

SRTR Center-Specific Reporting Tools: Posttransplant Outcomes

D. M. Dickinson^{a,*}, T. H. Shearon^b,
J. O’Keefe^c, H. -H. Wong^d, C. L. Berg^e,
J. D. Rosendale^c, F. L. Delmonico^f, R. L. Webb^b
and R. A. Wolfe^a

^aScientific Registry of Transplant Recipients, University
Renal Research and Education Association

^bScientific Registry of Transplant Recipients, University
of Michigan

^cOrgan Procurement and Transplantation Network/United
Network for Organ Sharing

^dDepartment of Health and Human Services, Health
Resources and Services Administration, Healthcare
Systems Bureau, Department of Transplantation

^eUniversity of Virginia

^fHarvard Medical School

*Corresponding author: David M. Dickinson,
dickinsn@urrea.org

Measuring and monitoring performance—be it waiting list and posttransplant outcomes by a transplant center, or organ donation success by an organ procurement organization and its partnering hospitals—is an important component of ensuring good care for people with end-stage organ failure. Many parties have an interest in examining these outcomes, from patients and their families to payers such as insurance companies or the Centers for Medicare and Medicaid Services; from primary caregivers providing patient counseling to government agencies charged with protecting patients.

The Scientific Registry of Transplant Recipients produces regular, public reports on the performance of transplant centers and organ procurement organizations. This article explains the statistical tools used to prepare these reports, with a focus on graft survival and patient survival rates of transplant centers—especially the methods used to fairly and usefully compare outcomes of centers that serve different populations. The article concludes with a practical application of these statistics—their use in screening transplant center performance to identify centers that may need remedial action by the OPTN/UNOS Membership and Professional Standards Committee.

Key words: Methodology, OPTN, patient survival, professional standards, risk adjustment, SRTR, transplant center performance, transplant data

Introduction

Many audiences, many reports

Reporting the results of transplant centers and organ procurement organizations (OPOs) is one of the many contract responsibilities of the Scientific Registry of Transplant Recipients (SRTR). These analyses have a wide range of intended audiences within the transplant community, each with different understandings of clinical and statistical concepts, and each with different goals:

- (i) Patients and families may use them to find a transplant program with good experience among similar patients.
- (ii) Transplant professionals, such as surgeons or administrators, may use them to help explain a patient’s prospects for recovery, or as a quality control mechanism for benchmarking against other programs.
- (iii) Insurance companies and other payers may use them to ensure good care for the patients they serve.
- (iv) Regulatory bodies both within and outside of the Organ Procurement and Transplantation Network (OPTN) may use them to help identify programs in need of remedial action or further study.

The publicly available transplant center-specific reports (CSRs) published on the SRTR website at www.ustransplant.org are the most widely used of a whole ‘family’ of tools for program-specific reporting produced by the SRTR at least every 6 months. Similar reports document the organ procurement activity within each donation service area (DSA). A quarterly report for the OPTN Membership and Professional Standards Committee (MPSC) helps that committee identify centers for performance review. A prescribed set of statistics is prepared as part of a ‘Standardized Request for Information’ and made available for centers to submit to insurers requesting information about center performance. All of these tools employ the same methodology for measuring outcomes; these are the methods discussed in this article.

The scope of questions addressed in these reports covers the entire spectrum of the transplant process. The organ procurement organization-specific reports (OSRs) examine the process of identifying and recovering donors. The CSRs begin by examining pretransplant activity and outcomes on the waiting list. These often-overlooked statistics, such as the mortality and transplant rates contained in Table 3 of

the CSRs, are an important component of the transplant process, as posttransplant outcomes are irrelevant to a patient who might die while still awaiting an organ. However, by far the most attention is focused on the graft and patient survival reported in Tables 10 and 11 of the CSRs. Therefore, we focus most of our explanation here on the techniques used for measuring these posttransplant outcomes, many of which are also applicable to other sections of the reports.

We conclude the article with a look at how one monitoring body, the OPTN MPSC, implements these statistics to help recommend changes for improving transplant center operations.

Advantages of a Standardized Calculation

Using SRTR-calculated center-specific statistics provides several advantages over having each center report its own statistics:

- (i) **Uniform statistical methodology:** The methods used by the SRTR are standard and accepted within the statistical and medical communities.
- (ii) **Uniform and required data collection:** Accurate submission of transplant data is required for participation in the OPTN organ allocation system. The United Network for Organ Sharing (UNOS), the contractor for the OPTN, works with help from the SRTR to ensure the accuracy and reliability of these data.
- (iii) **No duplication of effort by facilities:** Calculating these statistics can be a tedious task that is most efficiently programmed for all centers at the same time.
- (iv) **Extra ascertainment of mortality:** The SRTR helps find information about patients who become lost to follow-up. Outcomes for these patients may be very difficult or even impossible for transplanting centers to track and report. Extra ascertainment builds trust in the completeness of reporting.
- (v) **Risk-adjusted comparison points:** Comparison of outcomes should be based on risk-adjusted models that account for the types of patients treated. Without national data, it is impossible for centers to calculate risk-adjusted comparison points.

Interpreting Posttransplant Outcomes

Posttransplant outcome tables dominate the questions and concerns about the CSRs, and have figured prominently in the Conditions of Participation for funding transplant hospitals recently proposed by the Centers for Medicare and Medicaid Services (CMS) (1). The issues illustrated by these tables apply to many of the other statistics in the reports, such as risk-adjusted comparison of transplant and mortality rates from the waiting list, or risk-adjusted comparisons of donation rates for OPOs. We focus here on

posttransplant outcomes as the primary examples in our examination of CSRs, though waiting list outcomes are also raised as secondary examples.

Percentage surviving at the end of period: an interpretable result

Table 1 shows portions of CSR Table 11, Patient Survival after Transplant, published in the July 2005 release of the CSRs. We will call the example liver program shown here "Hospital A." Table 1 presents much information that is referred to throughout this article, but it is limited to results for 1 year following transplantation. Similar columns, produced for outcomes at 1 month and 3 years, are omitted.

The first panel of results, beginning at line 2, shows the percentage of patients surviving at the end of the period (in this case, 1 year). The percentage of patients surviving is intuitively understandable, and meaningful to a wide range of audiences—the reader, perhaps a patient, learns that in recent history, 87.78% of other patients who received a liver transplant at Hospital A were alive a full year after transplantation (line 3). Other measures, such as a rate per year at risk, may not be as intuitively understandable to most audiences.

The same patient, or perhaps a transplant administrator, may compare that survival percentage to the national average of 86.26%, also on line 2. While a conclusion that the center has above-average results compared to the national average is accurate at face value, we must look further to determine whether this is either:

- (i) because the center is 'above average' in its treatment practices, or
- (ii) because the types of patients treated by this center tend to have better outcomes no matter where they are treated (e.g. they are younger or start off with fewer complications than patients in other centers).

