Debate

Citation for published version:
O'Neill, S, Wigmore, SJ & Harrison, EM 2015, ‘Debate: should we use variable adjusted life displays (VLAD) to identify variations in performance in general surgery?’ Bmc surgery, vol 15, 102. DOI: 10.1186/s12893-015-0087-0

Digital Object Identifier (DOI):
10.1186/s12893-015-0087-0

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Bmc surgery

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Debate: should we use variable adjusted life displays (VLAD) to identify variations in performance in general surgery?

Stephen O’Neill, Stephen J. Wigmore and Ewen M. Harrison*

Abstract

Background: The recent push for the publication of individual surgeon outcomes underpins public interest in safer surgery. Conventional, retrospective assessment of surgical performance without continuous monitoring may lead to delays in identifying poor performance or recognition of practices that lead to be better than expected performance.

Discussion: The variable life adjusted display (VLAD) is not new, yet is not widely utilised in General Surgery. Its construction is simple and if caveats are appreciated the interpretation is straightforward, allowing for continuous surveillance of surgical performance.

Summary: While limitations in the detection of variations in performance are appreciated, the VLAD could represent a more useful tool for monitoring performance.

Background

The recent push for the publication of individual surgeon outcomes underpins public interest in safer surgery. Conventional, retrospective assessment of surgical performance without continuous monitoring may lead to delays in identifying poor performance or recognition of practices that lead to be better than expected performance. The variable life adjusted display (VLAD) is not new, yet is not widely utilised in General Surgery. Its construction is simple and if caveats are appreciated the interpretation is straightforward, allowing for continuous surveillance of surgical performance. While limitations in the detection of variations in performance are appreciated, the VLAD could represent a more useful tool for monitoring performance.

Discussion

VLAD

The VLAD was established by Lovegrove et al. [1] to demonstrate the difference between observed and expected mortality over a specified period of time in Cardiac Surgery. The VLAD is sometimes called the expected-observed cumulative sum (CuSum) plot [2]. It is a graph that plots the cumulative difference in observed mortality from expected mortality on the y-axis against individual cases in the chronological order that they occur on the x-axis. Therefore a VLAD for a mortality rate that is equal to what is expected will end at zero, while a VLAD for a mortality rate above what is expected is seen as a falling line, and vice versa. This easily interpretable visual summary explains why the VLAD is popular amongst clinicians [3]. However, this apparent strength of the VLAD, can also be viewed as a weakness due to the strong temptation to view observed minus expected outcomes as ‘lives saved’ or ‘lives lost’, which is inappropriate.

An example: expected mortality of 5 %

Consider an example in a surgical context where the probability of death for a given procedure is 0.05 or 5 % (Fig. 1a). Each consecutive procedure performed is assigned a binary value, which is 0 if there is no death and 1 if the patient died. A score is calculated from the predicted risk of death for that procedure, which in this example is 0.05. The VLAD score is calculated by subtracting the observed outcome (either 0 or 1) from the expected outcome (in this case 0.05). Therefore for a
procedure resulting in a death the score would be 0.05 minus 1, which is equal to a downward increment of 0.95. While for a procedure that resulted in survival, the score would be 0.05 minus 0, which is equal to a positive increment of 0.05. If 20 cases were performed whereby the expected mortality and observed mortality was equal, then in the 19 cases where there was no death, the surgeon would have 19 upward increments of 0.05, which is equal to 0.95. This would be balanced by the one expected death that is observed, which would result in a downward increment of 0.95 and the line on the VLAD would return to zero. Therefore, in summary:

\[ \text{VLAD} = \text{Cumulative} \left( \text{Expected outcome} - \text{observed outcome} \right) \]

Expected outcome is the probability of death e.g. 0.05

Observed outcome where survival = 0 and death = 1

Advantages of VLAD

A VLAD is simple to construct and can be easily generated without any specialist statistical knowledge or software [4]. The VLAD facilitates targeted and continuous real time outcome surveillance. This allows the VLAD to include a surgeon’s entire caseload, which provides a better perspective of overall performance. Compared with the practice of retrospective assessment, this continuous surveillance mechanism offers the opportunity to identify and address the causes of unexpected results at an earlier stage. This may mitigate on-going poor performance or highlight better than expected performance [5]. Funnel plots are not designed for real
time monitoring so the ability of the VLAD to be used as a continuous surveillance tool is a distinct advantage [3].

