

Data sources and structure

David M. Dickinson^{a*}, Mary D. Ellison^b and Randall L. Webb^a

^aScientific Registry of Transplant Recipients (SRTR)/University Renal Research and Education Association (URREA), Ann Arbor, MI

^bOrgan Procurement and Transplantation Network (OPTN)/United Network for Organ Sharing (UNOS), Richmond, VA

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

Key words: Data collection, data sources, data structure, death ascertainment, OPTN, SRTR, statistical analysis, transplantation, UNet

Received 17 September 2002, revised and accepted for publication 4 December 2002

Introduction

This article discusses a rich resource of data used to describe all aspects of transplantation, from donor and recipient characteristics to immunosuppression medications. These data are used by the SRTR, the OPTN, and a wide variety of other researchers as the basis for reporting on the state of transplantation in the United States, as well as answering a wide array of research questions. They are the source for the figures and tables in the OPTN/SRTR *Annual Report*. They form the basis for reporting on both OPTN and SRTR web sites, providing medical professionals and patients alike with the answers to such critical questions as: *How fast are waiting lists growing? Which center has experience serving patients like me? How quickly might I get an organ if I register at a different center, and are my prospects for survival after transplant there as good?* Finally, these data form the basis for extensive analyses in support of policy-setting by the Secretary's Advisory Committee on Transplantation (ACOT), OPTN/UNOS committees, and other government and nongovernment requesters: *Is a transplant candidate*

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 is no restriction 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 but are available online at <http://www.ustransplant.org>.

better off accepting an organ from a less-than-ideal candidate or staying on a waiting list? How do antigen matching rules affect racial distribution of organs, and how do they affect survival? What are the effects of allowing patients to be put on waiting lists at more than one transplant center? The many questions that may be asked of the transplantation data are to some degree controlled by how the data themselves are gathered and arranged.

It is the goal of this article to further understanding of the way transplantation data are collected and organized, in order to enable better interpretation of research results, more acute awareness of data limitations, and clearer concepts of how new analyses might proceed. This article is intended for an audience of researchers in the transplant community: both those who use existing research and those who create new analyses with these data. By examining the sources, quality, and organization of the different types of transplant data available, we hope to stimulate new exploratory initiatives and help researchers with study design—as well as improve the understanding of existing results.

A fundamental step in describing the data available for research on transplantation is to conceptualize the range of information available and to organize it into areas of research interest. The first section of this article previews the final research database by showing how the diverse collection of data are organized in records representing the different types of 'units of analysis' of interest to a researcher, saving a detailed discussion of the sources for each type of record for later sections. We describe how such a wide range of sources is reorganized from their original format suiting their original purposes—mostly organ allocation but also Medicare billing, Social Security Administration benefits, etc.—to a format better adapted to the support of research questions. Just as is the case in designing a research database, it is useful to begin describing this database by considering how the table organization will facilitate answering a series of interesting research questions.

The remaining sections of the article describe the sources of the underlying data, how they are collected, and how they fit into the framework outlined above. In the second section, we focus on data collected by the OPTN for the purposes of both organ allocation and transplantation research. This section focusses on the historical and technical development of these data collection systems, with an emphasis on how changes and quality control measures in these systems have improved the quality of data available. In this discussion we hope to acquaint researchers with the particular strengths and weaknesses

present in many of the primary data elements. We also point out the context in which these data were originally collected, which may be different from how they are used for research.

In the final section of the article, we describe in more detail the 'secondary' data sources incorporated into the research database used by the SRTR. These secondary sources are used to augment the primary data reported by OPTN members, both to improve the quality and to expand the scope of the data. We describe some of the sources available, and examine their impact on answering several research questions.

Further discussion of the types of analyses supported by these data can be found in 'Analytical Approaches for Transplant Research' (1), a companion article in this supplement, as well as in Appendix H of the *2002 OPTN/SRTR Annual Report*.

Organizing Data for Research

Data structure and units of analysis

This section describes the organization structure of many sources of data assembled for transplantation research. Though the examples here are taken directly from the SRTR, they are generic in application: They might resemble data organized for similar purposes by the OPTN or any other researcher who obtains these data from either the SRTR or the OPTN.

We should first review some terms used to describe data organization. Data are arranged into separate 'tables', often SAS or SPSS datasets or SQL tables. Each of these tables, which are 'relational' in that they may be linked one to another, contains a series of rows or records, each representing one item of interest such as a person, transplant recipient, or organ. Each column, known as a field or variable, represents a different characteristic of that record. In a table describing transplants, for example, these columns include such things as age at transplant, type of organ, and information about the transplant center; in a table describing each organ available from a deceased donor, these fields might include the eventual disposition of the organ, how many candidates refused the organ, or the reason that it was not recovered.

The roles in the 'relationship' between two tables are often described as 'parent' and 'child': for each record in a child table, there is a linked record in its parent table. There may, however, be some parent records with no child records, while other parent records have many child records. For example, in the relationship between a transplant (parent) and transplant follow-up (child), a transplant may have no follow-up forms filed, or one, or two, or 10; yet all follow-ups must be linked to one and only one

transplant. Extensive parent-child organization is useful for maintaining data integrity in applications that keep track of constantly changing values, such as the OPTN organ allocation procedures, though it may make research with these data computationally intensive.

Instead, when preparing analysis files, consideration is given to the 'unit of analysis' that may be of interest to the researcher. Different tables are organized for different research questions, using different units of analysis as rows in each table. More emphasis is placed on creating a table where a single record carries a wide variety of information about a record of inherent interest to the researcher, and less consideration is given to the efficiency of data storage, waiting list management, or allocation matches. Data from many sources and related tables may be summarized and attached to the record of interest. For example, many researchers want to examine transplants (unit of analysis) and their post-transplant survival, such as Tables X.9 in each organ-specific section of the data tables in the Annual Report. A table in which each row represents a transplant may be augmented with data summarized from the related tables of follow-up sources, such as each recipient's latest status as alive or dead and the date of that status. A table in which all of this information is summarized on a single record is easier to analyze than assembling information from multiple parent and child rows in multiple tables. However, for other purposes, such as counting immunosuppressive medications during follow-up periods, it may still be useful to use individual records for each follow-up period. Figure 1 shows a useful scheme of organizing these data into a 'record of interest', drawn from the example implemented for analyses by the SRTR. This figure also gives an idea of the breadth of commonly used units of analysis and the relationships between them.

One central organizing element in this structure is the Person Linking Table (PLT), in which each record is a person—perhaps a living donor, transplant candidate, or transplant recipient. The PLT facilitates a common patient identifier to be assigned to records in all other tables, linking persons on the basis of Social Security numbers (SSNs), names, dates of birth, and other person-level information, while accounting for many of the mistakes in entering these fields. The maintenance of this identification roster, with aggregated identification information compiled from all data sources, has two primary functions. First, it facilitates a system of matching to both external data sources and other records within OPTN data, such as for persons who receive multiple transplants or even for donors who later become recipients. Second, the common patient identifier provides an anonymous means of person identification for researchers without revealing names or SSNs. The matching system is described in greater detail below.

