Use of cumulative summation (CUSUM) as a tool for early feedback and monitoring in robot assisted radical prostatectomy outcomes and performance

Garson Chan[1];Stephen E. Pautler[1,2]
[1]Division of Urology, Department of Surgery; [2]Division of Surgical Oncology, Department of Oncology; Western University, London, ON, Canada

***Acknowledgements**: Thank you to Workhorse Consulting Group Andrew McClure and Norine Foley for statistical advice and consulting, as well as Dr. Marie Dion and Linda Nott for compiling and assistance with the exhaustive RARP database.*

Cite as: *Can Urol Assoc J* 2018 July 31; Epub ahead of print.
http://dx.doi.org/10.5489/cuaj.5350

Published online July 31, 2018

***

Abstract

Introduction: Today's surgical practice has evolved, with increasing emphasis on quality assurance. Many forms of quality-control monitoring have been suggested, but they are often impractical or difficult to implement. Cumulative summation (CUSUM) is a simple method to provide visual feedback before significant quality issues arise. We present our initial use and practical application of CUSUM in a surgical practice.
Methods: A retrospective analysis was applied to a prospectively collected database of 577 sequential patients who have undergone robot-assisted radical prostatectomy from a single surgeon over a 10-year period. Outcome measures were analyzed with CUSUM, which included a composite complication score, continence rates, length of hospital stay, biochemical recurrence, and need for adjuvant radiation. If any outcomes were out of control, they would cross the CUSUM failure line.
Results: CUSUM chart plotting for incontinence demonstrated an initial upward slope followed by trending to a new safety limit. Additionally, outcomes in complications and biochemical recurrence did not reach the established safety boundaries. Length of stay and radiation outcomes did initially cross the safety line, but were improved over time.
Conclusions: The use of CUSUM in clinical practice can fulfill the need for quality assurance. CUSUM plotting in our practice reflected the initial learning curve, followed by ongoing maintenance and improvement in performance. These changes were consistent with the implementation of changes in surgical techniques. Although this tool was used retrospectively, this strengthens our argument to implement this tool prospectively and assess real-time

refinement of surgeon skill. We have demonstrated that CUSUM can be appropriately used to assure quality control in a surgical practice.

Introduction

There is a shift in modern medicine with increasing importance on evidence based practice and quality assurance. This change has encouraged approaches to identify objective measures of quality, and accountability. However, there is a practical issue applying these standards and being able to evaluate clinical outcomes in real world practice.[1] The use of cumulative summation (CUSUM) has been suggested specifically for both surveillance and quality control.

A CUSUM chart or graph is a visual representation of a trend in the outcomes of a series of consecutive procedures over time. It is designed to quickly detect changes in performance. The measured outcome must be binary (i.e. did the event occur → yes or no?). This is then easily graphed and allows useful feedback. Every outcome is plotted and allows for early trends to be visualized. Other advantages of CUSUM charting include tracking learning curves, competency, and quality assurance. CUSUMs allow early detection of small changes to allow corrective action if necessary, thus it is a timely warning system. The binary system is easy to use, objective to record, and intuitive. It can allow simple comparisons between current and past performance, or between trainees, or consultants.[11]

CUSUM has been used in many different industries, and more recently emphasis has been placed on its application in medicine.[1-4] In fact, CUSUM has been recommended by the Royal Australasian College of Surgeons as a tool for self-analysis and audit.[5] Since then, more recent publications of CUSUM in urology have followed, but implementation remains limited. [6-9] We believe that CUSUM can be practically used in a surgical practice to assess quality outcomes. Our objective was to apply this CUSUM charting to analyze and evaluate our performance in robot-assisted radical prostatectomy (RARP) patient outcomes.

