Chapter VII
Innovations in the Assessment of Transplant Center Performance: Implications for Quality Improvement

Overview

- The transplant community has been a pioneer in periodic review of clinical outcomes to ensure the optimal use of limited donor organs.
- Through data collected from the OPTN and analyzed by the SRTR, transplantation professionals have intermittent access to specific, accurate, and clinically relevant data that provides information to improve transplantation.
- Statistical process control techniques including cumulative sum charts (CUSUM) are designed to provide continuous, real-time assessment of clinical outcomes.
- CUSUMs can be constructed that provide risk-adjusted program-specific data to inform quality improvement programs.
- When retrospectively compared to currently available data reporting, the CUSUM method was found to detect clinically significant changes in center performance more rapidly, which has the potential to inform center leadership and enhance quality improvement efforts.

Introduction

Continuous quality improvement programs have emerged as key components of high performing health care organizations. The collection and timely analysis of clinically relevant data is crucial in accomplishing quality improvement initiatives and ensuring the highest quality care for patients. Organ transplantation is unique among medical specialties in the quantity and quality of data collected on a national basis. Through data collected from the Organ Procurement and Transplantation Network (OPTN), accurate and clinically relevant data are available that can provide information to improve transplantation outcomes. Currently, however, these data are analyzed on an episodic basis and provided to centers semiannually through the Program-Specific Reports (PSRs). Consequently, recognition of clinically relevant changes in clinical outcomes may be delayed, limiting the success of quality improvement efforts.

Statistical process control charts were originally developed to study industrial processes in the 1930s by W. A. Shewhart and his colleagues at Bell Laboratories (1). These charts measure performance over time and "signal" if there is a deviation from accepted production standards. The CUSUM, or cumulative sum, chart was introduced in 1954 by Page and provided a very sensitive approach to monitoring a process and identifying changes in outcome (2). The purpose of these charts is to give timely and easily interpreted summaries of outcome data. The potential utility of CUSUMs in health care was recognized in the early 1970s when the paper, "Why don't doctors use CUSUM?" was published in the *Lancet* (3). Several years later, the *New England Journal of Medicine* published a manuscript which highlighted the value of CUSUM techniques in clinical applications (4). Broad acceptance of these techniques, however, was delayed initially by data collection limitations and subsequently by the inability to include meaningful risk adjustment.

Recent high profile events, including a cluster of heart transplant deaths in Britain (5) and the Institute of Medicine's report, "Crossing the Quality Chasm" (6), have contributed to a heightened awareness of the need to monitor surgical outcomes. Public interest and recent improvements in the analytic methods have led to rapid increases in the utilization of CUSUMs to track surgical outcomes. In a 2007 review, Biau and colleagues identified 31 studies which utilized CUSUMs to track surgical outcomes in cardiac, general, and ENT surgery (7). Other work on CUSUM methods and applications in medical studies can be found in various statistical and medical journals (8, 9, 10).

Application of CUSUMs to the management of transplant centers offers physician leaders the opportunity to track outcomes in a real-time, risk-adjusted manner. A previous retrospective analysis demonstrated that CUSUMs identified changes in clinical practice sooner and with higher sensitivity than current center monitoring techniques (11). Recent improvements in chart construction are based on survival analysis techniques and allow the incorporation of outcomes as they occur, rather than after passage of a specific time period (e.g., one year post-transplant) as had previously been the case (12). Furthermore, these CUSUM charts are risk adjusted using comprehensive models currently employed by the SRTR to adjust outcomes for patient and donor characteristics and are adjusted for patient mix. In general, the CUSUM compares observed outcomes with expected results;

it increases in value as graft failures or patient deaths occur and decreases during periods with no failures. If too many failures occur over time (compared to what would be expected) the value of the CUSUM will exceed a pre-determined threshold value and "signal" that a process review should be initiated.

This chapter provides a brief overview of the construction and application of CUSUM charts for transplant professionals. We begin with a brief summary of the construction and interpretation of CUSUM charts, including a number of examples. Following this, we briefly review the methods currently used by the SRTR to assess transplant programs in the-Program Specific Reports. Next, we present a retrospective comparison of CUSUM monitoring with the techniques employed in the PSRs. Finally, we address the strengths, as well as potential difficulties and risks, of a broad application of this technique.

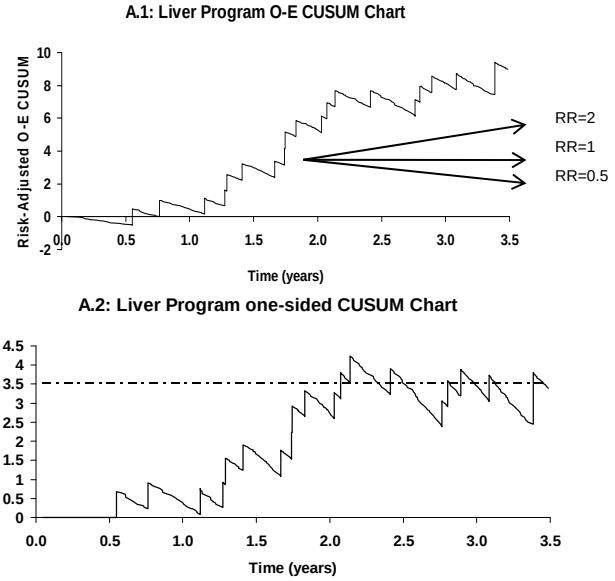
Defining CUSUMs

A CUSUM chart for a given program or center presents a simple graphical comparison of observed and expected numbers of events over time. In their initial description of a clinically relevant, risk-adjusted CUSUM, Steiner et al. and Grigg et al. described methods for assessing performance of a clinical system that produced a binary outcome (e.g., death following cardiac surgery) (10, 13, 14). In these methods, the CUSUM is increased or decreased by a variable degree depending upon the observed and expected outcomes from the process. Axelrod et al. applied this approach to transplant data and developed CUSUM charts to monitor one-year post-transplant survival for liver transplant recipients and one-year allograft survival for kidney transplant recipients (11). Based on the method of Steiner et al. (13), a logistic regression model, which included several donor and recipient factors, was utilized for risk adjustment.