This distinction is addressed by the concept of 'expected survival.'

Expected survival

The notion of expected survival addresses the critical question, 'What rate would be expected for the patients at this center if they had outcomes comparable to the typical national experience for similar patients?'

Line 4 of Table 1 ('Expected, based on national experience') allows the reader to examine whether a center's performance is itself above average, or whether the center starts off with healthier patients. In Hospital A, from Table 1, 89.41% of 'similar' patients, nationwide, were alive 1 year after transplant. Two conclusions can be made:

- (i) Because expected survival is higher for this center than the national average, the case-mix of patients treated by this center may be easier to treat than average patients throughout the country.

Table 1: Center-specific report Table 11—patient survival after transplantation, sample liver center 'Hospital A'

	Center 1 year	National 1 year
Adult (Age 18+)		
1	Transplants (n = number)	10 781
2	Percentage (%) of patients surviving at the end of period	
3	Observed at this Center	86.26
4	Expected, based on national experience	
5	Deaths during follow-up period	
6	Observed at this center	1392
7	Expected, based on national experience	1392
8	Ratio: observed to expected (O/E)	1.00
9	(95% Confidence Interval)	(0.65–2.32)
10	P-value (2-sided), observed v. expected	0.469
11	How does this center's survival compare to what is expected for similar patients?	Not significantly different
12	Percent retransplanted	4.4
13	Follow-up days reported by center (%)	93.9
14	Maximum days of follow-up (n)	365

Source: SRTR Center-Specific Reports, www.ustransplant.org, July 2005 Release.

(ii) While the survival rate observed at this center is above the national average for all liver transplant recipients, it is in fact below what would be expected for the type of patients treated by the center.

These conclusions rely on the notion of 'similar' patients—those with characteristics in common that may influence the waiting list or posttransplant outcome. The characteristics used to define 'similarity' include characteristics that are associated with survival in the general population, such as age; and disease-specific factors, such as specific etiology of disease and measures of severity of illness. We discuss how this list of factors is determined in the section 'Calculation of Models.'

Table 2 illustrates how adjustment works and why it is needed. In this table, we assume that the nation consists of only two kinds of patients: half are 'older' (with 80% 1-year survival) and half are 'younger' (92% survival), for an overall national average survival of 86%. At example Hos-

pital B, 24 of the 25 younger patients survived until 1 year (96%), as did 61 of the 75 older patients (81%). Within each age group, the center's survival rate compares favorably to the nation's, even though the center's 85% overall survival is lower than the national average. The center's expected rate of survival is 83%—80% for the 75 older patients, and 92% for the younger 25 patients. Unlike the comparison to the national average, the favorable comparison of the center's overall survival rate to this expected rate is consistent with the findings specific to each age group.

Many other important differences besides age exist among patients and organs. To simultaneously adjust for a long list of factors in the same way that age is controlled for in this example, the SRTR uses the Cox regression model (2). This semi-parametric model is very flexible in the types of data, event rate patterns and covariates it can incorporate. More details about the models, including lists of covariates, can be found in the technical documentation to the CSRs at www.ustransplant.org/srtr_resources.aspx.

Table 2: Simplified age-based risk adjustment

Age group	National		At Hospital B		Center vs. nation comparison
	Percentage in group	Percent survival	No. alive	Percent survival	
0–44	50%	92%	24 of 25	96%	96 > 92: better
45+	50%	80%	61 of 75	81%	81 > 80: better
	Average		Expected	Observed	
Calculation	.5 × 92% + .5 × 80%		.25 × 92% + .75 × 80%		
Result	86%		83%		(24 + 61)/100 = 85% 85 < 86: worse (wrong); 85 > 83: better (correct)

Source: SRTR.

Table 3: Effect of expanded criteria donor definition components on kidney graft survival

Factor	Hazard ratio
Hypertension	1.23
Creatinine > 1.5	1.13
Donor age: 65+ (ref. = 35–49)	1.46
COD Stroke (vs. head trauma)	1.30
'ECD' classification	1.21

Calculated as exp (Beta) from 1-year kidney graft survival model, CSRs released 01/11/2005. Source: SRTR.

The Cox model allows us to calculate the effect on outcome for each characteristic of the recipient and donor, which can be taken together to calculate the expected outcome for each patient. This effect is how each factor is 'weighted' in the risk-adjustment process. For example, many programs use expanded criteria donor (ECD) kidneys for recipients whose expected waiting time for a better kidney increases their risk of dying before receiving a transplant. To ensure that a lower survival rate for transplant programs using ECD kidneys does not, on its own, indicate poor performance, we incorporate these donor factors into the models for expected survival. Table 3 shows many of the factors used in identifying an ECD kidney and their separate effects on 1-year graft survival. Not all ECD donors are characterized by all of these factors. A kidney from a donor with a history of hypertension, whether classified as ECD or not, carries with it a risk of graft failure of 1.23 times, or 23% higher, than that of an organ from a donor without hypertension (Table 3). If that same donor was also older than 65, the kidney would be another 1.46 times as likely to fail, for a total elevated risk of $1.23 \times 1.46 = 1.80$. By multiplying the hazard ratios listed, note that a kidney from a donor with all of the characteristics listed in Table 3 represents a graft failure risk more than three times that of a kidney from a donor with none of these characteristics.

Adjusting for the case-mix of patients is extremely important in interpreting posttransplant outcomes. Table 4

shows the range of expected 1-year survival for different organs, suggesting that the mix of patients transplanted varies tremendously among centers. For example, even though the national average 1-year liver graft survival was 82.1%, centers' expected survival ranged from 61.0% to 87.4%. The second panel of the table shows that this wide variation is not limited to smaller centers that may treat just a few particularly difficult (or easy) cases. Especially for centers at the far ends of these ranges of expected survival, a comparison to the national average survival could be quite misleading.

Viewpoints on posttransplant outcomes

To return to the analyses shown in Table 1 for Hospital A, is the difference we see between the observed survival of 87.78% and the expected rate of 89.41% large enough to be meaningful? The answer may depend on the user's perspective. Table 5 shows three different ways of looking at the same comparison of outcomes.

The percentage surviving at 1 year is only 2% lower than expected, an apparently small difference. However, the same difference appears more consequential when comparing the percentage that died, a full 15% higher than expected. Finally, for the 90 transplants performed over 2.5 years, the count of deaths observed during follow-up was 30% higher than expected, accounting for 2.5 excess deaths.

