

Analytical approaches for transplant research, 2004

Douglas E. Schaubel^{a,*}, Dawn M. Dykstra^b,
Susan Murray^c, Valarie B. Ashby^a, Keith P.
McCullough^b, David M. Dickinson^b, Tempie E.
Hulbert-Shearon^a, Randall L. Webb^a
and Robert A. Wolfe^a

^aScientific Registry of Transplant Recipients, University of Michigan, Ann Arbor, MI

^bScientific Registry of Transplant Recipients, University Renal Research and Education Association, Ann Arbor, MI

*Corresponding author: Douglas E. Schaubel, deschau@umich.edu

This article provides detailed explanations of the methods frequently employed in outcomes analyses performed by the Scientific Registry of Transplant Recipients (SRTR). All aspects of the analytical process are discussed, including cohort selection, post-transplant follow-up analysis, outcome definition, ascertainment of events, censoring, and adjustments. The methods employed for descriptive analyses are described, such as unadjusted mortality rates and survival probabilities, and the estimation of covariant effects through regression modeling. A section on transplant waiting time focuses on the kidney and liver waiting lists, pointing out the different considerations each list requires and the larger questions that such analyses raise. Additionally, this article describes specialized modeling strategies recently designed by the SRTR and aimed at specific organ allocation issues. The article concludes with a description of simulated allocation modeling (SAM), which has been developed by the SRTR for three organ systems: liver, thoracic organs, and kidney-pancreas. SAMs are particularly useful for comparing outcomes for proposed national allo-

Note on sources: The articles in this report are based on the reference tables in the 2004 OPTN/SRTR Annual Report, which are not included in this publication. Tables from the Annual Report that serve as the basis for this article include the following: Tables 1.3, 1.5, 1.7, 5.2, 5.8, 5.9, 5.10, 5.11, 6.2, 6.8, 6.9, 6.10, 6.11, 7.2, 7.8, 7.7, 7.8, 7.9, 7.10, 7.11, 8.2, 8.8, 8.9, 8.10, 8.11, 9.8, 9.9, 9.10, 9.11, 10.2, 10.8, 10.9, 10.10, 10.11, 11.8, 11.9, 11.10, 11.11, 12.2, 12.8, 12.9, 12.10, 12.11, 13.2, 13.8, 13.9, 13.10, and 13.11. All of these tables may be found online at <http://www.ustransplant.org>.

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

cation policies. The use of SAMs has already helped in the development and implementation of a new policy for liver candidates with high MELD scores to be offered organs regionally before the organs are offered to candidates with low MELD scores locally.

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

Introduction

This article reviews many of the analytical approaches frequently used by the Scientific Registry of Transplant Recipients (SRTR), including those used in the 2004 OPTN/SRTR Annual Report, the Center-Specific Reports (CSRs) published at www.ustransplant.org, and analyses pertaining to various data requests from Organ Procurement and Transplantation Network (OPTN) committees and the Secretary's Advisory Committee on Organ Transplantation (ACOT). The SRTR research team both develops and uses existing analysis methods that are appropriate to the quality, timeliness and completeness of the data available. The chosen statistical method for any particular analysis depends strongly on the nature of the research question.

Data collected by transplant centers and organ procurement organizations (OPOs) and submitted to the OPTN are primarily designed to facilitate the efficient allocation of organs to transplant registrants 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 expanded the spectrum of addressable research questions on transplant outcomes by linking data from the OPTN to several other data sources, as described in 'Transplant data: sources, collection, and research considerations, 2004', an accompanying article in this report (1).

The nature of each SRTR analysis is structured to address the needs and interests of its intended audience. Determining the most appropriate analytic method is often challenging due to the complex nature of organ failure data. SRTR analyses often involve time-to-event data, which are inherently incomplete, since the event of interest (e.g. transplantation, death, graft failure) is not observed for all patients under study. Moreover, the characteristics of a patient may change over time; for example, a wait-listed liver

candidate's model for end-stage liver disease (MELD) score may increase while the patient is awaiting transplantation. Each method described in this report requires careful consideration of the sequence of events for each individual organ and patient.

