Impact of the Anesthesiologist and Surgeon on Cardiac Surgical Outcomes ☆

Olympia Papachristofi, MSc,* Jonathan H. Mackay, FRCA,† Sarah J. Powell, BA,‡ Samer A.M. Nashef, FRCS,§ and Linda Sharples, PhD||

Objective: **To determine the impact of anesthesiologists, surgeons, and their monthly caseload volume on mortality after cardiac surgery.**

Design: **Ten-year audit of prospectively collected cardiac surgical data.**

Setting: **Large adult cardiothoracic hospital.**

Participants: **A total of 18,569 cardiac surgical patients in the decade from April 2002 through March 2012, plus 21 consultant surgeons and 29 consultant anesthesiologists.**

Interventions: **Major risk-stratified cardiac surgical operations.**

Methods: **The primary outcome was in-hospital death. Random intercept models for the surgeon and anesthesiologist cluster, respectively, were fitted, achieving risk-adjustment through the logistic EuroSCORE. The intraclass correlation coefficient (ICC) subsequently was used to measure the amount of outcome variation due to clustering.**

Measurements and Main Results: **After exclusions (duplicates, very-short-term appointments, and cases performed by more than one consultant), there were 18,426 patients with 581 (3.15%) in-hospital deaths. The overwhelming factor associated with outcome variation was the patient risk profile, accounting for 97.14% of the variation. The impact of the surgeon was small (ICC = 2.78%), and the impact of the anesthesiologist was negligible (ICC = 0.08%). Low monthly surgeon volume of surgery, adjusted for average case mix, was associated with higher risk-adjusted mortality (odds ratio = 0.93, 95% CI 0.87-0.98).**

Conclusions: **Outcome was determined primarily by the patient. There were small but significant differences in outcome between surgeons. The attending anesthesiologist did not affect patient outcome in this institution. Low average monthly surgeon volume was a significant risk factor. In contrast, low average monthly anesthesiologist volume had no effect.**

KEY WORDS: anesthesiologist, cardiac anesthesia, cardiac surgery, EuroSCORE, mortality, surgical volume, surgeon.

CARDIAC SURGICAL OPERATIONS are performed by a team that includes, among others, surgeons and anesthesiologists. Risk-adjusted mortality for cardiac surgery undoubtedly is affected by which surgeon performs the operation.[1,2] This topic has received much attention recently, and surgeon-specific risk-adjusted mortality is available in the public domain in several countries, including the United Kingdom (http://www.scts.org/patients/hospitals/). In contrast, the impact of the anesthesiologist is less well known. Two relatively small studies (published more than 20 years ago) have suggested a potential impact of the anesthesiologist as a risk factor for cardiac surgical outcome.[3,4] The topic has since received scant attention.

In addition, there is some evidence that the volume of a surgeon's caseload is inversely related to risk-adjusted mortality,[1,2] but the impact of anesthetic caseload volume on outcomes is unknown. The increasing subspecialization within cardiothoracic anesthesia and the development of cardiothoracic intensive care into a specialty in its own right have raised questions about whether there should be a set minimum surgical caseload that an anesthesiologist or an intensivist-anesthesiologist should undertake to deliver and maintain safe care.

The motivation behind this work stems from the hypotheses that, firstly, cardiac surgical outcomes may vary among anesthesiologists and, secondly, that caseload volume may affect patient outcomes. This study aimed to quantify the variation in risk-adjusted mortality between cardiac anesthesiologists, to compare any such variation with that among surgeons, and to investigate the impact of caseload volume on outcomes.

METHODS

Using a large cohort of more than 18,000 cardiac surgical patients operated over a ten-year period (2002-2012), the authors studied the variation in outcome in relation to three factors: Patient risk, surgeon, and anesthesiologist. The data were collected prospectively and included all major cardiac operations at Papworth Hospital, a large specialist cardiothoracic hospital. Cardiac transplants, pulmonary endarterectomy procedures, and other procedures for which EuroSCORE is not appropriate were excluded. There were 18,662 patients treated by 21 surgeons and 29 anesthesiologists. There were 93 reoperations during the same hospital admission; in such cases, the patient information for the first procedure was linked to the final outcome after the second procedure, so that the authors considered second procedures as a consequence of the first. In the remaining 18,569 patients, those with more than one independent admission during the ten-year study period were treated as independent episodes; there was one patient with four, four patients with three, and 195 patients with two separate admissions. The only missing data, for which records could not be retrieved, were five cases in which the anesthesiologist was not recorded, one case in which the surgeon was not recorded, and one case of unknown outcome; these were removed from the data set. There were two patients with unknown gender, but this was adjusted for indirectly through the logistic EuroSCORE and did not affect the analysis.

