Quality assurance in congenital heart surgery

Nicholas Kang, Victor T. Tsang, Steve Gallivan, Chris Sherlaw-Johnson, Timothy J. Cole, Martin J. Elliott, Marc R. de Leval

Department of Cardiothoracic Surgery, Great Ormond Street Hospital for Children, London WC1N 3JH, UK
Clinical Operational Research Unit, Department of Mathematics, University College London, UK
Centre for Biostatistics and Epidemiology, Institute of Child Health, London, UK

Abstract

Objective: The aim of this study was to develop a graphical method of risk-stratified outcome analysis in paediatric cardiac surgery to provide a means of continuous, prospective performance monitoring and allow real-time detection of change in outcomes. Methods: Risk-adjusted survival following open-heart surgery was prospectively measured over a 15-month period (n = 460). Outcomes were charted using variable life-adjusted display (VLAD) charts, which indicate the cumulative difference in observed minus expected survival against the cumulative number of cases performed. Risk stratification was based on RACHS-1 (risk adjustment in congenital heart surgery) risk category and age at surgery, using our previously published risk model. The probability of deviation in performance from the expected baseline level was determined using a mathematical model. Results: By the end of the series, observed survival (443/460 = 96.3%) exceeded that predicted by the risk model (434.5/460 = 94.5%), equivalent to a one-third reduction in expected mortality. Mathematical modelling indicated a 1—5% likelihood that this difference would have occurred by random variation alone, suggesting the outcomes represented genuine improvement. Conclusions: VLAD charts provide an effective, easily visualised display of surgical performance and can be applied to paediatric cardiac surgery. Early detection of change, whether improvement or deterioration, is important for ongoing quality assurance within a cardiac surgery programme.

1. Introduction

The traditional design of most surgical audits is the retrospective analysis of outcome data, with formal hypothesis testing to look for differences in outcomes of statistical significance. This is the method currently employed in the United Kingdom by the Central Cardiac Audit Database (CCAD), a national database funded by the Department of Health, which looks for differences in mortality rates amongst institutions performing paediatric cardiac surgery [1].

Whilst this customary auditing process is an appropriate method of confirming outlying performance, it is only when the discrepancy in results has become of sufficient magnitude that firm conclusions can be drawn. Yet it is critical to the process of quality assurance in any surgical programme that methods are in place for the detection of deteriorating performance at an early stage, so that appropriate measures can be taken to rectify deficiencies before results become truly outlying.

A dilemma arises in trying to meet these two competing aims: the need to avoid failure to recognise poor surgical performance versus a desire to avoid raising false alarms. In order to address this dilemma, we sought to develop a method of prospective risk-stratified performance monitoring to allow early detection of divergent performance, while at the same time being able to quantify the degree of certainty that changes in outcomes had occurred other than by chance variation.

2. Methods

We applied and combined methodology from several earlier research efforts to develop our current performance monitoring system. The first step was the formulation of a risk-stratification model for congenital heart surgery. The risk adjustment in congenital heart surgery (RACHS-1) system [2], proposed in 2002, was the starting point for the development of our own risk model. We modified and refined the RACHS-1 system to develop a logistic regression equation, which provides estimates of postoperative mortality for any given...
operation involving cardiopulmonary bypass, based on patient age and RACHS-1 risk category. Details of this risk model are given in our earlier report [3], and the regression equation is provided in Appendix A.

The second step was the development of a graphical display of performance. Surgical performance charts, also known as control charts, exist in various forms including cumulative sum charts (CUSUM), cumulative risk-adjusted mortality charts (CRAM), and sequential probability ratio test charts (SPRT) [4—7]. We developed a form of control chart monitoring known as variable life-adjusted display (VLAD) [8—10]. These charts plot the cumulative difference in observed minus expected survival against the cumulative number of operations. Observed survival for any individual operation is equal to either 1 or 0, depending on whether the individual patient survives or dies, respectively. Expected survival for any individual operation is calculated as \( \frac{1}{C0} \) expected mortality. For example, if expected mortality is 5%, then expected survival equals 0.95. The expected mortality for any individual operation in this study was that derived from the regression equation, as detailed in Appendix A.

Using this method, the VLAD chart rises when patients survive, but falls when patients die. The degree to which these rises and falls occur is proportional to the risk of each case. For example, if a patient with an estimated risk of 25% survives, the chart moves up by 0.25. But if the same patient dies, the plot moves down by 0.75. In contrast, if a patient with estimated risk of 1% survives, the chart moves up by only 0.01, but moves down by 0.99 if the patient dies. In this way, high-risk cases carry the greatest 'reward' in the event of success but the least 'penalty' in the event of failure, whereas low-risk cases carry the least reward for success but incur the greatest penalty for failure.