Results from SRTR analyses are widely used and quoted and have great potential to influence decision-making and practice patterns (2–4). SRTR analyses are used to various extents in many phases of the organ transplant process by policymakers, administrators, physicians and patients. Results often must be interpreted with care due to the wide variation in medical practices, organs, registrants and recipients being studied.

The types of analyses conducted by the SRTR can be broadly classified as either descriptive or comparative. Descriptive analyses often involve the computation of unadjusted rates (i.e. events over patient time at risk) and Kaplan–Meier survival curves. Comparative analyses focus on the relative importance of various factors influencing survival time (i.e. time to event), rather than the survival times themselves.

Statistical methods overview

This section provides an overview of the methods of analyses frequently conducted by the SRTR, as well as the issues underlying those methods. The focus is on time-to-event analysis, which is highly pertinent to the analysis of organ failure and transplantation. The time-to-event analysis of waiting list and transplant outcomes must appropriately combine data from different cohorts of patients who have been followed for different lengths of time. A variety of statistical methods have been designed to address these features, including actuarial methods, the Kaplan–Meier estimator and Cox regression (5,6). Many of these were described in the 2003 *SRTR Report on the State of Transplantation* (7).

Cohort selection

A cohort is a group of patients followed over time, either prospectively or retrospectively. In selecting a cohort, one must consider both the size of the group to be analyzed and its ability to reflect the research population of interest. The number of observed events among patients in the cohort strongly affects the power to detect differences among subgroups and the precision of estimates; hence, the size and event rate of the cohort must be sufficient to make reliable inferences about the research question. Additional features that bear upon the size of the cohort to be analyzed include the variability and distinctiveness of subgroup event rates to be analyzed, the likely delays in data reporting and the variability of follow-up.

One may often reduce the size of the cohort necessary for analysis by increasing the length of follow-up and, hence, the number of observed outcomes from which to base

inferences. However, longer follow-up times come from patient information that is more dated. Since investigators often want to predict the future prognosis of current patients, and since improvements in medical practices and changes in organ allocation policy occur rapidly, it is desirable to use the most recent data available that are relevant to the research question. Time-to-event analyses making long-term predictions (e.g. wait-listing to transplant, transplant to death, transplant to graft failure) must by their nature use less recent data. For example, estimation of transplant failure rates during the fifth post-transplant year requires a cohort selection that includes patients who received a transplant at least 5 years ago. In cases where less-recent cohorts are included in making predictions for short-term outcome studies, one must carefully consider the trade-off between improving the precision and retaining the relevance of an analysis.

Follow-up period

In addition to the cohort selection issues described above, the choice of follow-up period must allow time needed for the outcomes to be reported and recorded. Ideally, we would like to estimate center-specific, 1-year survival probability, based on patients transplanted during the most recent year. However, 1-year follow-up can only be observed among patients transplanted at least 1 year previously. Variability in recording delays unaccounted for in the selection of the follow-up period would also affect the reliability of the survival analysis. Based on OPTN policy, centers are to submit follow-up reports within 60 days after the 1-year transplant anniversary. Some time must also be allowed for late reporting, for data to flow from the OPTN to the SRTR, and for supplemental data sources to be incorporated. To accommodate these anticipated delays, SRTR time-to-event analyses explicitly allow for reporting time lags of various lengths. For example, for the CSRs, a 4-month reporting time lag after each transplant anniversary is incorporated, based on observed patterns of data submission discussed in the accompanying article on data sources in this report (1).

Post-transplant follow-up reports are completed annually by each center; for abdominal organs, the first report is due 6 months following transplantation and annually thereafter. The OPTN requires that a follow-up form be filed within 14 days of a post-transplant death. However, unless the transplant center still sees the patient regularly, the center may not learn of a death until it prepares to complete its next annual follow-up report. To analyze 2-year survival, one must allow time for the 2-year follow-up reports to be filed; a 3-year follow-up report is needed to analyze the 2.5-year follow-up, since there is no report scheduled between 2 and 3 years following transplantation. The established SRTR protocol for choosing a follow-up period allows extra time for events to be recorded and available data sources to be merged appropriately, yet censors follow-up at the most recently reported transplant anniversary, beyond which only adverse events may be reported.