Methods

All patients had a RARP at Western University in London, Ontario by a single fellowship trained uro-oncologist surgeon. Patients between April 2005 and August 2015 were prospectively included in the present study (REB#13086E). A prospective database of 577 patients was collected that included patient demographics, as well as postoperative factors, and postoperative complications. We retrospectively looked at several outcomes including a composite complication score (CCS), continence rates, length of stay (LOS), use of radiation, and biochemical recurrence (BCR). The CCS compiles all post-operative RARP outcomes within the first year. This includes transfusions, peritonitis, urinary tract complications, need for gastrointestinal stoma creation, strictures or contractures, fistula formation, hospital readmission, and death. The score is based on a recent 2014 Ontario health initiative working group looking at all prostatectomy outcomes in Ontario (Table 1).[12]

Statistical analysis was performed and CUSUM graphs were generated retrospectively. CUSUM charting was used to assess surgeon performance and outcomes in comparison to acceptable standards from the literature. These charts were constructed with the methodology described by Rogers.[13] If performance is unacceptable, the CUSUM slopes upward. In this way, CUSUM graphs represent cumulative failure charting. Any sustained slope changes are signals providing early identification that a quality issue might exist. Type I and II error rates were set at 0.05%, in keeping with previous literature. These rates are displayed as control lines which act as boundaries for our curves. If the graph rises above the 'unacceptable' line then performance is said to be unacceptable, but if the graph remains between the two control lines then monitoring should continue.

The difficulty in constructing CUSUMs is in setting acceptable and unacceptable rates. However, this can also be seen as an advantage because what might be acceptable in one patient population might be unacceptable in another. We chose limits based on a combination of recent literature, expert opinion and local consensus. It is important to realize that an acceptable rate is not solely a level that a surgeon aims to but rather is also a level at which a surgeon has reached a standard the surgeon believes is safe. Consideration can then be given by the surgeon to resetting the graph so they do not get into 'credit' for their previously low rates.

We chose boundaries for post RARP incontinence (PRI) at one year with ≥10% as an unacceptable range, based on >1 pad per day.[14] We defined BCR at >25% based on the AUA definition of having confirmatory PSA value ≥0.2 ng/ml.[15,16] We also looked at radiation with >25% post RARP.[17-19] We considered LOS based on <3 days, starting from patient entering the hospital to discharge and a CCS ≥10%. Again, these were binary outcomes (i.e. yes or no).

Results

CUSUM graphs were generated for 577 patients on outcomes of LOS, CCS, BCR, radiation, and PRI. The graphs illustrate simply the various changes in surgical outcomes for our patient population over a consecutive 10 year period. Figure 1 demonstrates the curves for length of stay with the red lines acting as the upper failure boundary. Figure 1A illustrates the absolute curve, with Figure 1B plotted on a cumulative log-likelihood ratio test. Similarly Figures 2-5 are illustrated with red acting as the failure line for the outcome. When the line is crossed as in Figures 1, 4 and 5, the process is out of control. This is the warning marker and signals the need for corrective action. The CUSUM graph in Figure 1 approached and crossed the upper boundary near patient 150 and stayed elevated until after patient 200. Figures 2 and 3 for CCS and BCR did not cross the upper boundary. The radiation CUSUM graph shown in Figure 4 did cross the red boundary quite early near patient 20, and did not decline until after patient 140. Most interestingly the PRI CUSUM chart in Figure 5 showed an upper boundary crossing near patient 120, with an abrupt drop near patient 150, and another sharp decline around patient 375.

Discussion

Surgical quality and quality assurance have become increasingly important in urology. The surgeon is responsible for balancing the needs of the individual patient, the hospital, as well as cost implications when introducing a new technique or surgical technology. The use of tools for self appraisal allows for objective measurements to assess individual ability, and progression or deterioration. This performance feedback is invaluable for physicians doing technically demanding and complex procedures. However, this is an area that unfortunately receives little attention during surgical training and continued medical education. This area is becoming increasingly important in our professional development as the system we serve demands more tangible methods to allocate scarce resources. In addition, this technique identifies areas for improvement enhances self-esteem and develops self-awareness for the surgeon. Importantly, it allows users to identify reasons for discrepancies and may become a positive platform for career promotion and progression. For regulatory bodies, this may supplement supervisory systems that lack regular direct supervision. In these numerous ways, CUSUMs are a substantially beneficial tool that is underutilized.

