Transplant Center Quality Assessment Using a Continuously Updatable, Risk-Adjusted Technique (CUSUM)

D. A. Axelrod[a], M. K. Guidinger[b], R. A. Metzger[c], R. H. Wiesner[d], R. L. Webb[e] and R. M. Merion[e,f],*

[a]*Dartmouth Medical School, Lebanon, NH, USA*
[b]*Scientific Registry of Transplant Recipients (SRTR)/University Renal Research and Education Association (URREA), Ann Arbor, MI, USA*
[c]*TransLife-Florida Hospital Medical Center, Orlando, FL, USA*
[d]*Mayo Clinic College of Medicine, Rochester, MN, USA*
[e]*SRTR/University of Michigan, Ann Arbor, MI, USA*
[f]*Division of Transplant Surgery, University of Michigan, Ann Arbor, MI, USA*
Corresponding author: R. M. Merion, merionb@umich.edu

Access to timely, risk-adjusted measures of transplant center outcomes is crucial for program quality improvement. The cumulative summation technique (CUSUM) has been proposed as a sensitive tool to detect persistent, clinically relevant changes in transplant center performance over time. Scientific Registry of Transplant Recipients data for adult kidney and liver transplants (1/97 to 12/01) were examined using logistic regression models to predict risk of graft failure (kidney) and death (liver) at 1 year. Risk-adjusted CUSUM charts were constructed for each center and compared with results from the semi-annual method of the Organ Procurement and Transplantation Network (OPTN). Transplant centers (N = 258) performed 59 650 kidney transplants, with a 9.2% 1-year graft failure rate. The CUSUM method identified centers with a period of significantly improving (N = 92) or declining (N = 52) performance. Transplant centers (N = 114) performed 18 277 liver transplants, with a 13.9% 1-year mortality rate. The CUSUM method demonstrated improving performance at 48 centers and declining performance at 24 centers. The CUSUM technique also identified the majority of centers flagged by the current OPTN method (20/22 kidney and 8/11 liver). CUSUM monitoring may be a useful technique for quality improvement, allowing center directors to identify clinically important, risk-adjusted changes in transplant center outcome.

Key words: Outcome analysis, risk adjustment, statistical process control

Introduction

The provision of timely, risk-adjusted outcome information is crucial to improving clinical care processes. Frequent, real time monitoring of surgical outcomes allows physician leaders to validate clinical process improvements or to identify potentially correctable practice patterns. While standard statistical techniques, including average mortality, risk-adjusted average mortality and multivariate modeling, can be used to identify changing levels of performance at a national level, they have been found to be relatively insensitive to small changes in outcomes that occur at the hospital level (1–4). Furthermore, over time, these methods are likely to produce false positive results due to the need for multiple comparisons of the same data.

The cumulative summation method (CUSUM) is a technique of continuous monitoring derived from industrial statistical process control techniques. CUSUM monitoring has been shown to be a sensitive method to identify persistent deviations from expected results. Recently, Steiner et al. developed a risk-adjustment method permitting clinical implementation of CUSUM in a diverse patient population (5). CUSUM has the potential to be used by transplant centers as a quality improvement and monitoring tool to determine the impact of changing clinical practice (e.g. the use of induction therapy) on transplant center outcomes adjusted for patient and donor risk factors. The CUSUM methodology uses currently collected data and will 'signal' when risk-adjusted, charted clinical outcomes reach a pre-determined threshold value. This signal may initiate a comprehensive review by the program to determine if CUSUM has generated a false positive signal or if the signal represents a true improvement or decline from previous performance.

This study was designed to demonstrate the utility of the CUSUM method for tracking and analyzing center outcomes using a cohort of transplanted patients at multiple centers. Data were collected for all patients who underwent kidney and liver transplantation during a recent 5-year period. CUSUM charts were constructed for all transplant centers, blinded to center identity; transplant centers with significant improvement or deterioration in performance were identified. As a validity test, transplant centers flagged for declining performance were compared with transplant centers

identified using the existing evaluation methods of the Scientific Registry of Transplant Recipients (SRTR) and the Organ Procurement and Transplantation Network (OPTN).

Methods

Data sources

Data from the SRTR were analyzed for all kidney and liver transplants performed between January 1997 and December 2001. The SRTR database includes information on all wait-listed candidates and transplant recipients in the United States, supplemented by mortality information from the Social Security Death Master File (6). Primary outcomes of interest were death at 1 year for liver recipients and graft failure, including death with a functioning graft, at 1 year for kidney recipients.

Multivariable logistic regression risk-adjustment model construction

Descriptive statistics were compiled and analyzed to assess the relationship between the available covariates and the outcomes of interest, using the Student's *t*-test and chi-square analyses, as appropriate. We used a stepwise procedure to select covariates that were significantly associated with mortality or graft failure ($p < 0.05$). Endpoints for the regression models were death at 1 year post-transplant (liver) and graft failure including death at 1 year post-transplant (kidney).

Multivariable models were adjusted for the following covariates: donor characteristics [living or deceased donor source, age, race, expanded criteria donor status (kidney), donor-to-recipient weight ratio [kidney], ethnicity [liver], deceased donor cause of death, weight [liver], anti-cytomegalovirus status [liver], history of cancer [liver], donation after cardiac death [liver], liver biopsy, deceased donor history of hypertension [kidney], deceased donor serum creatinine > 1.5 mg/dL [kidney]); recipient characteristics (age, ethnicity, race, cause of end-stage organ failure, HLA mismatches [kidney], panel reactive antibody (PRA) level [kidney], previous transplants [kidney], dialysis modality [kidney], body mass index [BMI; kidney], history of symptomatic peripheral vascular disease [kidney], angina pectoris [kidney], previous transfusions, medical condition at transplant, time on dialysis [kidney], drug-treated hypertension [kidney], insulin-dependent diabetes mellitus [liver], symptomatic cerebrovascular disease [liver], height [liver], serum creatinine [liver], uncontrollable variceal bleeding [liver], ascites [liver], incidental tumor found at transplant [liver], previous upper abdominal surgery [liver], inotropes for blood pressure support [liver], portal vein thrombosis [liver], split liver transplant, geography—local, regional, or national [liver]); and cold ischemia time [kidney] (7). All data were blinded with regard to patient and transplant center identity. During the development of the renal transplant model, separate models were constructed for deceased donor and living donor kidney transplants. In this analysis, the receiver operating characteristic (ROC) curves for these models were nearly identical for the individual models and the combined (0.66 and 0.64, respectively, vs. 0.68). Thus, in this initial investigation the combined model was used for further analyses. In addition, because this was a retrospective analysis, values were missing for at least one significant variable in a sizable minority of the patients. The variables with the highest rate of missing values were donor to recipient weight ratio (missing in 33%) and recipient BMI (missing in 30%). Instead of dropping these cases from the analysis, binary indicators for missingness for given variables were included in the model as additional covariates to preserve these observations for CUSUM chart construction.