For instance, the SRTR's post-transplant death rate tables and patient survival tables use OPTN, Social Security Death Master File (SSDMF), and, additionally for kidney recipients, Centers for Medicare and Medicaid Services (CMS) end-stage renal disease outcome data sources. Using these three sources of mortality records, we expect to have nearly complete death ascertainment for anyone who received a transplant. Patients are assumed to be alive if none of these data sources reports a death during periods when we would expect to learn of a death from all sources. The accompanying article on data sources in this report provides more detail on multiple sources of follow-up and cohort choice (1).

Since the length of available follow-up and the nature of reporting delays differ between data sources, a sensible strategy for selecting an appropriate censoring time for analysis is required. In addition to reporting delays, censoring time is constrained by the follow-up schedule, which prompts OPTN members to report patient and graft status on transplant anniversaries. The multiple-source follow-up or censoring date is calculated in two steps. First, a database cutoff date is set to allow for delay in reporting before the current database snapshot date. This lag time—3–7 months, depending on the analysis—allows time for the reporting delays for the OPTN, CMS and 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. All sources of outcomes data are considered complete through this date of last expected follow-up.

Events and follow-up time reported after this date are disregarded. While the SSDMF and CMS are continuously updated with new death events, deaths are inconsistently reported to the OPTN: some reports occur near the time of death, others wait for the next transplant anniversary. Because we can no longer assume, after this anniversary, that the OPTN follow-up forms are an unbiased and complete source of mortality, it is likely that including these deaths and corresponding time at risk for all patients would result in a biased sample of outcomes. Although reducing precision, since fewer events are counted, these exclusions are necessary to correctly address the research question of interest. The accompanying article on data sources in this report presents a more detailed discussion of this topic (1).

Completeness of follow-up forms relative to follow-up period

There are considerable differences in compliance with OPTN data submission requirements across the various transplant centers. The actuarial method of measuring survival allows for use of cases with incomplete follow-up forms if the reason behind the missing forms is unrelated to the study outcome. But as the level of completion decreases, the potential for completeness of forms to re-

fect the underlying health status of the patients increases, so that unadjusted analyses could give biased results. For this reason, the SRTR computes a measure of completeness of follow-up information available from the centers for the CSRs.

The 'percent follow-up days reported by center' reports the percentage of days actually reported with follow-up forms relative to the number of days targeted for inclusion during the follow-up period. For patients who do 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 the 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. The 'percent follow-up days reported by center' is a measure of the completeness of the data rather than a strict measure of compliance, since even 100% compliant centers may have some short period of unreported patient status after the time their last form was submitted. For instance, a patient's most recent 1-year follow-up form may report his or her status at 10 months rather than 12 months due to scheduling issues unrelated to the study outcome. In this case, only 305 reported days out of 365 days are actually available for analysis.

With the inclusion of SSDMF and CMS data, the number of days of follow-up covered by any source equals the targeted number of days for all patients, regardless of death, and always equals 100%. However, because ascertainment of survival depends on multiple sources of mortality information, the completion of follow-up days reported by each center is still a valuable measure for evaluating the validity of the data. Therefore, even after the incorporation of the SSDMF and CMS data into the CSRs follow-up, the number of follow-up days is still reported in the CSRs, based only on center-reported data.

Analysis of transplant waiting times

There is an increasing shortage of donor organs relative to the number of registrants awaiting transplantation. This holds for each type of organ failure, with the gap between demand and supply widening each year. A variety of organ allocation methods are being developed to address this shortage, each appropriate to the treatment options available for that type of organ. Kidney transplants, which represented 46% of all deceased donor solid organ transplants in 2003, are allocated primarily on the basis of human leukocyte antigen (HLA) mismatch and waiting time (WT). Liver transplants are allocated primarily on the basis of registrant medical condition, with acute liver failure registrants given priority over patients with chronic liver failure, the latter listed on the waiting list in decreasing order of MELD score. Allocation of hearts is based on registrant medical condition and status. The lung allocation system will soon change from a system based upon WT to one based upon medical urgency and the net benefit of transplantation—with net benefit defined as the expected days of life gained

by transplant relative to remaining on the waiting list—during the first year following transplantation. Allocation of organs to registrants according to medical urgency and net benefit balances the value of avoiding imminent death due to organ failure while also avoiding futile transplants. Evaluation of the expected life gained by transplant recipients also provides useful information to registrants about the potential risks and benefits of transplantation. Analytical considerations for two organ waiting lists, kidney and liver, are discussed in greater detail below. Many of the issues raised are applicable to analyses of other waiting lists, but some are particular to these organs.

