American Journal of Transplantation 2003; 3 (Suppl. 4): 103–113
Blackwell Munksgaard

Blackwell Munksgaard 2003
ISSN 1601-2577

Analytical approaches for transplant research

Robert A. Wolfe[a]*, Randall L. Webb[b], David M. Dickinson[b], Valarie B. Ashby[a], Dawn M. Dykstra[b], Tempie E. Hulbert-Shearon[a] and Keith P. McCullough[b]

[a] *Scientific Registry of Transplant Recipients (SRTR)/ University of Michigan, Ann Arbor, MI*
[b] *SRTR/University Renal Research and Education Association (URREA), Ann Arbor, MI*
**Corresponding author: Robert A. Wolfe, bobwolfe@umich.edu*

Key words: Death ascertainment, OPTN, SRTR, statistical analysis, survival analysis, transplant research

Received 17 September 2002, revised and accepted for publication 20 November 2002

Introduction

This article is intended to provide a description of the analytical approaches used in various Scientific Registry of Transplant Recipients (SRTR) analyses, including the *2002 OPTN/SRTR Annual Report*, the center-specific reports (CSRs), and analyses for committees of the Organ Procurement and Transplantation Network (OPTN) and Advisory Committee on Organ Transplantation (ACOT). It is about 'choosing the right method to answer the right question' and the issues involved in selecting from among alternative approaches to data analysis.

Those who are likely to use the results in this report form the audience for this article, including physicians, patients, policy makers, and administrators. We attempt to address many of the issues that are raised frequently by these audiences as they use SRTR analyses.

Funding: The Scientific Registry of Transplant Recipients (SRTR) is funded by contract #231-00-0116 from the Health Resources and Services Administration (HRSA). The views expressed herein are those of the authors and not necessarily those of the US Government.
This is a US Government-funded work. There are no restrictions on its use.

Note on Sources: The articles in this supplement are based on the reference tables in the *2002 OPTN/SRTR Annual Report*, which are not included in this publication. Many relevant data appear in figures and tables directly referred to in the article; other tables from the *Annual Report* that serve as the basis for this article include the following: Tables 1.5, 1.6, 1.12-1.14, 5.2, 5.8, 5.9, 6.2, 6.8, 6.9, 7.2, 7.8, 7.9, 8.2, 8.8, 8.9, 9.2, 9.8, 9.9, 10.2, 10.8, 10.9, 11.2, 11.8, 11.9, 12.2, 12.8, 12.9, 13.2, 13.8, and 13.9. All of these tables are also available online at http://www.ustransplant.org.

The data collected by transplant centers and organ procurement organizations (OPOs) and submitted to the OPTN are designed primarily to facilitate the efficient allocation of organs to candidates and to allow limited evaluation of the outcomes of this process. These data have become an increasingly rich source of information about the practice and outcomes of solid organ transplantation in the United States. The SRTR has recently augmented the OPTN data by linking them with other data sources. See 'Data Sources and Structure', a companion article in this report, for details on these sources and linking (1).

The use of appropriate analysis methods is especially important for transplant data because of the complex longitudinal (time to event) nature of the data and the wide variation in medical practices, organs, candidates, and recipients present in the data. Careful linking and accounting of the sequence of events for individuals are fundamental to all of the methods described here. The SRTR uses a variety of methods for different types of data, but in this article we focus on methods for the analysis of time to event data. Standard statistical methods are used to aggregate data over time, including actuarial methods for the calculation of average rates, Kaplan–Meier survival curves, and Cox models. These statistical methods are designed to yield useful and interpretable results when data are combined from groups of individuals with different characteristics and lengths of follow-up, and with some incomplete data. However, it is also necessary to consider the limitations of the data in order to qualify the types of conclusions that can be reached from them.

Defining Salient Features of the Transplant Process

When summarizing the transplantation process, a surprising number of issues arise that involve deciding what to count and how to count them. Many of these issues are discussed in the 'Data Sources and Structure' article in this supplement. Appendix A in the *Annual Report* gives detailed definitions and methods for counting deceased donors, living donors, organs recovered, waiting list registrations, transplant candidates, time spent on the waiting list, waiting list offers of organs, transplant operations, organs transplanted, waiting list deaths, graft failures, and post-transplant mortality. A few of the most important distinctions are listed below.

A *donor* is any person, living or deceased at the time of organ removal, from whom an organ is procured with the intention of transplant, whether that organ is eventually

transplanted or not. *Deceased donors,* also known as cadaveric donors, are persons who have organs removed for transplantation after death. *Living donors* may include individuals donating single kidneys, liver segments or domino livers, lung lobes, or pancreas segments. A living donor may also contribute a healthy heart when receiving a simultaneous heart–lung transplant.

Transplanted organ counts may differ from the number of *transplants*. A kidney and pancreas transplanted from the same donor to the same recipient count as one kidney–pancreas transplant, but two transplanted organs; many other multiple-organ combinations exist. A single liver may be split into two segments for transplants in two different recipients, leading to one organ recovered, two separately coded organ dispositions, and two separate transplants. A single *recipient* may have more than one transplant operation, such as a pancreas after kidney transplant, or a retransplant following graft failure. Even among organs from living donors, the number of transplants may be different from the number of living donors. For example, a living donor might donate a kidney and pancreas segment, or two living donors might each donate a lung lobe for one transplant procedure.