Binary outcome CUSUM charts are very useful tools for monitoring situations in which the outcome is binary and rapidly ascertained; for example, in monitoring conversion rates (the percent of possible donors which actually result in transplantable organs), acceptance rates (the percent of organ offers which are accepted by a program), or mortality rates over a short, fixed period of time. They do have a disadvantage: in monitoring longer term survival outcomes such as one- or two-year mortality rates, the data on any given individual cannot be used until the corresponding period has elapsed.

In 2008, Biswas and Kalbfleisch developed a method to create risk-adjusted CUSUM charts that are based on a continuous time survival analysis approach and that are able to incorporate deaths or graft failures as they occur (12). These charts have a substantial advantage in the monitoring of longer term survival endpoints and are more consistent with the Cox model based risk-adjusted methods used by the SRTR. This method can be utilized to construct two types of CUSUM methods: a one-sided chart in which the value is restricted to non-negative sums and a two-sided or O-E chart, as described below; a more complete description of the calculation of the CUSUMs is included in the appendix.

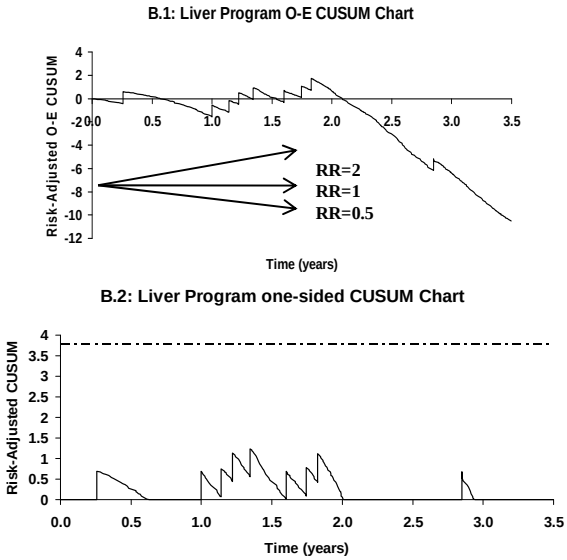
Figure VII-1: Panel A.1 (Top); O-E CUSUM Chart For One-year Mortality in the Liver Program at Center A. Panel A.2 (Bottom); One-sided CUSUM Chart For One-year Follow-Up in the Liver Program at Center A*



* 200 Liver Transplants Within a 3.5-year Period

The one-sided CUSUM is constructed principally to assess for a clinically significant excess of allograft failures or patient deaths. The CUSUM is restricted to positive values and so is bounded below at a value of zero. Thus, a center performing at, or better than, the expected performance (and thus having fewer observed failures than expected) would have a chart which tends to stay relatively close to zero. It would increase with any failure, and then return to zero in an ensuing period without failures. Conversely, if the center's outcomes are much poorer than the national average, the number of failures will lead to a substantial increase in the CUSUM and this will eventually result in a signal. The one-sided CUSUM signals when the plot line crosses a horizontal line, termed the control limit, which defines the signaling threshold. The height of this line (L) reflects the balance between rapid signaling that will very quickly identify centers with poor outcomes (sensitivity) and the desirability of avoiding false positive results (specificity) and signaling when the center's performance is actually consistent with the national average. The value of L can be adjusted for center volume to ensure that the sensitivity is kept at a suitably low level for all centers. Figures VII-1 to VII-3 (bottom panels) provide examples of the one-sided CUSUM. In our analysis, if a kidney transplant center's volume is 40 transplants/year or more and its true rate is twice the adjusted national average, the chance of a signal over a three-year period is 90% or more. If the center's rates are the same as the national average, a period of 30 years would be expected before the first false positive signal.

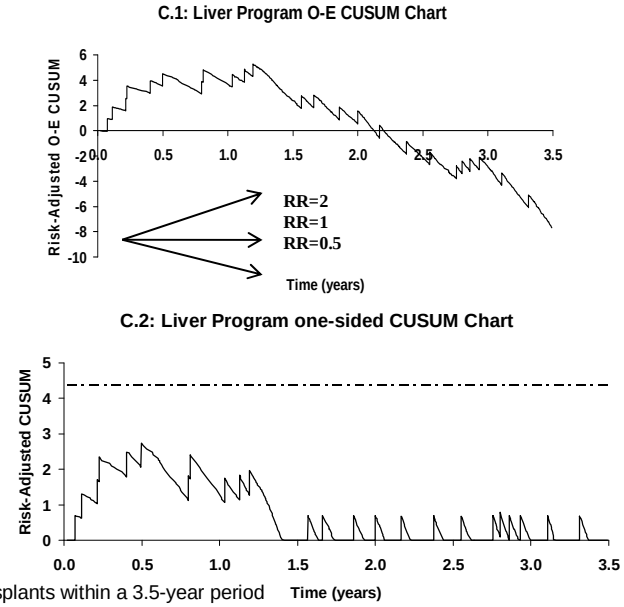
Figure VII-2: Panel B1 (top); O-E CUSUM Chart for the Liver Program at Center B. Panel B2 (bottom); One-sided CUSUM Chart for the Liver Program at Center B*