The other table entities in Figure 1 relate to a specific subject of interest for research: candidacies, donors,

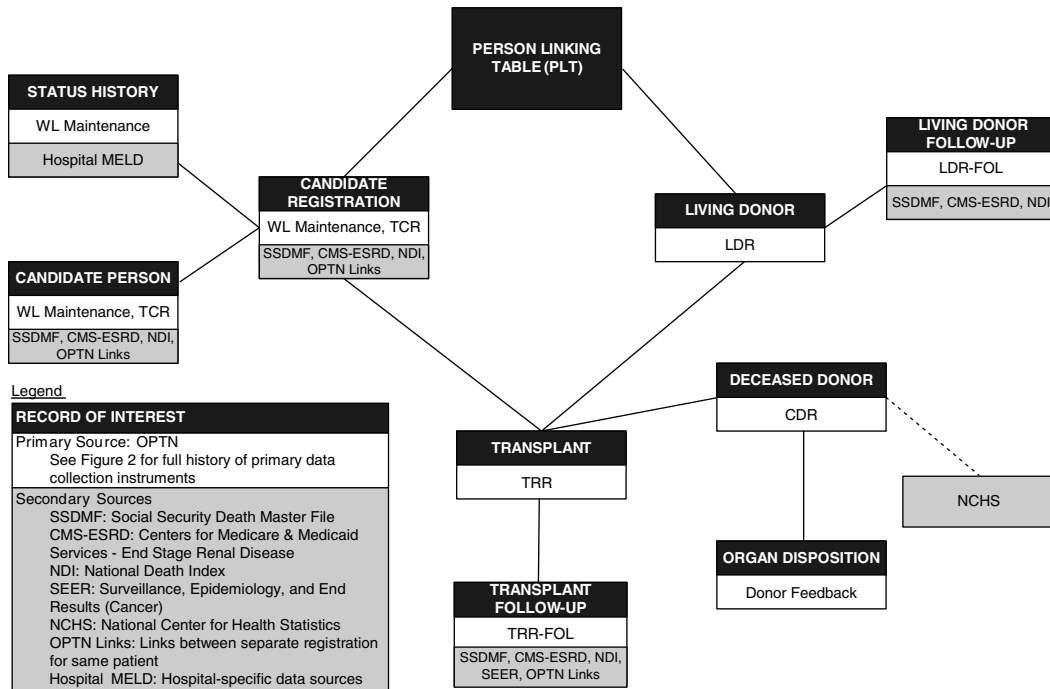


Figure 1: Transplantation research data organization, primary and secondary sources. Source: SRTR.

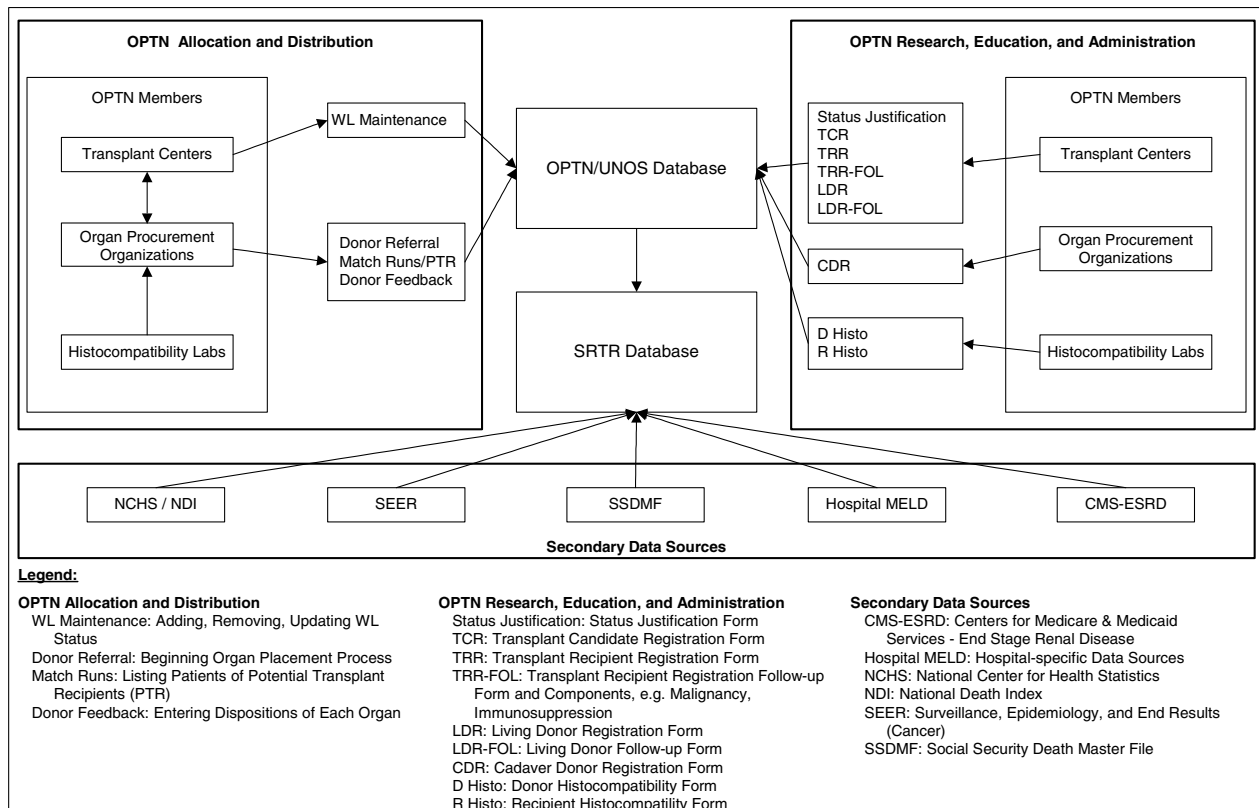


Figure 2: Data submission and data flow, primary and secondary sources. Sources: SRTR and OPTN.

transplants, and the components thereof. In addition, this figure documents some of the primary and secondary data sources which may contribute to each table. Further detail regarding the specific data collection instruments for the primary data collection by the OPTN is shown in Figure 2.

Analysis tables

Though the PLT is useful for keeping track of summarized person-level information and matching to newly incorporated records or data sources, it is often not directly useful for analyses. Instead, using a common patient identifier, data from this table are added to separate tables more closely based on 'units of interest' for analyses such as the following:

- transplant candidates (e.g. waiting time, mortality on the waiting list);
- transplant recipients (e.g. graft survival, complication rates, incidence of tumors);
- donors (e.g. donation rates, donor characteristics that influence graft survival, etc.).

Each of these, in turn, has its own child tables as well as both primary and secondary sources of data. Later in this article we discuss, in more detail, the primary data collection methods and secondary data sources for these tables.

Analysis tables: candidates

'Time to Transplant' tables, the second in each of the organ-specific data tables sections of the Annual Report, make use of a unit of analysis that represents a candidate registration. Such a table may also be useful for measuring mortality on the waiting list, either at the center-specific level or for comparison to post-transplant mortality in evaluating the efficacy of transplant for a given patient.

The 'candidate registration' table includes persons who are registered on the OPTN waiting list as well as additional candidates who have received a living donor organ, even if they have never been placed on the waiting list. The vast majority of candidate information comes from the candidate registration and waiting list information collected by the OPTN. This table presents information about candidates during the time they are waiting to receive an organ, such as the center at which they are listed, when they are listed, factors affecting organ allocation like blood group or medical urgency status, and when they are removed from the list and for what reason (death, transfer to another center, transplant, etc.). Often, fields in the operational data are transformed to more closely reflect events of interest for analyses. For example, to facilitate time to transplant analyses, waiting list removal dates for transplants are set to the transplant dates when the candidates cease to be eligible for allocation—though in practice a patient may be removed from the waiting list at any time from the date an organ is allocated until days after it has been transplanted.

As Figure 1 indicates, the candidate file is primarily based on data that are entered as part of waiting list maintenance by the transplant centers, as well as the Transplant Candidate Registration (TCR) Form. These data may also be augmented with data from other data sources as described below. Most notably, additional mortality sources are very important, because transplant programs are not required to track and report outcomes after removal from the waiting list (other than removal for transplant). These sources may include the Social Security Death Master File (SSDMF), Centers for Medicare and Medicaid Services data on End-Stage Renal Disease patients (CMS ESRD) for kidney and kidney-pancreas patients, and the National Death Index (NDI).

Some analyses make use of candidate data recorded at a narrower level than the usual record of interest. For example, the first relational 'child' table shown connected to the candidate file is the waiting list 'status history' table: for each registration on the waiting list, at least one record exists in the status history table. This table records characteristics that may change during the course of waiting list tenure, such as medical urgency status or Model for End-Stage Liver Disease (MELD) score for liver candidates. Each record in this table is associated with a time at which those characteristics began and ended. Such a file is useful for finding a patient's status on any given day, calculating the accumulated time at each status at any point in time, or examining how trends in a patient's MELD score might affect mortality. From an organ allocation perspective, on the other hand, only the current urgency status and a running total of time accumulated at (or above) each status are important. The status history analysis table is created by examining histories of changes to the operational waiting list that are recorded as part of the audit process in the operational organ-allocating database, noting all changes that involve status, and augmenting this file with nonoverlapping start- and end-dates for the span of each set of characteristics. Therefore, an analyst may move through this file in a temporal fashion for each patient, examining current status for each patient and facilitating a time-dependent model such as one that associates status on a given day with outcome (mortality or transplant) on the same day.

