Arab Journal of Urology

ISSN: (Print) 2090-598X (Online) Journal homepage: https://www.tandfonline.com/loi/taju20

Systematic methods for measuring outcomes: How they may be used to improve outcomes after Radical cystectomy

Khurram M. Siddiqui & Jonathan I. Izawa

To cite this article: Khurram M. Siddiqui & Jonathan I. Izawa (2015) Systematic methods for measuring outcomes: How they may be used to improve outcomes after Radical cystectomy, Arab Journal of Urology, 13:2, 122-127, DOI: 10.1016/j.aju.2015.02.008

To link to this article: https://doi.org/10.1016/j.aju.2015.02.008

Arab Journal of Urology (2015) 13, 122–127

Arab Journal of Urology
(Official Journal of the Arab Association of Urology)

www.sciencedirect.com

ONCOLOGY/RECONSTRUCTION
REVIEW

Systematic methods for measuring outcomes: How they may be used to improve outcomes after Radical cystectomy

Khurram M. Siddiqui *, Jonathan I. Izawa

Division of Urology, Department of Surgery, University of Western Ontario, London, Ontario, Canada

Received 23 January 2015, Received in revised form 24 February 2015, Accepted 28 February 2015
Available online 29 March 2015

KEYWORDS

Quality assurance;
Systematic methods;
Cumulative summation;
Charts

ABBREVIATIONS

RC, radical cystectomy;
M&M, mortality and morbidity;
POSSUM, (Portsmouth) Physiological and Operative Severity

Abstract In the era of managed healthcare, the measuring and reporting of surgical outcomes is a universal mandate. The outcomes should be monitored and reported in a timely manner. Methods for measuring surgical outcomes should be continuous, free of bias and accommodate variations in patient factors. The traditional methods of surgical audits are periodic, resource-intensive and have a potential for bias. These audits are typically annual and therefore there is a long time lag before any effective remedial action could be taken. To reduce this delay the manufacturing industry has long used statistical control-chart monitoring systems, as they offer continuous monitoring and are better suited to monitoring outcomes systematically and promptly. The healthcare industry is now embracing such systematic methods. Radical cystectomy (RC) is one of the most complex surgical procedures. Systematic methods for measuring outcomes after RC can identify areas of improvements on an ongoing basis, which can be used to initiate timely corrective measures. We review the available methods to improve the outcomes. Cumulative summation charts have the potential to be a robust method which can prompt early warnings and thus initiate an analysis of root causes. This early-warning system might help to resolve the issue promptly with no need to wait for the report of annual audits. This system

* Corresponding author at: Clinical Fellow, University of Western Ontario, Urology, 800 Commissioners Road, London, Ontario N5x4L4, Canada. Tel.: +1 2263772551; fax: +1 5196858511.
E-mail addresses: Khurram.Siddiqui@lhsc.on.ca, khurram_aku@yahoo.com (K.M. Siddiqui).

Peer review under responsibility of Arab Association of Urology.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.aju.2015.02.008

Score for the enUmeration of Mortality and morbidity;
VLAD, variable life-adjusted display;
(RA-)CUSUM, (risk-adjusted) cumulative summation;
RCA, root-cause analysis

can also be helpful for monitoring learning curves for individuals, both in training or when learning a new technology.

Introduction

The Institute of Medicine described the quality of care as '*the extent to which health services are provided to individuals and patient populations, to improve the desired health outcomes, the care should be based on the strongest clinical evidence and provided in a technically and culturally competent manner with good communication and shared decision making*' [1]. Individual surgeons and surgical departments have always taken pride in the quality of care they offer, and increasingly they are now mandated to report them to peer-review organisations and third-party payers. The objective measurement of quality is a routine practice in manufacturing industries, but is not an easy task, especially for a service industry such as healthcare. Methods for measuring outcomes are critically dependent on the quality of reporting systems. It is fundamental for these systems to be free of bias and be able to endure rigorous statistical analysis.

