Objective measure of learning curves for trainees in cardiac surgery via cumulative sum failure analysis

Elizabeth D. Krebs, MD, MSc, William Z. Chancellor, MD, Robert B. Hawkins, MD, MSc, Jared P. Beller, MD, J. Hunter Mehaffey, MD, MSc, Nicholas R. Teman, MD, Gorav Ailawadi, MD, and Leora T. Yarboro, MD

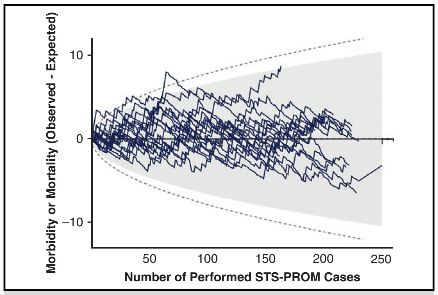
ABSTRACT

Objectives: Objective measures of cardiac surgery trainee progress are limited despite a push for competency-based assessments. We hypothesized that the cumulative sum failure technique could provide a risk-adjusted, quantitative measure of resident learning curves and competence.

Methods: Records of all coronary artery bypass grafting and valve operations performed by cardiac-track residents from 2007 to 2017 at a single institution were stratified by operative resident. Multivariable regression evaluated the association among resident, case number, and postoperative outcomes. To evaluate performance over time, risk-adjusted cumulative sum failure analysis was performed, taking into account institutional expected values and comparing residents with study-defined "early alert" and "concern" boundaries.

Results: A total of 3937 Society of Thoracic Surgeons Predicted Risk of Mortality cases were evaluated from 19 residents. Observed-to-expected ratios for mortality and combined morbidity-mortality were 0.66 and 0.72, respectively, and each individual resident exhibited better than predicted outcomes (all observed:expected ratios <1). When evaluating cumulative sum failure learning curves, residents exhibited an initial slight increase in complications, followed by improvement and better than expected performance. The "early alert" boundary was crossed by 36.8% of residents at any point in training, with 94.7% of residents under this boundary at the end of training. The higher "concern" boundary was crossed by 2 residents (10.5%), although all residents ended their training below this boundary.

Conclusions: Outcomes for trainee-performed cardiac surgery procedures were excellent, with no association between individual trainees and adverse events. Cumulative sum failure analysis based on postoperative outcomes is a potential tool for objective evaluation of resident proficiency. (J Thorac Cardiovasc Surg 2020;160:460-6)



Risk-adjusted learning curves of cardiac surgery residents.

CENTRAL MESSAGE

Trainee-performed cases have excellent results, with better than expected outcomes. Cusum analysis characterizes learning curves and is a promising tool for resident evaluation throughout training.

PERSPECTIVE

Although the safety of resident surgeons in cardiac surgery is well established, objective, outcome-based assessment measures of resident progress are lacking. Cusum analysis has been applied to senior surgeons adopting new techniques, but has not previously described resident learning curves.

See Commentaries on pages 467 and 469.

Providing excellent surgical training for residents and fellows without compromising patient outcomes is the paramount goal of academic cardiothoracic surgical programs.

Scanning this QR code will take you to the article title page to access supplementary information. To view the AATS Annual Meeting Webcast, see the URL next to the webcast thumbnail.

From the Department of Surgery, University of Virginia, Charlottesville, Va.

The National Institutes of Health supported research in this publication under award Numbers T32 HL007849 (Beller, Chancellor) and UM1 HL088925 (Krebs).

Read at the 99th Annual Meeting of The American Association for Thoracic Surgery, Toronto, Ontario, Canada, May 4-7, 2019.

Received for publication May 3, 2019; revisions received Sept 7, 2019; accepted for publication Sept 25, 2019; available ahead of print Oct 31, 2019.

Address for reprints: Leora T. Yarboro, MD, Division of Thoracic and Cardiovascular Surgery, PO Box 800679, Charlottesville, VA 22908-0709 (E-mail: edk9f@virginia.edu).

