arbor, MI

²Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC

³Department of Surgery, Children's Hospital of Philadelphia, Philadelphia, PA

⁴Division of Cardiac Surgery, Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD

⁵Division of Cardiovascular Surgery, Department of Surgery, Johns Hopkins All Children's Heart Institute, All Children's Hospital and Florida Hospital for Children, Saint Petersburg, Tampa, and Orlando, FL

⁶Department of Cardiac Surgery, University of Michigan Medical School, Ann Arbor, MI

⁷Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts

⁸Department of Cardiovascular Surgery, Boston Children's Hospital, Boston, MA

Abstract

Background—Performance assessment in congenital heart surgery is challenging due to wide heterogeneity of disease. We describe current case-mix across centers, evaluate methodology inclusive of all cardiac operations vs. the more homogeneous subset of STS benchmark operations, and describe implications regarding performance assessment.

Methods—Centers (n=119) participating in the STS Congenital Heart Surgery Database (2010–2014) were included. Index operation type and frequency across centers were described. Center performance (risk-adjusted operative mortality) was evaluated and classified when including the benchmark vs. all eligible operations.

Corresponding Author: Sara Pasquali MD MHS, University of Michigan C.S. Mott Children's Hospital, 1540 E. Hospital Drive, Ann Arbor, MI 48105, Phone: 734-232-8594, Fax: 734-936-9470, pasquali@med.umich.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Presented at the Fifty-second Annual Meeting of the Society of Thoracic Surgeons, Phoenix, AZ, Jan 23–27, 2016.

Results—Overall, 207 types of operations were performed during the study period (112,140 total cases). Few operations were performed across all centers; only 25% were performed at least once by 75% or more of centers. There was 7.9-fold variation across centers in the proportion of total cases comprised of high-complexity cases (STAT 5). In contrast, the benchmark operations comprised 36% of cases, and all but 2 were performed by 90% of centers. When evaluating performance based on benchmark vs. all operations, 15% of centers changed performance classification; 85% remained unchanged. Benchmark methodology was associated with lower power (35% vs. 78% of centers met sample size requirements).

Conclusions—There is wide variation in congenital heart surgery case-mix across centers. Metrics based on benchmark vs. all operations are associated with strengths (less heterogeneity) and weaknesses (lower power), and lead to differing performance classification for some centers. These findings have implications for ongoing efforts to optimize performance assessment, including choice of target population and appropriate interpretation of reported metrics.

Keywords

congenital heart disease; outcomes

Measures of center performance in congenital heart surgery are important to numerous stakeholders and initiatives aiming to improve outcomes. These include public reporting programs, multi-center collaborative quality improvement activities, and centers of excellence programs supported by various payers (1). As has been demonstrated previously, accurate performance measures are essential to the success of all of these initiatives (2). For example, it was recently shown that a wide-scale federal program in bariatric surgery limiting care to centers of excellence failed to improve outcomes, likely because the metrics underlying this designation did not accurately identify centers with the best outcomes providing the highest quality care (2).

Performance assessment in the field of congenital heart surgery is particularly challenging. One reason for this is the wide heterogeneity of congenital heart defects and expected outcomes, in addition to variation across centers in case-mix or the number and type of patients treated. Recent efforts have led to several important advances in methodology to address these challenges, including the development empiric techniques to better account for differences in case-mix (3). Current methodology is inclusive of nearly all types of congenital heart operations performed (3). While this maximizes sample size, the resulting heterogeneity among included operations can also have limitations, and the use of more homogeneous and circumscribed target populations for performance assessment has been recommended (4–6). A better understanding of current congenital heart surgery case-mix across centers, and strengths and limitations of various approaches to defining the target population would aid ongoing efforts to optimize evaluation, reporting, and interpretation of performance in the field.

In the present study utilizing the STS Congenital Heart Surgery Database (STS-CHSD), we assessed the current spectrum of case-mix across North American centers, evaluated strengths and limitations of methodology inclusive of the more homogeneous subset of STS

benchmark operations vs. the broad spectrum of all cardiac operations, and described potential implications of our findings on performance assessment.

Material and Methods

Data Source

The STS-CHSD represents >90% of US centers performing congenital heart surgery (7). Standard peri-operative data are collected on all patients undergoing pediatric and congenital heart surgery at participating centers (8). This study was approved by the Duke University and University of Michigan institutional review boards and was not considered human subjects research in accordance with the Common Rule (45 CFR 46.102(f)).

