Lesion-specific outcomes in neonates undergoing congenital heart surgery are related predominantly to patient and management factors rather than institution or surgeon experience: A Congenital Heart Surgeons Society Study

Tara Karamlou, MD,[a] Brian W. McCrindle, MD, MPH,[b] Eugene H. Blackstone, MD,[c] Sally Cai, MSc,[a] Richard A. Jonas, MD,[d] Scott M. Bradley, MD,[e] David A. Ashburn, MD, MSc,[f] Christopher A. Caldarone, MD,[a] and William G. Williams, MD[a]

Objective: To identify the role of institution and surgeon factors, including case volume and experience, on survival of neonates with complex congenital heart disease.

Methods: A total of 2421 neonates from 4 groups—transposition of the great arteries (n = 829), pulmonary atresia with intact ventricular septum (n = 408), Norwood (n = 710), and interrupted aortic arch (n = 474)—were prospectively enrolled from Congenital Heart Surgeons Society institutions. Multivariable analysis of risk-adjusted survival was performed for each group, entering each institution or surgeon into the multivariable analysis separately. Institutional performance was defined as [predicted survival – actual survival]. Neutralization of risk factors within each institution was evaluated using complex interaction terms. Institution and surgeon experience, defined by 5 domains (total case volume, total time each operation was performed, cases per year, rank-order of cases, case velocity), were also investigated.

Results: Institutional performance varied among all groups. Improved outcomes in Norwood and pulmonary atresia with intact ventricular septum were unrelated to any ''experience'' domains, whereas improved outcomes in transposition of the great arteries were significantly related to increased experience in most domains. No institution enrolling in all 4 studies ranked number 1 in performance for all groups. Neutralization of low birth weight as a risk factor contributed to decreased mortality after Norwood in one institution.

Conclusion: Survival of neonates with complex congenital heart disease is influenced more by patient and management factors than by institution or surgeon experience. Institutional excellence in managing some diagnostic groups does not indicate similar performance for all diagnostic groups. Weighted risk-adjusted comparisons could provide a mechanism to improve results in institutions with less than optimal outcomes. (J Thorac Cardiovasc Surg 2010;139:569-77)

Earn CME credits at
http://cme.ctsnetjournals.org

⌐ Supplemental material is available online.

There is considerable interest in establishing criteria for comparing surgical outcomes for congenital heart disease (CHD) among institutions and surgeons. Evidence from other surgical specialties, and recently within the field of CHD surgery, has suggested that higher case volume and increasing institution or surgeon ''experience'' may confer an outcome benefit.[1-10] However, the data are conflicting and often obtained from statistical models that are risk stratified for groups of patients with similar characteristics rather than risk adjusted for individual patients.[11] Inadequate risk adjustment is partly due to the source datasets from which statistical models are constructed (ie, registry or administrative databases that by definition are minimum datasets with variables identified by consensus rather than comprehensive academic datasets with data-derived risk factors[12]), and partly from a lack of consensus regarding the end point for comparison.[11,13-17]

Inferences regarding the importance of institution and surgeon experience on patient outcomes are therefore hampered by inadequate power, confounding influences of era and cumulative experience, shifts in the composition of surgical teams, differing case-mix among institutions, current low

Abbreviations and Acronyms
CHD = congenital heart disease
CHSS = Congenital Heart Surgeons Society
IAA = interrupted aortic arch
PAIVS = pulmonary atresia with intact ventricular
septum
TGA = transposition of the great arteries

operative risk, and the differential importance of patient and technical factors among diagnostic groups.[11,14,15,18-20]

We sought to use comprehensive diagnosis-based datasets (with the exception of the Norwood) derived from Congenital Heart Surgeons Society (CHSS) multi-institutional studies to overcome these problems and create risk-adjusted models to determine the role of institutional and surgeon experience on outcomes for neonates with complex CHD.

MATERIALS AND METHODS
Patients and Institutions
A total of 2421 neonates from 4 groups (transposition of the great arteries [TGA], Norwood, interrupted aortic arch [IAA], and pulmonary atresia with intact ventricular septum [PAIVS]) were prospectively enrolled from 24 to 33 CHSS institutions. Outcomes for all of these neonates, including the specific patient and procedural characteristics, have been reported.[15,21-24] Participation in the study and submission of patient information was voluntary and confidential. Parental consent was obtained consistent with individual institutional policies. Ethics approval for the CHSS Data Center is obtained annually from the Hospital for Sick Children in Toronto. Enrollment period, institution, and surgeon characteristics for each group are shown in Table 1.

Data Analysis
Data are presented as frequency, median with range, or mean ± standard deviation as appropriate, with the number of missing values indicated. Percentages, hazard functions, and competing risk estimates are presented with confidence limits equivalent to 1 standard error (68%). All data analyses were performed using SAS statistical software (version 9; SAS Institute Inc, Cary, NC). Multiphase parametric modeling of the hazard function[16,17] and competing risks methodology[22,25-29] were used to determine rates of transition to mutually exclusive time-related events and identify incremental risk factors associated with each transition rate as previously described.[16,17,22,25-29] Transformations to linearize variables and to optimize calibration to risk for continuous predictors were considered in all multivariable analyses. Variable selection was primarily by bootstrap bagging, with those variables having greater than 50% reliability included in the final models.[17,22]

Statistical Models
Overall risk-adjusted parametric mortality models were constructed on the basis of the total experience of all institutions, accounting for patient, demographic, and procedural factors that influence mortality for the entire diagnostic group. The overall mortality models for each group were specified a priori, meaning that we used the previously reported overall mortality model for each diagnostic group[15,21-24] (Table E1). However, the importance of risk factors affecting overall outcomes may or may not apply equally to each individual institution. We addressed this problem using the following methods to account for potential variation within each institution: (1) Each institution was forced into the overall model for each diagnostic group (ie, to determine whether that particular institution conferred an increased or decreased risk);