* 256 Liver Transplants within a 3.5-year period

The two-sided or O-E CUSUM. For a given center, this simplest of CUSUM plots, as a function of time, the difference between the observed number of deaths and the number of deaths that would be expected based on the risk adjusted national average. This CUSUM can be viewed as being updated daily by adding to its previous value the observed number of deaths on that day less the expected number. The expected number of deaths is estimated from a survival model based on national data and is adjusted for the particular patient mix at the center. Thus, the O-E chart traces out an approximately horizontal path (slope=0) if the death rate at the center is close to the risk-adjusted national average. An upward trend of the O-E plot over a specified time interval indicates that the center has worse outcomes than the adjusted national average. Conversely, a downward trend of the plot corresponds to better outcomes than the adjusted national average. Figures VII-1 to VII-3 (top panel) provide examples of O-E plots and are discussed further below.

Figure VII-3: Panel C1 (top); O-E CUSUM Chart for the Liver Program at Center C. Panel C2 (bottom); One-sided CUSUM Chart for the Liver Program at Center C*



*414 Liver Transplants within a 3.5-year period

As indicated by the arrows superimposed on the plots, the 'slope' of the plot gives an estimate of the approximate relative risk (RR), which is the ratio of the death rate at the center to that for the adjusted national average.

It is possible to use the two-sided CUSUM to provide a signal when there is statistical evidence that a center's outcomes are different from the national average. The two-sided CUSUM signals if the 'slope' of the plot exceeds a predetermined value over an extended period. The signal is obtained by systematically checking the slope of the plot at each successive time using a V-mask as introduced by Barnard (15,16) and discussed in the appendix. As in the height of L in the one-sided chart, the angle of slope which is considered significant in the two-sided chart can be adjusted to balance sensitivity and specificity.

The principal advantage of the O-E plot is that the slope of the plot over a given interval gives an immediate picture of the relative rate of outcomes within the center of interest compared to expected results. In the one-sided CUSUM, the slope of the chart is more difficult to interpret and immediate comparisons to the national average or expected results are more difficult to see. We find that both charts provide useful and complementary information, as the examples in the next subsection illustrate.

Examples of CUSUM Charts and Interpretation

A sample of CUSUM charts over a three-year period for liver transplant programs, labeled Center A, Center B, and Center C, are described here. For each center, the one-sided CUSUM and O-E CUSUM charts are presented for one-year patient survival. Similar charts could be constructed for graft survival or for other outcomes such as one-month or two-year survival.

Center A: From the O-E chart, the failure rate for the one-year survival in Center A is close to the national average for the first year, as is suggested by the nearly horizontal plot line (slope approximately 0) (Figure VII-1). For the second period, from one year to three and a half years, the death rate exceeded the national average, as illustrated by an increase in the slope of the O-E CUSUM. From the one-sided CUSUM chart (Figure VII-1), we see that these trends would have led to a signal at the end of the second year. If the chart had been in place, this signal suggests a review of center practices may be appropriate.

Center B: The O-E chart in Figure VII-2 illustrates that Center B experienced death rates very close to the national average (adjusted for patient mix) over the first two years of the CUSUM period. During the last year and a half, the center had considerably better one-year outcomes than the adjusted national average. Correspondingly, the one-sided CUSUM chart (Figure VII-2) did not signal.

Center C: From the O-E chart (Figure VII-3), we see that the one-year mortality at the center was higher than the national average for the first year or so. After that period, the one-year death rates were considerably lower than the adjusted national rates through the end of the three and a half year period. From the one-sided CUSUM (Figure VII-3), we see that the higher death rates observed early on were not sufficient to lead to a signal.

Current Methods of Risk Adjustment and Quality Assessment: SRTR Processes

Although there is no "gold standard" for performance assessment, the methods utilized in the PSRs prepared by the SRTR for the UNOS Membership and Professional Standards Committee provide an important benchmark for comparison (17). However, it is important to keep in mind a key difference between the CUSUM methods discussed here and the PSRs. The PSRs are a regulatory tool used to help ensure compliance with current performance standards; they *are not intended nor constructed to be used as a quality improvement instrument.* The PSRs supplied by the SRTR help the MPSC identify transplant programs or organ procurement organizations (OPOs) that might require site visits or case reviews to look more deeply into potential problems, whereas the CUSUM procedures can be used by the center directors themselves for real-time monitoring and quality improvement efforts.

Broadly speaking, the MPSC seeks to identify programs which experience significant deviations from expected performance measures related to the care of wait-listed and transplanted patients. Centers identified are characterized as needing further review if they meet the following criteria:

Clinical outcome failures that exceed predetermined thresholds relative to expected performance. Currently the thresholds include an excess death rate of at least 50% (Observed / Expected > 1.5);

A clinically important number of incidences, defined as an absolute number of excess deaths greater than 3;

Statistical confidence that the observed difference between observed and expected results is unlikely to have occurred by chance (one-tailed P-value < .05).

The SRTR utilizes a robust data set including all observed events during the time a patient is actually followed, either by the center and the OPTN or through other data sources including the Social Security Death Master File. Kidney graft survival data are supplemented by examination of claims data related to return to dialysis from the Centers for Medicare & Medicaid Services (CMS). This methodology essentially eliminates the possibility of un-captured recipient death or kidney allograft failure events.