Three surgeons and five anesthesiologists were excluded from the study because their workload during the decade was fewer than 30 cases (<1% total cases). They had retired just after the beginning of the study period, were appointed just before the end of the study

*From the *MRC Biostatistics Unit, Institute of Public Health; and Departments of †Anesthesia, ‡Clinical Audit, §Surgery, and ||Research and Development, Papworth Hospital, Cambridge, United Kingdom.*

Address reprint requests to Olympia Papachristofi, MRC Biostatistics Unit, Institute of Public Health, Cambridge, CB4 3EU United Kingdom. E-mail: olympia.papachristofi@mrc-bsu.cam.ac.uk

1053-0770/2601-0001$36.00/0
http://dx.doi.org/10.1053/j.jvca.2013.07.004

period, or held short-term contracts at the hospital. In addition, very-high-risk patients who required surgery by two or more consultant surgeons were also excluded (37 patients, <1% total cases). Thus, final analysis was performed on 18,426 patients treated by 18 surgeons and 24 anesthesiologists; a diagram depicting the inclusions and exclusions resulting to the final analysis data set is given in Figure 1.

The monthly caseload volume for surgeons and anesthesiologists was defined as the total workload divided by the number of months in active practice. Patient-related covariates were those required to calculate the logistic EuroSCORE[5] (EuroSCORE II[6] was not available at the beginning of this study). The primary outcome measure of interest was in-hospital death.

Statistical Analysis

The authors used logistic random effects regression analysis[7,8] to analyze the relationship between in-hospital death and potential covariates. Logistic regression is a method for relating binary outcomes (such as survived/died within 30 days) to covariates (such as surgeon and risk score). The structure of the data is grouped naturally (patients nested within surgeons/anesthesiologists); thus, the authors expected to have correlation among observations within a group (here represented by each surgeon/anesthesiologist). Usually in such models, it is assured all patients are independent. However, when there may be differences due to surgeon (or anesthesiologist), patients operated on by the same surgeon (anesthesiologist) may have more similar outcomes than patients who are operated on by different surgeons (anesthesiologists). To address this, the authors included terms in the model, called "random effects," that represented the surgeons (anesthesiologists). Additionally, the method assumed that the group of surgeons in the study was a random sample of all cardiac surgeons and that, had 18 other surgeons been chosen, the distribution of their results would have been similar, thus providing generalizable estimates. Failure to take this into account in the analysis can lead to bias in the estimated group and covariate effects and inaccuracy in the standard errors and p values for these effects. This approach also allowed the authors to delineate surgeon average effects from effects of their caseload volume.

The logistic EuroSCORE was included as a fixed effect in all models to standardize for different patient risk profiles; this was achieved by dividing the scores by 100 to transform them to probabilities and by further taking their logit transform. The authors fitted three different random intercept models, the first incorporating a random effects term for the surgeon, the second incorporating a random effects term for the anesthesiologist, and, finally, a model including random effects terms for both.

To investigate the effect of volume on outcome, the authors refitted the surgeon random intercept model, including the monthly average volume of cases per surgeon as a cluster-level covariate; this was replicated for the anesthesiologist random intercept model and the anesthesiologist monthly average. To examine the effect of volume on