(2) interaction terms were then created with each risk factor within the overall model to determine whether improved or reduced institutional performance could be explained by neutralization of known risk factors for the entire cohort; (3) institutional ''experience'' was then incorporated as an explanatory variable to define whether these factors improved performance discrimination; and (4) actual institutional performance at an individual institution was compared with the expected risk-adjusted mortality at that institution on the basis of all identified risk factors. Institutional ''experience'' was evaluated using the following 5 separate domains:

(1) total case volume (the total number of patients enrolled at each institution during the entire study enrollment period);
(2) total time period (years) in which each institution was performing a particular procedure (the difference in years between the date of the last procedure and the date of the first procedure during the study enrollment period; for example, if institution ''X'' enrolled their first patient with an arterial switch operation in 1985 and enrolled their last patient with an arterial switch in 1989, this value would be 4 years);
(3) cases per year (the total number of patients enrolled by each institution divided by the number of years in which that institution enrolled patients in each study);
(4) rank order of cases (the first case through the last case in consecutive order); and
(5) case velocity over time (the rank order of each case divided by the time interval (number 2 above) in which a particular procedure was performed).

Surgeon ''experience'' was likewise evaluated using the same 5 domains but was considered only in Norwood and TGA, the only 2 groups with available surgeon-level data. The effect of era was considered by incorporating the date at which each institution or surgeon enrolled their first patient undergoing a particular procedure as an explanatory variable.

RESULTS
Impact of Institutional Experience Factors on Mortality
Impact of institution ''experience'' factors on risk-adjusted overall 5-year mortality after admission is shown in Table 2. In complex procedures, such as the Norwood, there was no apparent survival benefit conferred by increased institution ''experience'' in any of the 5 domains investigated (Figure 1). Improved results in management of neonates with PAIVS were also not significantly related to increased institution experience (Table 2).

However, in TGA, improved 5-year survival is significantly associated with increasing institutional experience in 4 of the 5 domains investigated (Table 2). Lower annual TGA case volume, with a threshold value of less than 50 cases per year, resulted in significantly increased mortality (Figure 2, A). The improvement in mortality, though, was due to improved outcomes for the arterial switch (n = 516) rather than for the atrial switch (n = 276), in which the risk did not change significantly (Figure 2, B). For the arterial switch, increased case velocity over time was associated with decreased mortality (parameter estimate: −0.06 ± 0.13 per case; P < .001). Similarly, mortality was inversely related to increased total procedure time that each institution had with the arterial switch (parameter estimate: −0.24 ± 0.10 per year; P = .01). However, for the atrial switch, there was no significant survival benefit conferred by any of the experience domains considered.

Institutional Performance
There was considerable variation in institutional performance among the different lesion sets (Figure 3, A–D). No

TABLE 1. Institution and surgeon characteristics for each diagnostic group

Variable	Value
I. Norwood	
Enrollment period	1994–2000
Total no. of patients	710
No. of institutions	29
Total cases (median and range)	6 (1–155)
Cases per year (median and range)	14 (1–40)
Total No. of surgeons	56
No. of surgeons per institution (median and range)	1 (1–10)
Total cases	5 (1–130)
Cases per year (median and range)	3 (1–36)
II. TGA	
Enrollment period	1985–1989
Total no. of patients	829
No. of institutions	24
Total cases (median and range)	53 (1–169)
Arterial switch (median and range)	52 (1–146)
Atrial switch (median and range)	21 (1–31)
Cases per year (median and range)	12 (1–60)
Arterial switch per year (median and range)	17 (1–52)
Atrial switch per year (median and range)	5 (1–15)
No. of surgeons	50
No. of surgeons per institution (median and range)	1 (1–10)
Total cases per surgeon (median and range)	38 (1–87)
Cases per year per surgeon (median and range)	9 (1–28)
Arterial switch per year per surgeon	11 (1–28)
Atrial switch per year per surgeon	3 (1–12)
III. IAA	
Enrollment period	1987–1997
Total no. of patients	474
No. of institutions	33
Total cases (median and range)	9 (1–57)
Cases per year (median and range)	3 (1–10)
IV PAIVS	
Enrollment period	1987–1997
Total no. of patients	408
No. of institutions	33
Total cases (median and range)	17 (1–41)
Cases per year (median and range)	3 (1–9)

IAA, Interrupted aortic arch; *TGA,* transposition of the great arteries; *PAIVS,* pulmonary atresia with intact ventricular septum.

TABLE 2. Impact of institution "experience" factors on risk-adjusted mortality

Factor	Parameter estimate ± SE	P
Increased total case volume*		
Norwood (per case)	−0.002 ± 0.001	.17
IAA (per case)	−0.02 ± 0.01	<.001
PAIVS (per case)	0.02 ± 0.01	.07
TGA (per case)	−0.02 ± 0.01	<.001
Increased cases per year†		
Norwood (per case)	−0.005 ± 0.01	.38
IAA (per case)	−0.03 ± 0.03	.37
PAIVS (per case)	0.05 ± 0.04	.24
TGA (per case)	−0.07 ± 0.02	<.001
Increased rank order‡		
Norwood (per case)	−0.002 ± 0.002	.38
IAA (per case)	−0.02 ± 0.01	.06
PAIVS (per case)	0.02 ± 0.01	.16
TGA (per case)	−0.01 ± 0.003	<.001
Increased case velocity§		
Norwood (per case)	−0.002 ± 0.004	.15
IAA (per case)	0.01 ± 0.01	.39
PAIVS (per case)	0.001 ± 0.002	.76
TGA (per case)	0.002 ± 0.002	.09
Increased total procedure time‖		
Norwood (per year)	−0.04 ± 0.03	0.17
IAA (per year)	−0.04 ± 0.05	0.36
PAIVS (per year)	−0.07 ± 0.04	0.05
TGA (per year)	−0.24 ± 0.08	0.002