In medical science these stringent methods should not only be able to accommodate variations in patient factors, but should also be able to identify the disparities among different surgeons and institutions. In many situations this problem is compounded by a lack of consensus on defining 'a complication', and furthermore, there is little agreement on the comparative benchmarks. The traditional methods of surgical audits are periodic and for most procedures are typically annual. Therefore, there is a long time lag before any effective remedial action could be taken. To reduce this delay, manufacturing industry has used statistical control-chart monitoring systems, as they offer continuous monitoring and are better suited for monitoring outcomes systematically and promptly.

The healthcare industry is also beginning to use these methods to improve outcomes. The first report of the application of control charts in assessing surgical procedures was in cardiac surgery [2]. Radical cystectomy (RC) is one of the most complex procedures in urological surgery. During the course of treatment and recovery after RC, a patient uses many hospital services. An uncomplicated course is a reflection of the quality of all the services provided in an integrated and efficient manner, including the surgical technique, which is a critical component and is potentially responsible for many complications. The expertise in this procedure might be used to reflect on the surgical abilities and the learning curve of an individual surgeon. Systematic methods for measuring the outcomes after RC can identify areas of improvements on an ongoing basis, which can be used to initiate timely corrective measures. We review the currently available methods used to improve the outcomes after RC.

Mortality and morbidity (M&M) meetings

The proceedings of M&M meetings have been traditionally used to evaluate the quality of surgical care. These meetings determine the quality of management before, during and after surgery, the main objective being educational, but they lack a systematic process of initiating effective corrective measures to improve any system errors. High-risk surgical procedures like RC are often the focus of debate in these meetings. There are lengthy discussions on events around surgical operations, where negative or adverse outcomes are debated. Due to the sensitive nature of the discussion most of these meetings are not multidisciplinary. These meetings are often criticised as being 'incestuous' and hence are limited in the scope of the outcomes open for debate [3]. These meetings are also resource-intensive, as input is required from highly trained individuals. The value of this traditional M&M method is questionable in current times, as the bar of measuring 'quality' has been raised by other industries.

Surgical audits

Surgical audits determine the incidence of postoperative complications. These periodic audits are either initiated on demand or are ongoing, and they help to determine the 'point prevalence' of an outcome. The incidence of complications is used as a surrogate marker of quality. Sometimes there can be some disagreement on the definition of a 'complication' and these definitions might be variably used at different centres. In urology there are no standard guidelines or criteria to report outcomes [4]. To acquire credible evidence a standardised system is needed to objectively report outcomes, as this is the backbone of any systematic method. The Clavien–Dindo grading system for characterising and grading morbidity is a comprehensive system which has gained popularity in the reporting on quality of care [5]. It allows an objective comparison of surgeons and can be used to compare the performance of institutions. The graded complications can be used to illustrate to patients the anticipated course after a urological procedure [6].

In 2012, the European Association of Urology published guidelines on the reporting and grading of complications after urological procedures [7]. If these guidelines are implemented, with the help of statistical testing it will be possible to define factors affecting the outcomes. The frequency of adverse outcomes could also be used to define benchmarks for reporting the quality of care. However, this method is periodic and there is a long lag before any effective remedial actions could be taken.

Predictor equations: POSSUM

The main challenges in assessing the quality in healthcare are the variability in individual characteristics, risk factors, the ability to tolerate stress, and recuperation. However, the attempt is to produce similar outcomes for all. The crude morbidity and mortality frequency values have little meaning, as two different institutions or individuals might be providing care for patients with completely different risk factors. This variability was elaborated by Sommer et al. [8], who assessed the outcomes of two urologists, who were performing similar urological procedures, including RC. Although one of them had a higher crude morbidity rate, the risk-adjusted prediction was similar. This highlights the importance of risk stratification for assessing outcomes. One of the risk-adjusted predictor equations commonly used is termed 'POSSUM' (Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity). POSSUM was first described by Copeland et al. in 1991 [9], and uses 12 physiological and six operative variables to predict morbidity and the 30-day mortality. However, it was observed that POSSUM over-predicted postoperative mortality, particularly in patients at low risk, and this led to a revision by Whitely et al. [10], known as the Portsmouth modification (P-POSSUM). Although risk-adjusted equations are better than crude incidences, they have the same limitations of periodicity, which renders them less effective for the timely institution of corrective measures.

