

Monitoring Surgical Performance: Current Models and Limitations

Agostino Pozzi^{1*}, Laetitia F Colombo², Francesca De Stefano³, Alessandra Sala⁴,
Marco Bartolucci² and Fabio Villa⁵

¹Department of Gastrointestinal Surgery, San Raffaele Hospital, Milan, Italy.

²Department of Dermatology, San Raffaele Hospital, Vita-Salute San Raffaele University, Milan, Italy.

³Department of Gynaecology, San Raffaele Hospital, Milan.

⁴Department of Cardiothoracic Surgery, San Raffaele Hospital, Milan, Italy.

⁵The Fletcher School of Law and Diplomacy, Boston, MA, USA.

Received: 02 February, 2017; Accepted: 06 February, 2017; Published: 16 February, 2017

*Corresponding author: Fabio Villa, The Fletcher School of Law and Diplomacy, Boston, MA, USA, Telephone: 00393497853684; E-mail: fabiovilla210486@gmail.com

Abstract

Purpose: The article aims to evaluate methods currently used to monitor surgical performance. Different model's advantages and limitations are analyzed to possibly identify the best model of assessment.

Design/Methodology/Approach: An extensive online research took place, looking in particular for articles containing terms such as 'surgical performance monitoring', 'forced ranking', and 'surgical outcomes'. The search was independently carried out by three medical doctors. References were screened to find more articles on this particular topic. A conceptual review was then performed.

Findings: Among the current available methods, the Risk-Adjusted Bernoulli Cumulative Sum (RA-CUSUM), set on a real time prospective monitoring, is preferred and might improve surgical outcomes. This is especially true when the RA-CUSUM is compared with the Variable Life Adjusted Display (VLAD) retrospective model, although VLAD is easier to interpret and it is more frequently used by surgeons.

Practical Implications: Applying a standardized surgical performance assessment model identifies and manages practice variations in the same institution and among hospitals.

Originality/value: RA-CUSUM applied to surgical practice can be an effective tool to determine variations in performance.

Keywords: Key performance indicators; Improvement models; Outcomes, Resources efficiency; Staff satisfaction; Standards

differences in performance quality and the absence of an efficient assessment tool.

The absence of an efficient assessment tool and consistent differences among individual performance quality are major drawbacks when it comes to developing a performance evaluation system.

Numerous attempts to evaluate surgical outcomes have been carried out via forced ranking systems used in corporations, but their accuracy remains questionable. Despite their limitations, data suggests that performance assessment will improve surgical practice, outcomes and reduce costs.

Standardizing surgical care will permit to streamline procedures and eliminate negative implications comprising the working environment. Doing so will help reduce unnecessary costs and minimize complications. Further evaluations of the available models are needed to reach these goals.

In this article, we review the currently available evaluation models used in surgical practice to highlight their advantages and limitations

Analyzing in depth these evaluation models, we will be able to identify the appropriate method resulting in a better surgical outcome.

Discussion

Various strategies aiming at evaluating individual productivity relying on management science have proved rather controversial. Biometric systems vitality curve or forced ranking, are based upon presumably objective data, allowing to classify the best and least efficient performers.

Policies vary among companies; e.g., the former General Electric CEO, Jack Welch, proposes to reward the top 20 percent, to coach the middle 70 percent and disregard the bottom

Introduction

Crucial factors for every healthcare system comprise cost-effectiveness and patient satisfaction. Recently there has been a growing interest in evaluating and publishing individual outcomes related to surgical performance. Current drawbacks in developing an efficient performance evaluation system involve the various independent methods detecting consistent

10 percent performers [1]. Despite many attempts to apply proficiency assessment in surgery, a substantial disagreement exists regarding its use. Several questions remain: is it possible to define objective parameters to classify surgeons' abilities and increase surgical care effectiveness and efficiency? Is it possible to create a model able to include all the patient- and environment-related factors that may affect a surgeon's performance? Would using such a model be effective in ameliorating surgical outcomes? Unspecific methods that include feedback on the major complications related to a surgical operation have a strong impact on surgical performance and costs. However, the studies' non-randomized design and their limited number reduce the findings' significance [2].

The American Aggregate Physician societies and the United States government agencies identified useful criteria to standardize surgical performance evaluation and consequently healthcare quality improved. These guidelines were delineated in the National Quality Forum and mainly applied to cardiac surgery [3]. Despite these meritable efforts carried out by the National Quality Forum staff, many patient-centered aspects still require improvements. Surgical performance measurement ought to take into consideration: (i) patients are and just not mortality rates; and (ii) the surgical decision-making processes' value [4].

Another important point is to unify patient, institutional and scientific perspectives. One of the most crucial benefits from this system involves the reduction of complications. With regards to the methodology, mortality rate is among the most used measures. This may be misleading as it is based on the presumption that all mortality in surgery can be prevented, however proper risk adjustments are lacking [5]. Mortality as a result of a surgical procedure is rather rare. Focusing mainly on this aspect does not allow enough consideration for complications arising from variations in surgical techniques and surgical performance.

Another issue related to assessing mortality rates is the incongruence between expected and observed rates. Variable life adjusted display (VLAD) is a commonly used parameter, however it has its limitations with regards to timing features and its monitoring retrospective nature [6]. The data collection is based upon arbitrary time intervals and the negative trends in surgical performance. It could be identified with a relevant delay, causing inevitable inefficiency in quality improvement interventions.