The differences among these interpretations are stark. The first change from a 2% difference to a 15% difference reflects the change in denominator—a small percentage point difference is a much smaller fraction of survival (usually a large number at 1 year) than of mortality. Several years after transplant, when survival rates may be close to 50%, this contrast would not be as evident.

The difference between the percentage that died and the death count is subtler. The expected number of deaths is calculated according to the time that patients are followed and surviving after transplant, so the expected number of

Table 4: Range of expected 1-year graft survival rates, July 2005 center-specific reports

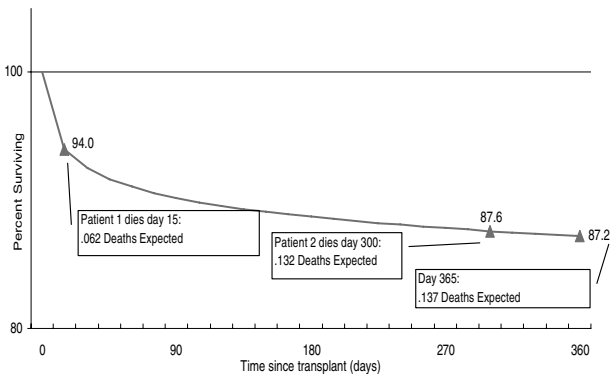
Organ	National rate	Range of center expected rates				
		Minimum	5th percentile	Median	95th percentile	Maximum
At all centers						
Heart	86.4	61.2	81.9	86.9	90.4	94.7
Lung	80.6	47.3	67.4	80.9	85.2	88.3
Kidney	91.5	84.7	88.2	91.8	94.6	96.8
Liver	82.1	61.0	76.0	82.8	86.8	87.4
At centers with 10 or more transplants in cohort						
Heart		79.6	82.8	86.9	90.3	91.2
Lung		52.6	68.1	81.1	84.9	85.8
Kidney		84.7	88.2	91.7	93.9	95.9
Liver		74.8	77.0	82.9	86.6	87.4

Source: SRTR calculations from CSRs released July 2005, www.ustransplant.org.

Table 5: Three interpretations comparing the same outcomes, example 'Hospital A'

	Expected	Observed	Ratio or relative risk	Interpretation
Percentage who survived after 1 year	89.41%	87.78%	0.98	2% lower
Percentage who died after 1 year	10.59%	12.22%	1.15	15% higher
Deaths during follow-up period	8.48	11	1.30	30% higher; 2.52 excess deaths

Source: SRTR.



Source: SRTR

Figure 1: Calculating expected deaths.

deaths for a patient whose follow-up ends—for any reason, including death—immediately after transplant is smaller than it would be if that follow-up extended longer. Therefore, this last statistic accounts for the difference between a patient who survives only briefly during follow-up, and one who survives nearly the entire period, patients who would be identical in the end-of-period accounting of 'percentage died.'

Figure 1, based on Table 6, illustrates this point. The curve shows the percentage surviving at each day after transplant for a given type of patient. It falls quickly from 100%, consistent with the immediate risk of surgery, before leveling out to reach a 1-year survival of 87.2%.

Fifteen days after transplantation, when Patient 1 died, we would have expected 0.062 deaths. (At any point in time t , the expected probability of death is calculated as $-\ln(S(t))$, where $S(t)$ is the survival percentage at that time. For survival percentages near 100%, this is closely approximated by 100 minus the survival percentage.) Visually, the expected probability of death is approximated by the vertical distance down from the horizontal line at 100% to the survival curve; this distance increases with the time since transplantation. For Patient 2, who died after 300 days, the vertical distance is larger and the expected number of deaths is 0.132. With this example survival curve, we assess a probability of death of 0.137 for any patient surviving until at least 365 days. Table 6 shows how observed and expected deaths would be counted and summed if a cen-

Table 6: Aggregating observed and expected events by center, example "Hospital C"

	Days followed	Observed death events	Expected death events	Ratio of observed to expected
Patient 1	15	1	0.062	16.10
Patient 2	300	1	0.132	7.56
Patient 3	365	0	0.137	0.00
—	—	—	—	—
Patient 15	365	0	0.137	0.00
Sum total		2	1.975	1.01 (Overall ratio)

Note: The 'sum total' line reflects the total for all lines, including the omitted lines for patients 4–14. Each omitted line has 0 observed death events and 0.137 expected death events.

Source: SRTR.

ter, Hospital C, transplanted 15 patients, including these 2 and 13 others who survived 1 year.

For both of the patients who died, the observed number of deaths (1) is far higher than expected, but more so for the patient who died on day 15 ($1/0.062 = 16$ -fold higher than expected) than for the patient who died on day 300 ($1/0.124 = 7.5$ -fold higher than expected). Each of the other patients has 0 observed and 0.137 expected deaths. For the 15 patients at Hospital C, the number of observed deaths (2) and number of expected deaths (1.975) compare quite closely: the ratio of 1.01 indicates that the center experienced about 1% more deaths than would be expected given this patient-risk group.

Note that different types of patients would have different curves, either higher (better survival) or lower (worse survival) than the one depicted in Figure 1. For illustration purposes we assume here that all patients are 'similar' and have the same expected survival curve; the actual CSR calculation of expected events takes into account the differences between patients by using a different survival curve for each patient.

Returning to Table 1 (CSR Table 11), the second panel (lines 5–10) focuses on these expected (8.48) and observed (11) deaths after transplant for Hospital A. The ratio's confidence interval suggests that while we estimate a ratio of observed to expected deaths of 1.3—or 30% more deaths than expected—there is a 95% chance that the 'true' ratio of observed to expected lies between 0.65 and 2.32. The

p-value measures the possibility that any discrepancy between observed and expected occurred by random chance alone: in this case, the p-value of 0.469 suggests that there is about a 47% chance that the difference occurred by random chance. Most statistical literature considers a p-value of less than 0.05 to indicate a 'statistically significant' finding; this is the significance threshold used in line 11 of Table 1.

This panel of CSR Table 11—observed and expected counts of deaths—is the most appropriate for use by those who want to identify centers that perform particularly well or particularly poorly, even though it may not be as intuitively interpretable as the percentage surviving 1 year after transplantation.

Considering pretransplant outcomes

Table 7 shows how the comparison between observed and expected rates carries over to waiting list outcomes. Hos-

pital D, shown in this representation of CSR Table 3, has a rate of 0.36 transplants per year that a patient spends on the waiting list, exactly the national average for 2004. The expected transplant rate for this program, only 0.27, suggests first that the types of patients served by this center typically wait longer or are more likely to die before transplant. The fact that the observed rate is higher than expected suggests that the program does a good job of achieving the goal of wait-listing (obtaining a transplant) for these types of patients—as long as it is not at the expense of accepting poor-quality organs. This trade-off is for one reason that it is important to consider both pre- and posttransplant outcomes.