CUSUM chart construction

A separate CUSUM chart was constructed for each liver and kidney transplant center that was active during the 5-year period. Next, the CUSUM chart was analyzed to determine if center performance exceeded the *a priori* control limit of 3.0 when tuned to detect a doubling of the expected rate of graft failure (kidney) or mortality (liver) at 1 year. Similarly, CUSUM charts were constructed with the goal of identifying centers with a 50% reduction in graft failure or mortality, when the CUSUM exceeded a threshold value of 5.0, to increase the specificity for center improvement. Centers in which the CUSUM chart signaled a significant deterioration in performance were identified and compared with centers identified using existing statistical methodology (8). A detailed description of the CUSUM methodology is provided in the Appendix.

Current SRTR/OPTN center-specific report methodology

The SRTR provides quarterly reports to the OPTN Membership and Professional Standards Committee (7). In these analyses, individual centers are flagged for review if 2-year, center-specific outcome meets the following criteria: observed-to-expected failure ratio exceeds 1.5, the difference between the observed and expected outcomes is statistically significant ($p < 0.05$), and the absolute number of excess deaths or graft failures exceeds three (1). Centers identified by the SRTR methodology were compared in a blinded fashion with those identified using the CUSUM technique.

Data analysis

Data analyses were conducted using SAS 9.1 (SAS Institute, Cary, NC, USA). The project was approved by the University of Michigan Medical School Institutional Review Board.

Results

Kidney transplants

A total of 59 650 kidney transplants were performed at 258 transplant centers during the period of study. Overall, 9.2% of renal allografts failed within 1 year (including death with a functioning graft). Donor and recipient characteristics are summarized in Tables 1 and 2. Among the 258 centers, there was a very large range in the rate of graft failures in the first year post-transplant (mean 9.0%; range 0–66.7%).

Multivariable logistic regression analysis included 25 donor and recipient characteristics (Table 3). For the overall cohort, the area under the ROC curve attributable to the model was 0.68. After risk adjustment using this model, the predicted probability of graft failure at 1 year for the average recipient ranged from 1.9% to 22.2% among the transplant centers studied.

Using a signaling control limit of 3.0 and a CUSUM tuned to detect a doubling of the risk-adjusted incidence of graft failure, 52 centers (20%) were flagged by CUSUM over the 5-year period. The CUSUM chart of a representative center with declining performance demonstrates a dramatic increase in the incidence of graft failure beginning at transplant number 230 (Figure 1). The CUSUM methodology also demonstrated significant improvement in 92 centers whose 1-year graft failure rate was reduced by at least 50%. This improvement may reflect a natural learning curve effects (Figure 2) or may be due to an abrupt change in clinical practice (Figure 3).

Table 1: Renal transplant donor characteristics

Variable	%
Living Donor	36.5
Deceased Donor Cause of Death	
Anoxia	6.4
Cerebrovascular/stroke	25.8
Head trauma	29.2
CNS tumor	0.7
Other	1.3
Missing	0.2
Donor age (years)	
<18	10.5
18–34	30.1
35–49	35.0
50–64	21.1
65+	3.0
Missing	0.3
Donor race	
White	84.7
African-American	11.7
Asian	2.3
Other (not White, African-American or Asian)	1.3
Deceased donor serum creatinine > 1.5 mg/dL	6.9
Deceased donor history of hypertension	12.4
Expanded criteria donor	9.9
Donor-to-recipient weight ratio	
Quartile 1 (0–0.75)	15.1
Quartile 2 (0.75–0.90)	12.5
Quartile 3 (0.90–1.15)	19.3
Quartile 4 (1.15+)	20.2
Weight ratio missing	33.0

Table 2: Renal transplant recipient characteristics

Variable	% (or Mean)
Recipient age (years)	
18–34	21.1
35–49	36.1
50–64	34.5
65+	8.3
Recipient ethnicity	
Hispanic	11.6
Non-Hispanic	86.6
Missing	1.8
Recipient race	
White	70.3
African-American	23.3
Asian	4.1
Other (not White, African-American or Asian)	2.2
Cause of end-stage renal disease	
Tubular and interstitial diseases	5.6
Polycystic kidneys	8.8
Congenital, rare familial and metabolic disorders	1.5
Diabetes	21.1
Renovascular and other vascular diseases	4.6
Neoplasms	0.3
Hypertensive nephrosclerosis	15.0
Retransplant/graft failure	9.2
Glomerular diseases	25.6
Other	7.9
Missing	0.5

Table 2: Continued

Variable	% (or Mean)
Number of A mismatches	
Zero	25.6
1	42.6
2	31.8
Number of B mismatches	
Zero	23.2
1	43.1
2	33.7
Number of DR mismatches	
Zero	30.9
1	46.6
2	22.5
Peak PRA	
0–9%	72.1
10–79%	20.1
80%+	6.9
PRA missing	0.9
Previous transplant	13.5
Dialysis status	
No dialysis	13.5
Peritoneal dialysis	17.0
Dialysis—unknown type was performed	1.7
Hemodialysis	67.8
Recipient BMI	
<20	6.3
20–24.9	25.0
25–29.9	22.8
30+	15.6
BMI Missing	30.3
Symptomatic peripheral vascular disease	3.8
Symptomatic peripheral vascular disease missing	9.1
Angina/coronary artery disease	9.6
Any previous transfusions	30.4
Previous transfusions unknown or missing	18.7
No previous transfusions	50.9
Medical condition hospitalized, in ICU or on life support	2.4
Time on dialysis (years)	3.56
Preemptive or date of first dialysis missing	13.2
Drug-treated systemic hypertension	82.2
Drug-treated systemic hypertension missing	7.1
Cold ischemia time (hours)	
0–12	33.3
13–18	15.2
19–24	15.7
25–30	9.7
31+	6.3
Missing	19.8