Variable life-adjusted display (VLAD)

The VLAD is a longitudinal and systematic method of continuous monitoring [11]. Complications can be monitored by displaying the difference between the expected and observed outcomes against the total number of patients. If wound infection is assessed as an outcome, then each observed outcome, i.e., wound infection, assumes the value of 1, or 0 for an uneventful outcome. The expected outcome (E) assumes the value of $p0$ and an adverse outcome is calculated as $(1 - p0)$. The chart will therefore try to balance a negative outcome with a positive one. The limitation of this system is that with this method all patients are assumed to have equal surgical and medical risks. VLAD does not incorporate a threshold and hence cannot trigger an alarm when unacceptable levels are approached.

Cumulative mortality plots

These plots are a simple continuous-monitoring chart system. The observed outcome is assumed to be 0 and death is scored as 1. If all outcomes are uneventful the graphical chart will show a flat line, and if all are 'mortality' then it will be a 45° curve [12]. These plots also lack the capacity to raise an alarm, as benchmarks are not incorporated.

Funnel plots

These plots use benchmarks defined by CIs, and are commonly used in meta-analyses. The CI is set for surgeons or centres and outcomes falling beyond the limits are depicted [13]. The construction of Funnel plots is time-consuming and their use in quality assurance is limited. They are more suited for periodic reviews.

Cumulative summation (CUSUM) charts

The CUSUM (cumulative sum control chart) is a statistical tool for quality control, designed to monitor a change in real time. This systematic method was first reported to be used in the healthcare industry in 1950 [14]. CUSUM has been used effectively for quality assurance in diagnostic laboratories. It was later introduced to monitor outcomes for surgical procedures, the first surgical application being in cardiothoracic surgery [15]. This system has also been used to monitor the learning curve for a new technology, e.g., robotic prostate surgery [16].

CUSUM charts can monitor any binary outcomes, and warning signals are initiated as they approach the predetermined benchmark. To construct CUSUM charts, the type 1 and type 2 error limits are set. For each outcome to be measured, benchmark thresholds for acceptable and unacceptable outcomes are determined. For example, RC has a mortality of 1–3% [17], so if this benchmark is breached the CUSUM generates a warning.

This continuous-monitoring system performs a statistical analysis after each procedure is recorded, the null hypothesis being that the outcome is as expected and the alternative hypothesis is that the adverse outcome has occurred. The odds ratio is calculated, control limits are set, and if they are breached an alarm is initiated. Simulation-based calculations are made for control average-run lengths. The out-of-control average run length is the average number of patients required to prompt CUSUM charts to initiate a signal when an unacceptable level of adverse outcome is reached by the calculation of the odds ratio. By fine tuning the control average-run lengths it is possible to increase the sensitivity to initiate a warning and detect even smaller differences.

A major limitation of this method is that all cases are treated as having the same risk of having complications, but the risk stratification can be used to modify the weight of adverse outcomes. However, due to the diversity of the covariates in a patient population, this risk-adapted CUSUM system is unable to assign weights for all outcomes.

Most of the outcomes that are used to assess the quality of care for RC are binary outcomes, and include wound infection, sepsis, wound dehiscence, myocardial infarction, pulmonary embolism, deep vein thrombosis, pneumonia, blood transfusion, unplanned re-operation, prolonged ileus, rectal or bowel injury, and anastomotic leaks, including the ileo-ileal and uretero-ileal anastomosis. Death is another binary outcome that can be easily monitored.