Other waiting list activity tables (CSR Tables 4 through 6) show outcomes that may be more interpretable from the point of view of a patient on the waiting list, helping the reader understand the likely waiting times and likelihood of different events at different times after listing.

Table 7: Center-specific report Table 3—Transplant and mortality rates among wait-listed patients, sample liver center 'Hospital D'

	This center		U.S.
	01/01/2003–12/31/2003	01/01/2004–12/31/2004	01/01/2004–12/31/2004
Waitlist registrations			
Sample			
Count on waitlist at start	197	236	17 061
Transplant rate			
Person Years	167.2	236.8	16 925.7
Living and deceased donors			
Removals for transplant	82	86	6144
Transplant rate (per year on waitlist)	0.49	0.36	0.36
Expected transplant rate	0.31	0.27	0.36
Ratio of observed to expected transplants	1.58	1.35	1.00
95% confidence interval			
Lower bound	1.26	1.08	NA
Upper bound	1.96	1.67	NA
p-value (2-sided)	<0.01	<0.01	NA
How do the rates at this center compare to those in the nation?	Statistically higher	Statistically higher	NA
Deceased donors only (similar content to above)			
Mortality rate after being placed on waitlist			
Person years	170.0	241.1	17 459.7
Number of deaths	25	25	2424
Death rate (per year on waitlist)	0.15	0.10	0.14
Expected Death rate	0.12	0.15	0.14
Ratio of observed to expected deaths	1.25	0.67	1.00
95% confidence interval			
Lower bound	0.81	0.44	NA
Upper bound	1.84	0.99	NA
p-value (2-sided)	0.321	0.045	NA
How do the rates at this center compare to those in the nation?	Not significantly different	Statistically lower	NA

Source: SRTR center-specific reports, www.ustransplant.org.

Table 8: Methods for addressing loss to follow-up

	Kaplan-Meier (KM)	Extra ascertainment of mortality
Method	Assume lost patients have similar outcomes to followed patients	Assume a patient is alive unless we know otherwise from any of many sources (Social Security, CMS, other transplant centers)
Advantages	Helps produce an interpretable 'percentage surviving at the end of period' Allows patients with incomplete data to contribute to the results Available for both graft and patient survival	Verifies center reporting with external sources Limits bias
Disadvantages	Subject to biases if lost patients are not similar to followed ones Small samples can be unstable	Extra sources unavailable for graft survival
Used in calculating	Graft survival (CSR Table 10) Percent surviving at end of period Handles time after last follow-up Patient survival (CSR Table 11) Percent surviving at end of period Handles time after last expected follow-up	Graft survival (CSR Table 10) All statistics: mortality is counted as graft failure at any time while the patient is followed Patient survival (CSR Table 11) All statistics

Source: SRTR.

Accounting for the Uncertainty of Loss to Follow-Up

Every transplant program is responsible, as a condition of its participation in the national organ allocation system, for reporting on outcomes, such as death and graft failure until (and sometimes beyond) the time that the transplant is no longer functioning. However, many patients are difficult to gather data on following transplantation—particularly kidney recipients, who have an alternative treatment (dialysis) that does not require them to return to a transplant center. Rates at which patients become 'lost to follow-up' are as high as 15% by the third year after kidney transplantation, but less than half as large for other organs (SRTR analysis).

To calculate estimates of survival for patients who become lost to follow-up, the SRTR employs both the Kaplan-Meier (KM) estimation and extra ascertainment of mortality from additional data sources (Table 8).

The KM method uses the experience of patients who are followed to estimate the outcomes of patients who are lost to follow-up (3). For example, if we last know a patient is alive 6 months after transplantation, the KM method uses the average outcomes of other patients also alive 6 months after transplantation to estimate what would likely happen to this patient. This method allows the calculation of the intuitively understood 'percentage surviving at the end of period' in Table 1, even when not all patients have been followed until the end of the period (either because they have been lost or because the transplant was too recent).

Table 9 shows a simple example of how 1-year survival is calculated for a cohort of patients of whom half (Group B)

are followed for only 6 months. For the 90 patients in Group B who are alive at 6 months but not followed thereafter, our best guess is that they will have outcomes similar to the 86 patients from Group A who also survived until 6 months. For both groups together, the survival rate during the first 6 months is 88%, yielding an estimated 1-year survival rate of 80%. Using this method allows us to include more recent transplants with only partial follow-up available in survival rates. In the case of the center in Table 9, this allows us to give credit for improved outcomes among more recent transplants.

The SRTR also accounts for outcomes among transplant recipients who become lost to follow-up by examining additional data sources beyond the transplanting center, including:

- (i) Waiting list additions or retransplants at other centers.
- (ii) The Social Security Administration, from death benefit and employment records.
- (iii) CMS billing records and benefit information for kidney patients.

A comparison with the National Death Index leads us to believe that by using all of these sources, we are able to capture more than 99% of the deaths among transplant recipients that occur during the time that these sources—as well as follow-up forms—are expected to be complete (4). This considerable certainty allows us to assume, for patient (not graft) survival analyses, that a patient is alive unless we know otherwise. Extending the calculations for patients who have become lost, by adding both death events and time at risk, a center's survival rate may improve or be lowered. In either case, these calculations are less subject to

Table 9: Simple Kaplan-Meier calculation

	Group A: followed 1 year	Group B: followed 6 months	Both Groups (A and B)
Transplants	100	100	200
Months 0–6			
At risk, start of period	100	100	200
Deaths	14	10	24
Percent survived	86%	90%	88%
Months 7–12			
At risk, start of period	86	90	176
Deaths	8	Not yet observed	
Percent survived	91%		Best guess: 91%
Full 1-year survival	$86\% \times 91\% = 78\%$	$90\% \times 91\% = 82\%$	$88\% \times 91\% = 80\%$

Note: the simple mean of the one-year survival estimates for groups A (78%) and B (82%) equals the overall survival only because the two groups match each other in size.
Source: SRTR.

Table 10: Excerpts from model description tables, analytic methods to the center-specific report

Graft survival model description
1 year (and 1 month) after transplant
Organ: heart
Adult (Age 18+)
90.0% graft functioning at 1 year when all covariates = 0.
95.4% graft functioning at 1 month when all covariates = 0.
The indexes of concordance are 63.5%, 65.3%, and 66.2%, respectively.