0022-5223/$36.00

https://doi.org/10.1016/j.jtcvs.2019.09.147

Abbreviations and Acronyms

CABG	= coronary artery bypass grafting
CI	= confidence interval
Cusum	= cumulative sum failure
STS-PROM	= Society of Thoracic Surgeons Predicted Risk of Mortality

The safety of training thoracic surgical residents is well established. Academic programs consistently demonstrate satisfactory short- and long-term outcomes for procedures such as coronary artery bypass grafting (CABG), mitral valve surgery, and aortic valve replacement.[1-5] However, objectively evaluating the safety and progression of an individual resident's skills remains a challenge across surgical disciplines.[6-8] Assessing operative skills in real-world environments is particularly difficult because of patient variability, safety, subjectivity of evaluator, and variability in assessment metrics.[6,9-11] Despite these barriers, surgical education is shifting to a competency-based training paradigm, further necessitating the development of objective metrics for trainee success.[6]

Cumulative sum failure (Cusum) analysis, initially used for quality control in the industrial sector, has been used to evaluate senior surgeons performing new cardiac surgical procedures.[12-15] This technique provides for real-time performance monitoring and is best for detecting small but persistent process deviations.[16] Risk-adjusted Cusum analysis evaluates observed outcomes compared with expected outcomes over time, using predefined boundaries to specify when a larger or smaller amount of adverse events are occurring than expected. Although the method has evaluated trainees performing simulated procedures in a variety of surgical disciplines, it has not to our knowledge been used to evaluate the impact of resident experience on outcomes after cardiac surgical procedures.

We aimed to evaluate the impact of individual cardiothoracic surgical residents' experience levels on postoperative outcomes. Given the Cusum technique's strength in providing real-time performance modeling, we hypothesized that this analysis would identify targetable areas for improvement in surgical resident education. Furthermore, we aimed to characterize the learning curve of residents performing standard procedures to evaluate if postoperative outcomes provide a reasonable means for evaluating resident improvement.

MATERIALS AND METHODS

Study Population

Patients undergoing cardiac operations with a Society of Thoracic Surgeons (STS) predicted risk score, including CABG, aortic valve replacement, mitral valve replacement (MVR), mitral valve repair, or a combination of CABG and valve surgery (STS Predicted Risk of Mortality [PROM] cases), between 2007 and 2017 were selected from an institutional STS database. These cases were chosen to represent standard, commonly performed operations, and because these are the operations for which standardized preoperative scores exist. The primary resident recorded for each case was abstracted and deidentified. To include only cardiac-track residents, cases performed by residents with fewer than 140 STS-PROM cases over the course of their fellowship were excluded (this number represents the minimum number of CABG and valve operations required, and a natural break in resident case volume). At our institution, fellows act as primary surgeon for all portions of standard procedures, from incision to closure. Primary outcome was the composite of operative mortality (composite deaths during hospitalization and within 30 days from surgery) and major morbidity (prolonged mechanical ventilation, renal failure, permanent stroke, reoperation, or deep sternal wound infection), as defined by standard STS definitions.[17] The study was approved by the institutional review board at the University of Virginia (IRB-HSR #19762).

Univariate and Regression Analysis

Univariate analysis evaluated overall demographics and outcomes for cases in the study, reporting characteristics as median (interquartile range), mean (standard deviation), or n (%). Analysis of residents evaluated the observed:expected ratios of composite morbidity and mortality for each resident. Regression analyses evaluated the impact of both case number and individual resident on composite morbidity and mortality, controlling for preoperative risk by incorporating the STS predicted risk of morbidity or mortality into the model. Statistical analysis was performed using SAS Version 9.4 (SAS Institute, Cary, NC).

Cusum Analysis

Risk-adjusted Cusum analysis was performed for the outcome of composite mortality or morbidity, according to methods described by Rogers and colleagues[16] and Grunkemeier and colleagues.[18] For each resident, the sum of observed minus expected failure (composite morbidity and mortality) rates were plotted along the y axis, with the number of STS-PROM cases performed along the x axis. Because institutional outcomes were better than expected by STS predictive risk scores alone, institutional expected values were derived by constructing a logistic regression model, accounting for STS predicted morbidity or mortality. Boundary lines were constructed according to a modified version of equations described by Grunkemeier and colleagues[18] and are intended solely to aid in interpretation and not for formal hypothesis testing (Appendix E1). These lines are drawn to approximate a 95% confidence interval (CI) ("early alert" boundary) and 98% CI ("concern" boundary.) Analysis was conducted using GraphPad Prism (GraphPad Software Inc, San Diego, Calif) and SAS Version 9.4 (SAS Institute, Inc, Cary, NC).