If, over a series of operations, there is no difference between observed and expected survival, the VLAD chart terminates level with the horizontal axis. The horizontal axis therefore represents a 'baseline' (expected) level of performance. Performance, which deviates from this baseline, may simply reflect random variation (chance), or more important divergence (good or bad performance). In order to distinguish between these possibilities, we developed a mathematical model to display performance in terms of probability of divergence from baseline. Boundary lines were mapped onto the VLAD chart to give an indication of the level of certainty that any divergence from the expected performance had occurred due to factors other than chance alone. These boundaries were displayed graphically using coloured zones, which were constructed by plotting the distribution of probabilities of all observed variation in outcome being due to chance. Details of the mathematical models used to construct these boundaries are given in an earlier report from our group [10].

Secondary analysis of possible causes for the observed change in performance was achieved by comparing observed and expected mortality rates in subgroups of patients using a chi-square test. Subgroups were studied by age (neonate, infant, child) and RACHS-1 categories ('low-risk' 1—3 and 'high-risk' 4—6).

3. Results

Fig. 1 displays the VLAD performance chart using a 3-year dataset of 1085 operations. These cases comprised the original dataset used to formulate our risk model. The plot is seen to oscillate about the horizontal axis and finishes very close to zero. This indicates that the difference between observed and expected mortality over the entire series was very small, implying the risk model was well calibrated. The plot stay mostly within the green zone, which represents the middle 50% distribution of probabilities that divergence of

Fig. 1. VLAD chart for open-heart operations from April 2000 to March 2003. Note that the plot terminates close to zero, indicating good calibration of the risk model which was formulated from these data. Coloured areas represent the distribution of probabilities that all outcomes could have occurred by chance variation.
this magnitude was due to chance alone. Some divergence into the yellow zone is observed in the first half of the series, indicating that poorer outcomes at this time were less likely to be due to chance, although there was still a 10–25% probability of this being random variation.

Fig. 2 displays more recent data for the next 460 open-heart (cardiopulmonary bypass) operations over a 15-month period. No patient was excluded from the analysis. Of note, however, 72 patients (16%) had procedures not classified by the RACHS-1 system. Using our risk model, these miscellaneous cases are coded along with RACHS-1 risk category three cases, in accordance with our previously published method [2]. Initially the performance remains similar to baseline, but over the last 150 cases the outcomes improve more noticeably and by the end of the series, the plot has entered the dark blue zone. Since there is only a 1–5% probability that this degree of divergence could have occurred by chance, there is a correspondingly greater certainty that this represents a genuine improvement in results.

Further analysis revealed improved survival in neonates and high-risk category cases, although this did not reach statistical significance. Neonatal open-heart mortality fell from 16.5% in the initial 3-year period to 13.4% in the more recent 15-month study period. Mortality fell in high-risk category cases (RACHS-1 categories 4, 5 and 6) from 18.5% to 11.8%. Note that these two subgroups (neonates and risk categories 4–6) are not mutually exclusive, as many neonates undergo operations that fall into the higher RACHS-1 risk categories. No individual subgroup of patients or operations could be identified as displaying improved survival to a level of statistical significance.

4. Discussion

The term ‘quality control’ was coined in the 1930s, when manufacturing industries started examining statistical methods by which productivity and outcomes could be measured and standards established. Juran and Shewhart, working at the Western Electric Company, a manufacturer of telephony hardware, pioneered the principles by which quality could be measured and improved [11]. Shewhart developed so-called ‘statistical control charts’, which were able to indicate when a process (such as the proportion of manufacturing defects on a production line) had deviated from expected standards [12]. In the 21st century, the healthcare industry has embraced another similar catchphrase termed ‘clinical governance’. This has been defined as ‘a system through which organisations are accountable for continuously improving the quality of their services and safeguarding high standards of care’ [13].

The performance monitoring technique (VLAD) described in this paper is an attempt to address several key issues pertinent to the concept of quality assurance. Firstly, a standard of care is defined. This is the baseline mortality rate, derived from a risk-stratification model. Secondly, outcomes are tracked over time so that the trends and direction in performance can be monitored. Thirdly, every outcome is risk-adjusted so that variation in casemix and complexity are accounted for. Finally, an indication is provided about the probability that fluctuations in performance could be due to chance alone.

The traditional paradigm of surgical outcome monitoring employs a retrospective approach to compare outcomes using rigid statistical testing. VLAD charts offer the following advantages over traditional auditing methods:

1. Results can be followed prospectively, allowing for ‘real-time’ examination of outcomes.
2. Trends in performance over time are readily visualised using the graphical display of data. Evidence of deteriorating performance can be used as a ‘trigger’ to modify practice.
3. Changes in performance can be detected at an early stage. Formal statistical tests to look for differences in outcomes can only identify poor performance at a late stage, when the discrepancy in results has reached a
The changes made in hypoplastic left heart syndrome at our institution have been effective and are helping to steer our performance in the right direction, not only for this specific disease entity but probably in the management of neonates and other complex cases generally.