IAA, Interrupted aortic arch; *TGA,* transposition of the great arteries; *PAIVS,* pulmonary atresia with intact ventricular septum. *Total number of patients enrolled at each institution during the entire study enrollment period. †Total number of patients enrolled by each institution divided by the number of years in which that institution enrolled patients in each study. ‡First case through the last case in consecutive order. §Rank order of each case divided by the time interval in which a particular procedure was performed. ‖Difference in years between the date of the last procedure and the date of the first procedure during the study enrollment period.

rank order of cases, indicating a potential learning curve (Figure 5).

Conversely, there was no influence of surgeon "experience" on outcomes after the Norwood procedure (Table 3). Risk-adjusted 5-year mortality was not significantly improved despite increasing surgeon annual case volume (Figure 6).

Neutralization of Risk: Management Paradigms as Reasons for Improved Outcomes

Institution-specific models using the interaction terms revealed one potential reason for improved institutional performance in the Norwood series. Risk factors for mortality identified from the overall model of all 710 neonates included the patient-specific factor lower birth weight.[15] However, one institution that enrolled 43 patients (total volume range for all institutions: 3–155) achieved the best performance because birth weight had little impact on survival within this institution. The favorable results in this institution may derive not from volume, but from their

clear relationship between volume and performance was evident, but wide overlapping confidence intervals indicate poor discrimination except between those institutions performing at both extremes. Institutional excellence in managing some groups of neonates did not translate into equally superior performance in their management of other groups (Figure 4).

Impact of Surgeon Factors and Learning Curve Dynamics

Increasing surgeon "experience" was related to improved survival for the TGA cohort (Table 3). Outcomes for the arterial switch improved significantly with increased
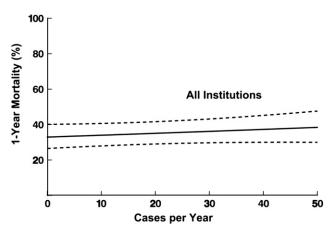
FIGURE 1. Risk-adjusted nomogram of 1-year mortality risk after the Norwood procedure (on the vertical axis) versus annual volume of cases (on the horizontal axis) with all other predictors from the overall model excluding institutions (14) set to the median value. Mortality for all institutions remains constant over the entire range of annual volume, indicating no survival benefit with higher annual case volume. Note, however, that the overall mortality during this era was approximately 40%. The continuous parametric point estimates (*solid lines*) are enclosed by 70% confidence intervals (*dashed lines*).

management of a patient characteristic, ''low birth weight,'' which was a risk factor for all other institutions, but has been neutralized in this institution (Figure 7).

Institutional protocols rather than volume or technical expertise per se are also implicated as the underlying reason for improved outcomes in the PAIVS cohort. The institution with the best results in PAIVS enrolled only 20 patients but used a balanced management protocol based on morphology. The morphologically driven protocol allowed this institution to achieve comparably excellent survival, with a high prevalence of biventricular repair and Fontan operation, and few patients remaining unrepaired at age 5 years (Figure 8).

DISCUSSION

Defining the impact of institution and surgeon experience on outcomes in neonatal CHD is crucial in the current climate of extensive public scrutiny.[30,31] The potential that parallel initiatives similar to those spearheaded by the Leapfrog Group[32] may be advocated in pediatric cardiac surgery further highlight the importance of the establishment of a precise and valid measure of performance. We report results from the first study to use an academic database with a comprehensive variable set to examine the influence of institution and surgeon experience on outcomes for neonates beyond hospital discharge. We have shown that risk adjustment with data-derived risk factors to ''level the playing field'' provides an accurate method to compare and explain institutional performance. Institution and surgeon experience are not the primary determinants of survival, but experience is more important in some diagnostic groups than others.
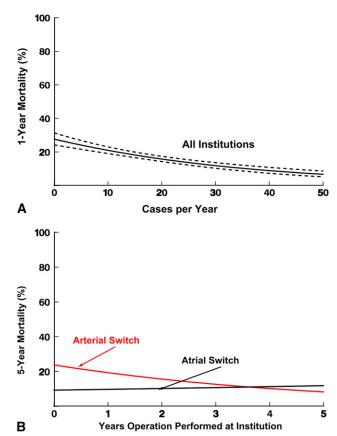

FIGURE 2. A, For the TGA cohort, there was a significant decrease in mortality with increasing values for the ''experience'' domains we investigated. Risk-adjusted nomogram of 1-year mortality risk versus annual volume of cases for the TGA cohort derived from the overall model excluding institutions (25) shows a significant association between decreased mortality and increasing annual case volume. Less then 50 TGA cases per year is associated with a rising mortality risk, indicating a potential threshold value. Continuous parametric point estimates (*solid line*) enclosed by 70% confidence intervals (*dashed lines*). B, Stratifying the nomogram by procedure type, either arterial switch (n = 516) or atrial switch (n = 276) demonstrates that the relationship between increased experience and better outcome is due to improved results with the arterial switch.