The number of patients is not necessarily the same as the number of candidates or waiting list registrations. A *waiting list registration* begins each time a patient is placed on a waiting list at a transplant program. An individual may have many registrations, occurring in sequence at different centers or the same center, overlapping as a 'multiple listing' for the same organ at different centers, or for more than one organ altogether. In most tables in the *Annual Report*, registrations are shown. *Candidates* include all registrations, as well as recipients of living donor transplants, who may not have been registered on any waiting list. An individual, counted only once for multiple listings or transplants is referred to as a *patient*.

Statistical Methods Overview

The analysis of time to event (wait-listing to transplant, transplant to death, or transplant to graft failure) is complicated by the fact that the reader often wants to know the prognosis for the current groups of patients, while complete outcomes data are available only for those who have been followed for several years. Since improvements in medical practices and changes in organ allocation policy are occurring rapidly, it is useful to utilize the most recent data available, but interest in long-term outcomes necessitates using older data, too. For example, transplant failure rates during the fifth year after transplant can only be estimated by using the experience of the cohort of patients who received a transplant at least 5 years ago. Even among those transplanted more than 5 years ago, some of the patients may have been lost to follow-up, which complicates the calculation of cumulative event rates. Another analytic goal is to improve the precision of estimates by including more subjects in a study, but this also occurs at the expense of using older data.

For the reasons above, the analysis of waiting list and transplant outcomes depends strongly on statistical methods that can combine data from different cohorts of patients that have been followed for different lengths of time. Actuarial methods meet this goal: they can combine data for various time intervals during follow-up, combine data from several cohorts of patients, and use data from patients who are lost to follow-up or who are dropped from analysis at selected time points.

Actuarial methods (2) use separate estimates of event rates during successive intervals of follow-up time. For each interval of time, only those subjects who are still being followed at the start of the interval are used to estimate the event rate during that interval. For example, subjects who transfer to another center during the first month after wait-listing are dropped from the calculation of transplantation rates during the second month after wait-listing. The rules used to drop persons at specific times from the analysis depend on the specific objectives of the analysis. A subject who is dropped from analysis at a particular time is identified as 'censored' at that time (i.e. the subject contributes to the survival analysis only up to that time point). The conventions used for censoring can be complicated, depending on the data and the analysis objectives.

Actuarial methods combine these separate estimates of event rates during successive intervals of time since start of follow-up (event rates during the first year, second year, etc.) to yield estimates of projected cumulative event rates based on the rates during successive intervals. These actuarial methods thus allow data for event rates from a specific interval of time to be combined from all of the various study cohorts that have been followed at least to the end of that time interval. For example, data from the most recent 2 years of transplant recipients could be used to estimate failure rates during the first year since transplant, while those transplanted 2 years ago can be used to estimate failure rates during the second year. Each group of patients yields data about failure rates during the successive intervals of time (first year, second year) that they were followed. Another important benefit of these methods is that they allow data to be used from recipients who were lost to follow-up, by using their data to estimate rates during each of the time intervals up until the time that they were lost to follow-up. The following examples show how the calculations work.

Following a single cohort with no loss to follow-up

If 10 of 100 organs, from a cohort transplanted 2–3 years ago, failed during the first year after transplantation, then 90% of the grafts survived to the end of the first year. If

18 of the 90 still functioning at the end of the first year failed *during* the second year, then 80% of the grafts surviving at the end of the first year subsequently survived to the end of the second year. Together, the survival fractions during each year can be multiplied to yield the result that 72% (i.e. 80% of 90%) of the original 100 grafts survive to the end of the second year. In this example, actuarial methods were not needed since it can be seen that 28 of the original 100 organs failed by the second year.

Combining cohorts

If 14 of 100 organs from a cohort transplanted 1–2 years ago failed during the first year after transplantation, then 86% of those grafts survived to the end of the first year. Combining this with the cohort in the example above yields an estimated failure rate during the first year of 12% (24 failures among 200 transplants). The resulting estimate for the first-year survival rate of 88% is more precise than the estimate based on a single cohort because the sample size has been doubled from 100 to 200. The actuarial estimate for the cumulative fraction surviving to the end of the second year is 70% (80% of 88%) and uses all data available from the most recent two cohorts to estimate failure rates during the first year ($n = 200$) and during the second year ($n = 100$).

Using incomplete data due to loss to follow-up

In the first example above, suppose that 10 patients had been lost to follow-up during the first year, so that only 80 were followed during the second year, with 16 failures. The failure rate during the second year would be estimated as 20% (16 of 80), and the survival rate during the second year would be 80%, as before. Ideally, those lost to follow-up have the same failure rates as those remaining in the study, as in this example, so that the resulting estimates are unchanged for each time interval.

The examples above show that actuarial calculations depend on dividing the follow-up period into successive intervals of time. The calculations are more accurate if shorter time intervals are used. The Kaplan–Meier method (3), which is used by the SRTR, uses successive time intervals 1 day in length. The Kaplan–Meier method is used in determining graft and patient survival rates and time-to-transplant calculations in the CSRs and the *Annual Report*.