Among the 52 centers identified as having a period in which graft failure rates were unexpectedly high, 20 (38%) were also flagged using the existing statistical methodology reported in national center-specific reports (2). Two centers (0.8% of all comparisons) were flagged by existing methods and not by CUSUM. Centers flagged by both the CUSUM method and current techniques had an average graft failure rate of 16.1% versus 7.6% at the centers not flagged by either method (p < 0.0001).

Table 3: Renal transplant risk-adjustment model

Variable	Adjusted odds ratio	p-value
Donor Cause of Death		
None—living donor	0.73	<0.0001
Anoxia	1.04	0.4898
Cerebrovascular/stroke	1.20	<0.0001
Head trauma	1.00	Ref
CNS tumor	0.69	0.0658
Other	1.11	0.3972
Missing	0.98	0.9380
Donor age (years)		
<18	0.99	0.8418
18–34	0.85	0.0001
35–49	1.00	Ref
50–64	1.20	<0.0001
65+	1.58	<0.0001
Missing	1.38	0.2966
Donor race		
White	1.00	Ref
African-American	1.17	0.0007
Asian	1.00	0.9773
Other (not White, African-American or Asian)	1.03	0.8440
Deceased donor serum creatinine > 1.5 mg/dL	1.20	0.0003
Deceased donor history of hypertension	1.21	<0.0001
Expanded criteria donor	1.18	0.0096
Donor-to-recipient weight ratio		
Quartile 1 (0–0.75)	1.19	0.0015
Quartile 2 (0.75–0.90)	1.19	0.0014
Quartile 3 (0.90–1.15)	0.99	0.7679
Quartile 4 (1.15+)	1.00	Ref
Weight ratio missing	1.23	0.0002
Recipient age (years)		
18–34	1.09	0.0547
35–49	1.00	Ref
50–64	1.35	<0.0001
65+	1.88	<0.0001
Recipient ethnicity		
Hispanic	0.74	<0.0001
Non-Hispanic	1.00	Ref
Missing	1.08	0.4793
Recipient race		
White	1.00	Ref
African-American	1.06	0.1398
Asian	0.69	<0.0001
Other (not White, African American or Asian)	0.90	0.3210
Cause of end-stage renal disease		
Tubular and interstitial diseases	1.21	0.0053
Polycystic kidneys	0.80	0.0008
Congenital, rare familial and metabolic disorders	1.03	0.8333
Diabetes	1.16	0.0010
Renovascular and other vascular diseases	1.16	0.0418
Neoplasms	1.20	0.4903
Hypertensive nephrosclerosis	1.19	0.0005

Table 3: Continued

Variable	Adjusted odds ratio	p-value
Retransplant/graft failure	1.20	0.0265
Glomerular diseases	1.00	Ref
Missing	1.24	0.2959
Other	1.15	0.0245
Number of A mismatches		
Zero	0.93	0.1176
1	0.94	0.0490
2	1.00	Ref
Number of B mismatches		
Zero	0.76	<0.0001
1	0.94	0.0591
2	1.00	Ref
Number of DR mismatches		
Zero	0.77	<0.0001
1	0.89	0.0012
2	1.00	Ref
Peak PRA		
0–9%	1.00	Ref
10–79%	1.15	0.0002
80%+	1.51	<0.0001
PRA Missing	0.85	0.4187
Previous transplant	1.19	0.0148
Dialysis status		
No dialysis	0.82	0.0019
Peritoneal dialysis	1.01	0.8031
Dialysis—unknown type was performed	1.03	0.7922
Hemodialysis	1.00	Ref
Recipient BMI		
<20	1.13	0.0869
20–24.9	1.00	Ref
25–29.9	1.10	0.0423
30+	1.24	<0.0001
BMI missing	1.18	0.0012
Symptomatic peripheral vascular disease	1.26	0.0018
Symptomatic peripheral vascular disease missing	1.01	0.9028
Angina/coronary artery disease	1.12	0.0146
Any previous transfusions	1.12	0.0010
Previous transfusions unknown or missing	1.06	0.1180
No previous transfusions	1.00	Ref
Medical condition hospitalized, in ICU or on life support	1.46	<0.0001
Time on dialysis (years)	1.01	0.0023
Preemptive or date of first dialysis missing	0.86	0.0250
Drug treated systemic hypertension	0.88	0.0011
Drug treated systemic hypertension missing	0.96	0.6484
Cold ischemia time (hours)		
0–12	1.00	Ref
13–18	1.06	0.2986
19–24	1.14	0.0116
25–30	1.19	0.0024
31+	1.32	<0.0001
Missing	1.15	0.0023