From our CUSUM results, we demonstrate visuals that are clear and simple. The graphs clearly depict changes in rates over the surgical series and are excellent markers of changes in surgical technique. Although the curve did approach and cross the unacceptable rate for LOS, radiation and PRI, ongoing technical improvements allowed the curves to once again return to acceptable rates. We believe part of the 'unacceptable' performance reflected a short learning curve for RARP. For instance in the first hundred cases there is an upward slope in LOS, radiation, and incontinence rates, seen in Figures 1, 4 and 5, with improvement over time and with increased caseloads. The CCS CUSUM is within limits and even better over time with an eventual plateau (Figure 2).

In Figure 1 LOS, the early upslope is likely related to the learning curve: robot and room set up time, operating time, and post operative course as patients were initially managed with the same post operative pathway as patients who underwent open radical prostatectomy. However over time, the pathway for RARP recovery changed as reflected in the improvement seen in the CUSUM graph. There are also slope changes that we discovered correlate accurately with resident and fellow rotation changes. This is not likely the only cause but highlights how this interesting application alerts us that a change happened and allows us to investigate further.

When we consider the CUSUM graphs for BCR and radiation (Figures 3 and 4) we notice an early rise noted in radiation (Figure 4), which may correspond to different practice patterns, treatments or patient preference over a decade ago. However, when compared to Figure 3 for BCR it is difficult to determine conclusively the effect radiation had on BCR.

Interestingly, the CUSUM for incontinence rates shows a marked decline around patients 150 and 375 seen in Figure 5. This first change may be attributable to the introduction of a new technique – introduction of a posterior reconstruction ("Rocco Stitch") – and the positive impact on outcomes. A second downward slope correlates again with another modification in technique. In this case, the change was the introduction of a barbed suture for the running anastomosis and a posterior reconstruction with simultaneous puboperineoplasty to create an autologous sling. These are just some examples of how CUSUM can help in self analysis and easy visualization to identify changes that may need attention.

Limitations include the retrospective nature of how the CUSUM values were applied. To be of greater value a prospective analysis in real-time would be applied. However, this paper illustrates the important clinical role that may promote the acceptance of these tools into daily practice. Additionally, these results validate the positive impact of subtle technical modifications to the RARP technique. Although this study was performed retrospectively, implementation of key changes was clearly evident during the CUSUM analysis. This further promotes the use of this tool prospectively, providing preventative and corrective analysis. With this quality framework, we will continue to prospectively collection information with confidence in using the CUSUM technique to analyze ongoing quality control and correct deviations early on if necessary. Our data affirms the next step in our usage of this tool and implementing real time monitoring and analysis. We encourage surgeons to join us and engage in this quality assurance measure.

CUSUM is objective and easily understood but is limited to only quantitative measures with binary outcomes. CUSUM will not show improvements that do not change the binary outcome. This discounts incalculable values such as patient care, patient satisfaction, communication and teamwork. It can be time consuming and difficult to completely compare CUSUM from different surgeons due to multiple factors; a heterogeneous patient population, health system factors (hospitals, staffing, etc) and other factors resulting in diverse outcomes. Thus, CUSUMs are best applied to individual surgeons' outcomes. CUSUMs unfortunately do not directly measure the contributions by trainees such as residents and fellows; where a new assistant might lead to significantly longer OR time, blood loss, etc. While these contributions should even out over time, the CUSUM calculation might be affected. In addition, while crossing the lower boundary is considered to be positive, the graph should be reset to bring the CUSUM back to baseline so that the surgeon will not get 'credit' for previous low rates of the measured outcome and they can continue to be accurately monitored. This 'reset' allows for continuous monitoring and improvement. As well, this tool was able to confirm desirable competency of surgeon performance and patient selection despite complex patients at a high volume tertiary referral centre. Lastly, there is sparse literature on individual surgeon performance using CUSUM analysis to compare our outcomes to in a meaningful manner.