The status history file is a highly focused accounting of the candidate file; the 'candidate-person' table, conversely, aggregates candidate records into a much wider view based on individual persons. In the first candidate table described, and for the purposes of organ allocation, a person is given a registration record each time he or she is entered onto a waiting list at a transplant center; a given person might have several registrations, either in sequence or concurrently. By using the common patient identifier, one can construct 'candidacies' that span registrations, separated for each person only by transplants. A candidacy in this file starts from the time a patient is first put on the waiting list at any center and ends when that

patient receives a transplant (from a living or deceased donor) from any center, is removed for the last time, or dies. A second candidacy might begin for the same patient when he or she is relisted after a failed transplant. The candidate-person table also has a status history sub-table, whose records have similar information and purpose to the status history child table of the registration-based candidacy file, but with the additional function of reconciling differences between status recorded for different listings and summarizing the number of concurrent listings at any point in time.

The candidate-person approach is consistent with an 'intent-to-treat' analysis. In such an approach, the original goal of any wait-listing is to transplant the patient, doing so at any center is a success for that patient, and the 'waiting time' that a patient cares about is the time from his or her first listing until transplant. By contrast, the registration-based candidacy table may be more relevant when evaluating a center's ability to move a patient through the waiting list process.

Analysis tables: transplants

A subset of the candidate registrations make their way into the transplant table, including persons who have received a transplant from the waiting list as well as those receiving a living donor transplant. The transplant file is used by analysts wishing to characterize trends in volume and characteristics of patients receiving transplants (Table 4 in the organ-specific data tables sections of the Annual Report), as well as analyses examining post-transplant survival (Tables 8 and 9, Graft and Patient Survival).

This table draws primarily upon information from the Transplant Recipient Registration (TRR) Form, filed by centers following each transplant. The table includes characteristics of the patient at the time of transplant and the transplant operation itself. For ease of analysis, characteristics of the donor are added, as well as donor-recipient interactions, such as calculated HLA mismatch scores, blood compatibilities, and whether the organ was 'shared', based on the relationship between the organ procurement organization (OPO) recovering the organ and the transplant center.

The primary transplant table also includes summarized information from the child table 'transplant follow-up'. Data in this table come from the post-transplant follow-up forms collected 6 months after transplant (except for thoracic organs) and then at each yearly anniversary. These follow-up forms contain items such as hospitalization, current lab values, functional status, and other developing medical conditions. This table, in turn, has specific sub-tables of its own, recording details of immunosuppression treatments and developing malignancies, for example.

The data gathered during the organ allocation process and follow-up forms are strengthened further: Similar to the candidates table, several secondary follow-up sources pertaining to death, graft failure, retransplant, and resumption of dialysis are summarized and added to the transplant table, as described in Figure 1. These important elements, and their ramifications for data completeness, are described later in this article.

Analysis tables: donors

Donor information is shown separately for living and deceased donors—not only because such different primary information is collected by the OPTN for each group, but also because each relates to its own set of secondary data elements and its own analyses. Indeed, much of the donor information that is common to both types of donors and useful for analysis of transplant outcomes has already been added to the transplant file itself. The donor files might be more frequently used for such things as analysis of organ disposition and reasons for nonrecovery of organs from deceased donors, or for examining the post-donation outcomes of living donors.

For each deceased donor, up to 11 whole organs or organ segments may be recovered (one heart, two kidneys and lungs, and up to two segments each for pancreas, intestine, and liver). This recovery information is stored in a sub-table of the deceased donor table, 'organ disposition', giving reasons for nonrecovery or nonconsent, and eventual disposition of each organ. The information from this table is taken directly from forms filed by OPOs. The third section of the data tables details disposition (e.g. local transplant, shared transplant, used for research), reasons for nonuse, and reasons for nonrecovery of organs. Analysts might also use such a table to glean additional information regarding unused organs or might wish to examine organ recovery data available in OPO-specific format.

In addition to organ disposition data, researchers may combine deceased donor information with external sources of mortality data for the general population such as information from the National Center for Health Statistics (NCHS). Combining such sources allows researchers to compare availability of potential donors in certain areas to the number of organs recovered, or to evaluate successful methods used to obtain family permission for organ recovery. The use of OPO forms and NCHS data is discussed below.

Living donors are also included in the PLT, to facilitate matching with internal and external data sources, and allow for additional ascertainment of events such as death, dialysis, or registration on a waiting list. For living donor follow-up, transplant centers are asked to report at 6 months and 1 year, though compliance and reliability are not as good as they are for recipient follow-up. Many centers submit follow-up forms for living donors as required, but are less likely to see these donors, who are

often healthier or live elsewhere and may therefore be more difficult to track. For living donors from 2000, though 90% have an appropriate 1-year follow-up form filed, 42% of these living donors are coded as 'lost to follow-up'—indicating that, even when complying with OPTN follow-up requirements, centers do not know what has happened to these patients.

Though possible secondary data sources are listed (SSDMF, CMS ESRD, NDI), lack of completeness and accuracy in living donor identification information jeopardizes the use of these external sources. Before April 1, 1994, SSNs were not collected for living donors. Since then, more than half of SSN matches to the SSDMF are highly improbable (based on review of names and implausible relationships among birth dates, death dates, and dates of organ recovery), indicating that there is probably significant inaccuracy in these identifiers even when they are available.

Primary Data: The OPTN Data Collection System

Data system components

The OPTN data collection system and database were developed in 1986 by the United Network for Organ Sharing (UNOS, the OPTN contractor) after the 1984 National Organ Transplant Act called for the creation of a national network for organ sharing and a scientific registry to monitor the clinical progress and effectiveness of transplantation. The information systems themselves have undergone many changes with regard to technology, data collection processes, and data content. The system consists of three components: the national transplant waiting list, the donor–recipient match process, and the data collection 'forms'. The first two together can be thought of as the allocation data, as these are the data essential for the day-to-day operation of distributing organs to potential recipients. The 'forms', collected with somewhat less urgency, are intended more for research and administration purposes.

Figure 2 shows data flow into both the OPTN and SRTR databases, focussing on the different mechanisms for submission of data by OPTN members to the OPTN database. The figure shows data separated into two types: that used for organ allocation, on the left, and that used for research, education, and administration, on the right. This figure also serves as a full list of the major data collection instruments in place for OPTN members. Copies of the forms may be found in Appendix I of the *2002 OPTN/SRTR Annual Report*.

The initial process of data collection, as well as organ allocation, begins with the waiting list. At the time a patient is placed on the waiting list, essential data are captured for donor matching and allocation. Such data have always included such variables as blood type and

medical urgency status. These data can and, in some cases, *must* be revised and updated by personnel authorized to access the waiting list. Although the transplant program controls the list for its patients, it may authorize the OPO or even the histocompatibility laboratory to perform maintenance on it. Adding a patient to the waiting list prompts the generation of the TCR, which is sent to the transplant program to collect additional information about the candidate that is used for purposes other than matching and allocation.

The donor–recipient match process begins when an OPO enters a donor into the system. Donor data essential for matching and allocation are captured and a 'match run' for each organ type available (e.g. kidney, heart, etc.) is generated. This programming accomplishes several functions simultaneously: it reflects organ distribution and allocation policies in place at the time of the match, identifies all patients that are clinically compatible with the donor, assesses their geographical appropriateness based on donor location, and assigns priority rankings. The product is a match run for each organ type, available electronically and in printable formats to the OPO for organ placement. If authorized by the OPO, a histocompatibility lab may run a match in lieu of the OPO.