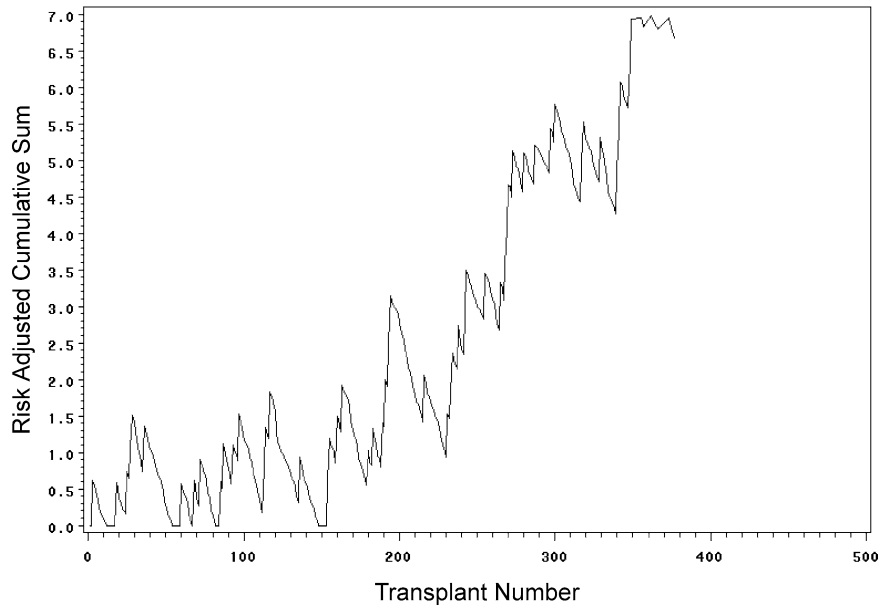## Kidney Transplant Center CUSUM Chart

Figure 1: CUSUM chart of renal transplant center with declining performance.

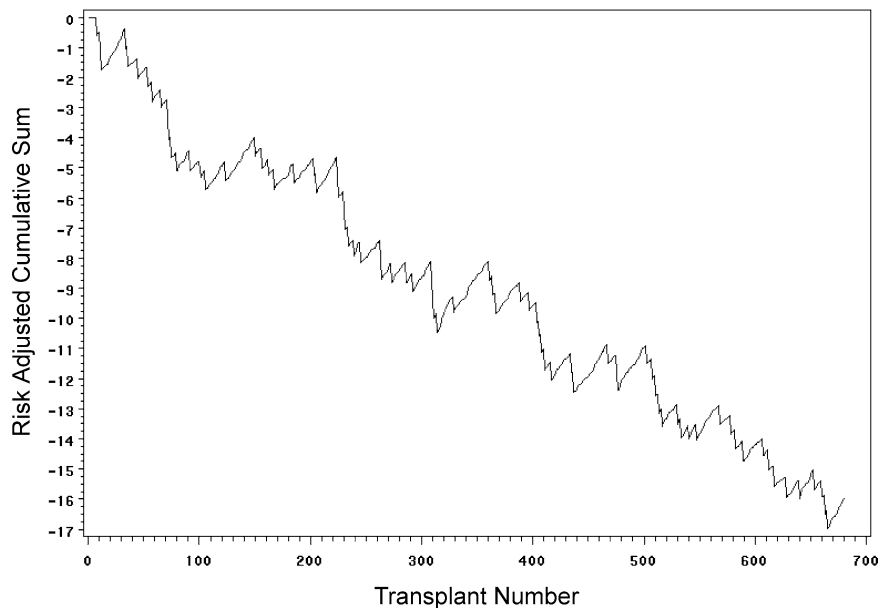## Kidney Transplant Center CUSUM Chart

Figure 2: CUSUM chart of renal transplant center with evidence of improvement consistent with a learning curve effect.

Liver transplants

Over the 5-year period of this analysis, 18 277 patients underwent liver transplantation at 114 transplant centers. Overall 1-year mortality was 13.9% and ranged from 0% to 50% among the centers studied. The 5-year transplant volume averaged 160 transplants per center (range 1–724). Demographic characteristics of the liver donors and recipients are summarized in Tables 4 and 5.

A multivariable logistic regression model was developed for risk adjustment. The model included 27 donor and recipient characteristics (Table 6). The overall area under the ROC curve was 0.66. Donor characteristics associated with significantly higher 1-year mortality rates included increased donor age and African-American race. Significant recipient factors included a history of diabetes mellitus, cerebrovascular disease, older age, a diagnosis of a malignant neoplasm and requiring intensive care unit admission pre-operatively. After risk adjustment, the overall predicted probability of death at 1 year for the average recipient was 12.1% and varied from 6.4% to 37.6%.
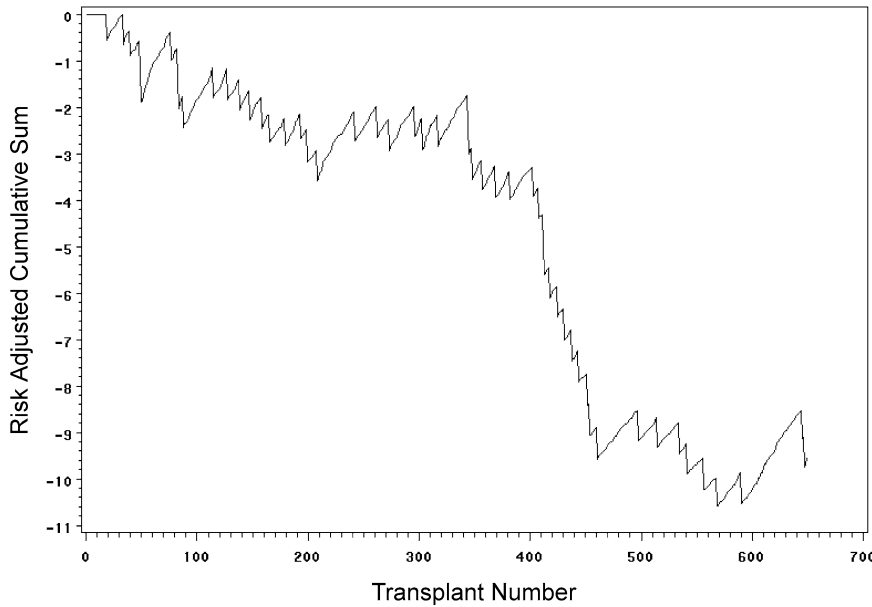
Kidney Transplant Center CUSUM Chart

Figure 3: CUSUM chart of kidney transplant center with a statistically significant improvement in outcome.