Characteristic covariates	Beta	Standard error	p-Value
Diagnosis: cardiomyopathy	-0.0826	0.0916	0.3674
Diagnosis: congenital heart disease	0.6432	0.2130	0.0025
Donor age: 0–17	-0.2828	0.1548	0.0677
Donor age: 18–34	-0.2554	0.0990	0.0099
Donor age: 50–64	0.1483	0.1335	0.2666
Donor age: 65+	-0.4201	1.0034	0.6754
Donor: deceased, COD cerebrovascular/stroke	0.0518	0.0998	0.6035
Ischemia time (hrs): linear (ref = average time*)	0.1595	0.0414	0.0001
Recipient creatinine: >1.5	0.6046	0.0865	<0.0001

Source: SRTR Center-Specific Reports, www.ustransplant.org, July 2005 Release.

biases caused by patients being lost and probably reflect actual outcomes more accurately. While the effect on national rates is quite small, it can be quite sizeable—in either direction—for some centers (5).

As described in Table 8, for graft survival analyses (CSR Table 10) the KM method is used to estimate survival percentage after patients are no longer followed by their center or when lag time prevents complete follow-up. For

Table 11: Range of indexes of concordance, July 2005 posttransplant graft and patient survival models

	Models	Average	Minimum	Maximum
Heart	8	60.3%	52.8%	63.5%
Kidney	12	67.4%	60.1%	76.1%
Liver	16	69.1%	61.8%	81.6%
Lung	8	63.2%	61.8%	64.7%

Source: SRTR Center-Specific Reports, www.ustransplant.org, July 2005 Release.

these graft survival statistics, extra ascertainment of mortality is used only when it indicates a death that occurred during this reported follow-up time. For patient survival (CSR Table 11), the KM method is used to estimate survival only after lag time prevents complete follow-up from any of the available sources. Portions of the cohorts used for 1-year survival are recent enough that only a 6-month follow-up form is reliably expected by the time the CSRs are calculated. The KM method estimates addresses the follow-up time after 6 months for these recent transplants.

Both methods are also used in several measures of waiting list outcomes. The KM method is used when patients transfer to other centers in a time-to-transplant analysis, assuming that if they had not transferred, their time until transplant would be similar to other patients at the same center who had waited as long. Note that in such an analysis, patients who die are not ‘censored’ in this way, as we are certain that they would not be transplanted. Extra ascertainment of mortality is used to identify unreported deaths before (or soon after) a patient is removed from the waiting list, but before any transplant event.

Selecting Model Covariates for the Center-Specific Reports

All of the methods discussed here rely on the concept of risk adjustment, or asking the question, ‘what result would

we expect for similar patients, according to the national experience?' What variables should be included when we decide which patients are similar?

Patient characteristics? Almost always. Adjusting for patient characteristics helps ensure that centers are not penalized for treating patients who are more likely to have poorer outcomes. For example, the age of the recipient is closely associated with outcomes, and not controlling for age might penalize centers that treat older patients.

Donor characteristics? Much of the time. The current move toward using more ECD kidneys provides an excellent example: by not controlling for these characteristics, which are known to result in elevated risk of graft failure, we would unfairly compare outcomes of ECD and non-ECD recipients, which might discourage the use of ECD organs. However, since choosing an appropriate donor is important, we may not want to adjust for all donor characteristics.

Transplant center characteristics? Usually not. Center volume is a good example of a characteristic that should not be included in these models even though it may be associated with better outcomes. In terms of performance, we want to give due credit to larger centers that perform well rather than adjusting away differences associated with volume.

The SRTR updates CSRs every 6 months, which allows ongoing adjustments to be made to the risk-adjustment models. At each report, the risk adjustment is recalculated, and each year the SRTR focuses on reviewing the entire set of risk-adjustment covariates for one or more organs. Models for kidney survival are being restructured in 2005 as lung and liver models were in 2004 and heart models were in 2003. The SRTR plans to continue this cycle.

Selection of model covariates is based on the entire body of analytical work performed by the SRTR for the OPTN committees and other groups. For each report, many separate models are estimated for each organ. Pediatric and adult transplants are evaluated with separate models because of different factors influencing pediatric survival (e.g. immune responsiveness and compliance with medications). Similarly, separate models are calculated for transplants from living and deceased donors, for patient and graft survival, and for different study endpoints (e.g. 1-month vs. 3-year outcomes). Separating models allows us to use covariates specific to each transplant type; it also allows their effects to vary.

Input from the organ-specific OPTN committees is particularly important when considering the clinical plausibility of each risk-adjustment model. The process for developing these models involves several steps repeated each time the models are updated.

Are the data available? The list of covariates that could be used in these models includes all the data elements collected by the OPTN during the cohort period. Characteristics that may be clinically significant cannot be included in the models unless they are collected consistently for all transplant patients in the country, creating some trade-off between full adjustment and data submission requirements for transplant centers.

What are the known predictors of survival? From the list of available covariates, we focus on those shown to be important in SRTR analyses or the medical literature. We usually start by including variables that often display p-values below or nearly below 0.10, even if they may not be significant at the 0.10 level in this particular model. In some cases, decisions must be made about which specific variables to use to incorporate certain factors into the model when there are several highly associated variables to choose from. These decisions are based on significance, interpretability of coefficients and data quality.

Are there additional factors that we know or suspect are clinically significant? Based on input from clinical experts from the SRTR and the OPTN organ-specific committees, additional variables are tested for inclusion in the model. Some of these are only added to the models if they reach a certain level of statistical significance; others may be included regardless of their statistical significance because they are widely believed to have an effect on survival.

Are we modeling each variable correctly? The proper form must be chosen for each covariate. Some variables may have a linear relationship with the outcome (e.g. cold ischemia time may be measured in effect per hour), while others use categories, allowing nonlinear relationships between the covariate and outcome. Often, categorical variables are chosen because of their versatility. In addition, interactions among variables in the model are examined.

Communication and documentation of the models

Each risk-adjustment model is published 1 month in advance of the CSRs. These models are presented as tables with the features described below; an excerpt from such a table appears in Table 10.

- (i) **The beta, or calculated coefficient**, shows the effect of each characteristic on expected risk of death or graft failure. Some users may be more familiar with the relative risk of each factor, which can be obtained by calculating $\exp(\beta)$.
- (ii) **The standard error and p-value** indicate how much random variance there was around this estimate, and our degree of certainty that the given characteristic has a real effect.
- (iii) **The index of concordance** measures the goodness of fit for each model. This measure shows the percentage of variation in the order of events (deaths or graft

Table 12: Percentage of centers flagged for adult patient survival by each review criterion, July 2005 center-specific reports

	Kidney	Liver	Heart	Lung	All
Number of programs	256	126	143	74	599
Percent flagged as:					
Actionable: O/E > 1.5	21.5%	17.5%	24.5%	24.3%	21.7%
Important: O-E > 3	12.5%	20.6%	8.4%	16.2%	13.7%
Significant: one-sided p < .05	7.0%	10.3%	11.2%	9.5%	9.0%
Overlap of flags:					
None	77.0%	74.6%	75.5%	71.6%	75.5%
Exactly one	11.7%	11.9%	12.6%	16.2%	12.5%
Exactly two	4.7%	4.0%	4.2%	2.7%	4.2%
All 3	6.6%	9.5%	7.7%	9.5%	7.9%

Source: SRTR.

failures) that is accurately predicted by the model. An index of concordance of 100% would suggest that the model perfectly predicts the order of events displayed in real life; 50% would suggest that the order is random with regard to predictors. Indexes of concordance are best for organs with many transplants in each cohort, such as liver and kidney for adult recipients. Table 11 shows the range of indexes of concordance for the July 2005 reports.