RESULTS

A total of 3937 STS predicted morbidity or mortality cases performed by 19 cardiac-track residents were included in the study. The most common case included was an isolated CABG (50.88%) followed by aortic valve replacement (22.07%). Patients had a mean PROM of 3.34% and a mean predicted morbidity or mortality of 20.14%, with an actual mortality rate of 2.39% and composite morbidity-mortality rate of 13.21%, giving observed-to-expected ratios of 0.66 and 0.72, respectively. Residents included in the study performed a mean of 207 ± 26 STS-PROM cases. Further case demographics are shown in Table 1.

Every resident exhibited an observed:expected ratio of composite morbidity and mortality less than 1 (Table 2),

TABLE 1. Demographics and outcomes of cases with a Society of Thoracic Surgeons predicted risk score performed by residents

	Total cases (n = 3937) Median [IQR], mean (standard deviation) or n (%)
Demographics	
Patient age	68 [59-76]
Sex (female)	1208 (30.68)
Peripheral arterial disease	651 (16.54)
Heart failure	1695 (43.05)
Hypertension	3209 (81.51)
Diabetes	1587 (40.31)
Stroke	67 (1.70)
Predicted risk of mortality	3.34% (4.61%)
Predicted morbidity or mortality	20.14% (14.50%)
Procedure type:	
AV replacement	869 (22.07)
AVR + CABG	472 (11.99)
Isolated CABG	2003 (50.88)
MVr	267 (6.78)
MVr + CABG	102 (2.59)
MVR + CABG	42 (1.07)
MVR	182 (4.62)
Outcomes	
Cardiopulmonary bypass time	95 [78-119]
Postoperative LOS	6 [5-8]
Renal failure	157 (3.99)
Prolonged ventilation	329 (8.36)
Deep sternal wound infection	9 (0.23)
Reoperation	110 (2.80)
Major morbidity	498 (12.65)
Operative mortality	94 (2.39)
Mortality or morbidity	520 (13.21)

IQR, Interquartile range; *AV*, aortic valve; *AVR*, aortic valve replacement; *CABG*, coronary artery bypass grafting; *MVr*, mitral valve repair; *MVR*, mitral valve replacement; *LOS*, length of stay.

TABLE 2. Observed and expected rates of composite morbidity and mortality for cases performed by cardiac-track thoracic surgery residents

Resident	Composite morbidity-mortality (%)	Expected rate of morbidity-mortality (%)*	O:E ratio
B	12.98%	17.62%	0.74
C	13.06%	21.77%	0.60
D	14.16%	21.59%	0.66
E	14.29%	20.94%	0.68
F	14.22%	20.84%	0.68
G	13.11%	21.44%	0.61
I	12.21%	20.08%	0.61
K	12.33%	21.11%	0.58
L	10.97%	21.30%	0.51
M	10.49%	18.22%	0.58
N	10.97%	19.06%	0.58
O	12.95%	18.99%	0.68
P	18.41%	19.57%	0.94
Q	14.00%	19.37%	0.72
S	13.93%	19.94%	0.70
T	14.06%	20.68%	0.68
W	12.75%	21.09%	0.60
Y	12.95%	18.77%	0.69
Z	13.57%	19.06%	0.71

O:E, Observed:expected. *Expected rate based on Society of Thoracic Surgeons predictive risk scores.

demonstrating that each resident had a less than expected overall rate of morbidity and mortality. On multivariable regression analysis, only 1 individual resident was associated with an increase in composite morbidity and mortality (Table 3). Case experience was not statistically associated with these risk-adjusted adverse outcomes; however, it did exhibit a trend toward being protective for composite morbidity and mortality (Table 3).

Risk-adjusted Cusum plots for all residents studied are shown in Figure 1. The "concern" boundary line (representing 98% CI) was crossed by 2 residents (10.5%) at any point in training, with 100% of residents below this boundary at the end of training. The "early alert" boundary line (approximating 95% CI) was crossed by 7 (36.8%) residents at any point in training; however, by the end of training all residents but 1 (94.7%) were under the "early alert" boundary. Figure 2 shows all Cusum curves, with those of residents crossing the alert boundary colored red.

The mean curve for all residents is shown in Figure 3. This curve features an initial upward slope, representing more observed than expected events initially, peaking at approximately 70 cases, followed by a gradual downward slope that levels out at approximately 0 (with observed morbidity matching expected at the institution), approximately 140 total STS-PROM cases.