There are several subsystems within the donor–recipient match process that collect data and generate donor forms. At the time of the match, the Potential Transplant Recipient (PTR) Form is created on the match run itself. This data form is made available for the OPO to record the refusal reasons (such as donor quality, recipient unavailability, or positive crossmatches) for potential recipients ranked higher on the list than the ultimate recipient(s). This information is provided by the OPO based on organ offer responses from transplant programs. Transplant centers may then validate, via UNet, the refusal reasons entered by the OPO during a 15-day period after the match is completed by the OPO.

The OPO is required to report the results of donor organ placement efforts through a process called 'donor feedback'. Upon completion of the donor feedback (itself a set of forms), the Cadaver Donor Registration (CDR) and Donor Histocompatibility (DH) Forms are generated to gather donor data for research and reporting purposes. OPO personnel complete and return the CDR forms, while the tissue-typing laboratory serving the donor hospital provides data requested on the DH form.

Transplant recipients must be removed from the waiting list within 24 h of receiving an organ. This completes the 'recipient feedback process', and two additional forms collect additional research and reporting information: the TRR Form, completed by the transplant program, and the Recipient Histocompatibility (RH) Form, submitted by the recipient center tissue-typing laboratory. When a hospital reports a living-donor transplant for a patient who was

not on the waiting list, a TCR, TRR, and RH form are generated for the recipient. A Living Donor Registration (LDR) and DH Form are generated for the donor. The transplant program completes the recipient forms and the tissue-typing laboratory completes the histocompatibility forms. The Living Donor Follow-up and Transplant Recipient Follow-up Forms are generated and transmitted to the transplant program 6 months after transplant and every 1-year anniversary of the transplant. If a post-transplant malignancy is reported on a follow-up form, a Post-transplant Malignancy Form is generated and sent to the transplant program. If a patient is reported as retransplanted, dead, or lost to follow-up, no further follow-up forms are generated for the specific transplant event. If a pancreas or kidney graft failure is reported, then follow-up on the patient is continued for only 2 more years. For all other organs, no further follow-up forms are generated.

Data collection forms submitted by transplant programs are completed by a variety of hospital employees, including nurses, clinical coordinators, clerks, and administrative assistants. A hospital's 'data coordinator' can be any of these types of personnel. Since the inception of the OPTN database in 1986, financial pressures on hospitals have increased, as has the volume of data forms for most hospitals. Some programs devote significant resources to OPTN data submission activity; others less so. The implications of budgetary pressures for data quality have been a primary concern of the OPTN/SRTR Data Working Group and the OPTN Data Advisory Committee, two new committees supporting data-related OPTN process and policy development. During a comprehensive 2-year process since the fall of 2000, these committees have been able to significantly streamline and reduce the amount of data to be collected by the OPTN. Final changes will be implemented once approved by the OPTN/UNOS Board and the Federal Office of Management and Budget. It is hoped that a lower data burden at the facilities will lead to higher quality for a smaller amount of data, focusing on the most scientifically relevant items.

History of the data collection system

Figure 3 shows some of the evolution of the OPTN/UNOS data systems. UNetSM, an Internet-based application for waiting list maintenance, donor-recipient matching, and forms-based data collection for research and administration, was implemented on October 25, 1999. Before UNet, the most significant modifications to the data system occurred in 1990 and 1994. In 1990, the waiting list and data forms systems were converted from a flat file data system to a relational database, making the data easier to manage with regard to both storage and analysis. Based on almost 7 years of use, analysis, and reporting by the OPTN/UNOS committee system, the UNOS staff, and the Federal Government, a large number of data elements were added to the data collection forms. These additions required a second database conversion in April 1994. All

OPTN/UNOS committees provided input during the forms revision process. Since that time, an incremental process of adding data fields has resulted in gradual increases in data volume.

In 1996, UNOS created a client-server application on a Lotus Notes platform called Tiedi[®] (transplant information electronic data interchange) to collect transplant data electronically rather than on paper. This was the first effort to transfer to OPTN members the ability to maintain the submission of their own data collection forms and to eliminate the need to mail paper forms to UNOS. With this system, as many as 50% of centers and laboratories were using Tiedi for data submission. Since its use was not a required method of data submission, some stayed with the existing paper-based submission and manual data entry system.

When UNOS implemented UNet in 1999, the degree of security increased significantly, requiring user-specific passwords and encryption for all patient-identified data transmission. An on-site UNet security administrator assigns access privileges and controls user access at each hospital. All data are now transmitted via the Internet with secure socket layer (SSL) technology and 128-bit encryption. As well, the utilization of electronic submission of data among members has increased greatly: currently, 3 years following the implementation of an Internet-based data system, 97% of OPTN centers, laboratories, and OPOs enter their research and administration forms electronically.

UNet has tightly integrated all three components of the data system. The waiting list and the forms databases were combined into a single longitudinal relational database (Microsoft SQL Server), and the data systems were no longer parallel and compartmentalized but seamlessly integrated. Within 6 months, the percentage of OPTN members using the new system to access and manage the waiting list increased from an estimated 40% to more than 90%. Currently, all waiting list management is performed on UNet (mostly by transplant center personnel) rather than by requesting changes by phone through the UNOS Organ Center.

For certain wait-listed patients, waiting list data management is not only available to the hospitals but in some cases is *required* in order to avoid automatic allocation status downgrades. For example, a patient listed as liver allocation Status 1 must be recertified weekly for that status by the hospital, on the basis of current laboratory data. Additionally, since July 8, 2002, status justification forms for liver and heart (Status 1A and 1B) must be submitted through the UNet system.

Before UNet, the donor-recipient match run yielded a computer-printed data list. With the implementation of UNet, the match list became a data file from which data

	1986–90	1990–94	1994–96	1996–99	1999 to Present
	Pre-OTIS		OTIS	OTIS + Tiedi®	UNet SM
Waiting List Management					
Communication	Phone to Organ Center with paper back up and validation. Some facilities use terminal emulation via modem			Member online (Web-based)	
Donor-Recipient Matching					
Communication	Terminal emulation and modem or phone to organ center and faxed to OPO			OPO generates online (Web-based)	
Data Collection Forms					
Mode of submission	Paper. Manual data entry at UNOS. Line prompt entry		Electronic forms added		Web-based submission. Paper forms phased out
Submission prompting	Member- initiated	Electronic events prompt form generation. Forms mailed by UNOS			Electronic events prompt blank web-form generation
Edit checks	Few		Checks added over this period, data verification reports by mail		All fields validated electronically. Verification reports by mail
System					
Storage system	VMS flat files	VMS relational database	VMS relational database, Lotus Notes		Microsoft SQL Server Relational Database
Component integration	None		Match and forms linked. WL addition initiates TCR		All systems completely integrated
Security	One password per center. No encryption during transmission			User-specific passwords. Full 128-bit encryption	

Figure 3: OPTN/UNOS data system evolution. OTIS = Organ Transplant Information System, Tiedi = Transplant Information Electronic Data Interchange, UNet = Internet-based data collection system. Source: OPTN.

variables could later be extracted for analysis. This change also allowed the OPTN to integrate PTR data with the match output. Another advantage of UNet is that matches for all organ types can now be run simultaneously—rather than serially—as was necessary on the previous mainframe computer system. The flexibility of match runs that are files rather than printed lists had allowed OPOs to view, print, and export matches as data files that can be stored in databases at the OPOs and in the OPTN data system.

With regard to data submitted on forms for research and administration (e.g. TCR, TRR, and TRF Forms), the transition to an Internet-based system has had a number of implications. Forms are generated and appear as 'expected forms' when the member is in UNet. The member can complete the electronic forms manually or import data from a local electronic records system. UNet forms include fewer text fields than the paper forms did, utilizing pick-lists and reducing the need for visual edit checks in these fields by UNOS data quality staff. Most other fields have programmed acceptable responses and standard data ranges. Immediate edit checks and cross-field edits for some variables reduce data errors by allowing the data collector to pay immediate attention to problems as the data are entered. Forms cannot be electronically marked as 'validated' (complete) until all fields have been entered and have passed a series of edit checks. Data quality has become largely the responsibility of the system and of OPTN members, and submission via UNet has eliminated the mailing back and forth of paper forms containing erroneous or incomplete data.