- (iv) **Models are repeated** for a series of three different cohorts of transplants, allowing a comparison of how stable the coefficients are across time.

To refer back to the earlier example of adjusting for ECD kidney donor characteristics, these tables allow us to see just how these factors are fitted in the model. Examining the kidney 1-year graft survival model, the fact that a patient received an ECD organ carries with it an increased risk of 20%; separately, the models also control for the components of the ECD definition—age, hypertension, high creatinine and stroke. By adjusting for all of these characteristics separately, we adjust for the fact that some ECD organs carry with them higher risk than others.

Using Center-Specific Outcomes to Select Centers for Review

The Membership and Professional Standards Committee (MPSC) of the OPTN works to ensure that member transplant centers remain in compliance with the criteria for OPTN membership. This role includes identifying centers that may not perform well, with the intention of helping them implement corrective action or reconsidering their membership. Because resources do not allow a close review of practices at all centers, the SRTR worked closely

with the MPSC to develop screening criteria to help identify and prioritize centers that are more likely to require attention. These criteria, along with the CSR calculations on which they are based, also figured prominently in the proposed Hospital Conditions of Participation for the Medicare program recently issued by CMS.

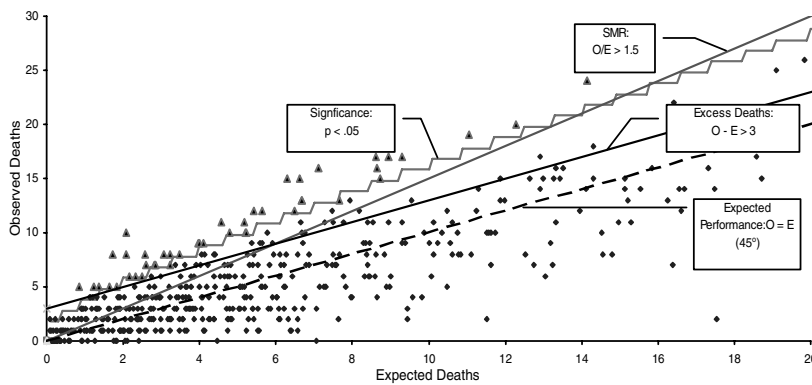
Concepts: actionable, important and significant

To be identified for further review by the MPSC, differences between observed and expected must meet all of the following criteria:

- (i) **Actionable:** a clinically significant pattern, suggesting a higher likelihood that practices contributing to poor outcomes might be identified, indicated by a high fraction of excess deaths
 - (a) Standardized Mortality Ratio (SMR) > 1.5; observed deaths divided by expected deaths greater than 1.5 (O/E > 1.5)
 - (b) Interpretation: there were more than 50% more deaths than expected
 - (c) CSR Tables 10, 11: line 8
- (ii) **Important:** the magnitude of the problem, in terms of potential lives saved, should be sufficient to take action and place the center near the top of the priority list
 - (a) 'Excess Deaths' of at least three; observed deaths minus expected deaths greater than 3 (O - E > 3)
 - (b) Interpretation: there were more than three deaths beyond what would be expected among the recipient cohort
 - (c) CSR Tables 10, 11: subtract line 7 from line 6
- (iii) **Significant:** it should be unlikely that the difference occurred by random chance alone
 - (a) One-sided p-value less than 0.05 (p < 0.05)
 - (b) Interpretation: there is less than a 5% chance that a poor (rather than different in either direction) outcome occurred by simple random variation
 - (c) CSR Tables 10, 11: line 10 shows a two-sided p-value; obtain a one-sided p-value by dividing these in half, for outcomes where O > E.

Each of these three thresholds is chosen with targeting facilities for review in mind. It might be possible, after several of the centers identified in this fashion have been reviewed, to 'lower' any of these criteria (using higher p-value or smaller differences between O and E), identifying additional centers. These criteria were designed to identify centers most in need of review.

In implementing these criteria, all comparisons should be based on observed and expected events during the time a patient is actually followed either by the center or, in the case of patient survival, by extra ascertainment (i.e. they should not be based on any results imputed by the KM method). These comparisons should also account for the difference in outcomes between a patient who dies in the



Source: SRTR Calculations from July 2005 CSRs. Each point represents a kidney, liver, heart, or lung center in the July 2005 CSRs. 33 facilities with expected deaths > 20 or observed deaths > 30 are not shown for scale.

Figure 2: Plot of centers flagged for adult patient survival by each review criterion, July 2005 center-specific reports.

first week versus the fifty-first week after transplantation. Therefore, these criteria are applied to the comparison of counts of observed and expected deaths as presented in 'Deaths during follow-up period', lines 6 and 7, in Table 1—the comparison described in the third row of Table 5, as well as to the graft failure equivalent of this outcome.

How many centers are affected, and by which flags?

Figure 2 shows how these three criteria affect actual centers. Each transplant center is plotted with observed deaths on the vertical axis and expected deaths on the horizontal axis (a few of the largest centers, with high expected deaths, are omitted for scale). The dotted line indicates where observed equals expected; centers that fall below and to the right of this line have fewer observed deaths than expected. Three other lines correspond to the MPSC criteria: (i) parallel to the dotted line, three observed deaths vertically above, is a line indicating the $O - E > 3$ threshold; (ii) rising more quickly from the origin with a slope of 1.5 is a line indicating the $O/E > 1.5$ threshold; (iii) the stair-stepped line indicates, for each number of expected deaths, the number of observed deaths necessary to achieve a one-sided p-value of < 0.05 .

To be flagged for review under MPSC (or CMS-proposed) criteria, a center must have enough observed deaths to fall above and to the left of all three of these lines. For most transplant centers, those with expected death counts between about 2 and 15, the stair-stepped p-value is the 'binding constraint,' or the highest of these lines. For some very small centers, the 'actionable' criteria (excess deaths) is the relevant binding constraint; for the very largest centers the 'important' criteria ($SMR > 1.5$) is the relevant line. While many facilities, particularly small ones, have an SMR above 1.5, very few of these meet either of the other criteria: many of the plotted dots in the lower left-hand corner are above the SMR line but below both others. For this reason, the MPSC and the SRTR are developing further methodology targeted at identifying smaller centers for review. In the meantime, the current methodology is more

likely to prioritize larger centers because of the 'important' constraint.