DISCUSSION

This retrospective study used regression and Cusum techniques to evaluate outcome-related proficiency for 19 residents performing approximately 4000 routine cardiac surgery procedures. Overall outcomes were superior, with far fewer adverse events than predicted based on STS risk scores. There was a trend toward improved outcomes throughout training, and only 1 individual resident was associated with increased morbidity or mortality compared with the others. Upon Cusum analysis, residents demonstrated a slight increase in adverse events at the beginning of training, peaking at 70 cases, followed by improvement

TABLE 3. Logistic regression model for risk-adjusted impact of residents and case experience on composite morbidity and mortality

Factor	OR (95% CI OR)	*P* value
Case No.	0.998 (0.997-1.000)	<.001
Resident B	1.09 (0.603-1.969)	.436
Resident C	0.824 (0.460-1.475)	.579
Resident D	0.912 (0.515-1.615)	.950
Resident E	0.964 (0.542-1.713)	.834
Resident F	0.982 (0.552-1.746)	.764
Resident G	0.797 (0.435-1.459)	.502
Resident I	0.882 (0.502-1.551)	.812
Resident K	0.803 (0.446-1.448)	.506
Resident L	0.694 (0.382-1.261)	.185
Resident M	0.785 (0.404-1.525)	.527
Resident N	0.758 (0.386-1.486)	.450
Resident P	1.449 (0.800-2.624)	.035
Resident Q	1.032 (0.570-1.871)	.604
Resident S	1.009 (0.561-1.814)	.673
Resident T	0.949 (0.523-1.724)	.899
Resident W	0.811 (0.444-1.481)	.552
Resident Y	1.015 (0.569-1.808)	.645
Resident Z	1.042 (0.578-1.877)	.567
PROMM score (logit)	2.649 (2.376-2.953)	<.0001

CI, Confidence interval; *PROMM*, Society of Thoracic Surgeons predicted risk of morbidity or mortality; logit = log (1/(1 – PROMM)).

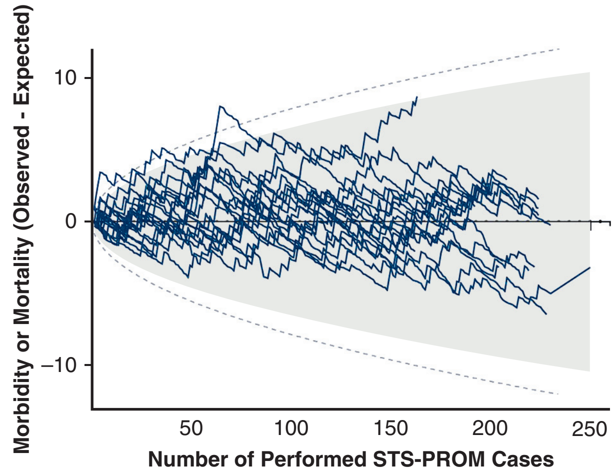


FIGURE 1. Learning curves of residents performing cardiac surgery procedures, risk adjusted by institutional predicted outcomes. The *shaded gray* boundary represents the 95% CI "early alert" boundary, and the *dashed boundary* represents the "concern" 98% CI boundary (Appendix E1). *STS-PROM*, Society of Thoracic Surgeons Predicted Risk of Mortality.

that ultimately resulted in better than expected outcomes by 140 cases, and well before the end of training. Although 7 residents crossed an "alarm" limit at some point during training, all were within predefined proficiency bounds at the end of training, demonstrating the method's value as an objective early warning assessment for trainees.

Overall outcomes were excellent, with each resident demonstrating observed morbidity and mortality rates lower than predicted based on STS predictive risk scores. Additionally, there was only 1 resident with a statistically greater rate of adverse events compared with the others. This is consistent with existing evidence demonstrating safety of trainees in cardiac surgery.[1-5] However, existing studies have also noted learning curves for trainees, with shorter operative and cardiopulmonary bypass time later in training.[5] Likewise, the present investigation noted an incremental improvement in outcomes later in training upon regression analysis, with Cusum analysis demonstrating improvement later in training. Thresholds in our study were noted at approximately 70 cases, when the learning curve peaked, with a decrease until 140 PROM cases, after which outcomes matched the institution's historical average. Coincidentally, the Accreditation Council for Graduate Medical Education minimum requirement for this cases (80 myocardial revascularization and 60 acquired valvular heart disease cases) is also 140. Although this may be solely coincidental, it is also possible that the Cusum analysis mirrors the current surgeon-determined gestalt for the minimum cases needed to achieve full competency.