A closely related methodology estimates the *average event rate per unit time* during the entire period of follow-up by dividing the total number of events observed during follow-up by the total follow-up time, until the event or censoring. The following example shows how the calculations work for 100 recipients followed during their first year until failure or censoring. Suppose that two transplants fail (at 3 months and at 9 months) and that three people are lost to follow-up (at 2, 3, and 4 months). The first step is to compute the total time of follow-up. The two persons with transplant failure were followed for a total of 1 year (3+9 months) and the three losses to follow-up were followed for 0.75 years (2+3+4 months). The 95 remaining persons were followed for 1 year each, for a total of 95 years of follow-up time. The total follow-up time for the entire group is thus 96.75 years, so the failure rate is estimated as 20.7 ($1000 \times 2/96.75$) failures per 1000 person-years. As with actuarial methods, follow-up can be censored when patients are lost to follow-up or dropped from analysis, when calculating average event rates. For the *Annual Report*, we use event rate analyses in Tables X.3 and X.7 of each organ-specific section, death rates on the waiting list and after transplant, respectively.

Both actuarial methods (also called survival analysis methods) and the average failure rate method account for censored data. Results from the actuarial methods are reported as percentages at selected time points during follow-up. Event rates are typically reported as event rates per patient-year during a period of time. The SRTR uses these methods for many aspects of transplantation research, as described below. In each case, the objective of the analysis helps to determine if and when a patient is to be censored, or dropped from the analysis.

The actuarial fraction without an event at the end of an interval is related to the average event rate during the interval by the exponential (inverse natural logarithm) function, as shown in the following example. If the annual event rate (e.g. death rate) is 15%, then 86% ($= \exp(-0.15)$) are expected to be event-free (e.g. alive) at the end of 1 year.

Time Waiting for a Transplant

There are several appropriate answers to the question 'How long is the wait for a transplant?' that involve answering the important components of this query: 'What fraction of people who waited received a transplant?'; 'How long did these successful candidates wait?'; and, 'How long did the unsuccessful candidates wait, and would they still be waiting if it weren't for some adverse event such as death?'

The median time from wait-listing to transplant among transplant recipients summarizes how long successful candidates had to wait for their transplant. However, time spent waiting by successful candidates does not account for time spent waiting by unsuccessful candidates, nor does it account for the fact that registrants who died or were removed from the waiting list will never receive a transplant. Different methods are described below for summarizing time-to-transplant data. Methods for all candidates are contrasted to those for

recipients only. Two methods for all candidates are described. One accounts only for the actual time spent waiting on the list (censoring at death or removal from the list). The other accounts for the fact that registrants who die or are removed from the list will never receive a transplant.

For the purposes of ranking different regions or groups of candidates, all of the methods described here typically yield similar results. The median time to transplant among recipients can be easily computed by counting recipients during a recent interval of time, without the need for actuarial methods. The other methods all use actuarial methods based on follow-up of a cohort of candidates after wait-listing. The ease of calculation and immediacy of the results for time to transplant among recipients makes such statistics useful for comparing time to transplant for different groups (e.g. different organ procurement organizations). However, the times to transplant among recipients do not give a realistic appraisal of the total time spent waiting by all transplant candidates, and may not convey an appropriate message to a new registrant about the prospects for receiving an organ. The median waiting time among transplant recipients is typically much shorter than the waiting time for all candidates looking forward to a possible transplant.

The outcomes for all wait-listed candidates may be summarized in detail by the fraction who receive a transplant, die without transplant, are removed for various reasons, are still surviving after removal from the list, and are still on the waiting list at various time points after wait-listing. Two examples of such statistics are described here. Among all registrants, the fraction transplanted (FT) is reported in the CSRs (CSR Table 5) at several time points after listing (30 days, 1 year, 2 years, and 3 years) for each transplant program. The FT is a simple fraction of all wait-listed candidates who are transplanted at each time point after wait-listing, regardless of the program at which the transplant is performed. This statistic is most useful to the candidate who wants to know the prognosis for transplantation.

The time to transplant (TT) is the time by which 50% of all wait-listed candidates receive a transplant. The TT is intended to provide a measure of the rate of transplantation at a particular program, so candidates who transfer to another program's waiting list or who are removed for reasons of good health are dropped (censored) at that time using actuarial methods, for the TT outcome. Candidates who die or are removed from the list for reasons of poor health are not censored and are counted as never receiving a transplant in the TT calculations (as for the FT calculation). Note that the TT would never be reached for groups in which more than 50% of candidates die or are removed for poor health, since these candidates are counted as never receiving a transplant. The TT calculation is designed to appraise a new registrant's chances and timeliness of receiving an organ from any source, from either a deceased or living donor.

Organ allocation rules are only in force for candidates while they are on the waiting list. Rates of transplantation among candidates on the waiting list are useful for evaluating and comparing the impact of allocation rules on different groups of candidates. Such rates can be estimated by censoring candidates from the analysis when they are removed from the waiting list for any reason. The *Annual Report* shows percentiles of waiting time (WT) before transplant based on rates of transplantation among all candidates during the time they actually spend waiting on the list. For such calculations, candidates are censored at removal from the list for any reason, including death, poor health, good health, or living donor transplant. This calculation reflects transplantation rates for all candidates while they are on the waiting list, but does not offer a realistic appraisal of the chances for a transplant, since some candidates never receive a transplant even though their waiting time is censored at death or removal. The WT estimates the time that would result for a hypothetical population with transplant rates identical to those observed, if no candidates were removed from the waitlist for any reasons other than transplantation.