All data provided to the MPSC are risk-adjusted using Cox proportional hazards regression models, including donor data, recipient characteristics, and transplant variables. For example, for the one-year deceased donor adult kidney patient survival model, the proportion of the variation explained by the model was 73% for the July 2008 release. It is important to note that the SRTR models used by the MPSC do not adjust for perioperative or post-transplant management practices (e.g., type of induction and immunosuppression). These factors, while clearly influencing outcomes, are within the control of the transplant center and contribute to the center's performance that is examined by the MPSC. The models are designed to control for differences in the underlying recipient severity of illness and donor quality, which are largely determined by the location of the center and the donor characteristics within the OPO.

The program review criteria were set by the MPSC in order to facilitate identification of programs for which interventions were likely to have a demonstrable clinical impact. The criteria are stringent to reduce the risk of a false positive finding in which outcomes vary from those expected by chance alone. Because smaller centers with poor outcomes may not meet the three thresholds despite poor outcomes, all centers performing nine or fewer transplants in a 2.5 year cohort are evaluated by a separate standard, whose only criterion is at least one adverse event (graft failure or death). The review thresholds for both small and large programs are reviewed periodically by the SRTR and the MPSC.

The MPSC examines quarterly data on a 2.5-year rolling average of outcomes. In order to have complete data on one-year survival, for example, the Program-Specific Reports for the MPSC and center directors are somewhat time delayed. Consequently, these analytic techniques are not well suited to day-to-day management of transplant center outcomes by center directors. Previous analyses have demonstrated that significant clusters of surgical failures may be missed utilizing only episodic analysis of average outcomes, particularly for large centers. Given the regulatory and legal implications of MPSC actions, criteria that are highly reliable and very specific are crucial. However, for the purpose of center management, a sensitive, continuous, up-to-date outcomes tracking tool would allow better assessment of center performance. CUSUM techniques may complement the PSRs, and would potentially allow center directors to determine trends in performance in an expedited manner.

Application to Liver and Kidney Outcome Assessment using the OPTN Database

To assess the potential value of CUSUM monitoring of transplant center performance, we performed retrospective analyses of one-year patient survival at liver and one-year graft survival at kidney transplant centers. Two separate charts were constructed to detect declining center performance: a one-sided continuous CUSUM chart and an O-E CUSUM chart with a V-mask (see Appendix). For each center, the value of L and the slope of the V-mask were defined so that they would provide a signal approximately 8% of the time in a 3.5-year period in centers of the same size whose one-year survival rates are identical to the overall national average. This “false positive rate” is equivalent to the false positive rate that would arise from the methods used to prepare the PSRs.

Data Sources

Data from the cohort of recipients of deceased donor transplants between July 1, 2004 and December 31, 2007 were reviewed for kidney and liver transplant programs in the OPTN database. The data included 11,957 liver transplant recipients transplanted at 67 centers which ranged in size from 1 to 185 liver transplants per year over the three and a half year period. We omitted 9 centers with less than 8 transplants per year, for which the CUSUMs would be expected to yield little power. The SRTR models for post-transplant survival used for the PSRs were used to determine the expected number of failures risk-adjusted to correspond to the center’s donor and recipient characteristics. The SRTR’s one-year survival model for liver transplants from a deceased donor adjusts for 23 donor and recipient characteristics, whereas the model for living donor liver transplants adjusts for 7 donor and recipient characteristics. This study was approved by HRSA's SRTR project officer. HRSA has determined that this study satisfies the criteria for the IRB exemption described in the "Public Benefit and Service Program" provisions of 45 CFR 46.101(b)(5) and HRSA Circular 03.

We also considered results from 113 kidney transplant programs that included 31,666 recipients transplanted between July 1, 2004 and December 31, 2007. These programs had average volumes of 1 to 278 deceased and living donor kidney transplants per year. We included the facilities with 8 to 220 transplants per year in our study and grouped them into six categories. This omitted 11 small programs with fewer than 8 transplants per year, and two large centers with 278 and 225 transplants per year. Again, we utilized the SRTR’s PSR models to represent national rates for graft failure; the model for deceased (living) donors includes 22 (13) donor and recipient characteristics. Details of both the liver and the kidney models can be found on the SRTR website (*www.ustransplant.org*).

For each program, the control limit L and V-mask were defined based on the center’s volume (see appendix) and it was determined whether the CUSUMs produced a signal (or flag). In addition, a benchmark was based on a review of the Program-Specific Reports to determine whether or not a flag occurred for each program in the study. The time to flagging for the PSR was taken as 3 years or 3.5 years depending on whether the flag occurred in the January 2008 or the July 2008 report.

Results

Results of this comparison are presented for five volume strata for liver transplant programs (Table VII-1). There is very close agreement between the CUSUM methods; in fact, they both identified exactly the same centers and at almost identical times. However, the results differed slightly from those obtained from the current PSR approach. There were two centers that were flagged by the CUSUMs that were not identified by the PSR approach, due to clusters of surgical failures which were not

Table VII-1: The number of facilities flagged and average time to flagging in the Program Specific Reports compared to methods based on one-sided and O-E CUSUMs for one-year survival in 58 liver transplant programs.