Cusum analysis has previously evaluated practicing cardiac surgeons adopting new techniques. The method has been used to evaluate a variety of procedures, including minimally invasive and off-pump CABG, thoracic aortic procedures with hypothermic circulatory arrest, and minimally invasive aortic valve replacement.[12,14,15,19] Although there is much variation, the resident learning curves in this intervention closely parallel the learning curves of experienced surgeons, with initial upslope of more complications, followed by eventual downslope and continued improvements. Similar to the present study, the majority of these evaluations have been retrospective, but have noted the technique's promise as a prospective quality improvement tool. One barrier is lack of user-friendly software for prospective analysis. For example, with accessible, web-based software, a surgeon could log the expected and actual outcomes for a series of cases and receive a real-time update of his/her progress.

Although the safety of resident surgeons is well established, upon Cusum analysis there were a number of residents identified to have crossed an "early alert" boundary at any point in training. This suggests that Cusum techniques provide a potential opportunity for early identification of trainees in need of increased guidance. By visually identifying outcome-related trends, mentors may be able to detect areas for improvement or provide targeted education earlier in training. Cardiac surgery centers have

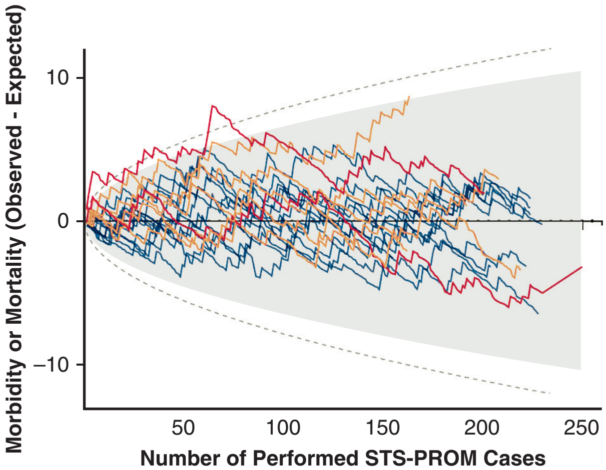


FIGURE 2. Learning curves of selected residents performing cardiac surgery procedures, risk adjusted by institutional predicted outcomes. The *shaded gray* boundary represents the 95% CI "early alert" boundary, and the *dashed boundary* represents the "concern" 98% CI boundary. Residents crossing an "early alert" boundary are *colored orange*, and those crossing a "concern" boundary are *colored red*. *STS-PROM*, Society of Thoracic Surgeons Predicted Risk of Mortality.

reported success using Cusum methods for departmental quality control, which could be similarly translated to trainees.[13] For instance, when noted to cross the boundary lines early in training, a trainee could receive increased guidance and supervision during cases, increased 1-on-1 skills training in a simulation environment, or targeted evaluation of postoperative management to identify deficiencies in clinical judgment. Cusum methods have been used to assess skills-related progress in general surgery trainees, evaluating simulation-based laparoscopic, open, and endoscopic proficiency.[20] These simulation-based studies suggest that Cusum analysis is more sensitive to assess competency than existing metrics, which may translate to Cusum assessment of postoperative outcomes.[20] Although the concern boundaries in this study were chosen to represent 95% and 98% CIs, individual programs could alter their confidence boundaries to more adequately identify residents in need of intervention. By using less lenient boundaries, more trainees may be identified earlier in their training. Because objective measures of trainee performance are lacking, Cusum analysis of outcomes for resident cases represents a promising tool for gauging resident progress during training.[6]

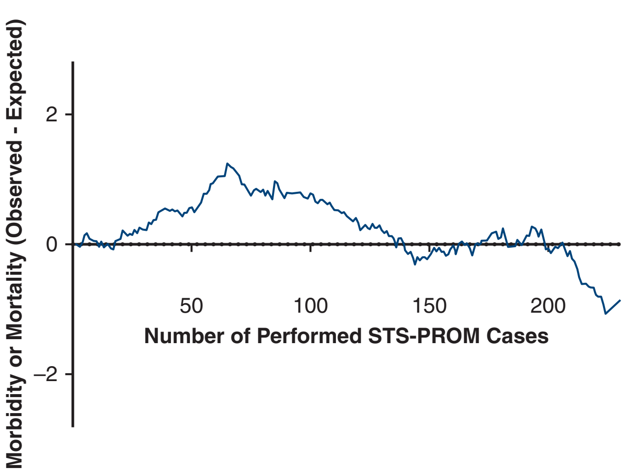


FIGURE 3. Mean learning curve of cardiac-track residents performing routine cardiac surgery procedures. *STS-PROM*, Society of Thoracic Surgeons Predicted Risk of Mortality.