The methods described above are all useful for describing waiting times for transplantation. The choice of method depends on the specific question or the purpose of the question. Evaluation of waiting time for transplantation is most relevant when waiting time is the primary criterion for getting a transplant while on the waiting list. The question 'How long is the wait?' is of crucial importance for kidney transplant candidates, since the allocation of kidney transplants is driven largely by waiting time. In contrast, liver organ allocation is driven primarily by medical urgency. An important question for liver candidates is 'What fraction of medically urgent candidates receives a transplant before they die?' and a secondary question is 'How long is the wait for candidates with less medical urgency?'

Mortality and Graft Failure Analysis

Is the glass half empty or half full? Actuarial methods use estimates of death rates to compute the corresponding survival rates during successive intervals. These success rates are multiplied to yield cumulative success rates. Depending on the question to be answered, the actuarial results are reported as either the fraction that died or the fraction still surviving. Post-transplant graft and patient outcomes are often reported as cumulative survival rates. These are reported as a Kaplan–Meier survival curve when they are calculated based on data from a single cohort and are shown at various time points after transplant. Results from different cohorts are sometimes shown at various time points after transplant, as in Table 1, but since these results are from different groups

Table 1: Patient survival at 3 months, 1 year, 3 years, 5 years, and 10 years—by organ, with and without addition of SSDMF death dates

	Follow-up period				
	3 months	1 year	3 years	5 years	10 years
Kidney: deceased donor					
Survival with SSDMF	97.3%	94.0%	88.4%	79.9%	59.4%
Survival using OPTN data only	97.4%	94.3%	89.4%	81.9%	62.4%
Transplants	13 717	13 717	13 404	13 115	11 782
Standard error	0.10%	0.20%	0.30%	0.40%	0.50%
Kidney: living donor					
Survival with SSDMF	99.0%	97.7%	94.7%	89.7%	79.4%
Survival using OPTN data only	99.0%	97.8%	95.0%	90.6%	81.5%
Transplants	8980	8980	7556	6397	3884
Standard error	0.10%	0.20%	0.30%	0.40%	0.60%
Kidney–pancreas					
Survival with SSDMF	97.1%	95.1%	89.2%	82.6%	60.8%
Survival using OPTN data only	97.0%	95.1%	89.6%	83.4%	62.6%
Transplants	1821	1821	1803	1749	905
Standard error	0.40%	0.50%	0.70%	0.90%	1.60%
Liver: deceased donor					
Survival with SSDMF	91.0%	86.4%	79.5%	72.4%	59.4%
Survival using OPTN data only	91.7%	87.2%	80.6%	73.9%	59.0%
Transplants	7911	7911	7343	6826	4424
Standard error	0.30%	0.40%	0.50%	0.50%	0.70%
Heart					
Survival with SSDMF	89.5%	85.1%	78.6%	69.8%	50.0%
Survival using OPTN data only	89.5%	85.2%	78.5%	69.7%	49.7%
Transplants	4173	4173	4410	4525	4052
Standard error	0.50%	0.60%	0.60%	0.70%	0.80%
Lung					
Survival with SSDMF	88.2%	77.3%	59.3%	42.5%	22.7%
Survival using OPTN data only	88.2%	77.1%	59.2%	42.4%	20.9%
Transplants	1763	1763	1692	1598	576
Standard error	0.80%	1.00%	1.20%	1.20%	1.70%

Source: SRTR Data Analysis, August 1, 2002. Other organs omitted because of small number of transplants and resulting large standard errors. Transplants and standard error are for survival with SSDMF and are virtually the same for survival using OPTN only.

of patients, the outcomes are not consistent across the years. For example, the 5-year survival for the 10-year cohort is not reported and should not be assumed to have been the same as the reported 5-year survival for the 5-year cohort. Several issues related to definitions for graft failure and for dealing with incomplete mortality ascertainment are discussed below.

Graft failure

What should be counted as a transplant failure? In order to evaluate the lifetime of a transplanted organ, both retransplant and death of the recipient are counted as transplant failures even if the death was unrelated to transplantation. For kidney transplant recipients, return to dialysis is also counted as organ failure. However, in order to understand the mechanisms that lead to transplant failure, it is sometimes useful to count only failures of the transplanted organ itself, while not counting deaths from other causes. In order to study such mechanisms, the actuarial methods described previously can be used by censoring the follow-up of an organ when a recipient dies without organ failure.

Death rates and loss to follow-up

Wait-listed registrants are not generally tracked for mortality after they are removed from the waiting list, and post-transplant mortality ascertainment stops when a recipient is lost to follow-up. Due to the incomplete follow-up available in the data, the actuarial methods described above must censor patients when they are lost to follow-up. If the failure rates after loss to follow-up are the same as the failure rates among those still being followed, then the actuarial method estimates are appropriate even though some observations were censored. However, if recipients at high risk for eventual failure are disproportionately lost to follow-up before they fail, then the estimated failure rates will underestimate the overall failure rates. When many subjects are lost to follow-up, it is crucial to know whether subjects lost to follow-up were at high or low risk for subsequent unreported events.

In order to answer these questions, external data sources were linked with OPTN data to yield more complete ascertainment of mortality (as detailed in the 'Data Sources and Structure' article). Failure rates and survival curves

computed with and without extra ascertainment are reported later in this article; they indicate that the actuarial estimates for post-transplant survival rates based on OPTN data alone were very accurate at the aggregate level, although results for several individual transplant programs changed substantially. In the next annual report, the impact on mortality rates on the waiting list will be similarly evaluated.