Risk adjustments
When using the VLAD, an appropriate adjustment for operative risk is critical for ensuring accurate assessments. Defining a risk of death specific to each individual may be more robust than defining the same risk of death for all individuals undergoing one procedure. Outcomes are therefore adjusted for risk by different models that estimate the risk of death for each patient based on their individual characteristics and comorbidities. However, caution must be observed in applying risk adjustments [3]. As surgical mortality rates decrease, risk scores need to be updated to represent the current standard of practice [3]. Tsang et al. [6] showed in paediatric cardiology how over a relatively short time period risk models could rapidly become out of date. No risk model is perfect and there may be inherent weaknesses in the method used to risk adjust. For example, the partial risk adjustment in surgery (PRAiS) model fails to adjust for certain co-morbid conditions and slightly underestimates risk for the highest risk patients. In a recent publication by Pagel et al. [7] this weakness in PRAiS led to a negative impression of performance in one UK centre that was involved in real time monitoring of risk-adjusted paediatric cardiac surgery outcomes using the VLAD.

Control limits
The VLAD lacks control limits, which can make it difficult to assess the possible contribution of random variation to performance [8]. It also means that identifying the appropriate time to take action based on observed results is not quantitatively determined. This has led to criticism that the VLAD is limited in its ability to identify mortality rate changes with adequate speed [3]. However, since VLADs show the change in outcomes over time, one may not wish to wait to hit ‘significance’ before reflecting on an apparent trend. This approach could lead to the loss of lives that might have been saved, and irretrievable damage (maybe wrongly) to a surgeon’s career when some insight or retraining may have helped [9]. As such, the VLAD should not be considered a statistical evaluation [1].

Despite this, control limits, which are sometimes called rocket tails, can often applied to the VLAD to act as alert thresholds [8]. Walter A. Shewhart, the inventor of the industrial control chart technique, used three standard deviations control limits but in healthcare these control limits are often set at the 5 % level. Although this cut-off is arbitrary it can be considered as the point when the probability that differences between expected and observed outcomes are unlikely to be due to chance alone [8]. Nevertheless, as with any control limit, if control limits are applied to the VLAD, care needs to be taken, as apparent variation in performance may be highlighted when control limits are crossed simply as a result of random variation [8]. An often-cited analogy is the use of metal detectors to screen passengers at airports. In this situation the sensitivity of the detector can be varied. Low sensitivity runs the risk that a prohibited metal item such as a gun will pass undetected. High sensitivity reduces the risk of failing to detect a gun, but increases the number of passengers who are not carrying metal who will be pulled by chance out of line. Where the limits of detection should be set depend on the circumstances of the outcome, its seriousness and the need to detect outliers.

Limitations of VLAD
Another criticism of the VLAD is that a good run of results may mask a subsequent poor run, which will mean that an excess of mortalities are needed to cross the control limit and trigger a review [5]. In Fig. 1c, surgeons are simulated with an actual mortality rate equivalent to that of the population mortality rate for 94 cases but then they have a poor run of 6 deaths in a row. Due to the previous good run, only 32 % of surgeons will cross the lower 95 % control interval control limit at this point.

There are also potential limitations with the VLAD for detecting more consistent changes in practice in an established system. This type of change may occur due
to surgeon performance but could also potentially occur secondary to any significant change in the healthcare environment (e.g. critical care provision). In Fig. 1d, surgeons are simulated with an actual mortality rate equivalent to that of the population rate for 100 cases. At case 100, the actual mortality rate changes to a higher level, but this new “change point” is not detected given the wider control limits at this time. These figures can also be explored further using the aforementioned web-application (http://www.datasurg.net/ VLAD).

One method to prevent good runs masking subsequent poor performance is to prevent the VLAD from becoming positive so that only runs of worsening outcome are examined but this may lead to excess triggering and unneeded reviews of performance [5].

Alternative plots such as the risk-adjusted CuSum and risk-adjusted exponentially weighted moving average also overcome these limitations but may be more complex to construct. The risk-adjusted CuSum plot utilizes a sequential sampling technique to test the hypothesis that the risk of death is increased and doesn’t allow for accumulation of credit for good performance as the statistical test is bounded by the lower limit of zero. The risk-adjusted exponentially weighted moving average plot is a running estimate of the mean output of a process, where the most recent observations are given exponentially more weight than historically distant observations [10].