Facility size		# and average time for PSR flagging		# and average time for one-sided CUSUM		# and average time for O-E CUSUM	
transplants/yr	*# facilities*	*#*	*time (yr)*	*#*	*time (yr)*	*#*	*Time (yr)*
8-20	4	1	3.50	1	1.76	1	1.76
20-50	20	2	3.00	2	0.68	2	0.68
50-100	25	6	3.25	8	2.06	8	2.01
100-140	7	1	3.00	1	3.27	1	3.27
140-185	2	0	NA	0	NA	0	NA

captured using the PSR methods. For the centers which were flagged as having below expected performance, the average time needed to reach the signal point was less than two years for the CUSUMs, compared to more than 3 years for the PSR approach. In this case, the CUSUMs signaled slightly more often than the SRTR's PSR approach (12 versus 10).

Similar results were generated when these methods were applied to the 100 kidney transplant programs using one-year graft survival as the outcome of interest (Table VII-

Table VII-2: The number of facilities flagged and average time to flagging in the Program Specific Reports compared to methods based on one-sided and O-E CUSUMs for one-year survival in 100 kidney transplant programs.

Facility size		# and average time for PSR flagging		# and average time for one-sided CUSUM		# and average time for O-E CUSUM	
transplants/yr	*# facilities*	*#*	*time (yr)*	*#*	*time (yr)*	*#*	*Time (yr)*
8-20	3	0	NA	1	1.00	1	1.00
20-50	27	5	3.20	8	1.51	8	1.52
50-100	37	2	3.00	5	1.92	5	1.91
100-140	17	0	NA	3	0.75	3	0.75
140-180	9	1	3.00	3	1.03	3	1.03
180-220	7	1	3.00	1	1.22	1	1.22

2). In this case, both CUSUMs signaled at the same 21 centers, compared to 9 centers that were flagged by the PSR. Here again, the average time to signal was much shorter for the two CUSUM approaches. In the 12 centers signaled in the CUSUMs but not in the PSR, 8 only slightly exceeded the control limit, 1 center signaled after 3 years when the PSRs are not fully complete, and the other 3 centers exceeded the control limit by a substantial amount.

These data suggest that there is general agreement between the CUSUMs and the PSRs in determining which centers have not achieved expected levels of performance, though the CUSUMs did signal somewhat more often than the PSRs using the signal limits established here, especially in the kidney outcomes. This is not surprising, perhaps, given that the aim of the CUSUMs is to be an early warning of potential difficulties. In general, the CUSUMs identified centers much more quickly, often one to two years earlier, than the PSRs and could have provided an opportunity for earlier intervention on the part of center personnel.

Conclusions and Discussion

The medical community is under increasing pressure to develop and implement effective strategies to assess and improve performance. The solid organ transplant community has already established its leadership in this area, by collecting and publicly reporting clinically valid, risk-adjusted center-specific outcome data. Currently, these data are primarily utilized in a regulatory fashion and are not particularly well designed for improving practice in an ongoing fashion. The implementation of a real-time, clinically relevant system of outcomes monitoring with CUSUMs may accelerate efforts to improve both transplant outcomes and organ utilization.

The CUSUMS discussed here depend on appropriate risk adjustment and require substantial analytical development to support. In the case of post-transplant graft and patient survival, there is substantial experience in risk evaluation and adjustment through the ongoing efforts of the SRTR and the OPTN organ-specific committees. As noted earlier, however, it is important to continually review these models to be sure that important baseline patient characteristics are included. If the center leadership misunderstands the CUSUM results, there is the potential of inappropriate reactions, such as a decision to limit access to high risk transplants. It should be noted, however, that appropriate adjustment takes these high risks into account in defining the CUSUMs and limiting access to high risk transplants would not necessarily result in any measured improvement in center performance.

The signaling thresholds in the one-sided and O-E CUSUM are designed to help the clinician understand when the differences are of sufficient magnitude that a review is suggested and help avoid premature or unjustified reaction to apparent trends. As proposed here, the CUSUM would not be a regulatory tool. Rather, as a quality improvement tool, it can be used to assess the clinical impact of changing practices (e.g., accepting higher risk donors). Thus, the CUSUM can help the center avoid coming to the attention of the MPSC and will not increase scrutiny. As noted, there is excellent

agreement between the methods in identifying which centers are at risk; the real difference in the two methods is the rate of ascertainment.

CUSUM charts have been clinically applied in a variety of surgical applications in single center series to assess outcomes of cardiac surgery (18-21), colorectal surgery (22), breast surgery (23), and other specialties (7). In general, CUSUM charts have been used for two purposes: first, to assess the learning curve inherent in the adoption of new surgical procedures (e.g., laparoscopic nephrectomy) (24), and second, to assess the outcome of a system of care within a single institution or across a group of providers. In the case of the learning curve, the CUSUM chart is followed until the number of observed events is consistently less than expected. The CUSUM chart can be effectively used to monitor processes for either clusters of failures or a steady change in outcomes (either positive or negative), providing that an appropriate one- or two-sided chart is used.

In this paper, we have concentrated on methods associated with monitoring mortality or graph survival outcomes with a CUSUM method based on the Cox regression model. This method is able to utilize all failures as they occur and avoids the waiting time required in the assessment of survival outcomes, using CUSUMs for binary outcomes as previously reported (10,11,13-15). We feel that these methods represent a major advance over previously published techniques. Because the risk-adjustment models are consistent with those used in the SRTR's PSRs, center directors can feel confident that the CUSUM will respond to clinically relevant changes in outcomes that may be under the center's control but will not penalize programs that transplant higher risk patients or utilize high risk organs. Furthermore, the utility of the CUSUM for ongoing management is enhanced by its ability to capture and include all patient deaths and graft failures immediately upon reporting rather than waiting, for example, for a full year post-transplant.