OPTN death ascertainment alone was used for computing death rates on the waiting list. Such follow-up stops when a candidate is removed from the waiting list, because organ allocation is not affected by events after removal from the waiting list. The actuarial method described above is used to censor patients at the time they are removed from the waiting list. However, the resulting death outcomes are difficult to interpret because candidates are often removed from the list if their health deteriorates to such a point that they are no longer suitable for a transplant. Thus, low death rates on a waiting list are likely to reflect an effective screening process for removing patients when their health deteriorates but are unlikely to reflect the survival prognosis for all wait-listed candidates. The SRTR plans to evaluate the utility of extra death ascertainment on waiting list death rates for the next annual report. This may be especially important when comparing pretransplant mortality (which includes time after removal from the waiting list) to post-transplant mortality.

Adjusted Outcomes

Many of the analyses performed by the SRTR involve comparisons of outcomes. For example, the CSRs compare mortality and graft failure rates at each transplant center to national mortality and graft failure rates. In order to make the comparisons more meaningful, they are adjusted so that the outcomes at each facility are compared to the outcomes that would be expected for the patient mix at that facility. For example, the death rate might be high at a facility that commonly performs transplants on high-risk patients, but still lower than expected for such high-risk patients. The unadjusted higher mortality can be explained by the large number of high-risk patients, but gives no indication that the facility actually has better outcomes than expected for such patients. In contrast, the adjusted comparison correctly identifies the facility as having good outcomes.

One method of adjustment, called 'indirect adjustment', uses results from the various subgroups of a standard population, often the national population, to evaluate what would be expected among a given group of patients, such as those transplanted at each transplant center. The subgroups generally are defined by patient age and other patient characteristics. For each patient, we look up the average event rate for the subgroup to which the patient belongs. Based on that event rate and how long each patient is followed, the expected number of events for that patient is computed as the average event rate for the subgroup that the patient is in multiplied by the length of follow-up for that patient. For example, a patient in a subgroup with a national annual event rate of 0.10 (10%) who is followed for 1.1 years would have 0.11 events expected during follow-up. These expected fractional counts for all of the patients from each transplant center are added together to yield the total expected for the patients at each center.

The SRTR uses another closely related adjustment method, based on regression equations, to compare the outcomes that would have resulted had the comparison groups been otherwise equivalent. Regression equations can be used to compute expected outcomes given a patient's characteristics. The proportional hazards Cox regression model (4) is commonly used for adjusted analyses of time to event data. Similar to the Kaplan–Meier estimates described above, the Cox regression model can yield survival curve estimates for two or more groups of patients, adjusted to show the comparison that would result if the groups were equivalent with regard to particular factors, such as age and diagnosis.

The results of a Cox model can be used to compare groups or to show a trend among groups, based on the ratio of average event rates in each group, adjusted for other differences. For example, an age- and diagnosis-adjusted relative risk (RR) of 1.59 for post-transplant mortality rates for deceased compared to living kidney donor recipients would indicate that on average, the death rate is 59% higher for recipients of deceased kidney donor organs compared to living kidney donor organs, who are of the same age and diagnosis. For example, an RR of 1.59 based on a 10% death rate means that 15.9 instead of 10 deaths would be expected, if all else was equal. An RR equal to 1.0 indicates no difference in adjusted event rates between the comparison groups.

The CSRs include comparisons of observed and expected outcomes (mortality and graft failure), based on follow-up of a cohort (recipients transplanted between January 1999 and June 2001 for 1-year rates, and between 1997 and 1998 for 3-year rates) during the most recent time period for which data were available. Survival percentages at 1 month, 1 year, and 3 years are reported for each center from both unadjusted (Kaplan–Meier) and adjusted (Cox) survival models. The statistical comparison reported in the p-value compares observed events to expected counts from the Cox models rather than these survival percentages. For example, if 14 events are observed in a facility during that time, while 9.2 would be expected, given the characteristics of the patients followed, then the event rate for the group is 52% higher than expected and the p-value reported indicates whether this difference is statistically significant.

Adjusted analyses, which are intended to make 'all else equal' when comparing outcomes among different groups, are used extensively by SRTR in CSRs and reports to committees. The choice of what to adjust for, or what to make 'equal' in the comparison groups, is an important one that is under constant review by the SRTR. In order to make meaningful adjustments, relevant data must be available, complete, and accurate. The choice of factors used when adjusting center-specific outcomes for the mix of characteristics at each center involves OPTN committees, as well as SRTR analysts. The CSR documentation (available at http://www.ustransplant.org/programs-report.html) includes detailed descriptions of the adjustment models used in the CSRs.

Cohorts Chosen for Analyses

The cohort is the group of patients included in a particular analysis. Selection of the cohort depends primarily on how much time must be allowed for the follow-up to be sufficiently complete and how many patients must be in the cohort to produce statistically reliable results. In addition, the variability of follow-up and the lags in reporting and transferring the data affects the selection of the cohort. Several issues related to the choice of a relevant cohort for analysis are summarized below.

Allowing sufficient follow-up time

In the CSRs, we would like to be able to answer the question, 'What is the 1-year survival for patients transplanted at this center in the past year?' However, full 1-year outcomes are only known for those transplanted at least 1 year previously.

Lag time in reporting

Based on OPTN policy, centers are to submit follow-up reports within 60 days after the transplant anniversary, and some time is required for the data to flow through the OPTN to the SRTR. For the CSRs, the SRTR allows a 4-month reporting time lag.

The completeness of follow-up

There is considerable variation among transplant centers in compliance with OPTN data submission requirements. The actuarial method of measuring survival allows us to use cases with incomplete follow-up, but as the level of completion goes down, the potential for biased results goes up. For this reason, the SRTR computes a measure of completeness of follow-up for the CSRs.

The 'percent follow-up days reported by center' is the percent of expected follow-up days that are actually reported. It is a measure of the completeness of the data rather than a measure of compliance. This measure reports the percentage of days that are targeted for inclusion during the follow-up period relative to the number of days that were actually reported with follow-up forms. For patients who did not die before the end of the period, the targeted number of days of follow-up is the entire period, such as 365 days for 1-year follow-up. For patients who die before the end of the period, the number of targeted days of follow-up is the number of days until death. A center can have 100% of expected *forms* completed, but less than 100% of expected *days*, since some completed forms may not cover the entire follow-up period. For example, when a center files a follow-up report, it reports the patient's last known status and the date of that status. Thus a 1-year follow-up form may report the patient's status at the patient's last visit, which was at 10 months. In this case, only 305 out of 365 days are actually reported on a report that is submitted on time.

With the inclusion of Social Security Death Master File (SSDMF) data, described in the 'Data Sources and Structure' article, the number of days of follow-up covered by any source is equal to the targeted number of days for all patients regardless of death, and is always equal to 100%. However, because ascertainment of survival depends on multiple sources of mortality information, the completion of follow-up days reported by the center is still a valuable measure for evaluating the validity of the data. Therefore, even after the incorporation of the SSDMF into the CSR follow-up, the number of follow-up days is still reported in the CSR, and is based on center-reported data only.

The 'percent of expected follow-up forms that have been completed' is another measure of completeness that is reported to OPTN committees. When we are measuring 1-year follow-up, we expect a 1-year follow-up report or a follow-up reporting death before 1 year. If a 1-year follow-up form has not been completed, we accept a 2-year or later report in lieu of the 1-year report because the later report confirms that the patient was alive at 1 year. This measure reflects the transplant center's compliance with data reporting requirements.

Follow-up time

Post-transplant follow-up reports are completed at hospital discharge, at 6 months for abdominal organs, at 1 year, and annually thereafter. Variability in the follow-up also constrains the survival analysis. For instance, to analyze 2-year survival, we must allow time for the 2-year follow-up reports to be filed for the latest transplants in the cohort, but in order to analyze 2.5-year follow-up, the 3-year follow-up report is needed. The OPTN requires that a follow-up form be filed within 14 days of a post-transplant death, but unless the transplant center still sees the patient regularly, the center may not learn of a death until it prepares to complete the next annual follow-up report. The SRTR has established a protocol for

determining the end of follow-up to address these and related problems, as described below.

The post-transplant death rate tables and the patient survival tables make use of multiple data sources to determine the last known follow-up date to determine a censoring time. Since the SRTR now uses both the OPTN and SSDMF data to find deaths, we now expect to have nearly complete death ascertainment for anyone receiving a transplant. As detailed in the 'Data Sources and Structure' article, the ascertainment of mortality using both OPTN and SSDMF data is very good. During time periods when we would expect to learn of a death from both sources, if no death is reported then we assume that the patient is alive.

Using multiple sources of death has implications for censoring in mortality analyses. If only follow-up forms returned to the OPTN were being used, censoring would occur when the patient is reported as lost to follow-up, or at the last follow-up form filed. With multiple sources of death data, a patient must be followed after being lost to follow-up, in order to account for time and events that are covered by other sources of mortality data. Therefore, the patient is followed as long as we would expect reporting from both sources; constraints include the schedule of follow-ups, which prompts OPTN members for follow-up on transplant anniversaries, and lag in reporting to each source. The multiple-source follow-up or censoring date is calculated in two steps. First, a database cutoff date is set to allow lag in reporting before the current database snapshot date (August 1, 2002, for the 2002 *Annual Report* tables). This lag time, 3–4 months depending on the analysis, allows time for the reporting lags in data from both the OPTN and the Social Security Administration. The multiple-source censoring date is moved back even farther, to the transplant anniversary (6 months, 1 year, 2 years, etc.) immediately preceding this database cutoff date. It is through this anniversary, when OPTN members are prompted for follow-up forms, that we expect both sources to be complete.

Events and follow-up time reported after this anniversary are disregarded since they are likely a biased sample of outcomes. This is because events such as death may be reported 'off schedule' from the regular expected follow-up forms. Patients who are alive will have follow-up status reported only when forms are due at 6 months (for non-thoracic organs), 1 year, 2 years, etc., after transplant. When a patient dies, however, the center can report that the patient died on an off-schedule follow-up form, creating additional reporting on a (biased) sample of patients who have died. Simply following patients until the last known OPTN follow-up date will include extra time for patients who die and have the follow-up form turned in early, but will not include this extra time for patients who are alive. To eliminate this bias in reporting deaths, we follow patients only until we expect to learn about all patients, both living and deceased. Even when not using additional sources of death ascertainment, it is important to consider this multiple source censoring date in analyses for this reason. We censor at the date of last *expected* follow-up, or the transplant anniversary, for all patients. In some cases, this date falls before reports of deaths filed to the OPTN by member centers and means that certain deaths and follow-up time will be excluded from analyses, but these exclusions are made in the interest of obtaining an unbiased sample.

Statistical significance

In order to increase the accuracy of a reported statistic, more patients can be included in an analysis by including older cohorts of patients. While increasing the accuracy, the inclusion of older cohorts carries the risk of yielding results that no longer represent the current experience. These opposing objectives must be balanced when choosing the most recent cohort of patients for analysis. The SRTR uses both p-values and confidence intervals, described below, to help in the evaluation of the accuracy of reported differences and statistics.

When making comparisons of outcomes, differences can occur due to non-replicable fluctuations that are due to chance or random causes. It is important to distinguish differences in outcomes that would likely recur upon replication of the study from differences that arise due to chance observations for a particular study group. Three major tools are widely used to help assess the influence of chance on a reported comparison. The *p-value* is a statistic that measures how likely it is that an observed difference might have occurred due to such random causes. The p-value is a probability, and a p-value less than 0.05 (5%) indicates 'statistical significance', since the corresponding result is unlikely to have occurred by chance. The *confidence interval* gives a range in which we can be confident that the true (replicable) difference is likely to be. For example, if 11 deaths were observed in a cohort where 9.2 were expected for similar patients, the RR is 1.52, which represents 52% higher mortality than expected. However, this difference is not significant (p-value $>$ 0.05) and the 95% confidence interval indicates that the observed mortality could represent as high as 138% excess mortality or as low as 8% reduced mortality compared to the expected. A third method for assessing the impact of random variation, based on Bayes methods, is seldom used by the SRTR because it is more difficult to convey this approach to clinicians.

Both the p-value and the confidence interval are related to the accuracy of an estimated comparison. The p-value depends on both the size of the difference and the size of the sample. Both a larger difference and a larger sample size tend to make a result more significant. The clinical importance of a comparison depends largely on the size of the estimated difference.

The Impact of Extra Ascertainment on Post-transplant Survival Rates

The 'Data Sources and Structure' article describes the following sources that can be used for extra ascertainment of death dates: linking within the OPTN data, the SSDMF, the Centers for Medicare and Medicaid Services (CMS) data primarily about patients with end-stage renal disease (CMS ESRD), and the National Death Index (NDI). While that article quantifies the additional deaths found in different sources, here we explore the effect of including these sources on post-transplant survival rates. Even though these additional sources of death inform us about a substantial number of additional deaths, actual survival rates may not decrease, and, in fact, could increase, because commensurate time at risk is also added for patients who are not reported as having died. The adjustment to the follow-up time described in the previous section allows for the possibility of the survival rates staying the same or increasing even when deaths are added.

To review the discussion in 'Data Sources and Structure', most deaths among patients in the SRTR database are identified by multiple sources. The relative contributions of the sources are evaluated by measuring the additional deaths contributed by each source as they are added in the following order: OPTN, SSDMF, CMS ESRD, and NDI. The contributions are expressed as percentages of the total number of deaths identified by any of the sources. For organs other than kidneys and pancreata, 96% of deaths are found in the OPTN data, but only 75% of the deaths for kidneys and pancreata (K/P) are found in the OPTN data. However, for deaths in the first year after transplant, the OPTN data included 99% of the non-K/P deaths and 95% of the K/P deaths. The SSDMF contributes almost all of the remaining deaths, with the contribution rising to about 10% of deaths 5 years or more after the transplant for non-K/P organs and to about 30% for K/P. The CMS ESRD and NDI contribute only 0.8% of the deaths. Thus, SRTR has evaluated the available sources, those sources agree on most of the deaths, and the last sources consulted add few additional deaths. While the additional sources do not give us a definitively complete set of death dates, it seems reasonable to assume that these sources identify almost all of the deaths.

Adding deaths from additional sources requires a change in the treatment of follow-up time. When using only the OPTN death data to compute death rates, the follow-up time (or time at risk) is computed until the date of last follow-up. In the actuarial method, we censor each patient at the time of last follow-up. When adding the deaths identified by the SSDMF to deaths identified by the OPTN data, the follow-up time must also be adjusted. Otherwise, information is added only about persons who die, and death rates would be overstated. The SRTR methodology makes the assumption that with the SSDMF data, virtually all of the deaths are known, and the corollary of that assumption is that post-transplant patients are alive unless known to be dead. Based on this assumption, patients are not censored at the last reported OPTN follow-up date, thus extending the follow-up time to the end of the study period.

Table 1 shows the effect on survival rates of adding the SSDMF death dates and extending the follow-up time as described above. The survival rates are almost equal, with the SSDMF augmented death rates often being slightly lower. Although the comparison is not statistically valid, it is interesting to note that in many cases the difference is less than the standard error. These small differences arise even though the SSDMF contributes 30% of K/P deaths and 10% of the deaths for other organs after 5 years post-transplant. It indicates that at the national level, the added follow-up balances the added deaths, and suggests that the loss of patients to follow-up over time is random and does not bias the actuarial results which were censored at the time of loss to follow-up by the OPTN.

While Table 1 suggests that the 1-year survival results are not biased at the national level, the SSDMF and data sources are very useful for center-specific survival even at 1 year. Articles in *Transplant News and Issues* (5) and the *Milwaukee Journal Sentinel* (6) have publicly questioned whether centers with low completion rates may have center-specific survival figures that are substantially different when additional sources are considered. Specifically, they cited substantial numbers of programs with low follow-up completion rates in the CSRs released in July 2001, wondering if the statistics derived for these centers were reliable.

To address these concerns, the CSRs released by the SRTR in July 2002 included the SSDMF data and extended follow-up time in mortality computations. Without identifying specific programs, summary data published concurrently with these reports examined the effect of this change on center-specific mortality measures (*http://www.ustransplant.org/accuracy.cgi*). Table 2, abstracted from this report, shows the frequency, direction, and magnitude of changes in 1-year survival for kidney and liver programs, which are representative of other organs. Figures are grouped separately by the level of completion of follow-up forms at the center.

Table 2 shows that the difference in center-specific mortality rates resulting from including extra ascertainment is very small, as it was in the national data presented in Table 1. In fact, average survival rates improve slightly for many facilities, and among kidney programs, twice as many exhibit *increases* in calculated survival as exhibit decreases. This suggests that, for many centers, the patients that are lost to follow-up are at least as healthy as those who are followed.

Table 2: Difference in 1-year survival: Center-Reported (CR[1]) vs. CR + SSDMF for kidney and liver transplant programs—transplants from 1/1/99 to 6/30/01

	Facility percent follow-up time reported[2]						
	All	100%	90–99%	80–89%	70–79%	50–69%	<50%
Kidney programs							
Number of programs	241	12	156	39	12	11	11
Average number of transplants	113	10	117	138	143	104	56
One-year survival rate average (weighted for n of transplants):							
CR only	95.5	94.3	95.9	94.9	94.7	94.2	94.4
CR + SSDMF	95.3	94.3	95.8	94.3	94.8	94.6	92.2
Center change	−0.23	0.00	−0.15	−0.66	0.16	0.44	−1.45
Number of programs where survival:							
Increases	117	0	83	20	6	4	4
Does not change	67	12	40	7	2	2	4
Decreases	57	0	33	12	4	5	3
Liver programs							
Number of programs	108	8	70	20	6	1	2
Average number of transplants	89	3	86	120	103	307	109
One-year survival rate average (weighted for n of transplants):							
CR only	86.9	81.0	87.5	86.1	85.4	86.7	83.8
CR + SSDMF	86.3	81.0	87.0	85.2	83.6	85.5	88.3
Center change	−0.65	0.00	−0.55	−1.06	−2.00	−1.48	5.95
Number of programs where survival:							
Increases	38	0	30	5	2	0	1
Does not change	21	8	10	2	0	0	0
Decreases	49	0	30	13	4	1	1

Source: SRTR CSRs, July, 2002.
[1]Center reported to the OPTN.
[2]Of days expected to be included in the SSDMF follow-up period, the percentage of days covered by follow-up reporting for these transplants. A low percent may indicate a non-representative sample from this facility for follow-up. This measures the possibility that events such as failure have occurred without being reported, and it is not a measure of compliance.

Prior to using the SSDMF for center-specific survival rates, the SRTR published 'completion percentages' intended to measure the probability that important patient events such as death might have gone unreported. These completion percentages provided a caveat for underreporting centers: that the survival rates based on low reporting might not be representative of all of the patients treated at the facility, and therefore might be misleading.

When the SSDMF was included, a small number of centers, with a range from poor to good follow-up, have mortality rates that appear to decline significantly with extra ascertainment. On average, programs with more missing data (more patients lost to follow up or unreturned transplant follow-up forms) tended to have poorer patient survival than programs with more complete data. Though it appears more likely for a facility with poor follow-up to have systematically missed reporting on deaths, it is also possible for facilities with apparently good follow-up to miss these types of patients as well.

This suggests that extra death ascertainment is a useful tool both for obtaining accurate data at the center level and for improving public confidence in the figures. Patient follow-up by transplant facilities continues to be extremely important, despite the availability of other data sources. Follow-up forms provide other valuable information about transplant recipients and the OPTN data do capture a significant number of post-transplant deaths that are not captured by the SSDMF.

Conclusion

The SRTR database is a rich source of information for the transplant community. Based largely on the data submitted by transplant programs and by OPOs to the OPTN and augmented with other data sources, it serves as the basis for a wide variety of analyses. Simple tabulations based on the SRTR data provide a description of the numbers of transplant donors, organs, candidates, and recipients that is important for assuring efficient and timely organ allocation efforts. Analyses of outcomes—including organ procurement rates, transplantation rates, graft failure rates, and mortality rates—require a thoughtful choice of analytical methods, particularly regarding

censoring. Such analyses are of value to the entire community, including patients, clinicians, and policy analysts.

References

1. Dickinson DM, Ellison MD, Webb RL. Data sources and structure. Am J Transplant 2003; 3(Suppl. 4): 13–28.
2. Allison PD, Survival analysis using the SAS system. A practical guide. Cary, NC. SAS Institute Inc., 1995.
3. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1972; 53: 457–481.
4. Cox DR. Regression models and life tables (with discussion). J R Stat Soc 1972; 34: 197–220.
5. Cooper L. Survival Data. Do the numbers really mean anything? Transplant News Issues 2001; 2(2): s9–s11, s13.
6. Marchione M. Transplant rate reports don't tell whole story. Milwaukee J Sentinel 2001; July 27: G1.