Table 12 shows the number of facilities that fall into each of these categories according to the July 2005 CSRs. For each organ shown, at least 20% of centers fall short of at least one criterion; 7–10% of centers, by organ, are flagged for review by all the three criteria. Many heart and lung centers, which tend to be small, fail the O/E criterion, consistent with the data depicted in Figure 2: for centers with few expected deaths (including small centers), a slight elevation in observed deaths may easily meet this criterion without bringing the center to the binding criterion for small centers, $O - E > 3$. The fact that the percent flagged on all the three criteria is higher than the percent flagged on exactly two confirms correlation among the criteria—centers with at least two flags are more likely to have all the three flags qualify.

Comparison to expected versus ranking centers

The comparisons and tests outlined above are intended to evaluate how well centers perform compared to risk-adjusted national averages; they are not intended for ranking centers relative to each other. While ordering a list of centers by observed survival rate is clearly incorrect (as survival rate may reflect either success or good patient case mix), even ordering by the SMR is problematic because of differences in the variance of the SMR estimate among centers. For example, such an order could imply that a center with an SMR of 0.8, but not significantly different than expected, performs better than a center with an SMR of 0.9 that is significantly better than expected; this is not necessarily true. No p-values or statistical tests presented measure a real difference between two centers. Users should be judicious when using or presenting data that might encourage false comparisons among centers.

Implementing the Screening Concepts

The MPSC continuously reviews program performance, as authorized by the National Organ Transplant Act (NOTA),

to oversee the quality of transplant services in the United States. The committee (made up of transplant professionals and recipient or donor family representatives) ensures that OPTN members, including clinical transplant programs, remain in compliance with OPTN criteria for institutional membership.

It is the goal of the MPSC review and audit process to ensure that patients receive quality transplant services and assist programs with improving their level of care. Programs that are identified as experiencing lower-than-expected outcomes first encouraged to implement corrective action, before any recommendations for adverse actions. However, the MPSC is ultimately responsible for the welfare of the patients at all centers, including those that appear to be offering transplant services with outcomes that are well below those anticipated.

Four times each year, the SRTR provides the MPSC with an updated report on all transplant programs, without any indication of transplant center name or location. The report provides much of the same information shown in Table 1: the number of transplants performed, the observed and expected numbers of graft failures and deaths, observed and expected survival rates and a one-sided p-value to measure statistical significance. These results, pertaining to 1-year survival, are shown for two recent and overlapping cohorts (in 2006, the MPSC will change from 2-year to 2.5-year cohorts to match the public CSRs). An earlier 5-year cohort of transplants is also included for historical reference. Each year, only one pair of transplant cohorts is examined by the MPSC; updated reports from the SRTR provide more recent and complete follow-up information, while the cohort of transplants examined moves forward only once per year.

Larger programs (10 or more transplants per cohort) that meet all the three criteria—actionable, important, and significant—for two consecutive cohorts, either for graft or patient survival, enter the MPSC audit process. Requiring programs to meet all the three criteria for two consecutive cohorts further ensures that programs are being appropriately identified for evaluation.

Using this methodology, smaller transplant programs (fewer than 10 transplants per cohort) are rarely flagged on all the three criteria. Therefore, the MPSC conducts separate reviews of these programs. The SRTR provides the MPSC with an annual report listing all small-volume programs that had at least one death or graft failure during the evaluation period. The committee then reviews data on patient outcomes for these centers, including transplant volume summaries, causes of death and graft failure, comparisons to national survival statistics, performance in years after the initial review period and survival rates based on a 5-year cohort. A program may enter the MPSC audit process if this review reveals concerns about its performance. The SRTR and MPSC are currently revising the

methodology for identifying possible underperformers among small programs.