Conclusion

CUSUM can be a simple yet dynamic and versatile technique that allows us to identify the trend of deteriorating performance early, prompting preventative and corrective measures. This is useful during the learning curve and for quality control in ongoing surgical performance. It also allows accountability and reassurance to our patients and health system. CUSUM methodology can be applied to RARP as the procedure is common, the outcome is well defined, and closely related to surgeon skill. Future studies will include a real time prospective CUSUM analysis of outcomes from heterogeneous groups of patients using a risk adjusted or difficulty adjusted CUSUM process, which is more time consuming and complex. However even this simple

CUSUM application is able to detect important changes, thereby imploring us to investigate further and stay proactive with respect to surgical quality assurance.
Evidence-based medicine and quality self assurance is guiding the future of our health care. CUSUM is a valuable tool that is easy to use for a variety of outcomes. In this current study RARP outcomes were analyzed to identify quality variances and identified favourable outcome responses to technical changes. This instrument encourages accountability to our patients and for the use of health resources. There are challenges to CUSUM implementation but this alone should not impede our usage of this valuable tool. Future real time use of these quality instruments will guide our quality assurance and accountability.

References

1. Steiner SH, Woodall WH. Debate: what is the best method to monitor surgical performance?. *BMC surgery* 2016 Apr 5;16:15.
2. Bolson S, Colon M. *Int J Health Care Qual Assur* 2000;12:433-8.
3. Institute Of Medicine. Crossing the quality chasm: a new health system for the 21st Century. *Institute of Medicine's Quality of Health Care in America project* 2001.
4. Biau DJ, Resche-Rigon M, Godiris-Petit G, et al. Quality control of surgical and interventional procedures: a review of the CUSUM. *Quality and Safety in Health Care* 2007 Jun 1;16:203-7.
5. Surgeons.org. The Royal Australasian College of Surgeons: Surgical audit and peer review, c2008. http://www.surgeons.org/media/66599/surgical_audit_peer_review.pdf Accessed June 6, 2017.
6. Siddiqui KM, Izawa JI. Systematic methods for measuring outcomes: How they may be used to improve outcomes after radical cystectomy. *Arab J Urol* 2015; 13:122-7.
7. Williams AK, Chalasani V, Martinez CH, 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.
8. Maguire T, Mayne CJ, Terry T, et al. Analysis of the surgical learning curve using the cumulative sum (CUSUM) method. *Neurourol Urodyn* 2013; 32:964-7.
9. Nabi G, McLornan L, Cook J, et al. MP-16.14: Endoscopic extrapritoneal radical prostatectomy: CUSUM analysis of learning curve. *Urol J* 2009; 74:S123-4.
10. Cho SY, Choo MS, Jung JH, et al. Cumulative sum analysis for experiences of a single-session retrograde intrarenal stone surgery and analysis of predictors for stone-free status. *PloS one* 2014 Jan 14; 9:e84878.
11. Young A, Miller J P, Azarow K. Establishing learning curves for surgical residents using Cumulative Summation (CUSUM) Analysis. *Curr Surg* 2005 Jun 30; 62:330-4.
12. Ontario Health Technology Advisory Committee (OHTAC). Robotic-assisted minimally invasive prostatectomy: OHTAC recommendation. Toronto: Queen's Printer for Ontario; 2014 January.
13. Rogers CA, Reeves BC, Caputo M, et al. Control chart methods for monitoring cardiac surgical performance and their interpretation. *J Thorac Cardiovasc Surg* 2004 Dec 31; 128:811-9.
14. Ficarra V, Novara G, Rosen RC, et al. Systematic review and meta-analysis of studies reporting urinary continence recovery after robot-assisted radical prostatectomy. *Eur Urol* 2012 Sep 30; 62:405-17.
15. Pound CR, Partin AW, Eisenberger MA, et al. Natural history of progression after PSA elevation following radical prostatectomy. *JAMA* 1999 May 5; 281:1591-7.