Table 4: Liver transplant donor characteristics

Variable	% (or Mean)
Donor Age (Years)	
<18	12.5
18–34	31.4
35–49	27.6
50–64	20.7
65+	7.6
Age missing	0.2
Donor ethnicity	
Hispanic	10.3
Donor race	
African-American	11.0
White	86.1
Other (not Black or White) or missing	2.8
Donor cause of death	
Stroke	41.0
Head trauma	43.4
Other	15.6
Log of donor weight	4.25
Donor weight missing	0.6
Donor anti-CMV Positive	59.0
Donor history of cancer	2.3
Donation after cardiac death	0.8
Donor liver biopsy	13.7

CUSUM charts were constructed for each liver transplant center and flagged for review at a control limit of 3.0. The CUSUM was initially tuned to detect a doubling in expected mortality. Over the 5-year period, 24 centers were flagged using the CUSUM method. A representative CUSUM chart for a center with declining performance is shown in Figure 4. The 1-year mortality rate for this center was 22.9% (1.65 times the overall national rate). This chart indicates higher than expected mortality throughout the period of examination. Subsequent analysis was conducted to identify centers with rapidly improving performance. During the period of study, improvement in performance was documented at 48 centers.

Among the 24 centers flagged by CUSUM for declining performance, eight were also identified using current methodology. The average 1-year mortality rates among patients transplanted at the eight centers flagged by both methods were significantly higher than at centers not flagged by either method (22.1% vs. 12.2%; p < 0.0001). Three centers identified by the SRTR methods were not flagged by CUSUM. The average yearly volume of transplants at these centers was 13.7 (range 12–14.8). Although these centers had higher than expected 1-year mortality rates (21.7–27.8% vs. 13.9% for all other centers), there was insufficient volume for the CUSUM to signal at the chosen threshold level. Review of the CUSUM charts suggested that performance was declining but failed to reach the control limit.

Among the 16 centers not flagged using current methodology, there were periods of declining performance that appear to have been corrected internally (Figure 5). Review of the CUSUM chart suggests that there was a decline in clinical outcomes between transplants 50 and 90 followed by an abrupt clinical improvement. Given the blinded nature of this data, correlation with changes in clinical practice was not possible in this examination.

Discussion

Our analysis has demonstrated the successful application of a continuously updated, risk-adjusted outcome

Table 5: Liver transplant recipient characteristics

Variable	% (or Mean)
Recipient Race	
White	86.8
African-American	7.2
Other or missing	6.0
Recipient ethnicity	
Hispanic	11.3
Missing	0.4
Recipient age (years)	
18–24	2.3
25–34	4.5
35–44	18.1
45–54	40.3
55–64	26.6
65+	8.1
Recipient insulin-dependent diabetes	7.4
Recipient symptomatic cerebrovascular disease	0.5
Recipient previous transfusions	32.9
Cause of end-stage liver disease	
Acute hepatic necrosis	7.9
Cholestatic liver disease/cirrhosis	12.7
Metabolic diseases	3.0
Malignant neoplasms	3.3
Non-cholestatic cirrhosis	69.8
Other	3.3
Recipient medical condition	
Recipient on life support at transplant	7.5
Recipient in ICU, not on life support at transplant	12.6
Recipient hospitalized, not in ICU at transplant	15.5
Recipient not hospitalized at transplant	64.4
Log of recipient height	5.14
Recipient height missing	3.4
Log of recipient creatinine (set to 4 mg/dL for patients on dialysis)	0.06
Complications of ESLD	
Variceal bleeding	5.9
Ascites	74.2
Ascites missing	4.9
Incidental tumor found at time of transplant	3.7
Recipient previous upper abdominal surgery	43.4
Recipient inotropes for blood pressure support	4.2
Recipient portal vein thrombosis	2.1
Partial or split liver transplant	6.3
Proximity	
Regional transplant	19.2
National transplant	9.7
Local transplant	71.1
Living donor	4.6

Table 6: Liver transplant risk-adjustment model

Variable	Adjusted odds ratio	p-value
Donor age (years)		
<18	0.99	0.8626
18–34	0.89	0.0608
35–49	1.00	Ref
50–64	1.18	0.0098
65+	1.29	0.0032
Missing	0.54	0.3048
Donor ethnicity		
Hispanic	1.05	0.5331
Donor race		
African-American	1.14	0.0697
White	1.00	Ref
Other (not African-American or White) or missing	1.13	0.3314
Donor cause of death		
Stroke	1.10	0.1153
Head trauma	1.00	Ref
Other	1.06	0.4237
Log of donor weight	0.79	0.0055
Donor weight missing	0.43	0.0889
Donor anti-CMV positive	1.05	0.3473
Donor history of cancer	0.97	0.8232
Donation after cardiac death	1.33	0.1977
Donor liver biopsy	1.09	0.1576
Recipient ethnicity		
Hispanic	0.90	0.1373
Missing	1.26	0.4793
Recipient race		
African-American	1.26	0.0052
White	1.00	Ref
Other race (not African-American or White) or missing	0.84	0.0694
Recipient insulin-dependent diabetes	1.26	0.0026
Recipient symptomatic cerebrovascular disease	1.99	0.0043
Recipient previous transfusions	1.14	0.0058
Recipient age (years)		
18–24	0.70	0.0761
25–34	1.00	Ref
35–44	1.02	0.8825
45–54	1.26	0.0541
55–64	1.54	0.0004
65+	2.02	<0.0001
Cause of end-stage liver disease		
Acute hepatic necrosis	1.16	0.0621
Cholestatic liver disease/cirrhosis	0.74	0.0001
Metabolic diseases	0.92	0.5322
Malignant neoplasms	1.41	0.0026
Other	1.33	0.0124
Non-cholestatic cirrhosis	1.00	Ref
Recipient medical condition		
Recipient on life support at transplant	2.12	<0.0001
Recipient in ICU, not on life support at transplant	1.38	<0.0001
Recipient hospitalized, not in ICU at transplant	1.44	<0.0001
Recipient not hospitalized at transplant	1.00	Ref
Log of recipient height	0.49	0.0032
Recipient height missing	1.15	0.2235

monitoring technique for the evaluation of transplant center performance. Over the 5-year study period, 52 renal transplant centers and 24 liver transplant centers with a period of higher-than-expected graft failure and 1-year mortality, respectively, were identified using the CUSUM method. The CUSUM method also identified 92 kidney and 48 liver transplant centers with markedly improving clinical outcomes. The CUSUM method appears to have face validity when compared with existing techniques for assessing transplant center performance. Within the limits of current risk-adjustment techniques, CUSUM charting appears to