Study Population

Patients undergoing any index cardiovascular operation (first operation of an admission) with or without cardiopulmonary bypass at North American centers participating in the STS-CHSD from 2010–2014 were included (n=119 centers; 118,142 operations). Infants <2.5 kg undergoing patent ductus arteriosus ligation (n= 6002) were excluded, leaving a final study population of n=112,140.

Data Collection and Outcomes

Patient characteristics and operative data were collected using standard STS definitions (8). The primary procedure code for the index operation was used to define the type of operation performed. Operations were also characterized by their Society of Thoracic Surgeons-European Association for Cardiothoracic Surgery (STAT) Mortality Category (9). The average annual surgical volume of STAT-classified cases during the study period was also collected. The primary outcome was operative mortality, defined in the STS-CHSD as any mortality occurring in-hospital or in any location within 30 days of surgery (8).

In the present study, one of our primary interests was comparing performance methodology inclusive of all eligible operations as described by the criteria above (the current standard for performance assessment in the field) to a more homogeneous subset. For the latter we chose the STS benchmark operations, as this cohort of operations is currently used for certain aspects of reporting by the STS on a national level, and by certain states (7,10). The benchmark operations span the spectrum of complexity and can be defined using standard criteria – they include a group of 10 types of operations comprised of 26 primary procedure codes, as displayed in Table 1 (e.g. Fontan operation includes the 6 codes for different types of Fontan) (7).

Analysis

Operative case-mix (number and type of operations performed) was described across centers using standard descriptive statistics, and included evaluation of all eligible cardiac operations captured in the database, the subset of benchmark operations, and STAT categories. The proportion of operations that were performed by specified percentages of centers was described, as well as the proportion of operations that accounted for various percentages of all cases and mortalities during the study period.

We then evaluated the impact on assessment of center performance of methodology that included the benchmark operations only vs. all operations, Center performance was assessed by the risk-adjusted operative mortality rate, as this is currently the most commonly utilized metric in the field. Standard STS statistical models were used, and as described previously, centers with >10% missing data for model variables were excluded (3). This left a study population of 95 centers (83,751 operations) for this portion of the analysis. Risk-adjusted mortality rates were calculated for each hospital, and models adjusted for key patient characteristics and operative case-mix using the STAT score for the index operation, as described previously (3). The only difference in the models was whether the benchmark operations only vs. all of a center's eligible operations were included. Based on the model outputs, centers were classified as having lower, higher or same-as-expected mortality if their 95% CI for risk-adjusted mortality fell entirely below, above, or overlapped the overall aggregate mortality rate for the study cohort, respectively (3). We evaluated the proportion of centers classified in these three performance groups when the benchmark operations vs. all eligible operations were included.

Finally, we performed a theoretical power evaluation using specified assumptions, and calculated the center-level sample size required for 80% power to detect a doubling in mortality rate (compared to the overall aggregate mortality rate in the sample) using a one-sample binomial test with $\alpha=0.05$. We determined the number of centers in the study cohort who met this sample size/volume threshold, both for all operations and the subset of benchmark operations. Analyses were performed using SAS Version 9.4 and R Version 3.2.1.

Results

Overall operative case-mix

Overall, 207 different types of operations were performed across the 119 included centers, comprising a total of 112,140 cases during the 5-year study period. Across the 207 types of operations, there was wide variability in the number of cases performed, ranging from 1–8,833 cases per operation. Many operations were performed infrequently with 92/207 (44%) performed <100 times in the 5-year study period. Overall, 19 of the 207 operations (9.2%) captured 50% of the total cases, 42 operations (20.3%) captured 75% of the total cases, and 74 operations (35.7%) captured 90% of the total cases. Data regarding the proportion of overall deaths accounted for by various operations are also reported in Table 2. Overall, approximately 1/3 of the operations (n~60–70 operations) captured the vast majority of both the total number of cases performed and the mortalities during the study period (Table 2). The rest of the operations were performed less frequently and/or were not commonly associated with mortality.

Case-mix across centers

On a center level, 95 (45.9%) of the 207 total operations were performed at least once during the 5-year study period by 50% or more of the centers, 52 operations (25.1%) were performed at least once by 75% or more of the centers, and 22 operations (10.6%) were performed at least once by 90% or more of the centers (Table 2). In other words, relatively

few of the 207 different types of operations were performed across all centers, and many operations were only performed at certain subsets of centers. Examples include heart transplant, which was performed at 60 centers (50.4%) during the 5-year study period, and the double switch operation for l-transposition of the great arteries, which was performed at 31 centers (26.1%).