Chalasani et al. [18] reported the outcomes for 150 consecutive RCs performed by one surgeon, using CUSUM charts. We used the method described by Rogers et al. [19] and show the results in both formats, i.e., cumulative-failure charting and the cumulative

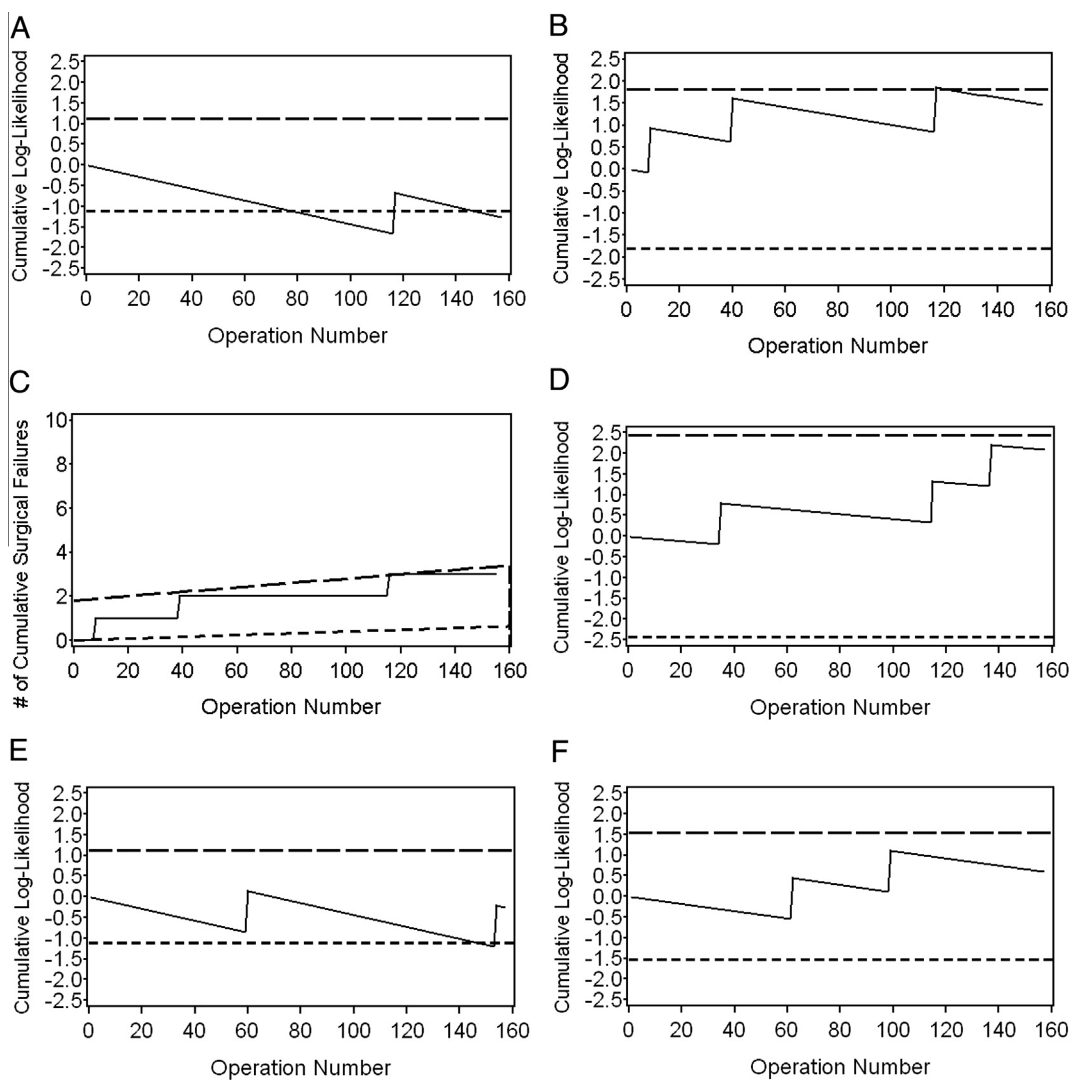

Figure 1 Cumulative log-likelihood ratio charts for (A) death within 30 days of RC, and (B) pulmonary embolus. (C) A cumulative failure chart for pulmonary embolus. (D–F) cumulative log-likelihood ratio charts for uretero-intestinal leak, unplanned re-operation and myocardial infarction, respectively.