Study Limitations

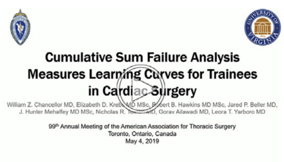
This study does have noted limitations. First, the Cusum methodology is designed primarily for process control rather than rigorous statistical testing, and as with all studies using Cusum methodology, the results should not be overinterpreted. Second, the analysis evaluates only cases with an STS risk score (CABG or left-sided valve procedures) and does not consider other cases that residents have performed throughout their training. Although this was necessary for risk adjustment and adding uniformity, the impact of these other cases is not known. The case mix was also variable, with CABG and aortic valve replacement cases overrepresented in the sample. Furthermore, there is likely some variability in resident role from case to case, although our institution does intend to have residents act as senior surgeon for the entirety of the case. Additionally, although not necessarily a limitation, it should be noted that for Cusum risk adjustment, we derived institution-specific risk scores, taking into account STS risk scores rather than using the STS risk scores alone. Although using STS risk scores would have shown all learning curves as continual better-than-expected downslopes, we thought deriving institution-specific expected values provided a better representation of true learning curves. Finally, this investigation is retrospective and observational in nature.

CONCLUSIONS

By analyzing approximately 4000 cases performed by 19 surgical residents, outcomes for routine cardiac surgery procedures were better than predicted, with no resident demonstrating a greater than expected rate of morbidity and mortality. However, trainees demonstrated a learning curve, with improved outcomes throughout training. Cusum analysis based on patient-level postoperative outcomes is a promising tool for objective evaluation of residents throughout training and may provide early identification of trainees in need of increased guidance to achieve proficiency. As surgical education shifts toward a competency-based education paradigm, the development of objective evaluation metrics will become increasingly important for ensuring success for the next generation of cardiac surgeons.

Webcast

You can watch a Webcast of this AATS meeting presentation by going to: https://aats.blob.core.windows.net/media/19%20AM/Monday_May6/206AC/206AC/S71%20-%20Teaching%20and%20learning%20cardiac%20surgery/S71_5_webcast_080034692.mp4.



Conflict of Interest Statement

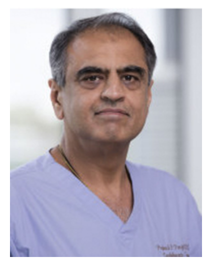
Dr Ailawadi is a consultant for Abbott, Edwards, Medtronic, Admedus, and Gore. All other authors have nothing to disclose with regard to commercial support.