16. Stephenson AJ, Kattan MW, Eastham JA, et al. Defining biochemical recurrence of prostate cancer after radical prostatectomy: a proposal for a standardized definition. *J Clin Oncol* 2006 Aug 20; 24:3973-8.
17. Zincke H, Oesterling JE, Blute ML, et al. Long-term (15 years) results after radical prostatectomy for clinically localized (stage T2c or lower) prostate cancer. *J Urol* 1994 Nov; 152:1850-7.
18. Stephenson AJ, Scardino PT, Kattan MW, et al. Predicting the outcome of salvage radiation therapy for recurrent prostate cancer after radical prostatectomy. *J Clin Oncol* 2007 May 20; 25:2035-41.
19. Bottke D, Wiegel T. Adjuvant radiotherapy after radical prostatectomy: indications, results and side effects. *Urol Int* 2007; 78:193-7.

Figures and Tables

Fig. 1A. Plot of cumulative failures for the outcome of length of hospital stay greater than 3 days (n=577 patients). The plotted line (black) is the cumulative failure. The acceptable and unacceptable failure rates were set at 10% (blue) and 30% (red), respectively.

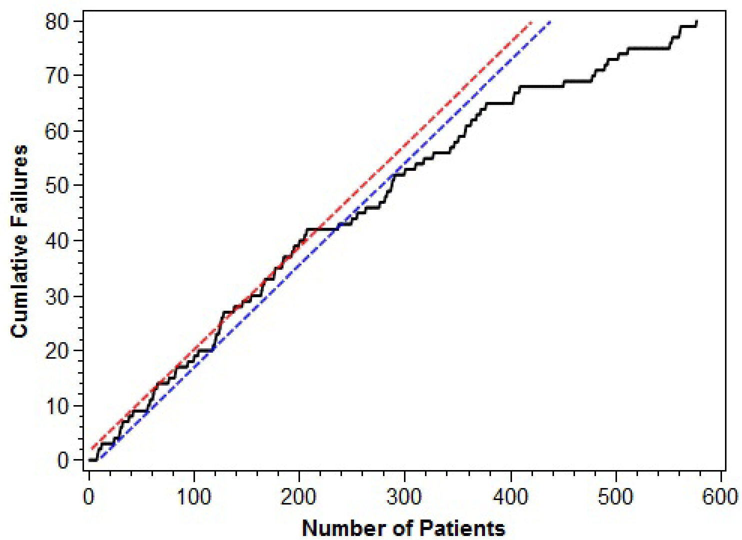

Fig. 1B. Plot of cumulative log likelihood ratio test for the outcome of length of hospital stay greater than 3 days (n=577 patients). The plotted line is the cumulative failure. The acceptable and unacceptable failure rates were set at 10% (blue) and 30% (red), respectively.

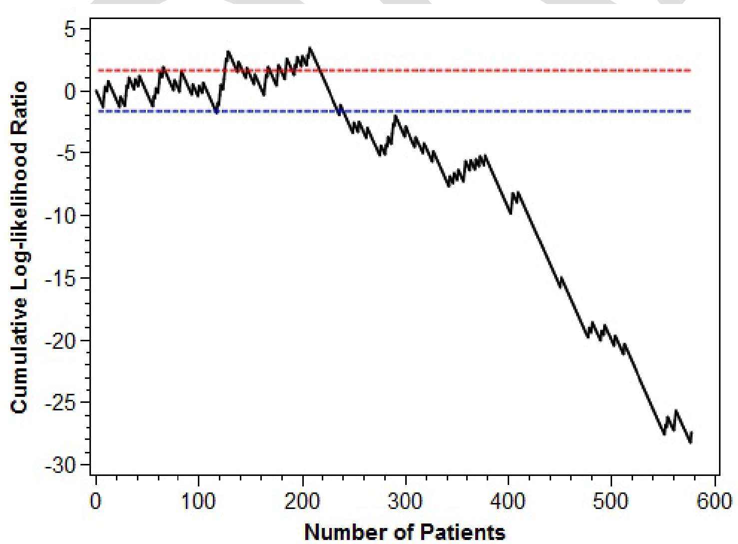

Fig. 2A. Plot of cumulative failures for the outcome of complications within 5 years of procedure (n=563 patients). The plotted line (back) is the cumulative failure. The acceptable and unacceptable failure rates were set at 5% (blue) and 10% (red), respectively.