Kidney transplantation

For kidney transplants, which are still allocated primarily according to WT and HLA mismatch, the SRTR uses several different measures to reflect WT, particularly for CSRs, which reflect the time until transplant rather than the probability of transplant versus other outcomes:

1. Among all registrants, what percentage received a transplant (or other outcome) within a particular time period (e.g. 6, 12 or 18 months)?
2. By what time after listing had 50% of registrants received a transplant?
3. What is the rate of transplantation per time period among actively listed registrants?

Answers to questions 1 and 2 are the most relevant to registrants because they reflect the raw probability of transplantation, accounting for all potential outcomes, including both inactive time and death without transplant. Question 3 is relevant for registrants who are actively listed and for evaluation of the allocation process, which involves only actively listed registrants. The first two questions can be answered directly by evaluating outcomes in different groups of registrants, while the third involves a measure of events per unit of patient time (e.g. patient years).

For the purposes of studying different regions or groups of registrants, all of the measures described above typically yield similar conclusions. The answer to a fourth, related question—What is the median WT among actual transplant recipients?—can be easily computed from recipients during a recent interval of time. This statistic is useful for comparing WTs among regions or among transplant programs. However, the average WT among recipients is not useful for patient counseling, since it does not factor in WTs from registered patients who have not received an organ, or from patients who died or were removed from the waiting list before receiving a donor organ. Although average time until transplant among recipients has little relevance to patients currently on the waiting list, the statistic may be meaningful for future prognosis of transplant patients; for example, increased time on dialysis is known to have a very strong influence on survival following renal transplantation and on developmental problems in pediatric recipients.

The outcomes for all wait-listed registrants are summarized by the fraction who receive a transplant, die without a 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, 1 year, 2 years and 3 years) for each transplant program (www.utransplant.org). The FT is a simple fraction of all wait-listed registrants who received a transplant, regardless of the program where the transplant was performed. The FT summarizes the time to transplantation at any program among all registrants in that transplant program.

The time to transplant (TT) is the time since listing by which 50% (or another stated fraction) of all wait-listed registrants receive a transplant. The TT calculation summarizes the time to transplantation at a transplant program or within a group, taking into account the possibility of not ever receiving an organ. The TT measures the rate of transplantation at a particular program, so registrants 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. Registrants 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 TT and the FT calculations. Note that the median TT would never be reached for groups in which more than 50% of the registrants die or are removed for poor health, since these registrants are counted as never receiving a transplant.

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

Liver transplantation

In the setting where organs are allocated based on waiting list survival probability, the seemingly simple question, 'How long do patients wait for a transplant?' is no longer so simple to answer. Taking liver failure as an example, organs are allocated to chronic liver failure patients based on

waiting list mortality, through a measure (MELD) that changes over time. Estimation of a patient's time until future transplant requires that the probability of potential future MELD pathways be quantified. Even if the probability of changing MELD categories is correctly specified, there is still the issue that changes in MELD score correspond to both changes in waiting list mortality and in transplant probability itself. In some regions of the country, registrants with very low risk of death might never be allocated an organ unless and until their condition worsens. Due to the difficulty in projecting WT until transplant, other important questions arise when considering the liver waiting list:

1. Among Status 1 registrants (acute liver failure), what fraction gets a transplant, what fraction dies and what fraction recovers?
2. Among chronic failure registrants, 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 registrant dies during the same time?