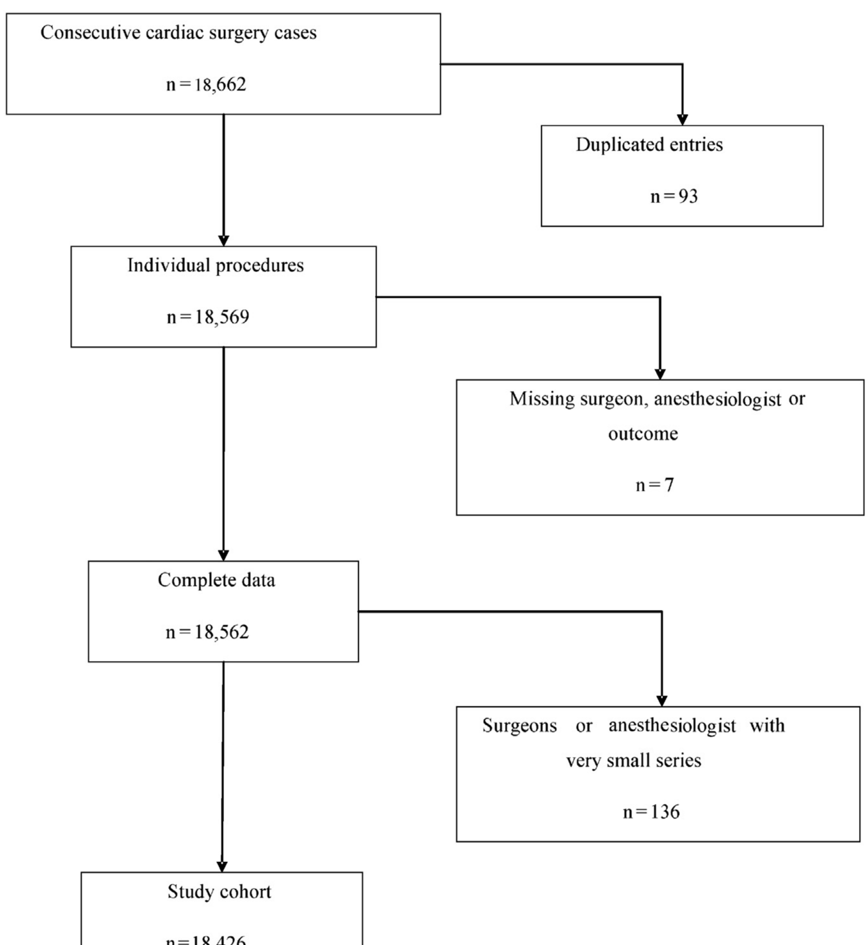

Fig 1. Inclusions and exclusions for the final data set.

Table 1. Patient and Operative Characteristics

Patient Characteristics	Category	Frequency (Percentage of n = 18,426)
Age at admission (years)	16-36	262 (1.4%)
	36-56	2,165 (11.7%)
Mean: 67.8 (11.4)*	56-66	4,363 (23.7%)
Median: 70	66-76	6,768 (36.7%)
IQR: (61,76)	76-86	4,491 (24.4%)
	86-96	376 (2.0%)
Gender	Male	13,346 (72.5%)
	Female	5,078 (27.6%)
	Unknown	2 (0.01%)
EuroSCORE (probability)	(0,0.1)	14,351 (77.9%)
Mean: 0.0807 (10.77)*	[0.1,0.2)	2,419 (13.1%)
Median: 0.0437	[0.2,0.3)	799 (4.3%)
IQR: (0.0225, 0.0901)	>0.3	857 (4.7%)
Operative Characteristics		
Priority	Elective	14,438 (78.4%)
	Urgent	3 034 (16.5%)
	Emergency	954 (5.2%)
Operation Type	CABG (isolated)	8,891 (48.3%)
	AVR (isolated)	2,498 (13.6%)
	MVR + other	2,118 (11.5%)
	CABG + AVR	1,979 (10.7%)
	CABG + other procedures	813 (4.4%)
	CABG + other valve	658 (3.6%)
	Other procedures	614 (3.3%)
	AVR + other procedures	613 (3.3%)
	CABG + AVR + other	242 (1.3%)

Abbreviations: AVR, aortic valve replacementl; CABG, coronary artery bypass graft; QR, interquartile range.

*For continuous variables, the mean (SD), median, and interquartile ranges are given.

outcome in relation to the average case mix treated by each surgeon, a model incorporating fixed-effects terms for both the average monthly volume and the average risk of case mix for each surgeon was fitted.