Operations were also grouped into STAT categories, and STAT case-mix was described across centers (Table 3). The greatest variation was seen for the higher-complexity operations, where there was 7.9-fold variation across centers in the proportion of total cases comprised of STAT 5 cases (Table 3).

Benchmark operations

The benchmark operations comprised 36% of total cases (40,545/112,140) and accounted for 33.5% of the mortalities during the study period (Table 4). The benchmark operations were all performed by 90% of centers during the 5-year study period, with the exception of arterial switch operation + ventricular septal defect repair (89% of centers), and truncus arteriosus repair (82% of centers). However, as displayed in Table 4, while most centers performed these operations, the number of cases performed varied widely across centers.

Center performance as assessed by benchmark vs. all operations

The median risk-adjusted operative mortality rate for all operations was 3.5% (range 0%–13.4% across centers). For benchmark operations, the median risk adjusted operative mortality rate was 3.1% (range 0%–20.4% across centers). When centers were classified into performance categories (lower, higher, or same-as-expected mortality) based on models which included the benchmark operations only vs. all operations, 81/95 centers (85%) did not change performance classification (Table 5). The other 14 centers (15%) changed by one category; no center changed by two categories.

Of the 14 centers that changed classification, 10 (71%) moved to a higher performance (lower mortality) category. Centers that changed classification compared to those who did not had no significant difference in the proportion of their cases comprised of benchmark operations (39% vs. 35%, $p=0.2$), or proportion of STAT 4/5 cases (23% vs. 24%, $p=0.9$), but had significantly higher average annual surgical volume (median 293 vs. 168 cases/year, $p=0.04$). This volume relationship was present both for those centers who moved to a higher performance category and those who moved to a lower performance category (median 269 cases/year and 316 cases/year, respectively, vs. 168 cases/year in those who did not change).

Power

Methodology inclusive of the benchmark operations only vs. all operations was associated with lower power - 35% vs. 78% of centers respectively met the volume threshold over a 4-year period needed to detect a doubling of mortality (Table 6). Despite this, the overall proportion of centers classified as a “statistical outlier” (either higher or lower than expected mortality), was similar regardless of the methodology [$n=23$ centers (24%) when including all operations, and $n=21$ centers (22%) when including benchmark operations only].

Comment

The current practice of congenital heart surgery consists of a broad spectrum of operations, many of which are performed infrequently, and wide variation in case-mix across centers. Performance metrics based on benchmark vs. all operations are associated with less heterogeneity across centers, but also lower power, and these two approaches lead to differing characterization of performance for some centers. These results have important implications for evaluation, reporting, and interpretation of performance in congenital heart surgery.

Healthcare performance measurement is a complex undertaking, and involves several important considerations, including: the data source, choice of performance metrics, target population, sample size/power, adjustment for patient risk-factors and differences in procedural case-mix, statistical methodology, classification of performance and outliers, and interpretation of indirectly standardized outcomes (5,11). The present study focused primarily on the aspects of case-mix and target population. It is generally recommended that the target population should be as homogeneous as possible in order to minimize potential biases related to differences in case-mix across centers (4–6). This is because it can be challenging for any type of adjustment (e.g. model-based extrapolation) to reliably account for the lack of data in an area of non-overlap in case-mix between centers (5,11). For example, in adult cardiac surgery, the STS reports center performance separately by operation type, such as isolated coronary artery bypass grafting (CABG), isolated aortic valve replacement, combined CABG/aortic valve replacement, etc (4).

In congenital heart surgery, the choice of target population is more challenging due to disease heterogeneity and sample size issues. Jacobs et al. previously demonstrated that across the individual benchmark operations, power to evaluate between-center variability in mortality was low for all except the Norwood operation, limiting the feasibility of performance assessment based solely on individual operations (13). Further, as demonstrated in this study and others, there is substantial variability in operative case-mix across centers. Due to these two factors, it is likely necessary to report performance based on a target population that involves some aggregate of operations (whether all operations or some subset), and to continue to utilize statistical methodologies to attempt to account for differences in case-mix across centers as best as possible (3).