Measures of internal data quality

The quality of the data within the OPTN database is affected by the timeliness, completeness, and accuracy of the data submitted by members. Also pertinent in any discussion of quality is whether the variables collected are sufficient and appropriate (and not superfluous) for the needs of the OPTN, the SRTR, the Federal Government, and the public. These measures of data quality are currently being evaluated by a new committee, the joint OPTN/SRTR Data Working Group. The most recent OPTN and SRTR contracts required that such a committee examine data quality in detail and advise the OPTN/UNOS Data Advisory Committee and Board on necessary revisions. Other aspects of OPTN data quality are addressed by activities of the SRTR, internal operations at UNOS, and the OPTN data submission policy compliance process.

Approaches to improve data timeliness and completeness

Until June 30, 2002, OPTN data submission policies required that 99% of data forms due from an OPTN member be submitted within a year of the dates they were expected. In most cases, the expected date for a form was 60 days after it was generated (e.g. transplant date or transplant anniversary). In an effort to improve the timeliness of data collected by the OPTN, the Health Resources and Services Administration (HRSA) of the Department of Health and Human Services included in the current OPTN contract a requirement that 100% of each program's data be complete within 6 months of the

Table 1: Transplant Recipient Follow-up (TRF) form submission at 1 year after form generation, by transplant program type and volume

Organ and program volume*	No. of programs	Percentage of TRF forms submitted within 1 year of expected date				
		0%	1–33%	34–66%	67–99%	100%
Heart						
Low (0–9)	44	0.0%	4.5%	6.8%	29.5%	59.1%
Medium (10–17)	44	0.0%	6.8%	4.5%	40.9%	47.7%
High (18+)	42	0.0%	11.9%	9.5%	57.1%	21.4%
Total	130	0.0%	7.7%	6.9%	42.3%	43.1%
Kidney						
Low (0–24)	78	1.3%	6.4%	7.7%	32.1%	52.6%
Medium (25–59)	83	1.2%	6.0%	6.0%	41.0%	45.8%
High (60+)	83	0.0%	6.0%	13.3%	45.8%	34.9%
Total	244	0.8%	6.1%	9.0%	39.8%	44.3%
Liver						
Low (0–21)	37	0.0%	5.4%	16.2%	35.1%	43.2%
Medium (22–45)	39	2.6%	7.7%	10.3%	35.9%	43.6%
High (46+)	39	2.6%	10.3%	2.6%	38.5%	46.2%
Total	115	1.7%	7.8%	9.6%	36.5%	44.3%

Source: OPTN database. *Transplants performed in 2000.

form's expected date. The OPTN/UNOS Board approved this policy change in November 2001.

Table 1 shows transplant centers' compliance with the follow-up data submission policy in place before June 30, 2002. These results are stratified by the volume of transplants performed at each program in the previous year. The data show a number of high-volume programs in compliance with the previous policy. For example, 83 high-volume kidney programs (35%) submitted their year 2000 follow-up forms within a year of their expected dates and, as such, had perfect compliance. Although some low-volume programs show poor compliance, there is a slight tendency for smaller programs to have better compliance with follow-up policies. In response to concerns that the accuracy of publicly reported program-specific survival rates may be affected by incomplete outcomes data in the OPTN database, the SRTR has undertaken an effort to obtain missing OPTN outcomes data from other sources, as described below. Overall, 87% of TRFs were submitted in compliance with policy, as shown in Table 2. In contrast, nearly 95% of RH forms generated in 2000 were submitted on time. OPTN member compliance with data submission policy is an area of increasing focus for the UNOS Policy Compliance Department and the OPTN/UNOS Membership and Professional Standards Committee, which is exploring more direct means to ensure compliance.

Approaches to improve data accuracy

Monitoring the accuracy of data in the database involves edit checks during the data entry process, internal processes at UNOS, and a collaborative effort of the OPTN and the SRTR. The UNOS Help Desk takes calls from members who find inaccuracies within fields that can

only be modified by UNOS staff (e.g. transplant dates and SSNs). UNOS also creates computer programs that search for inconsistencies in the database and generate discrepancy reports. For example, one program compares data entered into the age, height, and weight fields for each patient, looking for cross-field entries that seem unlikely or impossible. Other such programs compare entries for employment status, education, and age, and check patient functional status for consistency with medical urgency status. In addition, the SRTR delivers similar discrepancy reports to the OPTN each month to raise further data quality issues. When problems with records arise, data quality specialists resolve them through UNet and direct contact with transplant centers. Problems that affect a large number of records can sometimes be resolved through programmed edits, but other fields must be addressed individually. Fields in which UNet allows incorrect data entry are identified on an ongoing basis, and UNet edit checks are regularly revised to reduce opportunities for data entry errors. Recent efforts to detect database problems have included 22 different discrepancy reports and 38 different database checks. Some of these reports and checks are rerun on a regular basis to correct recurring errors. Others involve one-time projects to resolve problems, such as those related to previous database conversions or modifications.

Database checks performed to detect problems in the data have included checks among living-donor and recipient records for invalid SSNs (e.g. strings of 0s or 9s sometimes used when SSNs are unknown at the time of data entry) and checks for inconsistent entry of date of birth, race, gender, and blood type across records for patients wait-listed at multiple transplant programs. Other checks have included searches for persistent waiting list registrations when programs have reported

Table 2: Data submission compliance rates, by form type, for forms generated during the year 2000

Form type	No. expected	No. received in compliance	Compliance rate
Transplant Candidate Registration (TCR)	46 199	41 761	90.4%
Transplant Recipient Registration (TRR)	26 073	23 320	89.4%
Transplant Recipient Follow-Up (TRF)	172 448	150 014	87.0%
Living Donor Registration (LDR)	6 096	4 939	81.0%
Living Donor Follow-Up (LDF)	2 294	1 739	75.8%
Post-transplant Malignancy (TMR)	628	628	100.0%
Cadaver Donor Registration (CDR)	12 817	11 872	92.6%
Recipient Histocompatibility (RH)	23 004	21 815	94.8%
Donor Histocompatibility (DH)	13 316	12 454	93.5%

Source: OPTN database.

patients as having been transplanted and searches for transplant records when waiting list registrations have been removed for reason of transplant. With the addition of a number of new features and data entry checks in UNet, many types of database checks are no longer necessary. This has resulted in more efficient use of time for staff and in improved data quality.

In addition to a number of special discrepancy reports generated through the UNet application and sent via UNet to OPTN members for problem resolution, the OPTN also generates and prints a number of reports that it mails to each member. These mailings include a monthly summary of the member's overdue forms, a monthly list of the member's reported living-donor transplants, and semiannual confirmation reports of transplants, living-donors, and deceased donors. Each member also receives an annual report of its data submission compliance rates, according to form type. Some aspects of data accuracy cannot be addressed by electronic data entry edits, programmatic data checks, or efforts to ensure compliance with data submission policies. Experience with OPTN data suggests that certain variables within the database may be more reliable than others. In an effort to learn more about the difficulties of providing accurate data for certain fields, UNOS staff have conducted preliminary on-site transplant program audits using actual patient charts to check the accuracy of information provided to UNOS. Results of the audits suggest that data variables involving objective information readily available in medical charts and requiring little or no interpretation (e.g. race, age, and gender) tend to be highly accurate. Other types of information (e.g. patient education level, employment status, and functional status) are more difficult to find in the charts. Results of various serological tests of interest to the OPTN are largely available in the charts, but details regarding testing methods and timing of the test in relation to the transplant procedure, also of interest to the OPTN, are more difficult to interpret from the chart. Such observations by UNOS staff and OPTN members alike are being factored in as the Data Working Group and Data Advisory Committee consider data collection revisions.