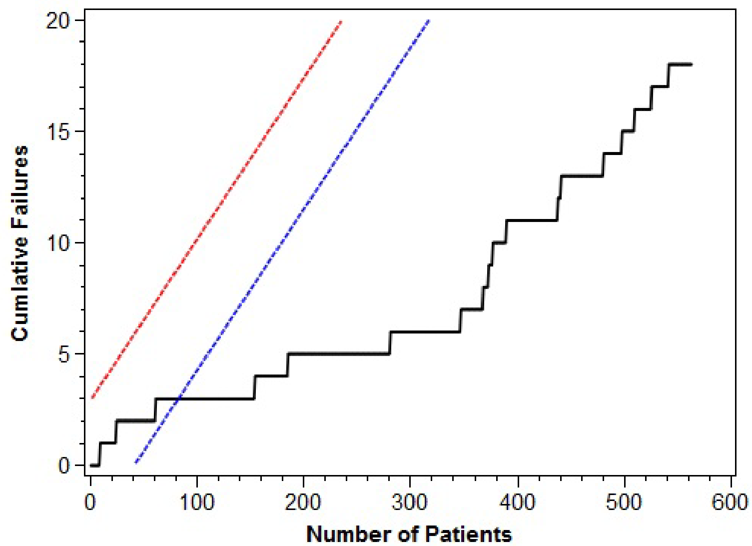

Fig. 2B. Plot of cumulative log likelihood ratio test for the outcome of complications within 5 years of procedure (n=563 patients). The plotted line (back) is the cumulative failure. The acceptable and unacceptable failure rates were set at 5% (blue) and 10% (red), respectively.

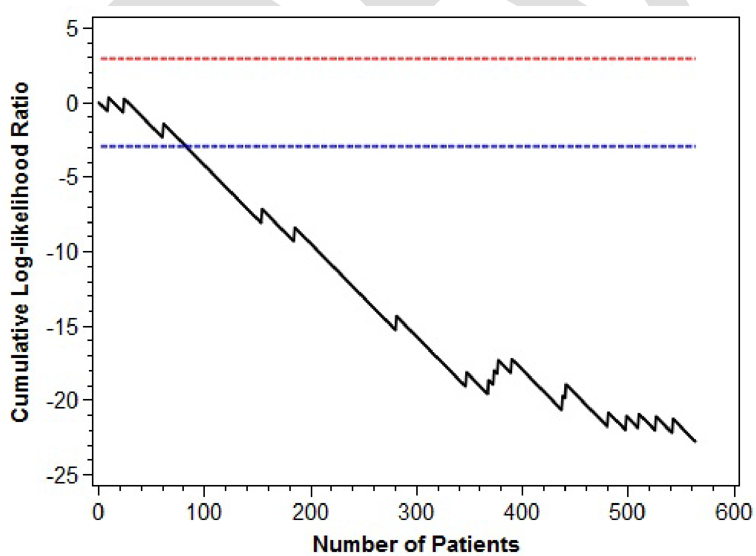

Fig. 3A. Plot of cumulative failures for the outcome of PSA >0.2x2 over 10 years (n=563 patients). The plotted line (back) is the cumulative failure. The acceptable and unacceptable failure rates were set at 5% (blue) and 25% (red), respectively.