Previous Models for Risk Stratification

Recent methods of risk stratification in CHD surgery, including the RACHS-1 and Aristotle Basic and Comprehensive scores, have been proposed as useful methods for comparing inter-institutional outcomes.[8,9] However, although these scores may be statistically useful when applied to a large group of patients with shared characteristics, they lack precision in estimating risk for individual patients.[7,33] Furthermore, candidate risk factors used in these systems have been consensus based rather than data derived.[11,33] Other problems with both the Aristotle and RACHS systems, including procedure ''groups'' that are not accommodated by the scoring systems, ambiguous or redundant nomenclature, and insufficient variables to permit
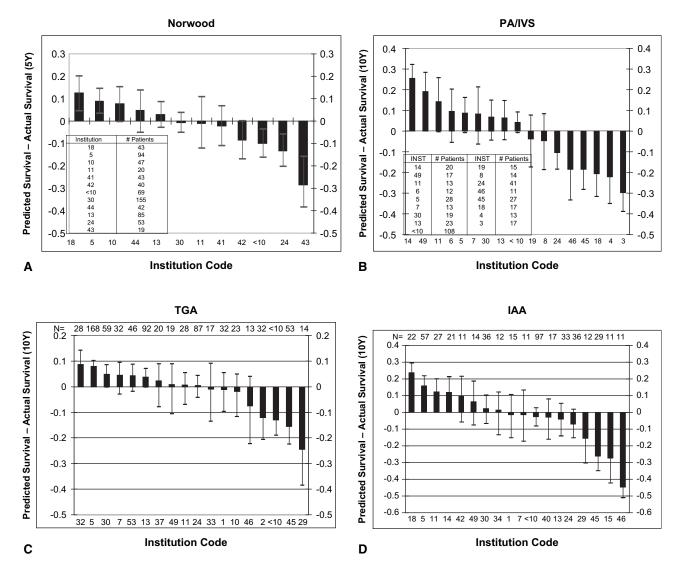
FIGURE 3. A, Institutional performance in the Norwood operation is shown as the mathematic difference between 5-year predicted survival after risk adjustment and actual 5-year survival (14). Institutions above the zero line are performing better than expected, whereas those below the zero line are performing worse than expected. Institution code is listed on the bottom, and the number of patients enrolled is listed in the enclosed box. Institutions enrolling less than 10 patients are grouped together as less than 10. There is no relationship between increasing patient volume (surrogated by patient enrollment) and improved outcomes. For example, institution 18, enrolling just 43 patients, performs better than institution 30 with 155 patients. B, Institutional performance for the PAIVS cohort. Note that performance in this diagnostic group also is independent of total patient volume. Those institutions that performed the best in the Norwood series (ie, institution 18) did not perform as well in this diagnostic group. C, Institutional performance for the TGA cohort measured at 15 years from admission to hospital. The number of patients enrolled per institution is listed at the top horizontal, and the institution code is listed at the bottom horizontal. Note that the highest volume institution is performing at the same level as an institution enrolling one eighth of the volume, but also notice that for this diagnostic group, those institutions with higher volume generally have performance that exceeds expected levels. D, Institutional performance for the IAA cohort. Note the overlapping confidence intervals. *PAIVS,* Pulmonary atresia with intact ventricular septum; *TGA,* transposition of the great arteries; *IAA,* interrupted aortic arch.

complete risk factor adjustment, have hampered widespread acceptance of either system.[3,11,14,33]

Previous studies evaluating outcomes have used ''early death'' in an attempt to simplify and standardize analysis. However, ''early'' is arbitrary, being variously defined as operative mortality (death within 30 days of operation and before hospital discharge), in-hospital mortality (death occurring after operation and during the same hospital admission as that operation), or 30-day mortality (death occurring within 30 days of operation regardless of hospital status).[12-14] None of the former recognize that the risk of death after operation changes over time. Typically the risk is high at the time of operation but decreases rapidly to a lower risk that is constant but not zero, and much later, the risk
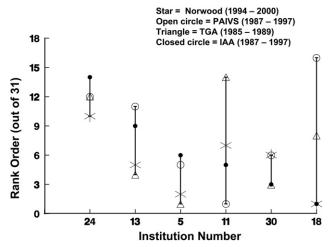
FIGURE 4. There were 6 CHSS institutions that contributed patients to all 4 diagnostic groups and had more than 10 patients within each group. These institutions were ranked out of a total number of 31 institutions based on the magnitude of difference between the predicted mortality at that institution after risk adjustment and the actual mortality at that institution. Institutions with the largest differential were ranked from lowest to highest, with a lower number indicating superior performance. The institution number is listed along the horizontal axis. Each geometric shape represents the individual institution's performance within a particular diagnostic group. Institutional excellence in managing certain diagnostic groups does not translate into excellence in managing all diagnostic groups. Note that institution 18 has excellent performance in the Norwood and IAA groups, but has the worst performance in the PAIVS group. *PAIVS,* Pulmonary atresia with intact ventricular septum; *TGA,* transposition of the great arteries; *IAA,* interrupted aortic arch.

increases in a third phase. Risk factors that influence outcome, either positively or negatively, in one phase may or may not affect outcome in any of the other phases. The early hazard phase of risk is unaffected by either the first

TABLE 3. Impact of surgeon "experience" factors on mortality

Factor	Parameter estimate ± SE	P
Norwood		
Increased total case volume	−0.001 ± 0.001	.40
Increased cases per year	−0.004 ± 0.007	.49
Increased rank order	−0.002 ± 0.002	.91
Increased case velocity	−0.006 ± 0.001	.70
Increased procedure time	−0.04 ± 0.03	.26
TGA		
Increased total case volume (per case)	−0.01 ± 0.004	.002
Increased cases per year (per case)	−0.04 ± 0.01	.002
Increased rank order (per case)	−0.01 ± 0.06	.009
Increased case velocity (per case)	0.004 ± 0.002	.09
Increased procedure time (per year)	−0.13 ± 0.08	.10
Arterial switch		
Increased rank order (per case)	−0.04 ± 0.01	.001
Atrial switch		
Increased rank order (per case)	−0.03 ± 0.03	.41

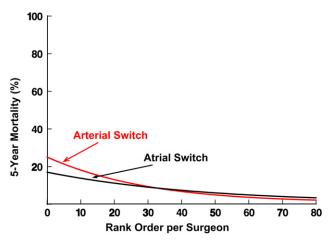*SE,* Standard error; *TGA,* transposition of the great arteries.