Secondary Data Sources

Reasons for additional sources

Other sources besides the data collected by the OPTN provide important information that may be linked to these data or used in conjunction with them. Additional data sources help determine the areas of weakness in compliance and accuracy of the data collection described above; they can also expand the scope of available research. For example, additional data sources can help researchers perform the following important tasks:

- Ensure complete ascertainment of mortality and graft failure, improving precision of analyses and answering questions about the quality of transplant data submitted by a transplant center.
- Expand measurement of events not collected by the OPTN, such as death after a candidate is removed from the waiting list.
- Provide additional ascertainment of other events, such as malignancies from local cancer registries across the country.
- Offer measures of potentially available donors for evaluating donation practice patterns.
- Establish correlations between measures not concurrently used in organ allocation, such as between the four medical urgency status groups used before 2002 (1, 2A, 2B, 3) and the more continuous computed MELD scores for liver recipients used since then.

The PLT and patient matching

The SRTR-ESRD PLT was developed by the SRTR to provide a central repository for patient identifying data from various sources and to provide a common patient identifier that can be used to link patient data across those sources. The records in the PLT include persons found in primary OPTN data, as well as those found in Medicare data about patients with ESRD. There is a large overlap in the population covered by these two databases, as kidneys account for about two-thirds of the transplant

candidates and recipients in the SRTR database. Because most ESRD patients qualify for Medicare benefits, most kidney transplant recipients and candidates (usually on dialysis) also are found in the CMS ESRD data.

The development of the PLT was a collaborative effort of University Renal Research and Education Association (URREA) and the Kidney Epidemiology and Cost Center (KECC) of the University of Michigan. HRSA, the agency that oversees the OPTN and SRTR, and the Centers for Medicare and Medicaid Services (CMS) have an inter-agency agreement for sharing organ transplantation data. Under this agreement, CMS discontinued its separate collection of kidney transplant data, the OPTN became CMS's source for transplant data, and HRSA gained access to the Medicare ESRD data. Since 1988, first as the United States Renal Data System (USRDS) and then under various CMS contracts, KECC has developed and maintained a database that integrates almost all of the CMS data on ESRD patients. URREA and KECC, as the SRTR, have integrated transplant patients into this common database by matching to existing patients where applicable, and by adding records for transplant patients not already in the ESRD portion of the database.

The PLT data are organized around people, rather than around organs, diseases, or events. These people are the set of donors and candidates; some candidates become transplant recipients, and some donors may become candidates themselves. At the basic unit of a person, the SRTR assembles information from a variety of sources:

- candidate, donor, and transplant information (including follow-up) collected by the OPTN;
- mortality and dialysis information from CMS for ESRD patients abstracted from institutional and physician/supplier claims, medical evidence forms, and death notifications;
- death information from the SSDMF;
- death information from the NDI.

To handle incomplete or erroneous identifiers in the diverse data sources used, patients are added to the PLT using a 'fuzzy' matching system that considers SSNs, names and nicknames, dates of birth, and other identifying information (e.g. gender, transplant dates, and death dates)—all with allowances for common coding mistakes such as transpositions or entry of the wrong birth year. For example, the first two records listed in Table 3 would be linked as the same person because of the similar name and SSN, along with date-of-birth evidence. However, the third person, perhaps a family member using the same Medicare billing number, receives a distinct patient identifier on the basis of conflicting evidence, despite having the same SSN and last name.

Table 3: Sample records for sorting into PLT

Name	SSN	Date of birth	Source	Person_ID
Joe Smith	123-45-6789	03-06-1968	KI candidate	1
Joseph Smyth	123-46-5789	03-06-1968	Living KI recip.	1
Lynda Smith	123-45-6789	05-12-1965	KI donor	2

Source: SRTR.

Ascertainment of graft and patient survival

The most important use of additional data sources has been in investigating the completeness of mortality data reported by transplant centers to the OPTN. In recent years, the reliability of such figures as the center-specific post-transplant survival calculations published by the SRTR has been called into question (2,3) because some centers have had poor return rates for post-transplant follow-up forms. Complete ascertainment of mortality is imperative for comparing post-transplant outcomes to the outcomes of those on dialysis or the waiting list. It is important to use multiple sources because no single data source is complete by itself, and because data submitted directly by transplant centers are subject to bias in reporting, either toward or away from sicker patients. For example, a center might have more contact with sicker patients, thus making it easier to report on them; on the other hand, it is possible that some centers could lose track of these patients more easily—or some might even attempt to 'fool the system' by underreporting patients with poor outcomes.

Linking within OPTN data

Although this is not a data source that is 'external' to the OPTN data collection system, a modification to the data from the structure established for organ allocation can be useful for research. This modification is made possible by the central organization of a patient record and common patient identifier in the PLT table. Within the organ allocation and data collection database, each waiting list registration and transplant is treated as a separate entity, as linkage is not necessary for allocation.

For patients with multiple waiting-list candidacies or multiple transplants, crucial data such as the patient's death date may be reported for only the last candidacy or transplant. The common patient identifier allows data for the same patient to be linked together within the SRTR database. For a patient with multiple transplants or candidacies, this allows a death date reported in follow-up for the last transplant or reported on a final candidacy after graft failure to be made available when analyzing any of the previous transplants. Within the OPTN database, linkage across multiple listings or transplants is accomplished primarily through SSN.

SSDMF

The SSDMF, publicly available from the Social Security Administration (SSA), contains over 70 million records

created from reports of death to the SSA. Records are reported for both beneficiaries and nonbeneficiaries; 90% are reported by family members and funeral homes, the remainder are reported by state and federal agencies, banking institutions, postal authorities, etc. This file includes the following information on each decedent: SSN, name, date of birth, date of death, ZIP code of last residence, and ZIP code of lump sum payment. Because it may miss some nonbeneficiaries, the absence of a particular person in this file does not prove the person is alive, and the deaths of children are more likely to be missing. Of the deaths included in the SSDMF, more than 98% are complete by the end of the third month after a death date.

Every month, the SRTR adds new information from the SSDMF into the patient table. For each patient in the PLT, the SRTR looks up the SSN for that patient in the SSDMF. When found, the names and birth dates are checked before the SSDMF death date is recorded in the patient table.

CMS ESRD database

Medicare data, described above in relation to the PLT file, provide an additional source of death data for ESRD patients. They also can provide pretransplant dialysis history and a source for inferring graft failure from return to dialysis. Because of Medicare rules, most of these data center on ESRD patients, though data can also be obtained for any patients in the PLT with failure of other organs who appear in the Medicare data.

The Renal Beneficiary and Utilization System (REBUS) system at CMS is the primary CMS ESRD database, and includes data from a number of sources that are useful in organ transplantation research. REBUS obtains death dates for beneficiaries from the Medicare Enrollment Database (EDB), as well as from the ESRD Death Notification Form, which includes the cause of death. As a source of dialysis history, the ESRD Medical Evidence Report is filed for all patients starting dialysis, certifying that a patient has ESRD and indicating the cause of ESRD and the date of first dialysis. This form may also indicate the date of a transplant, the date of return to dialysis after a transplant, and the date of death. Since 1995, dialysis facilities have been required to complete this form for all new dialysis patients, not just those eligible for Medicare.

In addition to these forms, detailed Medicare claims data are obtained separately from REBUS and are updated annually. These claims data are another source of date of death, date of first dialysis, and the date of return to dialysis after a transplant.

NDI

Compiled by the National Center for Health Statistics (NCHS), the NDI contains data from death certificate

information submitted by state vital statistics agencies. Researchers may use this file to determine whether subjects have died and to facilitate obtaining actual death certificates from the state agencies. Researchers may submit a list of subjects to NCHS, which in turn matches with the NDI using a 'fuzzy' matching algorithm similar to that described above for the PLT. Resulting match possibilities are returned to the researcher, who makes the final decision about the quality of each match.

While the NDI is the most complete source of death data used by the SRTR (missing approximately 5% of deaths in the United States), it has a number of significant limitations. First, the NDI is updated only annually. Taking into account the time for NCHS to process the death certificates and run matches, the reporting time lag is 1–2 years after the death date. Fees for NDI matching are also substantial.