Table 6: Continued

Variable	Adjusted odds ratio	p-value
Log of recipient creatinine (set to 4 mg/dl for patients on dialysis)	1.54	<0.0001
Complications of ESLD		
Variceal bleeding	0.98	0.8258
Ascites	1.11	0.0892
Ascites missing	1.50	0.0001
Incidental tumor found at time of transplant	1.34	0.0063
Recipient previous upper abdominal surgery	1.22	<0.0001
Recipient inotropes for blood pressure support	1.17	0.1163
Recipient portal vein thrombosis	1.30	0.0632
Partial or split liver transplant	1.19	0.2621
Proximity		
Regional transplant	1.10	0.0868
National transplant	1.28	0.0069
Local transplant	1.00	Ref
Living donor	0.95	0.8258

offer a useful tool for transplant center management and quality improvement.

Center-specific transplant outcomes have been associated with a variety of factors, including center volume, case mix, immunosuppression management, donor quality and technical expertise (9–11). High volume has been correlated with improved outcomes for liver, kidney and heart transplant. In addition, there are important differences between centers related to case mix. Centers with higher num-

bers of African-Americans, older patients, repeat transplants and patients with lower socio-economic status are often noted to have worse outcomes. In addition, variation in donor supply and recipient demand may affect outcome. As waiting lists increase in size, liver transplants are performed for patients with more advanced disease, contributing to generally worse post-transplant outcome (11).

Therefore, effective performance assessment and improvement tools should include a risk-adjustment methodology. The CUSUM technique reported here includes the results of two logistic regression analyses designed to predict post-transplant outcome using donor and recipient data. Unfortunately, post-transplant survival remains very difficult to predict, as it reflects the intersection of donor, recipient, operative and post-operative factors (12). The ROC curves for the models in the current study (0.68 kidney and 0.66 liver) compare favorably with the index of concordance for the Cox regression models currently used to prepare the SRTR center-specific reports (0.67 for kidney and 0.70 for liver) (7). As is true with any risk-adjustment technique, these models remain imperfect tools for making comparisons between centers. However, within any single transplant center, the risk adjustments using various patient and donor characteristics that are available do improve the results provided by CUSUM charting.

Regular reporting of center-specific outcomes has been performed primarily by the SRTR, under provisions of the National Organ Transplant Act (13). The current method used to generate center-specific outcome reports uses Cox regression models and the indirect adjustment method; this provides expected outcomes by adjusting the national

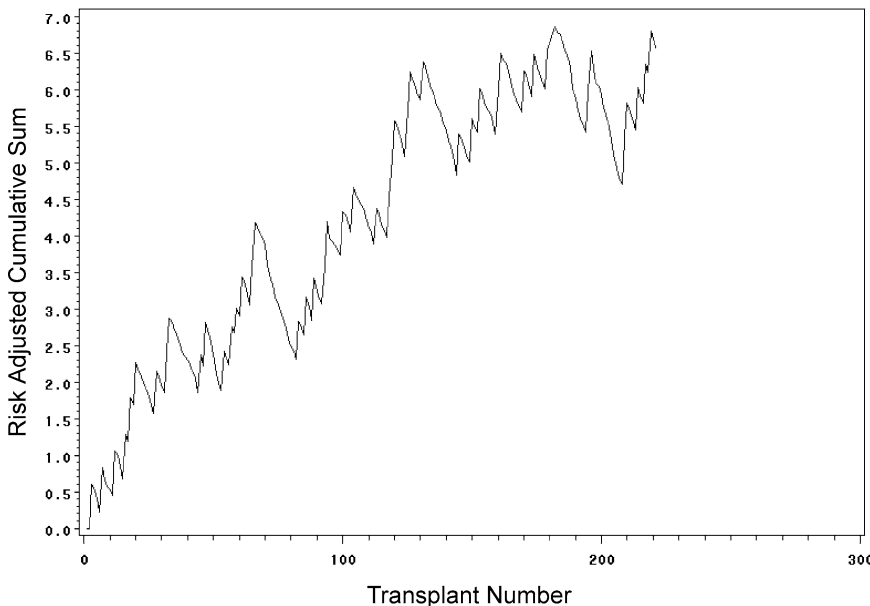**Liver Transplant Center CUSUM Chart**

Figure 4: CUSUM chart of a liver transplant center with a significant decline in risk-adjusted outcome.

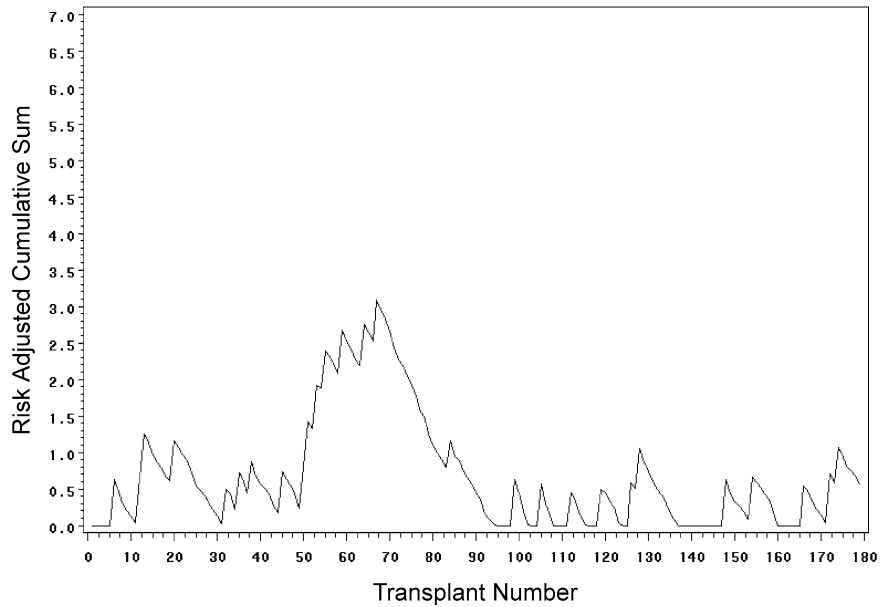## Liver Transplant Center CUSUM Chart

Figure 5: CUSUM chart of liver transplant center flagged by CUSUM for declining performance, which was followed by improvement.

sample to mirror the center's case mix for relevant patient and donor factors. Observed results are compared with expected outcomes for each center twice per year, which can result in flagging due to chance (false positive), as a result of repeated comparisons (1,14). For center review purposes by the OPTN, this risk is minimized though the use of three assessment measures, including the absolute number of excess failures and the ratio of observed-to-expected failures, in addition to the test of statistical significance. However, for internal center management, this approach may miss early trends that could be identified through statistical process control techniques like CUSUM before externally applied measures of performance are triggered.