To assess the amount of variation between clusters for each model, the random effects variance was calculated; the intraclass correlation coefficient (ICC),[9] interpreted as the proportion of the total variation explained by clustering, also was determined. The p values determining the significance of the fixed-effects terms were calculated using the likelihood ratio test. Analyses were implemented using R (version 2.15; R Core Team, R Foundation for Statistical Computing, Vienna, Austria) and STATA (12.0; StataCorp LP, College Station, Texas).[10]

RESULTS

All baseline characteristics for the study cohort are summarized in Table 1. Briefly, mean (SD) age was 67.8 (11.4) years (range 16-96). Most were male (72.5%) and the majority were elective operations, with 16.5% urgent and 5.2% emergency procedures. The median EuroSCORE was 4.4% (IQR 2.25%- 9.01%) with the mean at 8.1% (range 0.9%-97.3%).

Overall, 581 (3.2%) of 18,426 patients died in hospital. The cases performed by each surgeon and anesthesiologist are summarized in the appendix, along with the death rates, the average monthly volume, and the average EuroSCORE per operator. In both the surgeon and the anesthesiologist models, the logistic EuroSCORE was a significant covariate (coefficient 0.90, 95% CI 0.840-0.970 and coefficient 0.91, 95% CI 0.851-0.978, respectively; p value < 0.0001 for both). Note that the estimated coefficients are close to, but less than, one indicating that the logistic EuroSCORE correctly orders risk although it is known to overestimate risk.

Using the estimated surgeon-only model, the authors could predict the probability of an in-hospital death for each surgeon if they operated on a patient with mean EuroSCORE (estimated as 8.1%). Figure 2 shows the surgeon-specific probability of in-hospital death with its 95% confidence interval for a patient with average EuroSCORE. The horizontal line shows the overall probability of in-hospital death. Three surgeons had an estimated probability of death that was significantly greater than the average risk for Papworth surgeons, with their 95% confidence interval lying wholly above the average probability of death, and these could be investigated locally for factors not represented in the original logistic EuroSCORE, using standard audit practice.[11]

For the anesthesiologist-only model that adjusts for patient risk, the anesthesiologist random-effects variance was very small, with $ICC_{anesthesiologist} = 0.000506$, which was negligible. Using this model, the authors plotted probabilities of in-hospital death for a patient of average EuroSCORE risk for each anesthesiologist (Fig 3). This showed that there was almost no variability in outcome due to the anesthesiologists.

The authors also fitted a model including two random intercepts to investigate whether specific surgeon-anesthesiologist pairs affected outcome, but no pair effects could be identified. Table 2 is based on a model containing both surgeon and anesthesiologist effects and shows that the greatest source of variation in outcome was the patient's risk profile, accounting for 97.14% of the variation. The second largest effect could be ascribed to the surgeon, who accounted for only 2.78% of the variation. In comparison, any effect of the anesthesiologist was negligible (0.08%).

For both surgeons and anesthesiologists, the results suggested a weak association between higher monthly volume of cases done and reduction in risk of in-hospital death, with the estimated odds ratios of 0.94 (95% CI 0.88-1.01) and 0.99 (95% CI 0.97-1.01) respectively; however, these were not significant at the 5% level ($p = 0.103$ and $p = 0.208$, respectively).

When both the average monthly volume and the average risk of case mix for each surgeon were included in the model, giving odds ratios of 0.93 (95% CI 0.87-0.99) and 2.04 (95% CI 0.93-4.50) for in-hospital death, respectively, average monthly volume was related significantly to in-hospital death ($p = 0.0464$ and $p = 0.112$). The authors repeated this

Table 2. Sources of Variation in Outcome From the Model With Both Surgeon and Anesthesiologist Random Effects

Variation in outcome attributed to:	Estimate
Surgeon	2.78%
Anesthesiologist	0.08%
Patient	97.14%