FIGURE 5. There was a significant positive association between increased surgeon experience and improved outcomes for the TGA cohort that was specific for the arterial switch operation. This nomogram shows a decline in risk-adjusted 5-year mortality as a function of increasing case rank-order per surgeon stratified by procedure type. Although mortality declines for both procedures, there was a significant decrease in mortality only for the arterial switch.

30-day period after operation or by hospital discharge. The CHSS report of outcomes in 710 neonates after the Norwood operation, for example, documented an early phase mortality that persisted beyond both hospital discharge and the 30-day definition.[15] Other CHSS studies, using techniques described by Blackstone and others,[15-24] used parametric hazard function analysis to define both the phases of risk and those factors that influence risk during a given phase of follow-up.

Impact of Institutional and Surgeon Factors on Outcome

We have shown that survival of neonates with complex CHD is influenced more by patient and management factors than institution or surgeon experience. Despite similar complexity, however, experience is clearly more important in some lesions and procedures (the arterial switch for TGA) than others (the Norwood procedure). In agreement with our finding, Checchia and colleagues[1] also found that outcomes after the Norwood operation were independent of surgeon volume. What reasons underlie this disparate influence? First, outcomes may be more dependent on experience for the arterial switch because operative technique is the defining factor for the arterial switch operation, whereas for the Norwood operation, the technical issues are overshadowed by the comparatively more complicated preoperative and postoperative care. A patient who has had a technically successful arterial switch operation, therefore, does well with little requirement for critical postoperative care. In contrast, an arterial switch operation that is technically compromised (eg, coronary stenosis) results in a high risk of death irrespective of postoperative care. Second,
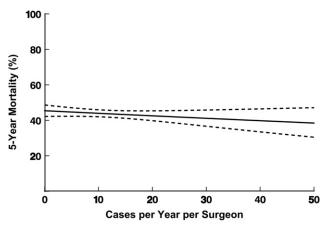
FIGURE 6. Outcomes after the Norwood operation were not associated with increased surgeon experience. Risk-adjusted nomogram shows that 5-year mortality after the Norwood operation is independent of the number of cases done per year (annual case volume) per surgeon. Note the widening confidence intervals once the annual volume reaches more than 30.

FIGURE 8. Risk-adjusted competing-risks nomogram adapted from Ashburn and colleagues[22] for 1 institution on the basis of the morphologic spectrum of PAIVS. This institution is one of two whose adoption of a morphologically driven protocol allowed them to achieve comparably excellent survival, with a higher prevalence of 2-ventricle and Fontan repair, with few patients remaining unrepaired at 5 years.

the relatively narrow morphologic spectrum of TGA compared with the wider morphologic spectrum of hypoplastic left heart syndrome is also likely to increase the relative importance of experience. Third, the TGA cohort circumscribed an earlier time period (1980s vs late 1990s for the Norwood series) that included the introduction of a new technique. It is probable that the strong influence of increased experience in TGA outcomes is due to capturing a learning curve for the arterial switch, which was missed in the Norwood cohort.
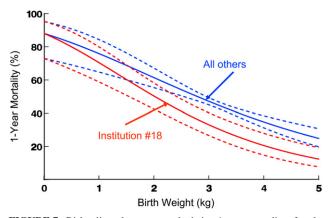

FIGURE 7. Risk-adjusted nomogram depicting 1-year mortality after the Norwood operation as a function of increasing birth weight stratified by whether a neonate was treated in institution 18 (the institution with the best results) compared with all others. A 2.5-kg neonate undergoing the Norwood operation in institution 18 fares significantly better than the same neonate treated at any other institution. Favorable results in this particular institution may derive from their management of a patient characteristic, low birth weight, which was a risk factor for all other institutions, but has been neutralized by institution 18. Continuous parametric point estimates (*solid lines*) enclosed by 70% confidence limits (*dashed lines*).
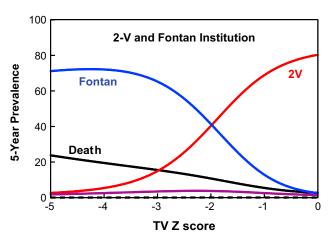
Concordant with these findings, Birkmeyer and colleagues[2] found that the importance of hospital volume varied markedly according to procedure type. They found that hospital mortality for pancreatectomy and esophagectomy was highly correlated with increasing caseload, whereas volume had little impact on mortality after coronary artery bypass grafting or carotid endarterectomy.