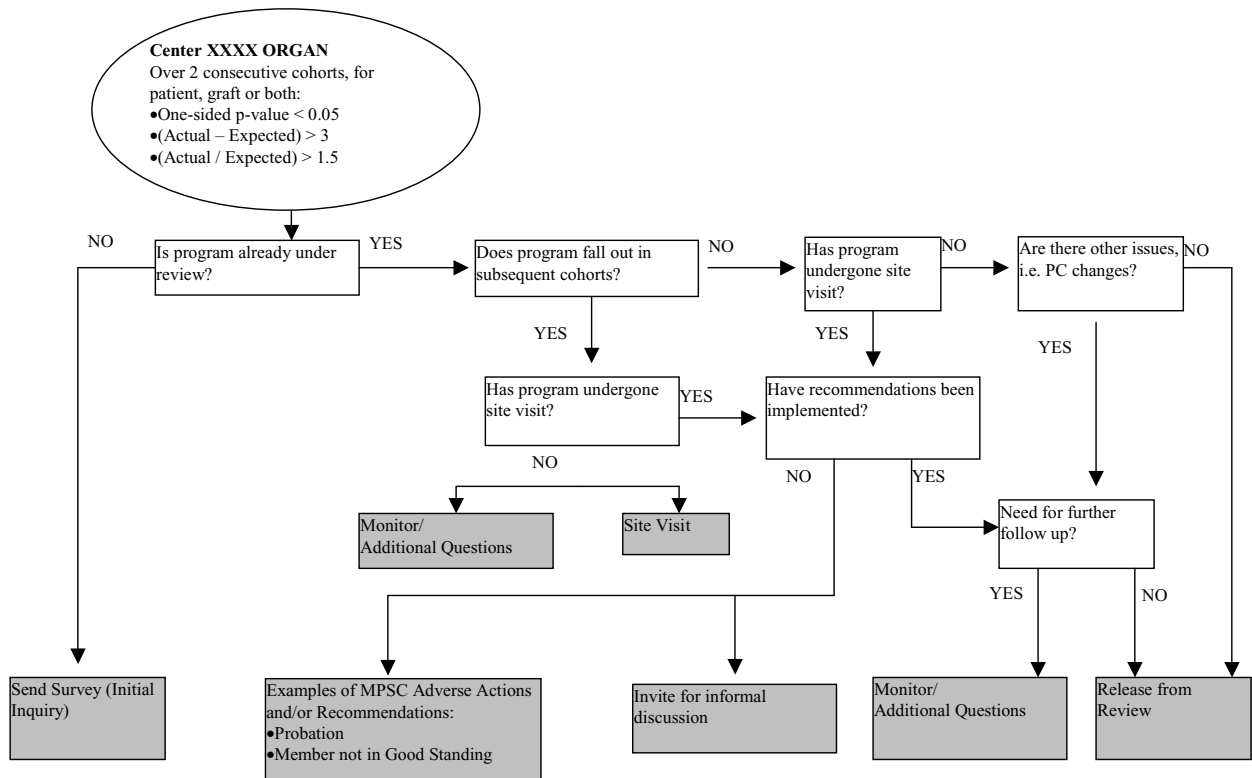
MPSC audit process

Figure 3 provides an overview of the course of action for those programs identified for comprehensive MPSC audits. Once a program, either small or large, enters the MPSC audit process, it is sent an initial survey to validate the data submitted into UNet, upon which screening criteria were based. This survey requests additional information on program activity, such as the number of patients evaluated for listing during a designated period, and provides an opportunity for the program to inform the MPSC of unique clinical aspects that may have influenced the observed survival rates. A synopsis of the deaths and graft failures that occurred within 1 year of transplantation is also requested for MPSC review. The MPSC considers changes in key personnel, as well as the causes of graft failure and death, in determining which programs require further study.

During the audit process, the MPSC may release the program from review if the committee is satisfied that the issues that led to the lower-than-expected outcomes have been addressed by the program, or if the survival rates in subsequent years have improved. Alternatively, the MPSC may continue to monitor the program by following outcomes in successive recipient cohorts, or it may recommend corrective or adverse actions.

If the MPSC has concerns about the performance of a transplant program and its ability to improve outcomes on its own, the committee may offer the program the opportunity to undergo a site visit from a team usually including a transplant surgeon, transplant physician, an administrator and UNOS/OPTN staff. For 2 days, the team interviews key personnel, conducts in-depth reviews of relevant patient charts, and reviews hospital facilities. At the conclusion of the visit, a preliminary summary of findings is given to the center, with a formal report submitted to the MPSC for issuance to the program. The program must submit an action plan, current data and progress reports in response to the committee's recommendations. The MPSC's recommendations for corrective action may include revision and standardization of protocols, such as for immunosuppression or ECD donors; additional staff such as social workers, nephrologists, or posttransplant coordinators; implementation of clinical practice guidelines; or allocation of resources for continuing education for a range of staff.

The MPSC continues to monitor the program's progress in implementing the site visit recommendations as well as changes in its subsequent outcomes. During monitoring, the committee may also invite program staff for an informal discussion of current outcomes and activities; these discussions do not, in themselves, constitute an adverse action.



Source: OPTN.

Figure 3: Overview of MPSC review and corrective action process.

If the MPSC concludes that the program has not taken appropriate steps to improve its outcomes, such as submitting and complying with a corrective action plan, the committee may recommend to the OPTN Board of Directors that an adverse action be taken against the program. Recommended actions could include placing the member on probation, withdrawing the transplant program from OPTN membership, or making it a Member (of the OPTN) Not in Good Standing. Any program recommended for adverse action is offered due process, including the opportunity to participate in an interview and present new information, after which the MPSC may make a recommendation to sustain its previous recommendation, rescind the recommendation, alter the recommendation, or hold the recommendation in abeyance. If the recommendation is sustained, the program may participate in a formal, in-person, hearing with the MPSC. Adverse recommendations sustained at this point may be challenged by appeal to the OPTN Board of Directors for review.

In an appellate review, programs appear in person and discuss their challenge to the MPSC recommendation directly with the OPTN Board. The Board may sustain, alter, or rescind the MPSC recommendation. Further appeal may be directed, in writing, to the Secretary of Health and Human Services.

Further detail regarding the appeals process may be found on the OPTN website, at <http://www.optn.org/policiesAndBylaws/bylaws.asp> (see Bylaws Appendix A—Application and Hearing Procedures).

The consequences of being a transplant hospital 'Member Not in Good Standing' may include withdrawal of voting privileges in OPTN/UNOS affairs, or suspension of the program's personnel from OPTN committees and Board of Directors. A formal notification of the Member Not in Good Standing status is made to the OPTN Membership, UNOS, state health commissioner or other appropriate state representative, patients and the general public in the program's area, and the Secretary of the Department of Health and Human Services (HHS).

Since 1999, 261 programs have been reviewed for outcomes by the MPSC.

Conclusion

Measuring and monitoring performance—be it posttransplant and waiting list outcomes by a transplant center, or organ donation success by an OPO and its partnering hospitals—is an important component of ensuring good

care for people with end-stage organ failure. Many parties have an interest in examining these outcomes, from patients and their families to payers such as insurance companies or CMS; from primary caregivers providing patient counseling to government agencies charged with protecting these patients. It is important for all of these users to have at their disposal the best statistical methods, computed consistently for all transplant providers, based on the most reliable and complete data available. Moreover, it is important that these readers understand the central concepts important to using these statistics.

In this article, we have used the example of graft and patient survival to explain these important concepts. It should be well understood, though, that graft and patient survival are only a piece of the puzzle constituting good patient care, and that similar measures are available and pertinent for waiting list outcomes such as mortality or transplant rates. All of these measures rely on the concepts described here: the risk adjustment that allows fair comparison despite differences among patients treated, methodology for dealing with incomplete data and a basic understanding of how to interpret the magnitude and direction of these outcomes. We provide a detailed primer on these concepts that will enable readers to use these statistics wisely, as well as provide background to some of the statistical methods used in many other analyses comparing outcomes or performance, such as the OPO-specific reports. Finally, we have offered an example of the effective use of these posttransplant outcome statistics for screening transplant center performance to identify centers that may need remedial ac-

tion by the OPTN Membership and Professional Standards Committee.

Acknowledgment

The Scientific Registry of Transplant Recipients is funded by contract number 231-00-0116 from the Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services. The views expressed herein are those of the authors and not necessarily those of the U.S. Government. This is a U.S. Government-sponsored work. There are no restrictions on its use.

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.

References

1. Centers for Medicare and Medicaid, Department of Health and Human Services. Medicare Program; Hospital Conditions of Participation: Requirements for Approval and Re-Approval of Transplant Centers To Perform Organ Transplants; Proposed Rule. In: Federal Register 42 CFR Parts 405, 482, and 488; February 4, 2005: 6140–6182.
2. Cox DR. Regression models and life tables (with discussion). *J Roy Stat Soc Series B* 1972: 197–220.
3. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53: 457–481.
4. Dickinson DM, Ellison MD, Webb RL. Data sources and structure. *Am J Transplant* 2003; 3(Suppl 4): 13–28.
5. Dickinson DM, Bryant PC, Williams MC et al. Transplant data: Sources, collection, and caveats. *Am J Transplant* 2004; 4(Suppl 9): 13–26.