In the present study, we evaluated the strengths and weaknesses of methodology inclusive of all eligible congenital heart operations compared vs. the more homogeneous subset of benchmark operations. As expected, a strength of the benchmark approach is the increased homogeneity - all but 2 of the benchmark operations were performed by 90% of the centers. However, the benchmark approach was associated with lower power to detect differences between centers, due to the smaller sample size. However, we found that even with this lower power, the overall number of outliers detected using benchmark vs. all operations approach was similar, suggesting that the current magnitude of variation in outcomes across centers may offset some of the potential power issues. A second limitation is that even though overall the benchmark operation approach may allow more homogeneous assessments across centers since nearly all centers perform these operations, our analysis

demonstrated that there is still substantial variation across centers in the frequency with which individual benchmark operations are performed. Thus, differences in case-mix across centers persist even with this approach.

Overall we found that while the majority of centers remained in the same performance category regardless of which methodology was used, 15% of centers did change performance classification, emphasizing that these methodologic issues can have important implications for a subgroup of centers. The majority of centers that changed performance classification moved to a higher performance category when limiting to the more homogeneous subset of benchmark operations, and centers who changed classification had higher average annual volumes. These are similar to findings described previously in adult cardiac surgery, where it has been hypothesized that when heterogeneous case types are analyzed together, results from high volume tertiary centers performing a greater proportion of high-complexity cases may appear inferior to programs more frequently performing lower complexity cases (4,12). In addition, due to limited statistical power, true performance whether “good” or “bad” in general can be more difficult to discriminate for low volume centers - confidence intervals tend to be wide and overlapping with the line of unity regardless of the type of methodology used for case inclusion, hence these centers are less likely to change performance categories (14).

Examining the results of our study within the context of the key aspects of performance assessment, several implications are apparent (4–6). With regard to our findings regarding the target population, it is clear that there are strengths and limitations to both the approach of including all operations and the subset of benchmark operations. They may be viewed as complementary, and reported together providing different windows into performance, or utilized preferentially depending on the goals of various initiatives, with limitations acknowledged. For example, if the goal is to optimize power to identify outliers, methods inclusive of all operations may be favored. If the goal is to enable the most homogeneous comparisons across sites, methods inclusive of the benchmark operations may be favored.

In addition, our findings regarding case-mix have implications for the reporting and interpretation of performance metrics in the field. As previously described,, methodology known as indirect standardization is most often used in calculating “risk-adjusted” healthcare performance metrics, and is currently utilized for public reporting involving the STS-CHSD (5,6,11). This allows assessment of a center’s observed outcomes in relation to what would be expected if patients with a similar case-mix to theirs had been cared for at an “average” center in the reference population. Because indirectly standardized outcomes are estimated only for the patients a center actually treated, results only apply to their particular case-mix. The results derived using indirect standardization cannot be used to directly compare two hospitals unless their case-mix has been demonstrated to be similar, and it cannot be assumed that a center achieving better than average results in a generally low-risk population could do the same in a population of higher-risk patients (5,11). Given our findings demonstrating wide variation in case-mix across centers, it is particularly important to understand and acknowledge these nuances when reporting and interpreting congenital heart surgery performance metrics. The development of additional tools to facilitate

understanding and reporting of centers' case-mix may be useful in supporting the most appropriate interpretation of reported performance metrics.

Finally, in order to address the power issues identified in this study and others, adding other types of common operations to the cohort of benchmark operations for performance analysis could be considered. An empiric approach could be used to balance competing goals of maximizing sample size/power while preserving homogeneity across centers. In addition, the use of composite performance metrics, which combine information across several performance domains (e.g. both morbidity and mortality) and increase the effective event rate, may also aid in increasing power (5,14).

Limitations

We compared methodology inclusive of the subset of benchmark operations vs. all eligible cardiac operations (the current standard for performance assessment in the field). Because “performance” and “quality” are abstract concepts without concrete definitions, there is no “gold standard” for comparison, as is the case in all analyses of this nature. This study focused primarily on aspects of the target population and case-mix; further study of other areas related to performance assessment and continued refinement of methods related to statistical modeling and adjustment for different patient factors may also prove useful. Further efforts are also needed to move beyond the early post-operative period and to develop performance metrics based on longer-term outcomes.

Conclusions

Performance assessment in congenital heart surgery is a complex undertaking and further efforts are needed to optimize methodology and address the challenges identified in this study and others. This may include the development of methodology for reporting performance metrics in a more homogeneous subset of operations, ongoing work to develop a composite performance metric in the field (which among other goals will aid in addressing power issues), and efforts to support the appropriate interpretation of risk-adjusted performance metrics across centers.