log-likelihood ratio. Fig. 1A, shows the 30-day mortality for RC. The benchmarks are shown, and if the graph crosses the bottom control line then the performance is confirmed to be acceptable, and if it overshoots the upper limit then it is unacceptable. Another example, that of pulmonary embolism as an outcome, is shown Fig. 1B and C, where the information on pulmonary embolism is shown in both formats. There were three cases of pulmonary embolism in this series. The boundary lines were set with an acceptable pulmonary embolism rate of 0.4% and an unacceptable rate of 2%; as the plot has not yet crossed both the upper and lower limit, no conclusions can be drawn, and monitoring is recommended to continue. Fig. 1D–F shows CUSUM graphs for leakage of the uretero-ileal anastomosis, unplanned re-operations and postoperative myocardial infarctions. These clearly show that CUSUM charts can be effectively used for monitoring all outcomes after RC.

Once an outcome crosses the bottom line, as shown for the mortality rate in this report, it can be less rigorously followed. Similarly, surgeons with an overall morbidity rate approaching the upper limit might be selected for audits and resources diverted to an in-depth root-cause analysis (RCA) of their outcomes. These RCAs should include risk stratification of the patients, to avoid undue penalisation for surgeons operating on high-risk patients.

Another limitation of this method is that if complications are encountered earlier in the monitoring process, the CUSUM charts will raise the alarm of unacceptable complication rates. Like all other systems this requires careful interpretation.

Risk-adjusted CUSUM charts (RA-CUSUM)

The basic difference in the method of calculating the RA-CUSUM is that points are allocated to the high-risk group which are redeemed at the time the plots are constructed. For example, a patient is classified as having a high risk of death, e.g., 30%. If he or she survives, then 0.30 points are rewarded in addition. But if he or she dies, then instead of having a score of −1 for mortality, the risk-adjusted CUSUM subtracts 0.30 and calculates it as 0.70. This theoretically adjusts the slope of the curve.

Although the basic concept is not difficult, the construction of RA-CUSUM chart control limits involves many statistical nuances, making it a complex task. Various models for making these calculations have been proposed, but despite the potential advantage of the RA-CUSUM compared to the unadjusted analysis, the statistical model for risk adjustments is not perfect and cannot eliminate all confounding variables [12,20]. Forbes et al. [21] used a three-tier risk-stratification system, classifying the risk as low, intermediate or high, to compare the observed vs. predicted mortality. This type of system helps to choose control limits for each group. The use of a validated risk-stratification system is the cornerstone in constructing objective and unbiased control limits. In unpublished data from our centre, we used the surgical mortality probability model to calculate the risk of mortality, and the Gupta Score for calculating perioperative cardiac events [22,23]. We analysed 50 patients undergoing RC, and found no difference in true event rates and mean calculated risks. In our experience, the RA-CUSUM is an efficient and effective technique for the dynamic monitoring of outcomes after RC.

Conclusion

There is no disagreement on the need for effective systematic methods for measuring the outcomes after a high-risk surgical procedure such as RC. Systematic methods have the potential to address gaps in the management of bladder cancer in particular and all urological procedures in general. We reviewed previous reports of the currently available systematic and non-systematic approaches. CUSUM charts have been shown to longitudinally monitor the outcomes after RC. This is a robust method which can be designed to monitor any outcome of interest. Alarms are set taking into account the established benchmarks, and if they are violated a timely warning is generated to initiate a RCA. The early detection of unwanted outcomes might help to resolve these issues in a timely manner, with no need to wait for the report of an annual audit. This system can also be helpful in monitoring the learning curves of individuals, both in training or when learning a new technology.