The modified CUSUM approach proposed by Steiner and colleagues provides a method of assessing surgical outcome using a risk-adjusted, continuously updated, easily interpretable chart (5). As applied in this analysis, clinical outcomes were weighted using a risk-adjustment model to assess the probability of 1-year mortality (liver) or 1-year graft failure (kidney). A pre-set signal level was then determined based on clinical judgment and the severity of the outcome of interest. Once the CUSUM chart reached this threshold, a signal was reported. CUSUM signaling does not necessarily prove that a clinically important decline or improvement in clinical quality has occurred. Rather, the signal suggests that closer examination by a program's quality improvement team may be indicated.

The CUSUM method has several distinct advantages over other techniques. First, the CUSUM technique explicitly ac-

counts for the impact of time as a variable. Thus, multiple graft failures within a short time are more likely to generate a signal than if the same number of graft failures were randomly spread across a longer period of analysis (15). Second, empirical assessment of risk-adjusted CUSUM charts, also known as CRAM charts, demonstrated the greatest sensitivity in detecting clinically important deterioration in performance when compared with other statistical techniques (1). Third, unlike standard modeling techniques, the CUSUM charting method does not suffer from the problem of multiple statistical examinations of highly overlapping data, which is likely to produce false positive alarms. Finally, and perhaps most importantly, given the graphical nature of the CUSUM, the slope of the chart can provide valuable insight into transplant center performance even before a signal occurs. Thus, the CUSUM can be used as a real time tool to assist transplant centers in improving clinical practice and patient management protocols.

Like other statistical techniques, CUSUM entails a tradeoff between the desire to rapidly recognize changes in clinical practice (sensitivity) and the necessity of limiting the number of false alarms (specificity). In CUSUM analysis, this tradeoff is reflected in the average run length (ARL), which reflects the average number of transplants that would occur prior to the CUSUM signaling by chance if there was no true change in outcome. Ideally, an effective CUSUM analysis will have a long ARL when the process is in control and a short ARL when the process deviates from the expected results. In the case of transplant center outcomes, it is perhaps better to err on the side of over-signaling rather than to miss a clinically important deterioration in outcome. In

our study, an *a priori* control limit of 3.0 was set. However, with experience, the CUSUM can be adjusted to improve the ratio of true positives to false positives to allow efficient application of quality improvement procedures (Appendix). Furthermore, the CUSUM chart requires that a specific point be chosen as the endpoint of interest (e.g. 1-year graft survival). Consideration of additional endpoints (e.g. 3-year graft survival) requires construction of alternative CUSUM charts and, potentially, a revised risk-adjustment model. Utilization of a short-term outcome is important if the CUSUM is to be successfully used as an active transplant center management tool.

This study represents the first report of CUSUM analysis for multiple centers using an endogenously derived risk-adjustment model. Previous examples of CUSUM monitoring in surgical analysis have been limited principally to single center studies to illustrate the learning curve inherent in new techniques. Novick and colleagues have applied CUSUM in a variety of analyses for outcomes following telerobotic cardiac surgery (3) and off-pump coronary bypass procedures (2). The authors reported that standard statistical methods failed to identify the reduction in the rate of complications over the course of the center's learning experience with telerobotic cardiac surgery that was clearly evident in CUSUM analysis. Similarly, Forbes and colleagues examined the outcome of endovascular aneurysm repair (4). Using CUSUM analysis, they demonstrated that the learning curve for endovascular repair was nearly 60 patients, which was much longer than expected. Finally, the United Kingdom Transplant Service has begun supplying its members with CUSUM charts designed to assess 30-day renal graft failure (D. Collett, Director of Statistics and Audit, UK Transplant, personal communication April 1, 2005). These reports have assisted transplant center directors in determining whether or not a series of graft failures is likely to be due to chance or is a reflection of transplant center modifiable issues of medical and/or surgical management.

There are several important limitations in our analysis. First, to protect center confidentiality, we did not validate CUSUM findings with changes in clinical management or process at the blinded centers. Therefore, it was not possible to determine whether centers flagged by CUSUM analysis had experienced a genuine change in outcome that resulted from a change in clinical practice or, by chance, had experienced several adverse outcomes in rapid succession. Prospective application of the CUSUM technique is necessary to accurately assess the ratio of true positive to false positive signals. Because CUSUM can be set up as a very sensitive method of detecting changes in clinical outcomes, its best application would appear to be as a quality improvement and management tool at the transplant center level, rather than as a replacement or addition to existing OPTN methodology. Second, the risk-adjustment methodology remains imperfect. This analysis made use

of all available clinical data elements; further refinement in the risk-adjustment model may require the collection of more detailed data. Furthermore, the presence of missing data may have affected the model, and future implementation should require that data from all relevant fields are included. As with any measure of outcome, a careful review of all relevant data must be conducted to eliminate the possibility that there has been a false positive signal prior to changing clinical practice. Finally, because this analysis is retrospective and other OPTN-based center outcome measurements were already being performed, it was not possible to determine the incremental benefit of concurrent CUSUM reporting on center outcomes. A prospective trial of this method in which CUSUM results are returned regularly to program directors would be useful to assess the potential utility of the technique for transplant center quality improvement.