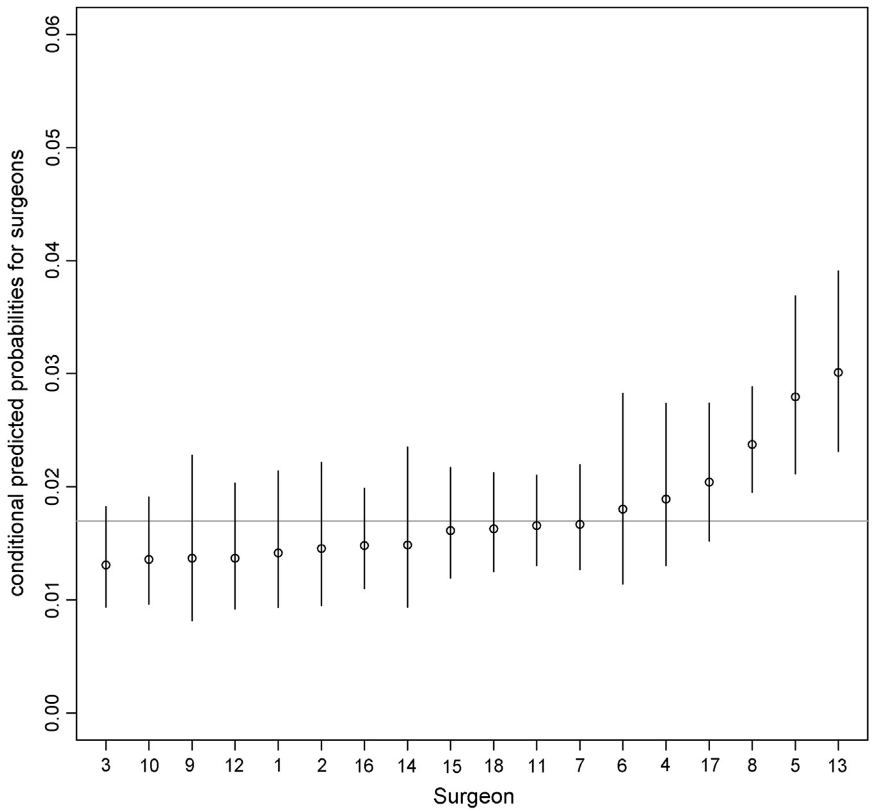

Fig 2. Surgeon-specific probability of an in-hospital death for a patient with average EuroSCORE risk; the horizontal line represents the average in-hospital mortality at Papworth Hospital for a patient with an average EuroSCORE of 8.1%.

analysis for anesthesiologists, but there were no significant relationships.

There was a mild correlation between average case mix and volume for surgeons (Fig 4). Two of the three surgeons whose in-hospital death rate was higher than the Papworth average (Fig 2, surgeons 5 and 13), had both higher-than-average risk and lower-than-average volume.

DISCUSSION

Surgery is a complex intervention comprising a combination of components that can act independently and interdependently. Among these are the different health professionals involved in the delivery of surgical care. In this study, the authors aimed to establish the individual effect on patient outcome of both the surgeon and anesthesiologist involved in the operation. A systematic review identified no recent publications that assessed the effect of the individual anesthesiologist on the most commonly recorded surgical outcome of in-hospital mortality.

The authors found that most of the risk of in-hospital death ($>97\%$) was driven by the patient risk profile; the individual surgeon had a small but measurable effect on outcome, but no effect was found for the individual anesthesiologist.

As other studies have indicated, the individual surgeon has an impact on outcomes. In this study, this factor accounted for no more than 2.78% of the observed variability. All surgeons performed within an acceptable margin from the average expected performance, and overall risk-adjusted mortality was low for each individual surgeon. The variability in outcomes was not surprising, but clinical governance requires that, regardless of any such variability, the performance of all clinicians should fall within an acceptable range, consistent with what is achieved in current practice. Regular monitoring of risk-adjusted outcomes is useful in ensuring that clinical results are acceptable and in identifying rogue performance regardless of cause.[12]

Two characteristics of the delivery of anesthesia are relevant to the Papworth Hospital practice. The first is that all surgeons work with all anesthesiologists and there are no systematic "pairings". The second is that the delivery of cardiac anesthesia is largely protocol-driven, with standardized methods of preoperative anesthetic assessment, induction, maintenance, and monitoring of anesthesia, including the choice of anesthetic agents, vasoactive drugs, invasive monitoring methods, and the use of transesophageal echocardiography. This study showed that, under such circumstances, the standard of care in cardiac anesthesia was remarkably and reassuringly consistent. These findings may not apply to anesthetic departments elsewhere. The methodology used in this study had the ability to detect underperformance and to offer the potential to correct it. The authors recommend such methodology to other institutions with an interest in the robust monitoring of surgical outcomes and in the performance of professionals such as surgeons and anesthesiologists.