References

1. Bakaeen FG, Sethi G, Wagner TH, Kelly R, Lee K, Upadhyay A, et al. Coronary artery bypass graft patency: residents versus attending surgeons. *Ann Thorac Surg*. 2012;94:482-8.
2. Caputo M, Chamberlain MH, Ozalp F, Underwood MJ, Ciulli F, Angelini GD. Off-pump coronary operations can be safely taught to cardiothoracic trainees. *Ann Thorac Surg*. 2001;71:1215-9.
3. Oo AY, Grayson AD, Rashid A. Effect of training on outcomes following coronary artery bypass graft surgery. *Eur J Cardiothorac Surg*. 2004;25:591-6.
4. Saxena A, Virk SA, Bowman SRA, Jeremy R, Bannon PG. Heart valve surgery performed by trainee surgeons: meta-analysis of clinical outcomes. *Heart Lung Circ*. 2018;27:420-6.
5. Yount KW, Yarboro LT, Narahari AK, Ghanta RK, Tribble CG, Kron IL, et al. Outcomes of trainees performing coronary artery bypass grafting: does resident experience matter? *Ann Thorac Surg*. 2017;103:975-81.
6. Vaporciyan AA, Yang SC, Baker CJ, Fann JI, Verrier ED. Cardiothoracic surgery residency training: past, present, and future. *J Thorac Cardiovasc Surg*. 2013;146:759-67.
7. Dijksterhuis MG, Voorhuis M, Teunissen PW, Schuwirth LW, ten Cate OT, Braat DD, et al. Assessment of competence and progressive independence in postgraduate clinical training. *Med Educ*. 2009;43:1156-65.
8. George BC, Dunnington GL, DaRosa DA. Trainee autonomy and patient safety. *Ann Surg*. 2018;267:820-2.
9. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg*. 2010;97:972-87.
10. Reznick RK, MacRae H. Teaching surgical skills–changes in the wind. *N Engl J Med*. 2006;355:2664-9.
11. DaRosa DA, Zwischenberger JB, Meyerson SL, George BC, Teitelbaum EN, Soper NJ, et al. A theory-based model for teaching and assessing residents in the operating room. *J Surg Educ*. 2013;70:24-30.
12. Murzi M, Cerillo AG, Gilmanov D, Concistre G, Farneti P, Glauber M, et al. Exploring the learning curve for minimally invasive sutureless aortic valve replacement. *J Thorac Cardiovasc Surg*. 2016;152:1537-46.e1531.
13. Murzi M, Cerillo AG, Bevilacqua S, Gasbarri T, Kallushi E, Farneti P, et al. Enhancing departmental quality control in minimally invasive mitral valve surgery: a single-institution experience. *Eur J Cardiothorac Surg*. 2012;42:500-6.
14. Novick RJ, Fox SA, Stitt LW, Kiaii BB, Swinamer SA, Rayman R, et al. Assessing the learning curve in off-pump coronary artery surgery via CUSUM failure analysis. *Ann Thorac Surg*. 2002;73:S358-62.
15. Mazine A, Stevens LM, Ghoneim A, Chung J, Ouzounian M, Dagenais F, et al. Developing skills for thoracic aortic surgery with hypothermic circulatory arrest. *J Thorac Cardiovasc Surg*. 2019;157:1360-8.e1368.
16. Rogers CA, Reeves BC, Caputo M, Ganesh JS, Bonser RS, Angelini GD. Control chart methods for monitoring cardiac surgical performance and their interpretation. *J Thorac Cardiovasc Surg*. 2004;128:811-9.
17. Shahian DM, O'brien SM, Filardo G, Ferraris VA, Haan CK, Rich JB, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1—coronary artery bypass grafting surgery. *Ann Thorac Surg*. 2009;88:S2-22.
18. Grunkemeier GL, Wu YX, Furnary AP. Cumulative sum techniques for assessing surgical results. *Ann Thorac Surg*. 2003;76:663-7.
19. Holzhey DM, Jacobs S, Walther T, Mochalski M, Mohr FW, Falk V. Cumulative sum failure analysis for eight surgeons performing minimally invasive direct coronary artery bypass. *J Thorac Cardiovasc Surg*. 2007;134:663-9.
20. Hu Y, Jolissaint JS, Ramirez A, Gordon R, Yang Z, Sawyer RG. Cumulative sum: a proficiency metric for basic endoscopic training. *J Surg Res*. 2014;192:62-7.

Key Words: education, training, fellowship, AVR, CABG

Discussion



Dr Prakash Punjabi *(London, United Kingdom)*. In the United Kingdom, public reporting of results has been going on for several years with the associated issues in terms of surgeon-specific mortality data, and we do that with the use of a variable life-adjusted display plot, which is done for institutions as well as surgeons.

My first question regards the case mix and validity of the model. As you are aware, there is a significant evidence base that subspecialty surgery, such as aortic and mitral, have superior outcomes when done in higher volume. In your model, 73% of the surgery is isolated aortic valve replacement or CABG, with only 40% being mitral valve surgery. On a practical perspective, this is 593 cases done by 19 residents, or 31 cases per resident over a 2-, 3-year period. If a surgeon has poor outcomes in one particular subspecialist field (eg, mitral valve surgery), does this model have the risk of being buried in a larger number of other cases, and so should this model be used for the same operation. What I am trying to say is, should you be doing specialist per case rather than using as a whole case mix to look at the trainee experience?

Dr W. Zachary Chancellor *(Charlottesville, Va)*. We chose to use both because it is a higher number and risk adjusted. We thought that the risk adjustment would account for any variation between the cases. However, I do see some utility in a world where we want to track our outcomes and training that looking at them individually is certainly possible and could be worthwhile.