In summary, CUSUM techniques can be used to assist transplant centers in assessing their outcomes using a real time, risk-adjusted process. Further analysis and prospective studies are needed to assess the impact of CUSUM monitoring on center outcomes and to fine tune the analysis to balance the need for early identification of genuine changes in care with the desire to reduce unnecessary scrutiny.

Acknowledgments

References

1. 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.
2. Novick RJ, Fox SA, Stitt LW et al. Cumulative sum failure analysis of a policy change from on-pump to off-pump coronary artery bypass grafting. Ann Thor Surg 2001; 72: S1016–S1021.
3. Novick RJ, Fox SA, Kiaii BB et al. Analysis of the learning curve in telerobotic, beating heart coronary artery bypass grafting: a 90 patient experience. Ann Thorac Surg 2003; 76: 749–753.
4. Forbes TL, DeRose G, Kribs SW, Harris KA. Cumulative sum failure analysis of the learning curve with endovascular abdominal aortic aneurysm repair. J Vasc Surg 2004; 39: 102–108.
5. Steiner SH, Cook RJ, Farewell VT. Risk-adjusted monitoring of binary surgical outcomes. Med Decis Making 2001; 21: 163–169.
6. Social Security Administration Death Master File. Springfield, VA: Federal Computer Products Center, National Technical Information Service, U.S. Department of Commerce; 2003.
7. Scientific Registry of Transplant Recipients: Center-Specific Reports. Available at http://www.ustransplant.org/csr/. Accessed September 23, 2005.

8. Wolfe RA, Schaubel DE, Webb RL et al. Analytical approaches for transplant research. Am J Transplant 2004; 4(Suppl 9): 106–113.
9. Port FK, Bragg-Gresham JL, Metzger RA et al. Donor characteristics associated with reduced graft survival: an approach to expanding the pool of kidney donors. Transplantation 2002; 74: 1281–1286.
10. Axelrod DA, Guidinger MK, Leichtman AB, Punch JD, Merion RM. Association of center volume with outcome after liver and kidney transplantation. Am J Transplant 2004; 4: 920–927.
11. Trotter JF, Osgood MJ. MELD scores of liver transplant recipients according to size of waiting list. JAMA 2004; 291: 1871–1874.
12. Ghobrial RM, Gornbein J, Steadman R et al. Pretransplant model to predict posttransplant survival in liver transplant patients. Ann Surg 2002; 236: 315–322.
13. Pub L No. 98-507. National Organ Transplant Act, 42 USC §273
14. McPherson K. Statistics: the problem of examining accumulating data more than once. N Engl J Med 1974; 290: 501–502.
15. Altman, DG, Royston SP. The hidden effect of time. Stat Med 1988; 7: 629–637.

APPENDIX: Statistical Methods for Quality Monitoring Using a CUSUM Chart

Background:

The CUSUM method accumulates evidence about an ongoing process (e.g. a series of surgeries) to identify a clinically important change in performance. Mathematically, the technique involves plotting a continuous sum of the scores (outcomes) versus time. If the process remains 'in control', resulting in expected surgical outcome, the CUSUM will remain near zero. If the surgical results deteriorate, the CUSUM will increase. When the score exceeds a pre-set value, the chart is said to signal.

Calculation of the CUSUM:

As proposed by Steiner and colleagues, the CUSUM (X_t) is calculated as a continuous sum of w_t, a weighted score that depends upon the patient's predicted pre-operative risk (5). As shown in Equation 1, three factors are included in the calculation of w_t:

p_t: the patient's predicted risk of death as derived from a logistic regression model.

OR_A: a pre-determined ratio of expected versus actual outcomes. To detect a doubling of expected mortality, OR_A would be set to 2.

y: an indicator variable equals 1 if there is a transplant failure and zero if the transplant is successful.

The CUSUM (X_t) is calculated using Equation 2. Based upon Equation 1, a transplant failure results in a positive value that increases the CUSUM by a risk-adjusted amount, while a success reduces the score. The CUSUM score is restricted to non-negative values to increase the sensitivity to detect clinical failures.

Equation 1:

$$W_t = \begin{cases} \ln\left[\dfrac{OR_A}{(1-p_t+OR_A p_t)}\right] & \text{if } y = 1 \\ \ln\left[\dfrac{1}{(1-p_t+OR_A p_t)}\right] & \text{if } y = 0 \end{cases}$$

Equation 2:

$$X_t = \max(0, X_{t-1} + w_t)$$

Signaling and Average Run Length

Because a transplant success results in a negative value, the CUSUM score will remain close to 0 for an in-control process. However, if a cluster of failures occurs, the CUSUM will rise. The CUSUM chart signals a change in performance when the sum of w_t reaches or exceeds a pre-set control limit (h). Choosing an appropriate control limit is empirical and reflects the trade off between sensitivity and specificity. A high control limit will reduce the number of false positive signals, but will require more failures prior to signaling. Conversely, a lower control limit will increase the sensitivity of CUSUM to clusters of surgical failures at the expense of false alarms. The average length of time prior to signaling, referred to as the average run length (ARL), is a reflection of the choice control limit. Ideally, the ARL should be long for processes that are in control and short for processes out of control. Determination of the appropriate value of (h) is an empiric process. As shown below, the number of centers flagged for review varies considerably according to the chosen control limit (Table).

Control Limit (h)	Number of centers flagged (Kidney)	Ratio CUSUM flagged to SRTR flagged (Kidney)	Number of centers flagged (Liver)	Ratio CUSUM flagged to SRTR flagged (Liver)
2	110	5.00	51	4.64
3	52	2.36	24	2.18
4	32	1.45	12	1.09
5	24	1.09	5	0.45

Analysis of CUSUM data:

Following a CUSUM signal, an examination of the process under review should be undertaken. A signal is not proof of a failure in care. Rather, it represents the canary in the mineshaft, which may alert the transplant team to the potential need for a revision in care plans or technique. Furthermore, the CUSUM can also be adjusted to reflect improving care, and therefore demonstrate the positive impact of a change in care (e.g. the impact of induction therapy on graft outcome).