Answering such questions allows the evaluation and comparison of access to liver transplantation for the purposes of both policy development and registrant counseling. Similarly, for each organ that is allocated on the basis of medical condition, it is useful to report measures of transplantation rates separately for different categories of medical condition. Analogous methods can be used for registrants 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 registrants began in February 2002, along with rules for exceptions for registrants with other specific diseases, such as liver cancer (8). The SRTR reports relevant summary statistics and tables to summarize rates of liver transplantation according to status and MELD in the CSRs.

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

Analysis of pre- and post-transplant mortality and graft failure

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) post-transplant 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 *2004 Annual Report*. However, since these results are from different groups of patients, the outcomes are inconsistent across the years. For example, the 5-year survival for the 10-year cohort is not reported and should not be assumed to be the same as the 5-year survival that is reported for the 5-year cohort.

Mortality

Generally speaking, wait-listed registrants are not tracked by their former listing centers for mortality after removal from the waiting list. That is, 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 if they were at high or low risk for subsequent unobserved events, compared with patients under observation.

OPTN death ascertainment alone was used to compute death rates on the waiting list, as reported in each organ-specific section in the *2004 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 registrants are often removed from the list if their health deteriorates to such a point that they are no longer suitable for a transplant. See the accompanying article on data sources for a discussion of post-removal deaths (1). Thus, low death rates on a waiting list are likely to reflect an effective screening process that systematically removes patients when their health deteriorates. Rates based on patients not removed from the waiting list do not apply to registrants generally but to patients currently on (i.e. not removed from) the waiting list.

For the purposes of the CSRs, mortality rates on the waiting list include extra ascertainment for death after removal from the waiting list or, in some cases, before removal.

For these analyses, time at risk begins at either the latter of the start date of the period or the date of first wait-listing; time at risk continues until the earliest of the date of death, transplant (at any center), 60 days after removal for recovered organ function, transfer to another center or the end of the period.

For the purposes of computing expected lifetimes on the waiting list, the SRTR uses information on deaths from other data sources, such as the SSDMF. This is especially important when comparing pre-transplant mortality (which includes time after removal from the waiting list) to post-transplant mortality.

Graft failure

The analysis of graft failure is complicated by the potential for subjects to die. Death serves as a competing risk (6) in the sense that the time of graft failure cannot be observed among patients who die with a functioning graft. Death-censored graft failure estimates the 'cause-specific' rate of graft failure; i.e. the rate of graft failure among patients who have not yet died. This is an interpretable measure that is frequently used. However, cause-specific rates, such as those estimated in an analysis of death-censored graft-failure, can only be combined to produce a meaningful survival curve if the competing risks are independent, an untenable assumption in the context of death and graft failure.

In addition to competing risks, there is also the issue of exactly which events constitute graft 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.

Adjusted analyses

Adjusted analyses are intended to compare patient subgroups with 'all factors being equal'; that is, all factors other than the subgroup-defining factor of interest. Many of the analyses performed by the SRTR involve comparisons of outcomes. For example, the CSRs compare center-specific mortality rates with the national average; an analogous analysis is performed for graft failure. In order to make the comparisons more meaningful, they are adjusted so each facility-specific event rate is measured against the rate expected in light of the facility-specific case mix. 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 higher mortality, being unadjusted, is attributable to the large number

of high-risk patients and, as such, gives no indication that the facility actually has better outcomes than expected for such patients. In contrast, an adjusted comparison would correctly identify the facility as having good outcomes.

The adjustment method used by the SRTR is known as 'indirect standardization'. Essentially, the event count at each center is compared with the expected event count, the latter computed as a weighted average of subgroup-specific national event rates. The subgroups generally are defined by patient age and other prognostic factors, such as disease leading to organ failure. For each patient, the expected event count is the product of that patient's follow-up time (e.g. in years) and the pertinent subgroup-specific national event rate (e.g. deaths per year). 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 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 CSRs.

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 (5) 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 post-transplant mortality rates for deceased donor compared with living donor kidney recipients would indicate that the death rate is 59% higher for recipients of deceased donor kidneys than for recipients of living donor kidneys of the same age and diagnosis. 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 0.5 and 4 years prior to report release for 1-month and 1-year rates, and between 3.5 and 6 years prior for 3-year rates. These cohorts are chosen to reflect 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 with expected counts from the Cox models rather than these survival percentages.