It should be noted that CUSUM techniques could be used to monitor many other outcomes besides survival. The methods based on a binary outcome, for example, could be used to monitor such outcomes as conversion rates or acceptance rates, in addition to short term survival. In the industrial context, CUSUMs were initially developed to monitor normally distributed outcomes; normal CUSUMs could be used to monitor other transplant outcomes such as measures of quality of life or creatinine clearance.

Notably, there have been no prospective clinical trials in surgical fields demonstrating that CUSUM charting is effective in a multicenter context. Previous analyses, including our own (11), have all been retrospective. We hope that by increasing knowledge of the techniques and collecting prospective data we can demonstrate the advantages of this methodology and gain support from OPTN members to adopt it. We are currently undertaking a prospective study of these techniques with a limited number of transplant centers and this work may help to fill this void. From the retrospective studies, however, it seems likely that there are substantial gains to be had through the earlier identification of poorer center performance that the CUSUM techniques can help to affect.

There appear to be two major barriers to widespread adoption of CUSUM in the transplant community. First, there is limited familiarity and comfort with using this method of assessing outcomes. Second, some center directors may fear that CUSUM will be used as yet another regulatory tool to identify and censure poor performance. As

proposed here, however, CUSUM would be provided confidentially to individual centers as a management tool to help them improve their own performance. One danger with CUSUM charts is that users may react to small trends in the charts that arise by random variation and look to examine and revise the process when such change is unnecessary and perhaps even detrimental. This risk can be minimized through adequate education and appropriate determination of signaling thresholds. The signal from a CUSUM is a useful indicator, but not definitive proof, of a change in center performance. As a tool for quality improvement, the CUSUM chart both validates the success of practice changes and can trigger for a comprehensive review and examination if needed.

Organ transplantation offers an unparalleled opportunity to restore life through the provision of a precious resource. Patients, payers, and the press are all interested in systems of increased transparency of outcomes to ensure that the limited supply of organs is optimally used. To obviate any need for additional regulations, the transplant community would be well served to adopt state-of-the-art monitoring systems to improve performance and maximize the benefit of the limited supply of donor organs.

References

1. Shewart WA. Economic control of quality manufactured produce. New York: Van Nostrand, 1931.

2. Page E. Continuous inspection schemes. Biometrika. 1954;41:100-14.

3. de Saintonge DM, Vere Dw. Why don't doctors use cusums?. Lancet. 1974;1:120-1.

4. Wohl H. The cusum plot: its utility in the analysis of clinical data. N Engl J Med. 1977; 296: 1044-5.

5. Poloniecki J, Sismanidis C, Bland M. Jones P. Retrospective cohort study of false alarm rates associated with a series of heart operations: the case for hospital mortality monitoring groups. BMJ 2004;328:375-8.

6. Committee on Quality of Health Care in America, Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. National Academies Press. 2001.

7. Biau D, Resche-Rigon M, Godiris-Petit G, Nizard R, Procher R. Quality control of surgical and interventional procedures: a review of the CUSUM. Qual Saf Health Care. 2007;16:203-7.

8. Grigg O and Farewell V. An overview of risk-adjusted charts. J. R. Statist. Soc. 2004; A 167, 3, 523-539

9. Spiegelhalter DJ, Grigg OA, Kinsman R, and Treasure T. Sequential probability ratio tests (SPRTS) for monitoring risk-adjusted outcomes. Int. J. Qual. Hlth Care, 2003; 15, 1-7

10. Steiner S, Cook R, Farewell V et al. Monitoring surgical performance using risk-adjusted cumulative sum charts. Biostatistics, 2000; 1, 441-452

11. Axelrod DA, Guidinger MK, Metzger RA, Wiesner RH, Webb RL, Merion RM. Transplant center quality assessment using a continuously updatable, risk-adjusted technique (CUSUM). Am J Transplant. 2006 Feb;6(2):313-23.

12. Biswas P, Kalbfleisch JD. A risk-adjusted CUSUM in continuous time based on the Cox model. Stat Med. 2008 Jul 30;27(17):3382-406.

13. Steiner S, Cook R, Farewell V. Risk adjusted monitoring of surgical outcomes. Med Decis Making, 2001; 21, 163-169

14. Grigg OA, Farewell VT, Spiegelhaler DJ. Use of risk-adjusted CUSUM and RSPRTcharts for monitoring in medical contexts. Statistical Methods in Medical Research, 2003; 12, 2, 147-170

15. Barnard GA. Control charts and stochastic processes. J Roy Stat Soc Ser B. 1959;21:239-71.

16. Barnard GA. Sampling Inspection and Statistical Decisions. Journal of the Royal Statistical Society, 1954; B, 16, 151-174

17. Dickinson DM, Shearon TH, O'Keefe J, Wong H-H, Berg CL, Rosendale JD, et al. The 2005 SRTR Report on the State of Transplantation: SRTR Center-Specific Reporting Tools: Posttransplant Outcomes. Am J Transplant 2006; 6:1198-1211.