Our results are in contrast with those reported by others who have examined the impact of volume on in-hospital or 30-day mortality,[1-6,10,32,34-36] but they confirm the findings of other recently published studies using large administrative databases.[36-38] Jenkins and colleagues[4] reported significantly decreased in-hospital mortality among those centers performing greater than 300 cases annually in their analysis of 2833 congenital heart operations performed at 37 centers. Hannan and colleagues[3] reached similar conclusions whereby the odds ratio for hospital death was 1.42 in low-volume institutions (annual volume <100 cases) when compared with high-volume centers (annual volume ≥100 cases). They also found that increased annual surgeon caseload (>75) significantly improved survival to hospital discharge. Similarly, Allen and colleagues[35] found that evidenced-based referrals from a small volume (<50 cases per year) to large volume (>300 cases per year) resulted in a reduction in mortality after congenital heart surgery.

Different findings reported in our study are related to the different end point used for comparison and our analytic methods. We have used parametric time-related analysis for mortality rather than logistic regression models fit for an arbitrary end point: hospital or ''early'' death. Further, the statistical models were risk adjusted on the basis of data-derived variables derived from an academic database rather than risk stratified on the basis of consensus-based

scores.[8,9] Recognizing that volume may not be the singular best metric for ''experience,'' we used 5 different measures for both institution and surgeon experience. Finally, the relationship between volume and outcome may be different given the considerable improvement in surgical outcomes in the contemporary era.[2] By using a measure of memory loss termed ''exponentially weighted moving average,'' deLeval and colleagues[34] showed that mortality estimates can be influenced when remote experiences are down-weighted compared with more current experience.

Welke and colleagues[37-39] used a large administrative database, the Nationwide Inpatient Sample, to demonstrate the inherent flaws in using case volume alone as a discriminator of mortality. The relationship between hospital pediatric surgical volume and in-hospital mortality was examined using data from more than 55,000 pediatric cardiac surgical operations performed at more than 300 hospitals, using grouped volume categories.[37] These authors found that very small hospitals (those performing ≤20 cases per year) had equivalent mortality to that of large hospitals (those performing >200 cases per year). Model discrimination using volume was poor (c-statistic of 0.5), indicating that predictive ability was no better than a coin flip. A subsequent analysis[39] exploring volume–outcome relationships revealed, using simple power calculations, that an annual volume threshold of 525 operations would be required to detect a doubling of the baseline mortality within the Nationwide Inpatient Sample database (from 4.2% to 8.4%) with a 20% type II error rate.

We found considerable ranked performance variation among institutions within each of 4 groups, with most institutions falling within the 95% confidence intervals of predicted survival. Difficulty achieving reasonable performance discrimination between individual institutions because of wide confidence intervals was also noted by Stark and colleagues[6] in their review of all operations performed in the United Kingdom between 1999 and 1998.

Our study also found that variation in performance among different diagnostic groups occurred within an individual institution. Excellent performance in one diagnostic group did not indicate similar performance in all other groups. Implications for evidence-based referral highlight the importance of defining the institutional distribution of specific expertise.

Management Paradigms in Complex Congenital Heart Disease

Institution-specific management paradigms rather than increasing volume or other measures of ''experience'' may provide solutions to such critically important problems as the treatment of low birth weight neonates or refining the selection criteria for specific repair pathways. We showed that the institution with the best results in the Norwood operation[15] had neutralized an important risk factor that still confounds other institutions. Likewise, selection of neonates for a single or biventricular repair pathway in one institution based on a morphologic protocol led to superior results in PAIVS.[18] In support of this contention, Checchia and colleagues[40] demonstrated a significant 1-year survival benefit (50% vs 10%, $P < .0001$) for infants after the Norwood operation by following the adoption of a preoperative selection protocol that triaged infants to either Norwood or transplantation on the basis of a weighed risk score.

LIMITATIONS

Hierarchic models that account explicitly for within-subject correlation among patients treated at one institution and that introduce surgeon-level factors as random effects (general estimating equations and mixed models) can be easily applied for use in non-timed event analyses such as logistic regression. Hierarchic models may improve discrimination among institutions by reducing variance between each institution.[22] However, software that integrates fixed and random effects within multiphase parametric models is not presently available.

CHSS member institutions represent a select group of hospitals, and submission of data was voluntary for these studies. It is possible, therefore, that the lack of a clear volume–outcome relationship in this study may be due to selection bias, and the results may not be applicable to non-CHSS institutions.

We were unable to quantitate accurately the total operative experience of each individual surgeon because of several factors. These 4 CHSS studies spanned several decades in which migration of surgeons to different institutions over time presumably occurred. Further, formal certification in congenital heart surgery has not existed until 2007, making it impossible to determine with certainty a definitive end point for training.

CONCLUSIONS

Institution and surgeon experience is not the only factor influencing outcomes in complex CHD. Experience cannot be singularly defined as volume and should be evaluated as a composite measure. Considerable overlap in ranked performance exists among institutions, and discrimination is likely to further decline with the overall improvement in outcomes and the ''shrinking'' margin of error. Excellent institutional performance in managing one diagnostic group does not indicate similar performance for all other diagnoses. Valid comparisons of institution and surgeon outcomes could provide a mechanism to improve results in institutions with less favorable outcomes.

References

1. Checchia PA, McCollegan J, Daher N, Kolvos N, Levy F, Markovitz B. The effect of surgical case volume on outcome after Norwood procedure. *J Thorac Cardiovasc Surg.* 2005;129:754-9.

2. Birkmeyer JD, Siewers AE, Finlayson EV, Stukel TA, Lucas FL, Batista I, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002; 346:1128-37.

3. Hannan EL, Racz M, Kavey RE, Quaegebeur JM, Williams R. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. *Pediatrics*. 1998;101:963-9.

4. Jenkins K, Newburger J, Lock J, Davis R, Coffman G, Iezzoni L. In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. *Pediatrics*. 1995;95:323-30.