Job plans must be constructed to ensure that consultants are sufficiently experienced in the areas in which they work. Future revalidation and credentials requirements in the United Kingdom

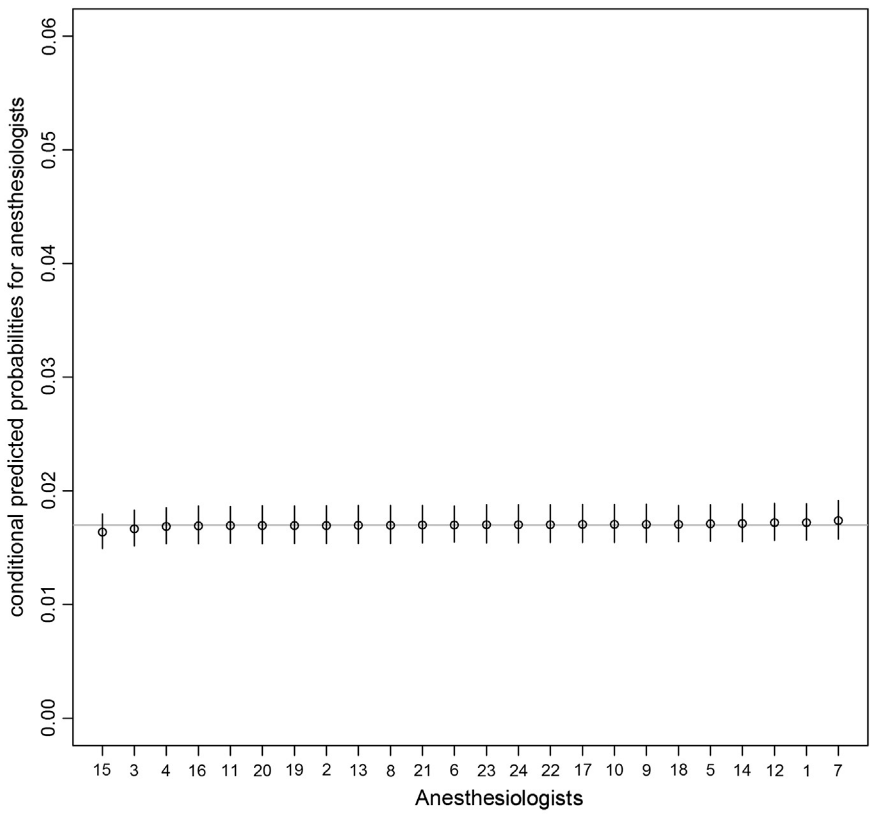

Fig 3. Anesthesiologist-specific probability of an in-hospital death for a patient with average EuroSCORE risk; the horizontal line represents the average in-hospital mortality at Papworth Hospital for a patient with an average EuroSCORE of 8.1%.

and North America may set a minimum number of cardiac cases to be performed per year for anesthesiologists to develop and maintain expertise in cardiac anesthesia. What is the minimum cardiac caseload needed to revalidate as a cardiac anesthesiologist? The United Kingdom Association of Cardiothoracic Anesthesiologists (ACTA) view, based on expert consensus, or level C evidence, is that consultants should undertake a minimum of 50 major heart operations per annum. Although one Papworth anesthesiologist performed fewer than 50 cases per annum, most of the clinical workload was in the setting of the cardiothoracic intensive care unit. This study provided reassurance that this number is effective for anesthesiologists spending most of their working week in cardiothoracic practice. Increasing subspecialization, which means that some cardiothoracic anesthesiologists will devote a substantial part of their time to intensive care, catheter-based procedures, thoracic anesthesia, and other activities, may result in some anesthesiologists undertaking fewer than 25 cardiac cases per annum in the future. The methodology used in this study has a potential role in monitoring the effect (safe or otherwise) of this major change in practice.

Surgeon-related factors (surgeon age, experience, and subspecialization) were beyond the scope of the manuscript and were not addressed, because the aim was to determine potential surgeon effects and the effects of their caseload volume and not necessarily to ascribe them to particular surgeon characteristics. Moreover, further data on these surgeon covariates were not available, and results obtained from this data would be of low power because the authors had a sample of only 18 surgeons.