Conflict of interest

None.

Source of funding

None.

References

[1] Pelletier LR, Beaudin CL. *Q Solutions: Essential Resources for the Healthcare Quality Professional.* 2nd ed, 3. Glenview, IL: National Association for Healthcare Quality; 2008.

[2] 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**:914–24.

[3] Campbell WB. Surgical morbidity and mortality meetings. *Ann R Coll Surg Engl* 1988;**70**:363–5.

[4] Gakis G, Efstathiou J, Lerner SP, Cookson MS, Keegan KA, Guru KA, et al. ICUD-EAU International Consultation on Bladder Cancer 2012. Radical cystectomy and bladder

preservation for muscle-invasive urothelial carcinoma of the bladder. *Eur Urol* 2013;**63**:45–57.

[5] Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004;**240**:205–13.

[6] Yoon PD, Chalasani V, Woo HH. Use of Clavien–Dindo classification in reporting and grading complications after urological surgical procedures: analysis of 2010–2012. *J Urol* 2013;**190**:1271–4.

[7] Mitropoulos D, Artibani W, Graefen M, Remzi M, Rouprêt M, Truss M. European Association of Urology Guidelines Panel. Reporting and grading of complications after urologic surgical procedures: an ad hoc EAU guidelines panel assessment and recommendations. *Eur Urol* 2012;**61**:341–9.

[8] Sommer F, Ehsan A, Klotz T, Haupt G, Caspers HP, Engelmann U. Comparison of individual urologists' performance. *Eur Urol* 2001;**39**:369–74.

[9] Copeland GP, Jones D. Walters. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991;**78**:355–60.

[10] Whiteley MS, Prytherch DR, Higgins B, Weaver PC, Prout WG. An evaluation of the POSSUM surgical scoring system. *Br J Surg* 1996;**83**:812–5.

[11] 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**:1128–30.

[12] Noyez L. Control charts, CUSUM techniques and funnel plots. A review of methods for monitoring performance in healthcare. *Interact Cardiovasc Thorac Surg* 2009;**9**:494–9.

[13] Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med* 2005;**24**:1185–202.

[14] Levey S, Jennings ER. The use of control charts in the clinical chemical laboratory. *Am J Clin Pathol* 1950;**20**:1059–66.

[15] Williams SM, Parry BR. Schlup MMT. Quality control. An application of the CUSUM. *Br Med J* 1992;**304**, 1359–136.

[16] Williams AK, Chalasani V, Martínez CH, Osbourne E, Stitt L, Izawa JI, et al. Cumulative summation graphs are a useful tool for monitoring positive surgical margin rates in robot-assisted radical prostatectomy. *BJU Int* 2011;**107**:1648–52.

[17] Wein AJ, Kavoussi LR, Novick AC, Partin AW, Peters CA, editors. *Campbell-Walsh Urology*. 10th ed. Philadelphia: WB Saunders; 2010.

[18] Chalasani V, Abdelhady M, Stitt L, Izawa JI. Quality assurance and benchmarking for radical cystectomy: monitoring early complications and mortality using cumulative summation charts. *J Urol* 2009;**181**:1581–6.

[19] 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.

[20] Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk adjusted cumulative sum charts. *Biostatistics* 2000;**1**(4):441–52.

[21] 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**:142–8.

[22] Glance LG, Lustik SJ, Hannan EL, Osler TM, Mukamel DB, Qian F, et al. The Surgical Mortality Probability Model. derivation and validation of a simple risk prediction rule for non-cardiac surgery. *Ann Surg* 2012;**255**:696–702.

[23] Gupta PK, Gupta H, Sundaram A, Kaushik M, Fang X, Miller WJ, et al. Development and validation of a risk calculator for prediction of cardiac risk after surgery. *Circulation* 2011;**124**:381–7.