Outputs support evidence over time. Poor surgical performance or surgical errors that hinder the final result is most likely to arise when it lies close to the control range extremities, rather than when the value is adjacent to the average. Although VLAD provides an easily understandable display, the model's limitations lead to significance and accuracy loss [7].

An alternative evaluation method is the risk-adjusted Bernoulli cumulative sum (RA-CUSUM), which is set on a real-time prospective observation, preventing the use of retrospective methods [8]. The RA-CUSUM is an attempt to apply the cumulative sum graphical method, used for quality monitoring in industrial settings, to the surgical field. It becomes significant across multiple cases and adopts run length distribution to evaluate performance

changes. A small cohort may increase the discrepancy between the observed and the expected mortality rate [9].

Contrarily to VLAD, which is a mortality-scoring system (i.e., a penalty is assigned for every death) based on perioperative death risk, the RA-CUSUM method identifies patients with improved surgical outcomes with respect to the expected ones. With this method, any failure in surgical performance can be reliably detected when compared to previous successful ones. Any deterioration in surgical performance is expressed as a positive slope, sending a signal when values above the upper control limit appear [10]. Risk adjustment is obtained by a model confronting the outputs with a statistical mortality risk and adverse events in the same operation, according to patient comorbidities and ana graphical characteristics (age, sex), and other factors related to an increase in operative complications [7,11]. Adverse events appertain to an improvised risk score that involves many variables, however it is not based on a single-case, representing a rough approximation to assess real risk related to the individual patient. It is not wise to assume that risk data extracted from literature fully supports the prognosis in uncertain contexts, such as complications attributed to poor surgical performance or events occurring during hospitalization.

This approach also neglects the pre-operative, environmental and team-related features, as surgical outcome complexity does not necessarily show a linear correlation with procedural quality. Hence, the variability related to the learning curve and the need for continued monitoring makes this method less practical.

Forced ranking distribution methods are useful only if applied constructively to attenuate the subjective and discretionary performance evaluation in different surgical departments. An apparently safe tool for ranking surgeons and medical staff fails to grasp service complexity. It should not be the only method used to classify surgical skills.

Healthcare competition has existed for some time now and positively correlates with the rise in life expectancy. Forced ranking distribution system has the potential to incentivize best practice in surgery. However, potential drawbacks in this system could be present, for example there may be a tendency to deliberately avoid high-risk procedures to limit uncertain outcomes.

Further complications may be caused by tension between competing team members and a delay in recognizing adverse events occurring after the procedure. The above-mentioned factors are prone to delay the improvement process involving surgical service and productivity, which is the forced ranking system's goal. A vitality curve detractor, Michael Schrage argues that this system, leads to 'dishonest and unfair evaluations by management' [12].

Conclusion

Given measurement accuracy and the healthcare system's peculiarity in an industrial scenario, work performance monitoring strategies must be revised and rankings should be a complementary performance index. Among the current methods available,

prospective RA-CUSUM monitoring is preferred and can play a role in improving outcomes. A strong commitment and collaboration with surgeons is required to implement a monitoring system in an ethical and goal-directed fashion.

References

1. Murray A. Should I Rank My Employees? <http://guides.wsj.com/management/recruiting-hiring-and-firing/should-i-rank-my-employees/>. Accessed April 2016. 2009.
2. Maruthappu M, Trehan A, Barnett-Vanes A, McCulloch P, Carty MJ. The Impact of Feedback of Surgical Outcome Data on Surgical Performance: a Systematic Review. *World J Surg*. 2015;39(4):879-889.
3. Sherlaw-Johnson C, Gallivan S, Treasure T, Nashef SA. Computer Tools to Assist the Monitoring of Outcomes in Surgery. *Eur J Cardiothorac Surg*. 2004;26(5):1032-6.
4. Hyder JA, Roy N, Wakeam E, Hernandez R, Kim SP, Bader AM. Performance Measurement in Surgery through the National Quality Forum. *J Am Coll Surg*. 2014;219(5):1037-1046.
5. Kiernan F and Rahman F. Measuring Surgical Performance: A Risky Game? *Surgeon*. 2014;13(4):213-217.
6. Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the Results of Cardiac Surgery by Variable Life-adjusted Display. *Lancet*. 1997;350(9085):1128-1130.
7. Steiner SH, Woodall W.H. Debate: What is the Best Method to Monitor Surgical Performance? *BMC Surgery*. 2016;16:15.
8. Cook DA, Steiner SH, Cook RJ, Farewell VT, Morton AP. Monitoring the Evolutionary Process of Quality: Risk-adjusted Charting to Track Outcomes in Intensive Care. *Crit Care Med*. 2003;31(6):1676-1682.
9. Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring Surgical Performance Using Risk-adjusted Cumulative Sum Charts. *Biostatistics*. 2000;1(4):441-452.
10. Harris JR, Forbes TL, Steiner SH, Lawlor DK, Derosé G, Harris KA. Risk-adjusted Analysis of Early Mortality after Ruptured Abdominal Aortic Aneurysm Repair. *J Vasc Surg*. 2005;42(3):387-91.
11. Forbes TL, Steiner SH, Lawlor DK, DeRose G, Harris KA. Risk-adjusted Analysis of Outcomes Following Elective Open Abdominal Aortic Aneurysm Repair. *Ann Vasc Surg*. 2005;19(2):142-8.
12. Schrage M. How the Bell Curve Cheats You. http://archive.fortune.com/-magazines/fortune/fortune_archive/2000/02/21/273841/index.htm accessed April 2016. 2000.