5. Limitations

This study has only examined the single short-term outcome measure of in-hospital survival. It is increasingly recognised that this is only the 'tip of the iceberg' in monitoring quality [16], albeit a rather sharp tip. In-hospital survival is perhaps a reflection of 'surgical safety', and as such, is a critically important component to quality assurance. The CCAD has made an important leap forwards by tracking 1-year results [1], which adds meaningful information about what the real outcome of surgery has been. Other outcome measures (e.g. neurological outcomes) may be equally, or indeed, more relevant. However, such outcomes are much more difficult to quantify and therefore have less inter-observer reliability as a tool for monitoring surgical performance.

6. Conclusions

VLAD charts provide an effective, easily interpreted graphical display of surgical performance. Alarm lines can be constructed mathematically to signal more significant changes in outcomes. Early detection of change, whether improvement or deterioration, is important for ongoing quality assurance within a cardiac surgery programme.

Acknowledgement

Research at the Institute of Child Health and Great Ormond Street Hospital for Children NHS Trust benefits from Research and Development funding received from the NHS Executive.

References

[7] Sirianidhis C, Bland M, Poloniecki J. Properties of the cumulative risk-adjusted mortality (CRAM) chart, including the number of deaths before a

Appendix A

Regression equation for predicted risk of surgery in patients less than 18 years of age using cardiopulmonary bypass [2].

Predicted risk = \[
\frac{1}{1 + \exp(-\text{log odds risk})}
\]

where \( \text{log odds risk} = -5.378 + \frac{[0.6359 \times \text{RACHS}] + [3.115 \times \text{1/age in days} + 1]}{C12}
\)

Note: miscellaneous cases not defined by the RACHS-1 classification (e.g. cardiac transplantation) are coded along with RACHS-1 risk category 3 in the logistic regression model (see reference [3]).

Online performance monitoring in cardiac surgery

Kang et al. [1] have created an interesting manuscript, of interest not only to experts in quality control or congenital surgery but also to all of us wanting cardiac surgery to sustain in this challenging world.

Indeed we are active in labour-intensive environments where cost-containment and cost-control challenge any use of human or material resource; irrelevant if this challenge is imposed by hospital administrations or self-imposed. The second challenge is the increased expectation of total absence of peri-procedural risk as well by patient as society. Their perception of absence of risk is substantiated in legal persecutions and the concomitant explosion of insurance protection. Our third fundamental challenge is the improved results of nonsurgical interventional but also of noninterventional therapies.

This impacts the mandatory processes of evolution of our profession. John Kirklin proposed in the second part of the 20th century his concept of incremental improvements. Surgeons should try to improve gradually every step or detail of their medical production process. There was very little risk of underperformance in this minimal-difference evolution, because Kirklin added the registration of the surgical process descriptive and outcome variables. The proof of benefit of these improvements was very precise if one used the Blackstone methodologies but also very time-consuming. The visualisation of the effects of these incremental changes was very dissociated in time, sometimes months or even years.

In the 21st century, the industry has moved away from stability and structure as its main competitive force towards disruptive innovation. These disruptive innovations are realised not for the innovations themselves but for the possibility of exponential improvements. The surgical community has to innovate similarly in a disruptive fashion to obtain also exponential improvements, either in annihilating early risk, improving late benefit, reducing resource consumption, preferably all three combined.

We are therefore in need of rapid systems of quality control, corrected for variability, allowing real time process control, because disruptive routines have the possibility of disruptive worse results. In addition, we will have to understand the limitations of our tools.

The first element in these rapid systems is the scoring of variability. The authors use the RACHS system [2], but most systems can be used, as well for adult as for congenital cardiac surgery. The limitations of these models are multiple: the variability of the original dataset, the richness of their descriptive variables, their relationships with socio-economic environments, their questionable stability over time, finally an outcome’s interval covering hospital mortality but not the complete peri-procedural risk. Units with exceptional expertise can even use their own models and use the correct outcome’s interval, as there are 3 months for CABG, 6 months for valve surgery and even longer for congenital therapies. Within some countries, formal agreements have been made on adult cardiac surgery systems, often without understanding their limitations. The K.U. Leuven uses self-developed complex systems with scientifically correct observation intervals but also outside systems, often without understanding their limitations. The K.U. Leuven uses self-developed complex systems with scientifically correct observation intervals but also outside systems with incorrect intervals for intra- and interdepartmental performance monitoring of cardiac surgery.

The second element is the use of a graphical display. The authors have chosen for the VLAD system [3]. This system with high graphical power is a very easy to understand. It allows the use of any observation interval, as stated earlier, and can be created in any spreadsheet program. Many national systems use it to visualise interdepartmental differences. This is perfectly possible since their study samples are usually large enough. Indeed the limitations of this display were the absence of uncertainty intervals and the use in reduced samples. From a conceptual perspective one can understand that the minimal sample should have a cumulative risk of one, but these limitations were never formalised, since the practical use would probably demand...