Adjusted analyses are used extensively by SRTR in the CSRs and in analyses based on other data requests. 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. For example, in a comparison involving patient characteristics (e.g. mortality rates by ethnicity), it would be prudent to adjust for variables reflecting therapeutic regimen, if available. However, in an analysis comparing center-specific transplant mortality rates, therapeutic regimen reflects a center’s practices. To adjust for such entities amounts to adjusting away the difference that, if present, one wishes to discern. 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 documentation of CSRs (available at www.ustransplant.org/programs-report.html) includes detailed descriptions of the adjustment models they use.

Regression models

Since its development in the 1970s, the Cox regression model (5) has become the predominant method of analyzing survival data. The popularity of the Cox regression model is well-founded. The model is semi-parametric, in the sense that covariates are assumed to act multiplicatively on the baseline event rate (parametric), but that no functional form is assumed for that baseline event rate (non-parametric). Despite its utility and flexibility, limitations exist with respect to regression models used for survival analysis, including the Cox model. For example, residual plots are generally difficult to interpret and the identification of patterns is a subjective matter. The more sophisticated methods recently developed are computationally intensive to the point of not being feasible for data sets as large as those typically analyzed by the SRTR. In addition, global measures of fit are not available through any standard software packages and would be time-consuming and computationally demanding. Clearly, further development is needed with respect to regression diagnostics for survival models.

Simulated allocation modeling

The simulated allocation models (SAMs) developed by the SRTR are designed to simulate organ allocation and result-

tant patient outcomes in the United States. These models, which have been recognized as valuable by several OPTN committees, provide a method to compare relative outcomes under alternative allocation policies prior to implementation of these policies.

SAMs incorporate both deterministic and random factors. If the input data are fixed, then the initial waiting list, waiting list arrivals, status changes, organ arrivals and rules of organ allocation are all deterministic. The match run itself is determined entirely by the allocation rule specification determined by the user, the organ offered and the patients remaining on the waiting list who are available for that organ.

After the match run has determined the order in which an organ will be offered to candidates, the remaining events are determined randomly through various probability functions. These events include the probability of organ placement with each successive candidate in the match run, time from transplant to death, relisting events and relisting history. Their probability functions depend on candidate and organ characteristics. Organ placement is modeled using logistic regression, with adjustments for relevant candidate demographics, clinical factors, organ factors and factors based on the particular organ and candidate involved (e.g. HLA match, distance, and so on). Post-transplant mortality is predicted using Cox regression models for time from transplant to death, with adjustments for organ, recipient and compatibility factors. Figure 1 shows the time order in which events are processed in SAMs (9).

A family of organ-specific simulation models has been developed by the SRTR with input from the OPTN committees. These include the liver simulated allocation model (LSAM), the thoracic simulated allocation model (TSAM), and the kidney–pancreas simulated allocation model (KPSAM).

Each of these organ-specific SAMs has separate organ-specific components for inputs (candidate information, waiting list histories and donor organ information),

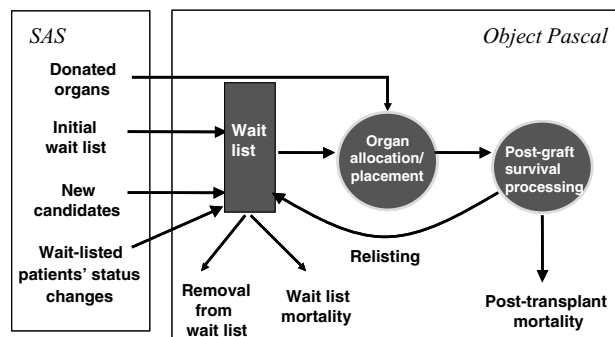


Figure 1: TSAM event-sequenced modeling processes events in time order. Source: SRTR.

allocation rule specifications, placement models and post-transplant events. SAMs are designed to compare the differences in outcomes expected between allocation policies if they were nationally enacted and all other behavior remained the same. Exact replication of actual outcomes for a given year is not a specific goal, due to the effects of physician judgment and local variations in the means of implementing national allocation policy. However, validation tests comparing results of these models with the actual results of particular calendar years has shown excellent agreement regarding those outcomes that are most relevant to the comparison of allocation rules. While certain proposed allocation systems require specific comparisons, the SRTR typically compares numbers of transplants, organ discards and patient deaths when examining sets of proposed allocation systems against current rules.