18. DeLeval MR, François K, Bull C, Brawn WB, Spiegelhalter DJ. Analysis of a cluster of surgical failures. J. Thor. Cardvasc. Surg., 1994; 104, 914-924

19. Lovegrove J, Sherlaw-Johnson C, Valencia O, Treasure T, Gallivan S. Monitoring the performance of cardiac surgeons. J. Oper. Res. Soc., 1999; 50, 684-689

20. 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, 1128-1130

21. Poloniecki J, Valencia O, Littlejohns P. Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery. Br. Med. J. 1998; 316, 1697-1700

22. Bowles TA, Watters DA.Time to CUSUM: simplified reporting of outcomes in colorectal surgery. ANZ J Surg. 2007 Jul;77(7):587-91

23. Lerch L, Donald JC, Olivotto IA, Lesperance M, van der Westhuizen N, Rusnak C, et al. Measuring surgeon performance of sentinel lymph node biopsy in breast cancer treatment by cumulative sum analysis. Am J Surg. 2007 May;193(5):556-9.

24. Bergman S, Feldman LS, Anidjar M, Demyttenaere SV, Carli F, Metrakos P, et al. "First, do no harm": monitoring outcomes during the transition from open to laparoscopic live donor nephrectomy in a Canadian centre.Can J Surg. 2008 Apr;51(2):103-10.

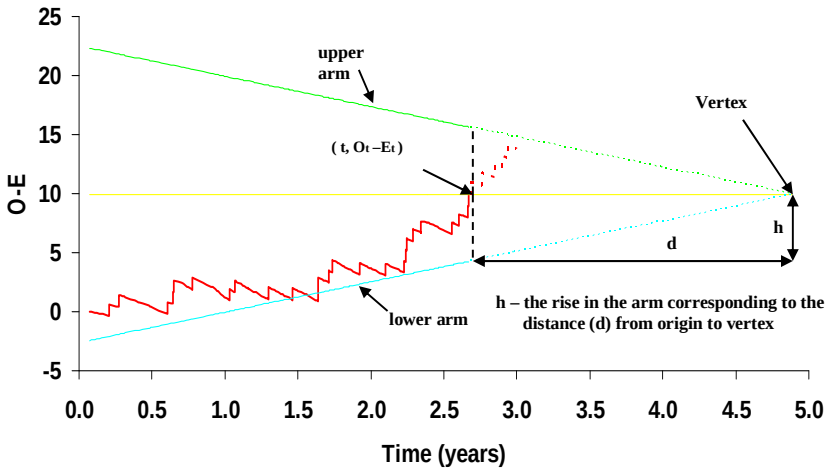
Statistical Methods Appendix

Calculation of the Continuous CUSUM

Utilizing national data, a Cox proportional hazards model is used to characterize the outcome of interest, such as patient survival, following liver transplant. The current SRTR models for post-transplant survival (or graft survival) can provide the basis for this, and by incorporating appropriate clinical and donor characteristics, an estimated probability of each patient's survival for each day following transplant can be determined. The CUSUM is then constructed by examining all the individuals who have been transplanted since the day the chart was initiated. The change in the CUSUM on day t depends on the observed number of patient deaths (d_t) and the risk adjusted expected number of patient deaths (e_t) on each day as derived from the model. The CUSUM is then recalculated daily, incorporating both longer survival times, as well as the experience of new transplants. This method can be used to track outcomes such as mortality over a specified period of time (e.g., one year) post-transplant. In this case, patients who are more than one year post-transplant would not be included in the CUSUM calculation.

One-sided CUSUM. This CUSUM is constructed by considering a test of the hypothesis that the actual mortality rate in the institution of interest is the same as that of the population in general (relative risk [RR]=1) versus an alternative in which the mortality rate is a multiple (RR_A>1) of the overall death rate. This relative risk RR_A is chosen to correspond to an increase in the mortality rate that would be considered clinically important. In what follows, we consider a relative risk of RR_A=2, or a doubling of the risk of death as the alternative of interest. In calculating the change in the CUSUM on day t, each death on day t increases the CUSUM by an amount (0.69) and this is reduced by subtracting e_t, the risk-adjusted number of deaths expected. Thus, the value of the CUSUM on day t is $U_t = U_{t-1} + 0.69\, d_t - e_t$ where d_t is the observed number of deaths. In the one-sided CUSUM, the sum is restricted so as to never become negative. Thus, if it becomes negative its value is replaced by 0. For example, suppose that 30 days into the CUSUM, one patient dies (d_t=1) whereas the model predicts that on average e_t =0.05 patients would die on that day. The CUSUM would increase by 0.69-0.05=0.64. On the other hand, if there were no deaths on that day (d_t=0), the CUSUM would decrease by the value 0.05. If the CUSUM became negative because of this adjustment, it would take the value 0. Once the CUSUM achieves the value zero it stays there until the next death. Thus, the one-sided CUSUM will increase when failures occur but will decrease only to zero with a long period of no deaths.

Figure VII-4. An O-E CUSUM and Associated V-Mask



The one-sided CUSUM signals that a process review is appropriate when the CUSUM (U_t) exceeds some pre-specified level L. An appropriate method for choosing L is still the subject of investigation and is largely an empirical process. On the one hand, we wish to choose L so that there will tend to be a long waiting time until a signal occurs if the institution has failure rates that are equal to or better than the national average. On the other hand, if the death rates of the facility are substantially in excess of the national average, we would like to identify the problem as quickly as possible. One approach is to choose L so that, given the facility size (average number of transplants per year), there would be a pre-specified probability of achieving a signal over a given period of time if, in fact, the facility death rates were exactly at the national rate. In this work, we adopted this approach and the level was chosen to achieve a "false positive rate" of about 8% over a three and a half year period. This is a comparable false positive rate to the screening methods currently in use by the SRTR and the MPSC. The average length of time prior to signaling, referred to as the average run length (ARL), is one way of assessing a specified control limit. Ideally, the ARL should be long for centers that are in control (i.e., operating at the national average or better) and short for centers that have high failure rates.