Acknowledgments

Funded in part by the National Heart, Lung, and Blood Institute (R01HL122261, PI: Pasquali). Dr. Pasquali receives support from the Janette Ferrantino Professorship.

References

1. Pasquali SK, Dimick JB, Ohye RG. Time for a more unified approach to pediatric healthcare policy? The case of congenital heart care. *JAMA*. 2015; 314:1689–1690. [PubMed: 26505593]
2. Dimick JB, Nicholas LH, Ryan AM, et al. Bariatric surgery complications before vs after implementation of a national policy restricting coverage to centers of excellence. *JAMA*. 2013; 309:792–799. [PubMed: 23443442]
3. O'Brien SM, Jacobs JP, Pasquali SK, et al. The STS Congenital Heart Surgery Database Mortality Risk Model: Part 1—Statistical Methodology. *Ann Thorac Surg*. 2015; 100:1054–1062. [PubMed: 26245502]
4. Shahian DM, He X, Jacobs JP, et al. Issues in quality measurement: Target population, risk adjustment, and ratings. *Ann Thorac Surg*. 2013; 96:718–726. [PubMed: 23816415]

5. Shahian DM, Normand S-LT. What is a performance outlier? *BMJ Qual Saf.* 2015; 24:95–99.
6. [Accessed 12/28/2015] Statistical issues in assessing hospital performance. Available at:<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf>
7. Jacobs JP, Shahian DM, Prager RL, et al. Introduction to the STS National Database Series: Outcomes Analysis, Quality Improvement, and Patient Safety. *Ann Thorac Surg.* 2015; 100:1992–2000. [PubMed: 26525868]
8. STS Database Specifications. [Accessed 12/28/2015] Available at:http://www.sts.org/sites/default/files/documents/pdf/CongenitalDataSpecificationsV3_0_20090904.pdf.
9. O'Brien SM, Clarke DR, Jacobs JP, et al. An empirically based tool for analyzing mortality associated with congenital heart surgery. *J Thorac Cardiovasc Surg.* 2009; 138:1139–1153. [PubMed: 19837218]
10. Pennsylvania Health Care Cost Containment Council. [Accessed 1/16/2016] Available at:<http://www.phc4.org/reports/cabg/pediatric/12/>.
11. Shahian DM, Normand S-LT. Comparison of “risk-adjusted” hospital outcomes. *Circulation.* 2008; 117:1955–1963. [PubMed: 18391106]
12. Shahian DM, Silverstein T, Lovett AF, et al. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation.* 2007; 115:1518–1527. [PubMed: 17353447]
13. Jacobs JP, O'Brien SM, Pasquali SK, et al. Variation in Outcomes for Benchmark Operations. *Ann Thorac Surg.* 2011; 92:2184–2192. [PubMed: 22115229]
14. Shahian MD, Normand S-LT. Low-volume coronary artery bypass surgery: Measuring and optimizing performance. *J Thorac Cardiovasc Surg.* 2008; 135:1202–1209. [PubMed: 18544354]

Table 1

Benchmark operations

Benchmark Operation	STS-CHSD Procedure Codes
VSD	110=VSD repair, Patch
TOF	350=TOF repair, No ventriculotomy
	360=TOF repair, Ventriculotomy, Nontransanular patch
	370=TOF repair, Ventriculotomy, Transanular patch
AVC	170=AVC repair, Complete
ASO	1110=ASO
ASO VSD	1120=ASO and VSD repair
Glenn/HemiFontan	1670=Bidirectional cavopulmonary anastomosis (bidirectional Glenn)
	1680=Glenn (unidirectional cavopulmonary anastomosis) (unidirectional Glenn)
	1690=Bilateral bidirectional cavopulmonary anastomosis (bilateral bidirectional Glenn)
	1700=HemiFontan
	2130=Superior Cavopulmonary anastomosis(es) + PA reconstruction
Fontan	970=Fontan, TCPC, Lateral tunnel, Fenestrated
	980=Fontan, TCPC, Lateral tunnel, Nonfenestrated
	1000=Fontan, TCPC, External conduit, Fenestrated
	1010=Fontan, TCPC, External conduit, Nonfenestrated
	2780=Fontan, TCPC, Intra/extracardiac conduit, Fenestrated
	2790=Fontan, TCPC, Intra/extracardiac conduit, Nonfenestrated
Truncus	230=Truncus arteriosus repair
Norwood	870=Norwood procedure
Off Bypass Coarctation ¹	1210=Coarctation repair, End to end
	1220=Coarctation repair, End to end, Extended
	1230=Coarctation repair, Subclavian flap
	1240=Coarctation repair, Patch aortoplasty
	1250=Coarctation repair, Interposition graft
	1280=Aortic arch repair