In support of OPTN committees charged with the development of national allocation policies, SAMs have been frequently used by the SRTR to assess the effect of proposed changes to allocation policies prior to implementation. For instance, TSAM was used to evaluate the effect of implementing the new lung allocation policy, which is based on waiting list urgency and transplant benefit, compared with the current system, which is based on WT. LSAM was useful for evaluating the effect of a new allocation system that involved regionally sharing livers for MELD and PELD scores above 15, the effect of changing the score calculated for adolescents age 12–17 years old from PELD to MELD, and the effect of requiring regional sharing of all pediatric donor livers to children ages 0–11 years old. KP-SAM was used to test the effects of increasing points for zero HLA DR mismatches for pediatric recipients of kidneys from donors <35 years old.

Developments and improvements of all simulation models are continuing efforts, including refinements of placement and post-transplant models and the addition of a patient generator. The first iteration of the SAMs used historical data (and model parameters) as inputs. The results from these SAMs showed how policy change would affect allocation and outcomes if there were no change in other factors. With the new data generator, SAMs can model the simultaneous effects of a hypothesized behavioral change together with rule changes. For example, in modeling a rule change that prioritizes a certain group of patients, the generator may be adjusted to reflect a possible increase in the number of patients wait-listed in that group; the generator might also be adjusted to raise the number of expanded criteria donors to be consistent with anticipated OPO focus in that direction.

In summary, SAMs can be used to analyze allocation effects in several ways: comparing outcomes with different allocation rules; generating realistic numbers of organ transplants and organ discards from the available pool of donor organs; approximating geographic distributions, organ type and status at transplant when current allocation rules are used; and enabling differential placement of organs with varying characteristics and compatibility (e.g. size and blood type). Results from the SAMs have been used by several OPTN committees in predicting the likely effects of changes in allocation rules before considering such rule changes for national policy.

Summary

The numerous methodologies described here are applied by the SRTR and are tailored to address specific questions. Statistical adjustments to make 'all else equal' for comparisons of variables of interest usually require clinical input and thoughtful consideration. Confounding and potential biases must always be evaluated. Simulated allocation modeling is particularly valuable when considering modifications of national policies.

References

1. Dickinson DM, Dykstra DM, Levine GM, Li S, Welch JC, Webb RL. Transplant data: sources, collection, and research considerations, 2004. *Am J Transplant* 2005; 5(Part 2): 850–861.
2. Wolfe RA, Ashby VB, Milford EL et al. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *N Engl J Med* 1999; 341: 1725–1730.
3. Ojo AO, Held PJ, Port FK et al. Chronic renal failure after transplantation of a nonrenal organ. *N Engl J Med* 2003; 349: 931–940.
4. Roberts JP, Wolfe RA, Bragg-Gresham JL et al. Effect of changing the priority for HLA matching on the rates and outcomes of kidney transplantation in minority groups. *N Engl J Med* 2004; 350: 545–551.
5. Cox DR. Regression models and life tables (with discussion). *J Royal Statistical Soc B* 1972; 34: 197–220.
6. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*, 2nd Edn. Hoboken, NJ: Wiley, 2002.
7. Wolfe RA, Schaubel DE, Webb RL et al. SRTR report on the state of transplantation: analytical approaches for transplant research. *Am J Transplant* 2004; 4(Suppl. 9): 106–113.
8. Freeman RB, Jr., Wiesner RH, Roberts JP, McDiarmid S, Dykstra DM, Merion RM. SRTR report on the state of transplantation: improving liver allocation—MELD and PELD. *Am J Transplant* 2004; 4(Suppl. 9): 114–131.
9. Thompson D, Waisanen L, Wolfe RA, Merion RM, McCullough K, Rodgers A. Simulating the allocation of organs for transplantation. *Health Care Manag Sci* 2004; 7: 331–338.