When assessing the effect of volume on outcome, there was no statistically significant evidence indicating a potential reduction in adverse outcomes with increased monthly average surgical volume. Nevertheless, there is some evidence suggesting that surgeons with low volume coupled with high average case-mix risk increased the chance of in-hospital death. Fig 4 shows that there were eight surgeons whose case-mix risk was higher than average (above the horizontal line of average risk), of whom five had higher-than-average monthly volume (to the right of the vertical line of average volume). Four out of five higher-volume surgeons managed to compensate for their higher case-mix risk, whereas of the lower volume surgeons, only surgeon 2 achieved this (note, as well, that surgeon 2's average case-mix risk was only just above average). This suggests that a high-volume surgeon can compensate for the potential high risk of his/her case mix, which in other circumstances would lead to an increased likelihood of poor outcome. These findings were of low power, because they were based on only 18 surgeons, and should be confirmed in other studies. Nevertheless, they lend some support to the view that surgeons with low volume should not take on very-high-risk patients.

The clustering within surgeons will have implications for the design of randomized clinical trials evaluating the performance of surgical interventions against the standard of care.[13] In

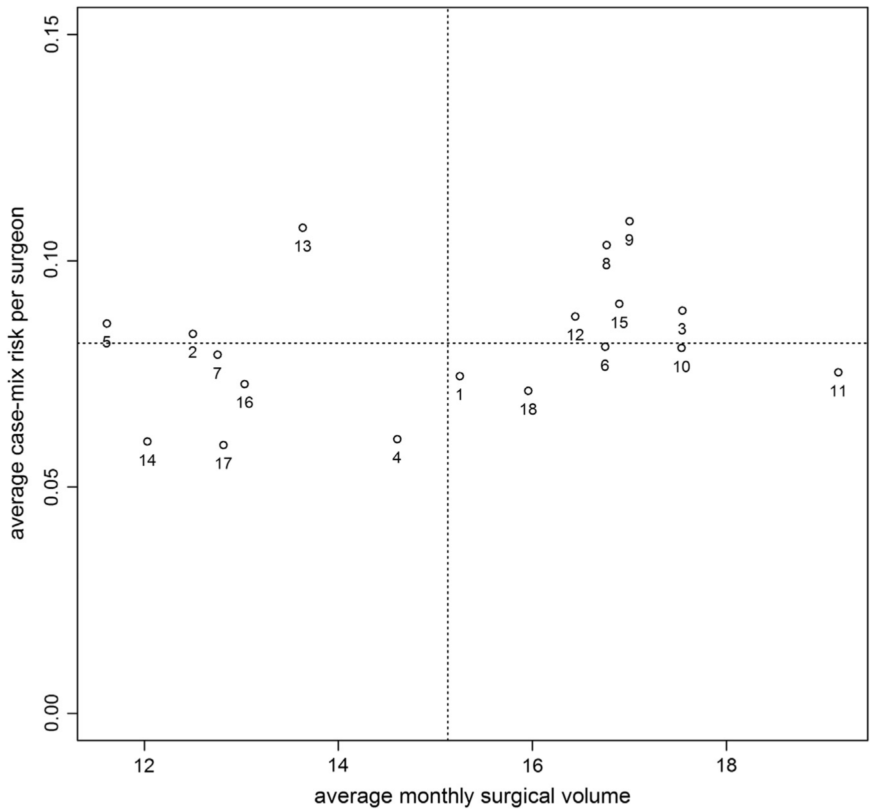

Fig 4. Average case-mix risk against average monthly volume per surgeon; the horizontal line represents the average case-mix risk for Papworth surgeons, and the vertical line represents the average monthly volume of Papworth surgeons.

the current study, the variation due to the surgeons was small, so there was not much impact on the power of a future planned study, whether the trial design is a stratified, individually randomized trial (each surgeon performing both procedures) or an expertise-based randomized trial (each surgeon performing the one procedure in which they have expertise).[14,15] The sample size for a stratified design would be reduced by a factor of (1-$ICC_{surgeon}$) compared with a standard nonstratified trial. For example, if an individually randomized clinical trial required a sample size of 100 to achieve adequate power, a trial stratified by surgeon would require a sample size of 97 (100% $\times$ [1$-$0.0278]). This design relies on surgeons being equally expert in all procedures and being in equipoise about which is the best treatment. However, when surgeons have reached a high level of competence in a procedure and believe that one approach is superior, an expertise-based trial, in which the sample size will be increased compared with a standard design, might be preferred. For example, assuming that the average surgeon performed 10 procedures during a trial, with an ICC of 0.0278, the sample size would need to be increased by 25%. Further, if the average number of procedures per surgeon was 20, the trial would need to be 53% bigger.