5. Chang RK, Klitzner TS. Can regionalization decrease the number of deaths for children who undergo cardiac surgery? A theoretical analysis. *Pediatrics*. 2002; 109:173-81.

6. Stark J, Gallivan S, Lovegrove J, Hamilton JR, Monro JL, Pollock JC, et al. Mortality rates after surgery for congenital heart defects in children and surgeons' performance. *Lancet*. 2000;355:1000-7.

7. Kang N, Cole T, Tsang V, Elliott M, de Leval M. Risk stratification in paediatric open-heart surgery. *Eur J Cardiothorac Surg*. 2004;26:3-11.

8. Jenkins KJ, Gavreau K, Newburger JW, Spray TL, Moller JH, Iezzoni LI. Consensus-based method for risk adjustment for surgery for congenital heart disease. *J Thorac Cardiovasc Surg*. 2002;123:110-8.

9. Lacour-Gayet F. Risk stratification scheme for congenital heart surgery. *Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu*. 2002;5:148-52.

10. Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empiric relation between surgical volume and mortality. *N Engl J Med*. 1979; 301:1364-9.

11. Williams WG. Surgical outcomes in congenital heart disease: expectations and realities. *Eur J Cardiothorac Surg*. 2005;27:937-44.

12. Williams WG, McCrindle BW. Practical experience with databases for congenital heart disease: a registry versus an academic database. *Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu*. 2002;5:132-42.

13. Jacobs JP, Mavroudis C, Jacobs ML, Maruszewski B, Tchervenkov CI, Lacour-Gayet FG, et al. What is operative mortality? Defining death in a surgical registry database. A report of the STS Congenital Database Taskforce and the Joint EACTS-STS Congenital Database Committee. *Ann Thorac Surg*. 2006;81: 1937-41.

14. Williams WG. Editorial re: What is operative mortality: defining death for a surgical registry database. *Ann Thorac Surg*. 2006:81:1941-2.

15. Ashburn DA, McCrindle BW, Tchervenkov CI, Jacobs ML, Lofland GK, Bove EL, et al. Outcomes after the Norwood operation in neonates with critical aortic stenosis or aortic valve atresia. *J Thorac Cardiovasc Surg*. 2003;125:1070-82.

16. Blackstone EH, Naftel DC, Turner ME Jr. The decomposition of time-varying hazard into phases, each incorporating a separate stream of concomitant information. *J Am Stat Assoc*. 1986;81:615-24.

17. Blackstone EH, Rice TW. Clinical-pathologic conference choice and use of statistical methods for the clinical study ''superficial adenocarcinoma of the esophagus.'' *J Thorac Cardiovasc Surg*. 2001;122:1063-76.

18. Tekkis PP, McCulloch P, Steger AC, Benjamin IS, Poloniecki JD. Mortality control charts for comparing performance of surgical units: validation study using hospital mortality data. *BMJ*. 2003;326:786-91.

19. DeLong ER, Peterson ED, DeLong DM, Muhlbaier LH, Hackett S, Mark DB. Comparing risk-adjustment methods for provider profiling. *Stat Med*. 1997;16: 2645-64.

20. Blackstone EH. Monitoring surgical performance. *J Thorac Cardiovasc Surg*. 2004;128:807-10.

21. McCrindle BW, Tchervenkov CI, Konstantinov IE, Neirotti RA, Jacobs ML, Blackstone EH. Risk factors associated with mortality and reinterventions in 474 neonates with interruption of the aortic arch: a Congenital Heart Surgeons Society Study. *J Thorac Cardiovasc Surg*. 2005;129:343-50.

22. Ashburn DA, Blackstone EH, Wells WJ, Jonas RA, Pigula FA, Lofland GK, et al. the Congenital Heart Surgeons Society. Determinants of mortality and type of repair in neonates with pulmonary atresia–intact ventricular septum. *J Thorac Cardiovasc Surg*. 2004;127:1000-8.

23. Williams WG, McCrindle BW, Ashburn DA, Jonas RA, Mavroudis C, Blackstone EH. members of the Congenital Heart Surgeon's Society. Outcomes of 829 neonates with complete transposition of the great arteries 12-17 years after repair. *Eur J Cardiothorac Surg*. 2003;24:1-10.

24. Culbert EL, Ashburn DA, Cullen-Dean G, Joseph JJ, Williams WG, Blackstone EH, et al. the Congenital Heart Surgeons Society. Quality of life of children after repair of transposition of the great arteries. *Circulation*. 2003; 108:857-62.

25. Karamlou T, Ashburn DA, Caldarone CA, Blackstone EH, Jonas RA, Jacobs ML, et al., for the members of the Congenital Heart Surgeons' Society. Matching procedure to morphology improves outcomes in neonates with tricuspid atresia. *J Thorac Cardiovasc Surg*. 2005;130:1503-10.

26. Konstantinov I, Karamlou T, Blackstone EH, Mosca RS, Lofland GK, Caldarone CA, et al. Truncus arteriosus associated with interrupted aortic arch in 50 neonates: a Congenital Heart Surgeons Society Study. *Ann Thorac Surg*. 2006;81:214-23.

27. Karamlou T, Jang K, Williams WG, Caldarone CA, VanArsdell G, Coles JG, et al. Outcomes and associated risk factors for aortic valve replacement in 160 children: a competing-risks analysis. *Circulation*. 2005;112:3462-9.

28. McGriffin DC, Naftel DC, Kirklin JK, Morrow WR, Towbin J, Shaddy R, et al, the Pediatric Heart Transplant Study Group. Predicting outcome after listing for heart transplantation in children: a comparison of Kaplan-Meier and parametric competing risk analysis. Pediatric Heart Transplant Study Group. *J Heart Lung Transplant*. 1997;16:713-22.