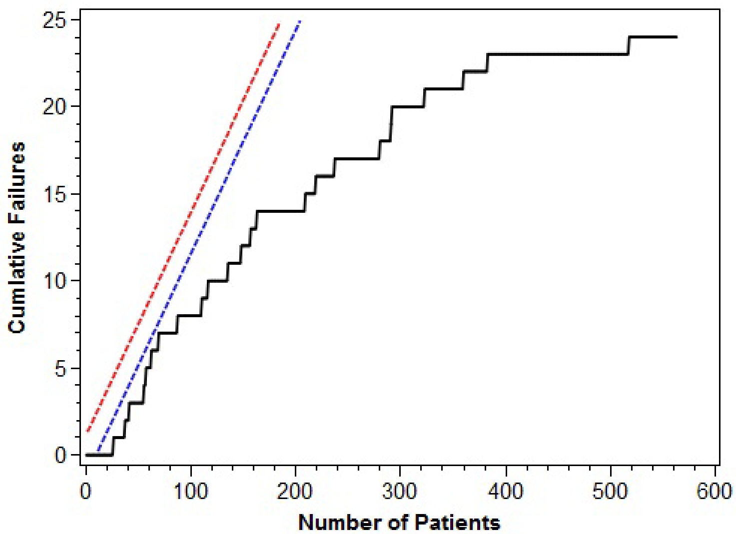

Fig. 3B. Plot of cumulative log likelihood ratio test for the outcome of PSA >0.2x2 over 10 years (n=563 patients). The plotted line (back) is the cumulative failure. The acceptable and unacceptable failure rates were set at 5% (blue) and 25% (red), respectively.

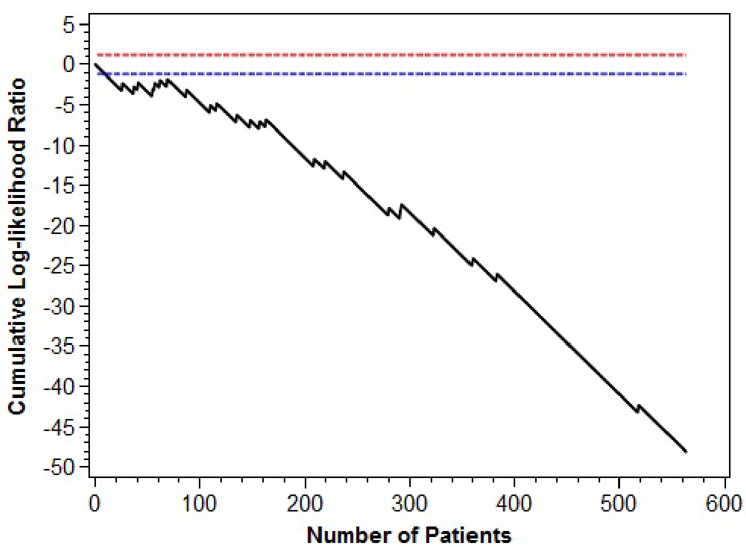

Fig. 4A. Plot of cumulative failures for the outcome of the need for adjuvant radiation therapy (n=563 patients). The plotted line (back) is the cumulative failure. The acceptable and unacceptable failure rates were set at 5% (blue) and 25% (red), respectively.