Handovers between anesthesiologists at Papworth sometimes occur at 5 p.m. (when the on-call anesthesiologist may take over) and 8 a.m. (when the on-call anesthesiologist may hand over). A few operations necessarily will cross these times and are ascribed to the primary anesthesiologist who is deemed responsible for the case. A limitation of this study was that the authors have not subanalyzed handovers, but there is no reason to believe that they occur in anything other than a random distribution. All anesthesiologists involved were consultants who had contributed to at least 30 procedures, so that results may not be generalizable to anesthesiologists in training. The setting was a specialist hospital where anesthetic practice is protocol-driven, which leaves less scope for variation in practice. Therefore, the findings might not apply in a different setting where practice differs substantially from the above. The authors could not find any directly relevant published information on the anesthesiologists' relationship with in-hospital death in a systematic review so that they could assess the consistency of the results with other centers.

CONCLUSIONS

In this institution, in-hospital mortality after cardiac surgery is primarily a result of patient risk, with the expertise and performance of the individual surgeon having a small but significant influence, especially when high risk coexists with low volume. The impact of the individual anesthesiologist was negligible. The authors encourage other institutions to use this methodology to evaluate their own outcomes.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1053/j.jvca.2013.07.004.

REFERENCES

1. Birkmeyer JD, Stukel TA, Siewers AE, et al: Surgeon volume and operative mortality in the United States. N Engl J Med 349: 2117-2127, 2003
2. Panageas KS, Schrag D, Riedel E, et al: The effect of clustering of outcomes on the association of procedure volume and surgical outcomes. Ann Intern Med 139:658-665, 2003
3. Merry AF, Ramage MC, Whitlock RM, et al: First-time coronary artery bypass grafting: The anaesthetist as a risk factor. Br J Anaesth 68:6-12, 1992
4. Slogoff S, Keats AS: Does perioperative myocardial ischemia lead to postoperative myocardial infarction? Anesthesiology 62: 107-114, 1985
5. Nashef SA, Roques F, Michel P, et al: European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardiothorac Surg 16:9-13, 1999
6. Michel P, Roques F, Nashef SA, et al: Logistic or additive EuroSCORE for high-risk patients? Eur J Cardiothorac Surg 23: 684-687, 2003
7. Lee KJ, Thompson SG: The use of random effects models to allow for clustering in individually randomized trials. Clin Trials 2:163-173, 2005
8. Omar RJ, Thompson SG: Analysis of a cluster randomized trial with binary outcome data using a multi-level model. Stat Med 19: 2675-2688, 2000
9. Rodriguez G, Elo I: Intra-class correlation in random-effects models for binary data. STATA Journal 3:32-46, 2003
10. Lee B, Lingsma H, Steyerberg EW, et al: Logistic random effects regression models: A comparison of statistical packages for binary and ordinal outcomes. BMC Med Res Methodol 11:77, 2011
11. Marshall CE, Spiegelhalter DJ: Reliability of league tables of in vitro fertilisation clinics: Retrospective analysis of live birth rates. BMJ 316:1701-1704, 1998
12. Arrowsmith JE, Powell SJ, Nashef SA: Local clinical quality monitoring for detection of excess operative deaths. Anaesthesia 61: 423-426, 2006
13. Cook JA, Bruckner T, MacLennan GS, et al: Clustering in surgical trials—Database of intracluster correlations. Trials 13:2, 2012
14. Vierron E, Giraudeau B: Design effect in multicenter studies: Gain or loss of power? BMC Med Res Methodol 9:39, 2009
15. Kerry SM, Bland MJ: The intracluster correlation coefficient in cluster randomisation. BMJ 316:1455, 1998