A second significant limitation is a restriction on how NDI data may be used. Agreements between the NCHS and the state agencies that collect the death certificates prohibit using the data for administrative or regulatory purposes. This means that while these data may be used for national mortality figures, they may not be reported back to transplant centers or be used for center-specific reports.

The OPTN and SRTR have carried out a test of the usefulness of the NDI for supplementing and benchmarking the completeness of the OPTN death data and other available sources of death data. The OPTN prepared a file of patients for whom the OPTN has no data since 1999 and who were alive at the last known time point. This file was matched against all years of the NDI. The results of this exercise are included in the discussion of all extra mortality sources below.

Implications of secondary sources for mortality

The OPTN data alone capture most of the deaths among patients in the SRTR database, and some deaths are captured only by the OPTN data, especially when multiple records within the OPTN data are linked and considered. The SSDMF and ESRD sources provide important additional coverage at low cost. The NDI provides some additional coverage, although at higher cost and with a longer time lag. Table 4 shows the frequency of update, usual reporting lag, and cost associated with these various sources of death ascertainment. Table 5 shows the contribution made by each of these sources to the ascertainment of deaths among transplanted patients. For most patients, death dates are found in more than one source; in these cases, the sources are checked in the order in which they appear (from left to right) in Table 5:

- *OPTN Primary* (death reported with the first transplant recorded);

- *OPTN Secondary* (death ascertained from a subsequent waiting list registration or transplant);
- SSDMF;
- CMS ESRD;
- *NDI* (searched last because it is the most difficult, restricted, and expensive source).

Any patient with an 'OPTN Primary' death date is classified with that source. If the patient does not have an 'OPTN Primary' death date, then the other sources are checked in the indicated order until a date is found, and each death is attributed to only one source. For example, if a person's death date is found only in the SSDMF and CMS ESRD, then the patient is classified with SSDMF. The 'contribution' of a source is the proportion of all

Table 4: Additional sources of transplant outcome data

Source of death data	Frequency of SRTR update	Reporting lag after death	Added cost	Used in 2002 Annual Report?
OPTN data	Monthly	1–15 months after death; may not be reported until next annual follow-up form	None	Yes
CMS ESRD data	Monthly	1–6 months	None	No
SSDMF	Monthly	3 months	Low	Yes
NDI	Yearly	1–2 years	High	No

Source: SRTR.

Table 5: Distribution of deaths from 1991 to 1999 among transplant recipients by source of death date, organ, survival time after first transplant, and patient age at death

	Deaths	Source of death date				
		Primary OPTN	Secondary OPTN	SSDMF	CMS ESRD	NDI
All	45561	77.3%	6.9%	14.3%	0.7%	0.8%
Kidney and pancreas (K/P)						
All	25859	68.9%	6.0%	22.9%	1.3%	0.9%
Kidney	24607	68.1%	6.1%	23.6%	1.4%	0.9%
Pancreas	69	60.9%	17.4%	17.4%	1.4%	2.9%
Kidney–pancreas	1183	87.2%	3.6%	8.8%	0.2%	0.3%
Non-K/P organs						
All	19702	88.3%	8.0%	3.0%	0.0%	0.6%
Liver	8412	81.5%	14.6%	3.1%	0.0%	0.8%
Intestine	173	82.1%	13.3%	2.9%	0.0%	1.7%
Heart	7649	93.0%	2.8%	3.6%	0.0%	0.6%
Lung	3147	94.7%	3.1%	1.8%	0.0%	0.3%
Heart–lung	321	94.1%	4.0%	1.2%	0.0%	0.6%
Kidney and pancreas						
Died within 1 year of transplant	4504	94.1%	1.4%	4.3%	0.1%	0.2%
Died 1–2 years after transplant	2157	82.2%	3.3%	13.4%	0.6%	0.6%
Died 2–3 years after transplant	2288	76.0%	4.3%	18.1%	0.9%	0.8%
Died 3–4 years after transplant	2459	71.6%	5.4%	20.9%	1.3%	0.9%
Died 4–5 years after transplant	2370	65.5%	6.8%	25.4%	1.0%	1.3%
Died ≥ 5 years after transplant	12081	56.0%	8.5%	32.4%	2.0%	1.1%
Non-K/P organs						
Died within 1 year of transplant	9651	90.3%	9.0%	0.5%	0.0%	0.1%
Died 1–2 years after transplant	2401	90.4%	7.4%	1.7%	0.0%	0.5%
Died 2–3 years after transplant	1737	88.7%	7.8%	2.8%	0.0%	0.7%
Died 3–4 years after transplant	1437	88.4%	6.4%	4.6%	0.0%	0.6%
Died 4–5 years after transplant	1177	86.5%	5.9%	5.6%	0.0%	2.0%
Died ≥ 5 years after transplant	3299	81.3%	7.1%	9.9%	0.0%	1.7%
Kidney and pancreas						
Age ≥ 21 years	25458	68.8%	5.9%	23.1%	1.3%	0.9%
Age < 21 years	401	76.8%	11.7%	9.7%	0.5%	1.2%
Non-K/P organs						
Age ≥ 21 years	17642	88.7%	7.3%	3.3%	0.0%	0.7%
Age < 21 years	2060	84.6%	14.0%	0.7%	0.0%	0.6%

Source: SRTR data analyses, August 2002.

deaths that were identified by this source 'first', and will depend on the order chosen. Therefore, a small contribution from a secondary source (the CMS ESRD data, for example) does not mean that that source identifies few deaths; it may simply identify the same deaths as sources searched earlier.

Table 5 reports on transplant recipient deaths identified by any of the sources and occurring from 1991 through 1999. This range of years was chosen because 1999 was the last year for which the NDI was searched. For patients who have had more than one transplant, transplant date (for computing survival time) and organ are determined from the first transplant. Statistics for kidney and pancreas patients are reported separately from those receiving other organs because the data differ substantially for the two groups. These differences are due to the existence of an alternative treatment (dialysis) for kidney failure, differences in data collection (e.g. OPTN-based follow-up only for 2 years after graft failure), and the availability of alternative sources of information (CMS ESRD).

The OPTN data provided information on only 75% of the deaths for kidneys and pancreata (K/P) but 96% of deaths for all other organs. However, for deaths in the first year after transplant, the OPTN data cover 99% of the non-K/P deaths and 95% of the K/P deaths. This explains in part the result reported in the 'Analytical Approaches' article that, for many transplant programs, center-specific survival is diminished little, if not improved, when SSDMF data are considered. Thus for 1-year survival, the OPTN data are quite good for the nation as a whole, but the remaining sources are particularly important for longer follow-up times.

The contribution of the SSDMF increases steadily as the survival period increases. For non-K/P organs, the contribution of the SSDMF rises to 6% after 5 years. For K/P organs, the increase is much more rapid, rising to 13% for 1–2 years following transplant and exceeding 32% for 5 and more years. The rise in the SSDMF contribution as survival time increases suggests that the transplant centers lose contact with patients as the time since transplant increases, and the higher percentages for K/P organs suggest that this happens even more rapidly for K/P recipients. This difference presumably occurs because dialysis is available as a treatment after a kidney graft failure, while transplantation is the only definitive treatment available for the failure of most other organs. Thus kidney recipients may be more likely to move out of the transplantation system and are less likely to be followed by a transplant center.

As expected, the CMS ESRD data contributed no deaths to the organs other than kidney and pancreas. Even with kidney and pancreas, the incremental contribution of the CMS ESRD data is only 1%. The NDI makes an even smaller contribution of 0.8%, or 352 deaths out of

45561. It thus appears that the combination of the OPTN data and the SSDMF does a very good job of identifying deaths.

Table 5 also shows results for two age groups, divided at age 21. For both organ groups, secondary OPTN sources contribute almost twice as much in the younger group than in the older group. This may be because younger patients are better candidates for a retransplant after a graft failure and thus are more likely to be relisted and retransplanted. The SSDMF has a much larger contribution in the older group than in the younger, although the SSDMF still contributes 9.7% of the deaths among the younger kidney and pancreas recipients. This may be a combined effect of the SSA covering more of the older patients and the OPTN data sources covering more of the younger patients.