VSD=ventricular septal defect, TOF=Tetralogy of Fallot, AVC=atrioventricular canal, ASO=arterial switch operation, PA=pulmonary artery, TCPC=total cavopulmonary connection

¹Includes procedures with OpType=No CPB (cardiopulmonary bypass) Cardiovascular

Table 2

Number of operations capturing specified proportions of centers, total cases, and mortalities

N (%) of 207 total operations performed by:		
50% of centers 95 (45.9%)	75% of centers 52 (25.1%)	90% of centers 22 (10.6%)
N (%) of 207 total operations capturing:		
50% of all cases 19 (9.2%)	75% of all cases 42 (20.3%)	90% of all cases 74 (35.7%)
N (%) of 207 total operations capturing:		
50% of all mortalities 12 (5.8%)	75% of all mortalities 33 (15.9%)	90% of all mortalities 63 (30.4%)

For reference, all but two of the benchmark operations are performed by 90% of centers, capture 36.2% of all cases, and 33.5% of mortalities

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Variation in STAT category case-mix across centers

STAT Category	Center-level percentiles for proportion of total case volume comprised of specified STAT categories					Magnitude of variation across centers
	10 th %	25 th %	50 th % (median)	75 th %	90 th %	Ratio 90 th %/10 th %
1	24.9%	29.0%	32.0%	36.8%	41.2%	1.7
2	26.1%	29.2%	33.1%	36.0%	38.8%	1.5
3	7.4%	9.0%	11.3%	12.9%	14.6%	2.0
4	14.1%	16.4%	19.2%	22.1%	25.7%	1.8
5	0.8%	2.8%	4.0%	5.2%	6.2%	7.9

The proportion of a center's total case volume comprised of various STAT categories and distribution across centers is displayed. The magnitude of variation across centers is described by the ratio of the 90th/10th percentile (excluding extreme outliers). For example, the proportion of total cases comprised of STAT 5 cases ranged from 0.8% (10th percentile) to 6.2% (90th percentile), or 7.9-fold variation across centers.

Benchmark operations

Table 4

Benchmark Operation	N (%) of total 112,140 cases (2010–2014)	Mortality Rate N (%)	N (%) of centers performing 1 of the specified operations (2010–2014)	Annual volume of specified operations across included centers (2014)		
				Min	Median	Max
VSD	8,633 (7.7%)	63 (0.7%)	119 (100%)	1	14	70
TOF	5,555 (5.0%)	58 (1.0%)	116 (98%)	1	8	46
Coarctation	4,893 (4.4%)	55 (1.1%)	115 (97%)	1	8	38
Fontan	5,180 (4.6%)	73 (1.4%)	113 (95%)	1	8	53
Glenn/HemiFontan	4,929 (4.4%)	108 (2.2%)	116 (98%)	1	7	41
ASO	2,282 (2.0%)	59 (2.6%)	110 (92%)	1	4	20
AVC	3,799 (3.4%)	117 (3.1%)	118 (99%)	1	6	38
ASO + VSD	1,003 (0.9%)	49 (4.9%)	106 (89%)	1	2	12
Truncus	743 (0.7%)	70 (9.4%)	98 (82%)	1	2	9
Norwood	3,528 (3.1%)	553(15.7%)	107 (90%)	1	6	28

Table 5

Change in performance categories: benchmark vs. all operations

		Benchmark Operations		
		Performance Category		
		Lower Than Expected	Same as Expected	Higher Than Expected
All Operations	Lower Than Expected	10	5	0
	Same as Expected	1	66	5
	Higher Than Expected	0	3	5

Cells depict the number of centers in each category when characterizing performance based on benchmark only vs. all operations. Those that fall along the diagonal (shaded cells) represent centers classified in the same category with either method. Overall, 81/95 (85%) centers did not change performance classification. Fourteen centers (15%) changed by one category; no centers changed by two categories.

Table 6

Power Calculation

	Overall unadjusted mortality rate	Center-level volume (N) needed to detect a doubling of mortality rate	N(%) of centers meeting volume threshold (1 year of data)	N (%) of centers meeting volume threshold (4 years of data)
All Operations	3.3%	256	23 (24%)	74 (78%)
Benchmark Operations	3.0%	285	0 (0%)	33 (35%)

See Methods for details

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript