CHAPTER IX
Analytical Approaches for Transplant Research

INTRODUCTION

This chapter describes the analytical approaches used in various Scientific Registry of Transplant Recipients (SRTR) analyses, including the *2003 OPTN/SRTR Annual Report*, the Center-Specific Reports (CSRs) published at www.ustransplant.org, and analyses for committees of the Organ Procurement and Transplantation Network (OPTN) and Advisory Committee on Organ Transplantation (ACOT). Different questions require different analytical methods, so a variety of methods are described here. Substantial attention is given to ensure that the analysis methods are appropriate to the quality, timeliness, and completeness of the data available.

SRTR analyses are widely used and quoted. Ensuring timely, accurate, and relevant analysis results is an ongoing challenge. Patients, physicians, policy makers, and administrators all use the results of SRTR analyses. The SRTR attempts to address the individual needs of each type of user, as well as to show appropriate new perspectives on the issues relevant to these diverse audiences.

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 augmented the OPTN data by linking them with other data sources. See "Transplant Data: Sources, Collection, and Caveats," Chapter II of this report, for details on these and other data sources.

The use of appropriate analysis methods is especially important for transplant data because of the complex longitudinal nature of the data and the wide variation in medical practices, organs, candidates, and recipients present in the data. All of the methods described here require careful linking and accounting of the sequence of events for each individual organ and patient. Many of the SRTR methods involve the analysis of time to event data. Standard statistical methods are used to aggregate data over time, including 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.

DEFINING SALIENT FEATURES OF THE TRANSPLANT PROCESS

When summarizing the transplantation process, many issues arise that involve deciding what to count and how to count them. The issues of availability of data are discussed elsewhere in this report. Additionally, technical notes in the *2003 OPTN/SRTR Annual Report* give 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 posttransplant mortality (1). 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 transplantation, 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.

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. **Candidate** counts 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**. In the *Annual Report*, most waiting list tables reflect only one record per person, even in the case of multiple listings.

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 future prognosis for current groups of patients, while complete outcomes data are available only for previous groups of patients who have been followed for one or more years. Since improvements in medical practices and changes in organ allocation policy are occurring rapidly, it is useful to use the most recent data available, but interest in long-term outcomes necessitates using less recent 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 five years ago. Even among those transplanted more than five 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 less recent 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. A variety of statistical methods have been designed to address these goals, including actuarial methods, the Kaplan-Meier estimator, Cox regression and Poisson regression. Many of these were described in the 2002 Report on the State of Transplantation (2).

Transplant Waiting Times

For each type of organ failure there is a shortage of organs compared with the number of candidates who could benefit from transplantation. A variety of methods of organ allocation, each appropriate to the treatment options available for that type of organ, are being developed to address this shortage. Kidney transplants, which represented slightly more than 46% of all deceased donor solid organ transplants in 2002, are allocated primarily on the basis of waiting time. Liver transplants are allocated primarily on the basis of medical condition, as indicated by chronic versus acute organ failure and by the MELD score among chronic liver failure patients. Allocation of hearts is based on medical condition and status. A recent proposal has been made to change the lung allocation system from one based upon waiting time to one based upon the net benefit of transplantation, or the extra years of life gained by transplant during some limited follow-up period (currently the proposal is for the first year following transplantation). Allocation of organs according to the net benefit balances the value of avoiding imminent death due to organ failure while also avoiding short-term failed transplants. Evaluation of the expected net years of life gained by transplant gives not only a criterion for prioritizing candidates for organ allocation but also provides useful information to candidates about the relative risks and benefits of transplantation.

Liver Transplantation. In the face of these varied allocation systems, the simple question, "How long is the wait for a transplant?" is no longer so simple to answer. For organs that are allocated on the basis of medical condition or net benefit, such as for liver transplantation, the medical condition of candidates is continually updated and candidates are reprioritized according to their current condition. In some regions of the country, candidates with very low risk of death might never be allocated an organ unless and until the time that their condition worsens.

Thus, instead of considering statistics about waiting times, more relevant statistics for liver candidates may be found in the answers to the following questions:

1. Among Status 1 candidates (acute liver failure), what fraction get a transplant, what fraction die, and what fraction recover?

2. Among chronic failure candidates, what is the rate of transplantation per month during the time that their MELD score has a particular value? What is the competing risk that the patient dies during the same time?

Such statistics will allow the evaluation and comparison of access to liver transplantation for the purposes of both policy development and patient counseling. Similarly, for each organ that is allocated on the basis of medical condition, it will be useful to report measures of transplantation rates separately for different categories of medical condition. Analogous methods can be used for candidates for other organ transplants, such as heart, if allocation rules are changed from a waiting-time basis to include death rates on the waiting list as a criterion.

The use of MELD to allocate livers among chronic liver failure candidates began in February 2002, along with rules for exceptions for candidates with other specific diseases, such as liver cancer. The SRTR is developing relevant summary statistics and tables to summarize rates of liver transplantation according to status and MELD, and expects that the data will have stabilized sufficiently following the change in allocation rules to allow such reporting in the next cycle of Center-Specific Reports.

Kidney Transplantation. For kidney transplants, which are still allocated primarily according to waiting time, the SRTR computes and reports several types of statistics, which answer the following questions:

1. Among all registrants, what fraction received a transplant within one year?

2. Among all registrants, how long did it take before 50% of them had received a transplant?

3. What is the rate of transplantation among actively listed candidates?

4. Among all recipients, how long did they wait on average?

Answers to questions 1 and 2 are the most relevant to registrants because they give the prognosis for transplantation accounting for all potential outcomes, including both inactive time and death without transplant. Question 3 is relevant for candidates who are actively listed and for evaluation of the allocation process, which involves only actively listed candidates. Question 4 is the least relevant to the transplant community but is the easiest to answer based upon recent data. Questions 1, 2, and 4 can be answered directly by evaluating outcomes in different groups of candidates, while question 3 involves a tabulation of person-years in a calculation of rates.

For the purposes of ranking different regions or groups of candidates, all of the questions above typically yield similar results. The median time to transplant among recipients can be easily computed by counting recipients during a recent interval of time. This statistic is useful for comparing waiting times among regions or among transplant programs. However, the average waiting time among recipients is not useful for patient counseling, since it gives an overly optimistic perspective compared with the prognosis among registrants by not accounting for the possibility that the patient might never receive an organ.

The outcomes for all wait-listed candidates are summarized 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 Table 5 of the CSRs at several time points after listing (30 days, one year, two years, and three years) for each transplant program (www.ustransplant.org). The FT is a simple fraction of all wait-listed candidates who have received a transplant, regardless of the program at which the transplant is performed. The FT summarizes the time to transplantation at any program among all registrants at a transplant program.

The time to transplant (TTS) is the time since listing by which 50% (or another stated fraction) of all wait-listed candidates receive a transplant *[Table 1.5 and Tables X.2 in the organ-specific sections]*. The TTS measures 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 TTS 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 both the TTS and the FT calculations. Note that the TTS 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 TTS calculation summarizes the time to transplantation at a transplant program or within a group, taking into account the possibility of not receiving an organ.

Different statistics are useful for the evaluation of organ allocation policies for deceased donor organs. For example, rates of transplantation among candidates on the waiting list are useful for evaluating and comparing the impact of allocation policies on different groups of candidates. Such policies only affect candidates while they are active on the waiting list. The *Annual Report* shows percentiles of waiting time (WT) based on rates of deceased donor transplantation among all candidates during the time from listing until removal from the list, excluding inactive time *[Table 1.6]*. For such calculations, time while inactive is excluded and candidates are censored at removal from the list for any reason, including death, poor health, good health, or living donor transplant. The WT estimates the time that would result for a hypothetical population with transplant rates identical to those observed, if all candidates remained active on the waiting list until transplant.

The various methods described above are all useful for describing waiting times for transplantation and each is appropriate for specific purposes. The choice of method depends on the specific question or the purpose of the question.

MORTALITY AND GRAFT FAILURE ANALYSIS

Actuarial methods use estimates of death rates to compute the corresponding survival rates during successive time intervals. The success rates for successive time intervals are multiplied to yield the cumulative success rate at the end of the final interval. Depending on the question to be answered, these actuarial results are reported as either the fraction that died, the fraction still surviving, or the expected years of life through the end of the last interval.

Unadjusted (crude) posttransplant graft and patient survival outcomes are reported as cumulative "success" rates. These are calculated by Kaplan-Meier survival curves when the analyses are based on data from a single cohort and they are shown at various time points after transplant. Results from different cohorts are sometimes shown at various time points after transplant, as in the Adjusted and Unadjusted Graft and Patient Survival tables in the Annual Report *[Tables X.8 through X.11]*. However, since these results are from different groups of patients, the outcomes are not consistent across the years. For example, the five-year survival for the 10-year cohort is not reported and should not be assumed to be the same as the five-year survival that is reported for the five-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 reported and 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. Generally, wait-listed registrants are not tracked for mortality after they are removed from the waiting list. That is, posttransplant mortality ascertainment stops when a recipient is lost to follow-up. Because of 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 important to know whether subjects lost to follow-up were at high or low risk for subsequent unreported events.

OPTN death ascertainment alone was used for computing death rates on the waiting list, as reported in each organ-specific section in the *2003 Annual Report*. 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 death rate per patient-year-at-risk method includes events and time only while on the waiting list and is not affected by events after removal. 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. (See Chapter II for a discussion of post-removal deaths.) 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.

For the purposes of computing expected lifetimes on the waiting list, the SRTR uses information on deaths from other data sources, such as the Social Security Death Master File. This is especially important when comparing pretransplant mortality (which includes time after removal from the waiting list) to posttransplant 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 with 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 with 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 an unadjusted

mortality statistic would give no indication that the facility actually has better outcomes than expected for such patients. In contrast, the adjusted comparison would correctly identify the facility as having good outcomes.

The SRTR adjustment method is called "indirect adjustment," which uses results from the various subgroups of a standard population, often the national population, to evaluate what would be expected for each patient among those transplanted at a particular transplant center. The subgroups generally are defined by patient age and other patient characteristics, such as disease, that may affect survival. We look up the average event rate for the subgroup to which each patient belongs. Based on that event rate and how long each patient is followed, the expected number of events for that patient is computed. The expected number is the product of 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 events for the patients at each center. The standardized ratio of the observed to the expected counts is reported in the Center-Specific Reports.

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 (3) 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 event rates in each group, adjusted for other differences. For example, an age- and diagnosis-adjusted relative risk (RR) of 1.59 for posttransplant mortality rates for deceased compared with living kidney donor recipients would indicate that the death rate is 59% higher for recipients of deceased kidney donor organs compared with recipients of living kidney donor organs of the same age and diagnosis. For example, an RR of 1.59 based on a 10% death rate would mean that 15.9 instead of 10 deaths would be expected, if all else were equal. An RR equal to 1.0 would indicate 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 of recipients transplanted between 1 and 3.5 years prior to report release for one-month and one-year rates, and between 3.5 and 5.5 years prior for three-year rates. These cohorts are chosen to reflect the most recent time period for which data were available. Survival percentages at one month, one year, and three 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 with 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 the probability that the difference is due to chance (in the CSRs, differences with $P<0.05$ are labeled as 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 and will differ according to the specific purpose of the analysis. 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 and SRTR analysts. The CSR documentation (available at www.ustransplant.org/programs-report.html) includes detailed descriptions of the adjustment models used in the CSRs.

COHORTS CHOSEN FOR ANALYSES

A cohort is a group of patients followed over time. Selection of the length of follow-up time for the cohort depends primarily on how much time must be allowed for the follow-up to be sufficiently complete, whereas the number of patients in the cohort depends on statistical considerations, such as event rates and power. In addition, the variability of follow-up and the lags in reporting and transferring the data affect 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 one-year survival for patients transplanted at this center in the past year?" However, full one-year outcomes are only known for those transplanted at least one year previously. Based on OPTN policy, centers are to submit follow-up reports within 60 days after the transplant anniversary, with some time allowed for late reporting and for the data to flow through the OPTN to the SRTR and for additional data sources to be incorporated. For the CSRs, the SRTR allows a four-month reporting time lag. Issues in choice of cohorts and follow-up patterns are discussed in detail in Chapter II.

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 decreases, the potential for biased results increases. For this reason, the SRTR computes a measure of completeness of follow-up for the CSRs.

The "percent follow-up days reported by center" reports the percentage of days that were actually reported with follow-up forms relative to the number of days that are targeted for inclusion during the follow-up period. It is a measure of the completeness of the data rather than a measure of compliance. 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 one-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 one-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, 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 one-year follow-up, we expect a one-year follow-up report or a follow-up reporting death before one year. If a one-year follow-up form has not been completed, we accept a two-year or later report in lieu of the one-year report because the later report confirms that the patient was alive at one year. This measure reflects the transplant center's compliance with data reporting requirements.

Follow-up Time. Posttransplant follow-up reports are completed at six months (for abdominal organs), at one year, and annually thereafter. Variability in follow-up also affects the reliability of the survival analysis. For instance, to analyze two-year survival, we must allow time for the two-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 three-year follow-up report is needed. The OPTN requires that a follow-up form be filed within 14 days of a posttransplant 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. Analysis presented in Chapter II, shows that this is often the case. The SRTR has established a protocol for determining the end of follow-up to address these and related problems, as described below.

The posttransplant 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 uses both the OPTN and SSDMF data to find deaths, we expect to have nearly complete death ascertainment for anyone receiving a transplant. During 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, 2003, for the 2003 *Annual Report* tables). This lag time, three to seven 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 (six months, one year, two 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 because 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 six months (for non-thoracic organs), one year, two 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 precision, 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 precision of reported differences and statistics.

When making comparisons of outcomes, differences can occur due to non-replicable fluctuations resulting from 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. Two 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 or greater difference might have occurred by chance alone when no difference actually exists. The *P*-value is a probability, and a *P*-value less than 0.05 (5%) is often used to establish "statistical significance." 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 with the expected.

Both the *P*-value and the confidence interval provide information about the accuracy of a comparison. The *P*-value depends on both the effect size and the sample size. Both a larger effect size and a larger sample size tend to make the *P*-value smaller. The clinical importance of a comparison depends largely on the size of the estimated difference.

COMPARING TREATMENT ALTERNATIVES

Many SRTR analyses are directed at the comparison of outcomes for alternative treatments. Such comparisons are complicated by the fact that patients often cross over from one type of treatment to another, so outcomes for pure treatment groups are not observed. Two types of analyses are described here, one based upon data for two observed treatment alternatives and another based upon the projected outcomes that would result for pure treatment groups using a competing risks methodology. Both of these methodologies have recognized limitations inherent in the nonexperimental nature of the observational data available for analysis.

The death rates with and without transplant are used to calculate the net benefit of transplant, in terms of extra years of life because of transplantation. The posttransplant death rates can be estimated based on the observed outcomes among recipients. However, the death rates without transplant are more difficult to estimate because the lifetime without transplant is not observed for those candidates whose lifetime without transplant is interrupted by receiving a transplant. The death rates among those without transplant can be estimated by actuarial competing risks analyses, which estimate death rates on the waiting list censored at the time of transplant (4). However, this methodology has recognized limitations for organs allocated on the basis

of waiting list mortality, such as livers. The death rates observed among those on the waiting list do not represent the death rates that would result if the transplant option were removed, as impending deaths are selectively removed from the waiting list follow-up group because of the priority given to high-risk patients. This leads to overrepresentation of low-risk patients among those remaining on the waiting list and underestimation of the death rates that would result if the transplant option were removed.

Two approaches have been use to compare outcomes after an exceptional transplant relative to a standard transplant. Examples of such analyses include comparison of outcomes for split versus whole livers, for expanded criteria donor (ECD) versus standard donor kidneys, for living donor versus deceased donor, and for dual- versus single-kidney transplants.