When examined in this order, the CMS ESRD and NDI sources each contribute less than 1% of the deaths: 336 CMS ESRD deaths and 352 NDI deaths out of a total of 45561. The largest contribution of the NDI for a subgroup is only 2%. We expected the NDI to make a larger contribution among younger patients on the assumption that the SSDMF would miss many younger patients, but the contribution in this group is minimal. It is not clear whether the other sources catch most of the deaths in this group or whether the NDI also is missing deaths among younger patients.

So far, we have shown that overall ascertainment of mortality looks good when all sources are considered. Next we address the question of whether all sources are necessary. Specifically, if we have good mortality data from secondary sources, how important is the OPTN membership as a data source for mortality?

Table 6 is similar to Table 5 but orders the death sources differently in order to show the deaths uniquely contributed by the OPTN data after deaths from the SSDMF and CMS ESRD data have been counted. When examined in this order, the OPTN data contribute 14% of the deaths. For kidney and pancreas, the OPTN data contribute only 5% for patients aged 21 and over, but they contribute 27% for patients under 21. For other organs, the OPTN data contribute 21% for the older group and 73% for the younger group. While these contributions decline with time, for deaths 5 or more years after transplant the percentages are still 17% for organs other than kidney and pancreas.

We conclude that at the national level, the OPTN data are very complete for 1-year survival, and that the SSDMF and CMS data are important for longer-term survival analyses, particularly for kidneys. Using the NDI is probably not worth the additional expense. While we do not know what proportion of actual deaths is missed by all these sources taken together, the fact that the two sources

Table 6: Distribution of deaths from 1991 to 1999 among transplant recipients by source of death date, time after first transplant, and patient age at death, with alternate ordering of sources to show unique contribution of OPTN data

	Deaths	Source of death date				
		SSDMF	CMS ESRD	Primary OPTN	Secondary OPTN	NDI
All	45 561	82.2%	2.8%	12.7%	1.5%	0.8%
Kidney and pancreas (K/P)	25 859	89.6%	4.8%	4.3%	0.4%	0.9%
Non-K/P organs	19 702	72.5%	0.2%	23.8%	2.9%	0.6%
Kidney and pancreas (K/P)						
Age ≥ 21 years	25 458	89.9%	4.8%	4.0%	0.4%	0.9%
Age < 21 years	401	71.1%	4.0%	20.7%	3.0%	1.2%
Non-K/P organs						
Age ≥ 21 years	17 642	77.8%	0.2%	19.2%	2.1%	0.7%
Age < 21 years	2060	26.7%	0.0%	62.8%	9.8%	0.6%
Kidney and pancreas (K/P)						
Died within 1 year of transplant	4504	89.8%	3.8%	5.6%	0.6%	0.2%
Died ≥ 5 years after transplant	12 081	88.7%	5.8%	4.0%	0.4%	1.1%
Non-K/P organs						
Died within 1 year of transplant	9651	67.9%	0.0%	28.3%	3.7%	0.1%
Died ≥ 5 years after transplant	3299	80.4%	0.6%	14.9%	2.3%	1.7%

Source: SRTR data analyses, August 2002.

added last contribute so few additional deaths suggests that a satisfactory fraction of deaths is found. and finally, because the SSDMF and OPTN each contribute a unique set of deaths, it is important to avoid relying on only one or the other.

The 'Analytical Approaches' article in the Annual Report discusses the use of the SSDMF in survival analyses. When deaths identified by the SSDMF are added to those identified by the OPTN data, we must also adjust the follow-up time for all patients. If information is only added about persons who die, then death rates will be overstated. The SRTR assumes that with the SSDMF data we know about virtually all of the deaths; a corollary of this approach is to assume that patients survive after transplant until the end of the study period during which we expect each source to capture deaths, unless we know otherwise. Therefore we do not censor patients at the last OPTN follow-up date, instead extending the follow-up time to the end of the study period. This adjustment results in almost no change in survival measures at the national level, even for 5- and 10-year survival. The lack of change even for these longer study periods—in which we have shown that many deaths are missing—suggests that the recipients actually followed by the transplant centers constitute an unbiased sample, and are similar to those patients who are lost to follow-up during the study period. However, at the transplant program level, some programs do show substantially different survival measures when the SSDMF data are added.

Other external sources and strategies

For measures other than mortality and graft failure, several additional data sources may also be incorporated with primary data sources for research on transplantation and

data validation. For example, the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute, one of the most complete sources of information on cancer incidence and survival in the United States, may be incorporated. After testing initial incorporation of SEER data from southeast Michigan, the SRTR hopes to make use of SEER's highly accurate cancer registries both for validating the post-transplant malignancy data reported on follow-up forms to the OPTN, and for gaining more complete information for time periods before the recent inception of malignancy ascertainment by the OPTN.

In some cases, data that are useful to correlate with each other are not collected by the OPTN at the same time. For example, in order to simulate the effects of the recent allocation rule change for livers from an urgency status-based one to MELD, it is necessary to have a period of data collection for which we know both the MELD score and the urgency status for each patient. There was a short period during which both measures were collected, but doing so was voluntary. Therefore, it has been useful to obtain hospital laboratory data for actual candidates on the waiting list, in order to associate an urgency status with a distribution of calculated MELDs. These data have also allowed an earlier look at associations between waiting list and post-transplant outcomes than might have been afforded by waiting for real allocation MELDs; they also allow a comparison of the associations between these outcomes and MELD to the former, more discrete, urgency status system. Going back to Figure 1, these data augment the candidate status history file.

Other external data sources do not necessarily require direct linking with primary source data in order to be

Dickinson et al.

useful. For example, the OPTN, SRTR, and other researchers have investigated methods to make associations between OPO practice patterns and donor procurement, considering the suitability for transplantation of deaths in hospitals served by each OPO. The NCHS provides files that can help tabulate numbers of 'evaluable' deaths (deaths that provide a suitable source of organs, given cause, circumstance, and location of death), as well as demographic data about the deceased.

Finally, the OPTN and SRTR are together investigating the possibility of sampling strategies to maintain or expand the scope of data collection while also decreasing the burden of data collection on the facilities. It is possible that certain research may not require data to be collected regarding all transplant recipients, and perhaps a subset of patients would be selected for an extended follow-up form to cover these areas.

Conclusions

We believe that researchers interested in any aspect of transplantation, from donor recovery to organ allocation to post-transplant survival, will find this article useful. We have shown that a tremendous effort has been in making these data high-quality and well-organized for research at the SRTR, OPTN, and among other researchers. Further, we have shown that these efforts have paid off. For many research questions, the data submitted to the OPTN are complete and of high quality; for other questions, secondary sources are easily integrated to improve data quality or expand data scope. These resources taken together pro-

vide a rich and accurate source of information about the transplant process.

Even the extensive effort of the OPTN and SRTR staff at ensuring high-quality and well-organized data for research pales in comparison to the resources devoted to data submission on the part of staff at transplant centers and OPOs. These OPTN members understand that improving patients' lives is an incremental process, the benefits of which may be long in being realized, and which often begins with ensuring that the information is available upon which to reach sound scientific conclusions. None of this rich source of data would be possible without these tireless efforts.

Acknowledgments

The authors gratefully acknowledge the insight and help of Paula Bryant, Jack Koller, and Chris Williams from OPTN/UNOS and Greg Levine, Shannon Li, and James Welch from SRTR/URREA.

References

- 1 Wolfe RA, Webb RL, Dickinson DM, Ashby VB, Dykstra DM, Hulbert-Shearon TE, McCullough KP. Analytical approaches for transplant transplantation. *Am J Transplant* 2003; 3(Suppl. 4): 103–113
- 2 Marchione M. Transplant rate reports don't tell whole story. *Milwaukee J Sentinel* 2001; July 27: G1.
- 3 Cooper L. Survival data: do the numbers really mean anything? *Transplant News Issues* 2001; 2(2): s9–s11, s13.