Two-sided or O-E CUSUM. This chart plots, as a function of time, the difference between the observed number of deaths O_t up to a given time t and E_t, the number of deaths that would be expected up to time t if the death rate at the center was exactly the same as for the overall national average. Thus, the O-E chart involves plotting $O_t - E_t$ versus t where $O_t = d_1 + d_2 + \ldots + d_t$ and $E_t = e_1 + e_2 + \ldots + e_t$ for $t=1, 2, \ldots$.

The V-mask, introduced by Barnard, provides a way of determining signals from a two-sided CUSUM (O-E plot) (15-16). This is a somewhat more complicated approach to determining signals than that available in the one-sided CUSUM. The V-mask is applied

at each time *t* and consists of constructing a horizontal V with vertex located a given distance *(h/k)* to the right of the current plotting position (t, O_t - E_t), where the arms of the V are chosen to pass through points *h* units above and *h* units below the current plotting position. The V-mask is defined by the choice of *h* and *k* and we have selected these values in order to achieve a false signal rate of approximately 8% over a three-and-a-half-year period, in the same manner as the choice of the control limit, *L*, in the one-sided CUSUM chart. An O-E CUSUM along with an associated V-mask is illustrated in Figure VII-4. The V-mask yields a signal at time *t* if the arms of the V intersect the previous path of the CUSUM. If the upper (lower) arm crosses the path, the rate of events at the corresponding center is significantly less (greater) than the national average. In the plot in Figure VII-4, the CUSUM has signaled at the indicated time *t* with an indication that the failure rates are significantly higher than the national average at this center. It is at this time that the V-mask signal would call for a review to determine whether there are correctable causes of the observed high rate of failures. The subsequent path shown in the figure was that observed without any such review.

Determination of an appropriate signaling threshold: To determine an appropriate level of L and shape of the V-mask, we designed a simulation study modeled on survival outcomes for liver transplant programs (Table VII-3). The simulation is constructed so that the rate of false alarms is set to be approximately 8% for a three-and-half-year period. This choice is similar to the criteria currently used by the SRTR, in which a significance test of the null hypothesis at the 5% level is conducted every six months for a moving window length of 2.5 years. The reported control limit is for a one-year death rate of 13.09%, which corresponds to the national average one-year mortality rate for liver transplants. The continuous CUSUM charts (both the one-sided and two-sided O-E V-mask charts) are designed to be sensitive to a relative risk of RR_A= 2, as compared to the national average. From Table VII-3, it is clear that the choice of a limit by this criterion is significantly affected by the size of the transplant program. By adjusting L, the potential for a false positive CUSUM signal can be made approximately equal regardless of program size. The column entitled 'Power' specifies the probability that a center whose relative risk is RR=2 would signal in the three year period. The ARLs in the

Table VII-3: Simulated Power and Average Run Length (ARL) for one-sided and O-E CUSUMs for a one-year survival, assuming a base national average one-year death rate of 13.1%*. The values of *L* and *h* and *d* (for the V-mask) are chosen to yield signals with probability 8% over 3.5 years for a facility of the given size operating at the national rate.

Facility size (transplants/ yr)	one-sided CUSUM			O-E CUSUM (with V-Mask)			
	L	*power*	*ARL (Yrs)*	*h*	*h/d*	*power*	*ARL (Yrs)*
10	2.4	0.63	24.6	3.4	0.44	0.59	27.1
40	3.8	0.94	33.5	5.2	0.44	0.94	30.1
80	4.5	0.99	35.1	6.4	0.44	0.99	36.1
120	4.9	1.00	30.9	7.0	0.44	1.00	33.5
160	5.2	1.00	30.4	7.4	0.44	1.00	30.1

* 13.1% is the overall national one-year death rate for liver transplants

table give the average number of years of follow-up before the first signal would occur in a center whose failure rates are at the national average (RR=1). In the review of liver transplant programs described in the main chapter, we utilize the control limits as obtained from this simulation.

Table VII-4 gives similar results from a simulation modeled on one-year graft survival in national kidney transplant programs, in which the rate of one-year graft failure is 7.24%. Here again the probability of false signal is set at 8% over the three and a half year period. The control limits from this simulation were used in setting control limits for the study of kidney transplant programs in the main chapter. We see from this table that the signal threshold (L) increases with program size to reduce the likelihood of inappropriate signaling. Once the volume is about 40 transplants per year, there is a 90% likelihood that a true signal would be detected, and a period of nearly 30 years expected between false positive signals (ARL).

Table VII-4: Simulated Power and Average Run Length (ARL) for one-sided and O-E CUSUMs for a one-year survival assuming a base national average one-year death rate of 7.2%*. The values of *L* and *h* and *d* (for the V-mask) are chosen to yield signals with probability 8% over 3.5 years for a facility of the given size operating at the national rate.

Facility size (transplants/yr)	one-sided CUSUM			O-E CUSUM (with V-Mask)			
	L	*power*	*ARL*	*L*	*power*	*power*	*L*
10	2.1	0.44	28.2	2.9	0.44	0.45	29.2
40	3.3	0.80	30.6	4.6	0.44	0.79	30.6
80	3.9	0.95	32.0	5.5	0.44	0.94	30.2
120	4.4	0.99	32.9	6.2	0.44	0.99	30.6
160	4.6	1.00	29.4	6.7	0.44	1.00	31.1
200	4.9	1.00	27.8	7.2	0.44	1.00	29.8

* 7.2% is the overall national one-year death rate for kidney transplants