Use of VLAD in General Surgery

Although it has taken time, examples of the use of VLADs in General Surgery are beginning to emerge. Collins et al. [3] retrospectively performed an analysis of the database of the Scottish Audit of Gastro-Oesophageal Cancer services using a VLAD. While Roberts et al. [5] recently published the first real-time, risk-adjusted VLAD of a single centre’s outcome after Ivor-Lewis oesophagectomy for oesophageal cancer. Guest et al. [4] applied the VLAD to single surgeon’s outcomes following oesophago gastric resections for cancer compared with those predicted by the Portsmouth predictor modification (P-POSSUM) score. Guest et al. [4] also went on to suggest that the VLAD was a potentially useful tool in the process of revalidation for surgeons. This could further extend the applicability of VLAD in the context in General Surgery, as could the use of the VLAD to monitor other performance outcomes such as post-operative complications. Even for the highest risk procedures in General Surgery (e.g. upper gastrointestinal cancer resection), the elective mortality rate is now on average <5 % [11]. Therefore other markers (e.g. failure to rescue, infection and anastomotic leak) could be particularly important in the General Surgical setting. However, before this can happen due consideration of data quality, definition of outcomes, case mix and institutional factors that affect outcome will be important.

Summary

In efforts to improve patient safety the monitoring of surgical performance is becoming more widespread. As general surgery data will be increasingly placed in the public domain it is important that general surgeons take an active role in this process. Different methods of monitoring surgical performance need to be examined by the general surgical community and the use of VLADs could contribute significantly to identifying variations in performance.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

SON drafting of the manuscript. SJW conception and design, and revision of the manuscript critically for important intellectual content. EMH conception and design, acquisition of data, analysis and interpretation of data, and revision of the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

Received: 8 June 2015 Accepted: 20 August 2015
Published online: 28 August 2015

References

1. Lovegrove J, Valencia Q, Treasure T, Sheralaw-Johnson C, Gallivan S: Monitoring the results of cardiac surgery by variable life-adjusted display. Lancet 1997, 350(9085):1128-1130.
2. Smith PC, Mossalos E, Papanicolas I. Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects: Cambridge University Press, 2010.
3. Collins GS, Jibawi A, McCulloch P: Control chart methods for monitoring surgical performance: a case study from gastro-oesophageal surgery. Eur J Surg Oncol 2011, 37(6):473-480.
4. Guest RV, Chandrabalan W, Murray GD, Auld CD: Application of Variable Life Adjusted Display (VLAD) to risk-adjusted mortality of esophageagastic cancer surgery. World J Surg 2012, 36(1):104-108.
5. Roberts G, Tang CB, Harvey M, Kadirkamanathan S: Real-time outcome monitoring followingoesophagectomy using cumulative sum techniques. World J Gastrointest Surg 2012, 4(10):234-237.
6. Tsang V, Brown KL, Synnergren MJ, Kang N, de Leval MR, Gallivan S, Utley M: Monitoring risk-adjusted outcomes in congenital heart surgery: does the appropriateness of a risk model change with time? Ann Thorac Surg 2009, 87(2):584-587.
7. Pagel C, Utley M, Crowe S, Witter T, Anderson D, Samson R, McLean A, Banks V, Tsang V, Brown K: Real time monitoring of risk-adjusted paediatric cardiac surgery outcomes using variable life-adjusted display. Implementation in three UK centres. Heart 2013, 99(19):1445-1450.
8. Sheralaw-Johnson C, Morton A, Robinson MB, Hall A: Real-time monitoring of coronary care mortality: a comparison and combination of two monitoring tools. Int J Cardiol 2005, 100(2):301-307.
9. de Leval MR, Francois K, Bull C, Brawn W, Spiegelhalter D: Analysis of a cluster of surgical failures. Application to a series of neonatal arterial switch operations. J Thorac Cardiovasc Surg 1994, 107(3):914-923; discussion 923-914.
10. Cook DA, Duke G, Hart GK, Pitcher D, Mullany D: Review of the application of risk-adjusted charts to analyse mortality outcomes in critical care. Critical care and resuscitation : journal of the Australasian Academy of Critical Care Medicine 2008, 10(3):239-251.
11. Health and Social Care Information Centre. National Oesophago-gastric Cancer Audit 2013.