An as-treated analysis estimates death rates after receipt of an exceptional organ and after receipt of a standard organ. In addition, death rates without transplant can be estimated using competing risks analysis of waiting list death rates censored at the time of transplant with either type of organ (4). All patients start in the waiting list group and can cross over to either the standard or the exceptional transplant groups. These three sets of death rates can then be compared with respect to relative risk, cumulative survival fractions, or expected lifetimes.

There are two major limitations to this as-treated methodology. First, death rates on the waiting list are likely to be underestimated, since high-risk patients are more likely to receive a transplant and thus be removed from the wait-list group than are low-risk patients. Second, the three treatment groups do not represent the actual choice facing a potential recipient of an exceptional organ, who has the option of accepting the exceptional organ or continuing to wait for a standard organ and the benefits of transplantation with a standard organ.

An alternative analysis addresses these two limitations by comparing two treatment groups: exceptional therapy, which is transplantation with an exceptional organ; and standard therapy, which involves continuing to wait for and possibly receive a standard organ. All patients start on the waiting list in the standard therapy group and can cross over to the exceptional therapy group.

The analysis of ECD kidneys demonstrates the differences in interpretation of these two methods. The as-treated analysis shows that death rates following ECD transplantation are lower than death rates among those remaining on the waiting list but higher than death rates after transplantation with a standard organ. This suggests that receiving an ECD organ is superior to remaining on the waiting list but inferior to getting a standard donor organ. However, at the time of offer of an expanded donor organ, the actual choice being made is not a three-way choice, but only a two-way choice between an ECD organ now versus continued time on the waiting list with a possible standard donor organ in the future. The two-group comparison shows very little difference in mortality between the expanded therapy option and the standard therapy option.

Organ Allocation

The comparison of treatment outcomes with and without transplant is an especially important consideration for organ allocation. Death rates of candidates on the waiting list with liver failure differ dramatically (over 100-fold) between Status 1 (acute failure) candidates and chronic liver failure candidates with MELD scores less than 10. With such disparities in death rates, it is of great value to those at high risk of death to receive high priority so that they will get a transplant before they die. At the same time, it may be useful to identify candidates whose prognosis after transplantation is poor and give them lower priority for transplantation. One approach toward balancing these two goals, of avoiding pretransplant mortality and avoiding early posttransplant mortality, is to rank candidates with regard to the difference in projected lifetimes with and without transplant, or the net extra years of life with transplant. This gives a utilitarian measure of the benefit of transplantation for each candidate in terms of extra years of life and could be considered along with other allocation goals, such as improving quality of life and assuring equity of access to transplantation.

Statistical models for projected lifetimes with and without transplant are used to calculate the net benefit, based on the characteristics of each candidate and donor. Pre- and posttransplant survival models are under constant development for all organs, with special attention currently being devoted to models for lung, liver, and heart since major changes in the allocation systems for these organs have been recently made or proposed.

REFERENCES

1. HHS/HRSA/OSP/DOT, UNOS, URREA. *2003 OPTN/SRTR Annual Report*: " Technical Notes and Analytic Methods."

2. Wolfe RA, Webb RL, Dickinson DM, et al. "SRTR Report on the State of Transplantation: Analytical Approaches for Transplant Research." *American Journal of Transplantation* 3(Suppl 4): 103-113, 2003.

3. Cox DR. Regression Models and life tables (with discussion). *J R Stat Soc* 1972, 34: 197-220.

4. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data, Second Edition*. Hoboken, NJ: Wiley, 2002.

CONTRIBUTORS

The following individuals prepared this chapter: Robert A. Wolfe, PhD[1]; Douglas E. Schaubel, PhD[1]; Randall L. Webb, BA[2]; David M. Dickinson, MS[2]; Valarie B. Ashby, MA[1]; Dawn M. Dykstra, BA[2]; Tempie E. Hulbert-Shearon, MS[1]; Keith P. McCullough, MS[2]. [1]Scientific Registry of Transplant Recipients / University of Michigan; [2]Scientific Registry of Transplant Recipients / University Renal Research and Education Association.