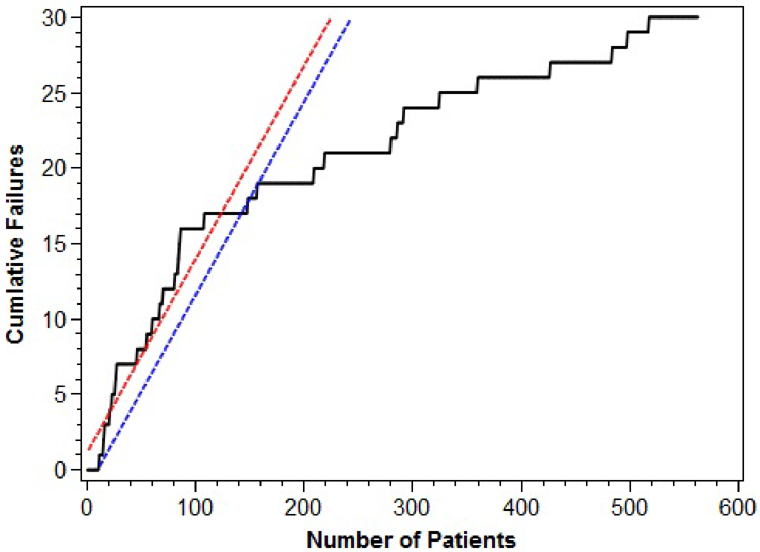

Fig. 4B. Plot of cumulative log likelihood ratio test for the outcome of the need for adjuvant radiation therapy (n=563 patients). The plotted line (back) is the cumulative failure. The acceptable and unacceptable failure rates were set at 5% (blue) and 25% (red), respectively.

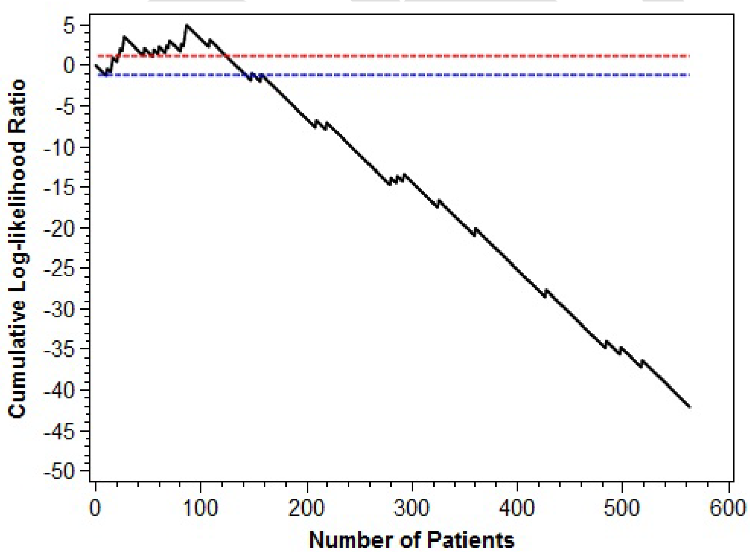

Fig. 5A. Plot of cumulative failures for the outcome of PPD ≤1 (n=488 patients). The plotted line (back) is the cumulative failure. The acceptable and unacceptable failure rates were set at 5% (blue) and 10% (red), respectively.

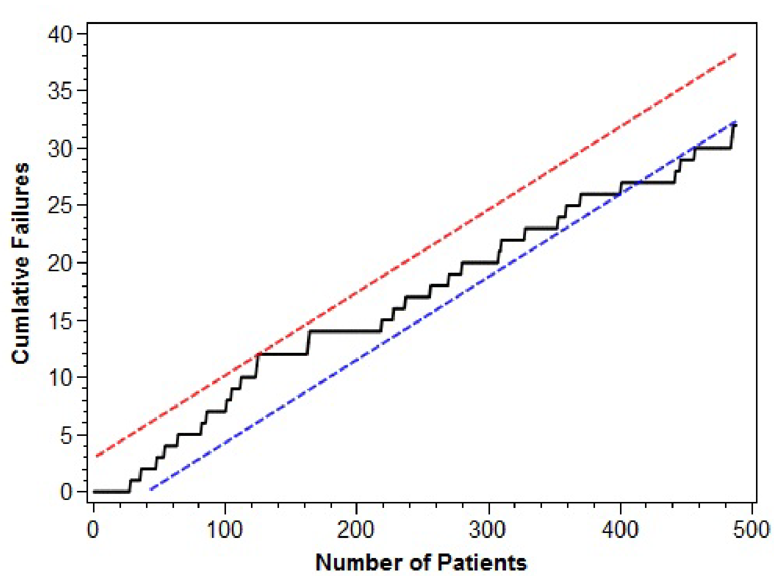

Fig. 5B. Plot of cumulative log likelihood ratio test for the outcome of PPD ≤1 (n=488 patients). The plotted line (back) is the cumulative failure. The acceptable and unacceptable failure rates were set at 5% (blue) and 10% (red), respectively.