29. Prentice RL, Kalbfleisch JD, Peterson AV Jr, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978;34:541-54.

30. Hannan EL, Kumar D, Racz M, Siu AL, Chassin MR. New York State's Cardiac Surgery Reporting System: four years later. *Ann Thorac Surg*. 1994;58:182-7.

31. Speigelhalter DJ, Aylin P, Bset NG, Evans SJW, Murray GD. Commissioned analysis of surgical performance by using routine data: lessons from Bristol inquiry. *J R Statist Soc (Ser A)*. 2002;165:1-31.

32. Birkmeyer JD, Finlayson EVA, Birkmeyer CM. Volume standards for high-risk surgical procedures: potential benefits of the Leapfrog initiative. *Surgery*. 2001; 130:415-22.

33. Blackstone EH. Let the data speak for themselves. *Semin Thorac Cardiovsc Surg Pediatr Card Surg Annu*. 2004;7:192-8.

34. deLeval MR, Francois K, Bull C, Brawn W, Spiegelhalter D. Analysis of a cluster of surgical failures: application to a series of neonatal arterial switch operation. *J Thorac Cardiovasc Surg*. 1994;107:914-24.

35. Allen SW, Gavreau K, Bloom BT, Jenkins KJ. Evidence-based referral results in significantly reduced mortality after congenital heart surgery. *Pediatrics*. 2003;112:24-8.

36. Hirsch JC, Gurney JG, Donohue JE, Gebremariam A, Bove EL, Ohye RG. Hospital mortality for Norwood and arterial switch operations as a function of institutional volume. *Pediatr Cardiol*. 2008;29:713-7.

37. Welke KF, Diggs BS, Karamlou T, Ungerleider RM. The relationship between hospital surgical case volumes and mortality rates in pediatric cardiac surgery: a national sample, 1988-2005. *Ann Thorac Surg*. 2008;86:889-96.

38. Welke KF, Diggs BS, Karamlou T, Ungerleider RM. Measurement of quality in pediatric cardiac surgery: understanding the threats to validity. *ASAIO J*. 2008;54: 447-50.

39. Welke KF, Diggs BS, Karamlou T, Ungerleider RM. The relationship between hospital surgical case volumes and mortality rates in pediatric cardiac surgery: a national sample, 1988-2005. *Ann Thorac Surg*. 2008;86: 889-96.

40. Checchia PA, Larsen R, Sehra R, Daher N, Gundry SR, Razzouk AJ, et al. Effect of a selection and postoperative care protocol on survival of infants with hypoplastic left heart syndrome. *Ann Thorac Surg*. 2004;77:477-83.

TABLE E1. Multivariable models for overall mortality

Norwood*
   Birth weight (kg)
   Age at operation (d)
   Circulatory arrest time (min)
   Ascending aortic dimension (mm)
   Reimplantation of the ascending aorta (yes/no)
   Shunt origin from the aorta (yes/no)
Interaction term between ascending aortic diameter
  and ascending aortic diameter
Interrupted aortic arch†
   Date of birth (squared transformation)
   Weight at repair (inverse transformation)
   Truncus arteriosus (yes/no)
   IAA type B or C (yes/no)
   VSD not large (yes/no)
   Absence of malalignment of the VSD (yes/no)
   No augmentation of the aortic arch at IAA repair (yes/no)
   Main pulmonary artery-aorta anastomosis performed at IAA repair
   (yes/no)
Pulmonary atresia with intact ventricular septum‡
   Graded spectrum of right ventricular size (square transformation)
   Tricuspid valve Z-score
Higher left ventricular pulse pressure (mm Hg)
Birth weight (kg)
Transposition of the great arteries§
   Birth weight (kg)
   Age at repair (d)
   Female gender (yes/no)
   Major associated cardiac anomalies (yes/no)
   Major associated noncardiac anomalies (yes/no)
   Number of pre-repair procedures
   Simple TGA (yes/no)
   Moderate or large VSD (yes/no)
   Multiple VSDs (yes/no)
   Degree of pulmonary stenosis
   Repair type
     Arterial switch
     Atrial switch
   Rastelli
   Use of circulatory arrest (yes/no)
   Circulatory arrest time (min)
   CrossClamp time (min)

IAA, Interrupted aortic arch; *VSD,* ventricular septal defect; *TGA,* transposition of the great arteries. *Norwood model from Ashburn DA, McCrindle BW, Tchervenkov CI, Jacobs ML, Lofland GK, Bove EL, et al. Outcomes after the Norwood operation in neonates with critical aortic stenosis or aortic valve atresia. *J Thorac Cardiovasc Surg.* 2003;125:1070-82. †IAA model from McCrindle BW, Tchervenkov CI, Konstantinov IE, Neirotti RA, Jacobs ML, Blackstone EH. Risk factors associated with mortality and reinterventions in 474 neonates with interruption of the aortic arch: a Congenital Heart Surgeons Society Study. *J Thorac Cardiovasc Surg.* 2005;129:343-50. ‡PAIVS model adapted from Ashburn DA, Blackstone EH, Wells WJ, Jonas RA, Pigula FA, Lofland GK, et al. Determinants of mortality and type of repair in neonates with pulmonary atresia-intact ventricular septum. *J Thorac Cardiovasc Surg.* 2004;127:1000-8. §TGA model from Williams WG, MsCrindle BW, Ashburn DA, Jonas RA, Mavroudis C, Blackstone EH. Outcomes of 829 neonates with complete transposition of the great arteries 12-17 years after repair. *Eur J Cardiothorac Surg.* 2003;24:1-10.