Dr Punjabi. Second, coming to you on the angle that seeks to bring some evidence based on a learning curve, training, and the number of cases required to become competent, as you note, the Accreditation Council for Graduate Medical Education requirements are for 140 cases, 80 revascularizations and 60 vascular, and we heard this morning about the different integrated ways of getting training. Figure 3 in your article suggested a learning curve of approximately 140 cases, no particular difference up to 200, and then potentially some improvement after that. In the United Kingdom, I and my colleagues have decided that a trainee needs to do between 200 and 250 cases before getting final certification to be a consultant. To play devil's advocate, does the further improvement beyond 200 cases suggest that the Accreditation Council for Graduate Medical Education requirement should be higher?

Dr Chancellor. I don't think that is what this analysis is showing. I can't speak to that based on this analysis. But the learning curve we saw does show the initial upslope, which shows higher than expected morbidity and mortality initially, but the downslope that levels off at approximately 140 cases indicates that they are actually doing better than expected. I think that one of the fallacies of overinterpreting these graphs is that a level line, no matter where it is on the graph, represents that the residents are doing well or doing as they are expected to do. They don't necessarily need to be downsloping all of the time.

Dr Punjabi. I take your point. I think you are right. It just brings out 2 other small points. One is the definition of ownership of a case by the trainee. As I am sure you will agree, training in different centers within the same country varies quite a lot in the definition of a case. However, my last comment is more about a practical aspect. As you know, a lot of our safety aspects come from the aviation industry. When a new pilot is learning to fly and the plane goes off course, it is taken over by the senior pilot. Presumably these operations and postoperative care were all done under the direct supervision of the attending surgeon, and sometimes the number of cases or the morbidity and the mortality cannot necessarily be blamed on the surgeon, on the training. What are the particular implications in terms of training to reassure the trainer and the institute that training can still be provided safe?

Dr Chancellor. That's an excellent question and one every institution struggles with and is one of the reasons that we used institution-specific data in this model, and every institution's curves are going to look a little different based on their approach. University of Virginia is all I can speak to personally, but I do know that residents do have quite a bit of operative autonomy; honestly, it's safe. The attending has oversight within the operating room and in the postoperative management. I do think that outcomes are reflective of the resident's management style. However, this particular analysis is not going to account for any oversight that the attending surgeon provides. I agree.

Dr Punjabi. I fully agree. Congratulations once again and look forward to further collaboration.

Dr Paul T. Sergeant *(Sint Joris Winge, Belgium).* I congratulate you for any work that is done on the learning process, but in fact what you have addressed is what is called in the science of learning operational learning or organizational learning, and that is preceded by all the different aspects of induced learning. The whole discussion about number of cases is, according to the science of learning, totally irrelevant. What is important is the process that precedes the operational and organizational learning.

A resident who does a case should, by definition, have the same result as the standard of that same unit. It cannot be accepted by society that a resident has a lower performance. You are absolutely right when you made the statement on line. We have been using exactly the method that you have described for more than 20 years in our residency training program, but on line, immediately, so that we can track immediately any deviation from the Cusum lines.

I congratulate you for your effort. There is only one additional limitation that you have not addressed. You are conceptually looking at maybe an incomplete perspective. The surgeon today is a member of a team. So you are only addressing the surgeon as the only member of the team. In fact, today we work as a team.

The crew resource management, as it is called, is an important matter on which we must evaluate our young residents. And so putting the young surgeon as responsible for a negative outcome is I think an outdated concept. It is the interaction with anesthesia, it is intensive care. It, of course, changes from country and from different socioeconomic environments. But we must get rid of that idea.

The consequence of the inappropriate surgeon-focused outcome monitoring is that some surgeons have had their career closed. Some very good surgeons had their career closed because of inappropriate public reporting. The reporting was catastrophic, but their performance was not.

APPENDIX E1

For prediction limits in Figures 2 and 3, we used a modified version of the equations described by Grunkemeier and colleagues.[18] Their equation describes the standard error (SE) of the risk-adjusted Cusum at time t as the square root of the cumulative sum of E(1-E) for all patients operated on up to time t, where E is the expected value at that time point. The 95% and 98% prediction limits are determined by multiplying the SE by 1.96 and 2.58, respectively. To estimate the curve for all residents, E was assumed to be the mean expected value of all operations (meanE), such that the SE at any point equaled the square root of t*meanE*(1-meanE).