Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Am J Transplant. Author manuscript; available in PMC 2015 April 07.

Published in final edited form as:

Am J Transplant. 2014 August ; 14(8): 1723–1730. doi:10.1111/ajt.12777.

Big Data in Organ Transplantation: Registries and Administrative Claims

Allan B. Massie, PhD, MHS[1,2], **Lauren Kucirka, ScM**[1,2], and **Dorry L. Segev, MD, PhD**[1,2]

Allan B. Massie: amassie@jhsph.edu

[1]Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD

[2]Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD

Abstract

The field of organ transplantation benefits from large, comprehensive, transplant-specific national datasets available to researchers. In addition to the widely-used OPTN-based registries (the UNOS and SRTR datasets) and USRDS datasets, there are other publicly available national datasets, not specific to transplantation, which have historically been underutilized in the field of transplantation. Of particular interest are the Nationwide Inpatient Sample (NIS) and State Inpatient Databases (SID), produced by the Agency for Healthcare Research and Quality (AHRQ). The United States Renal Data System (USRDS) database provides extensive data relevant to studies of kidney transplantation. Linkage of publicly available datasets to external data sources such as private claims or pharmacy data provides further resources for registry-based research. Although these resources can transcend some limitations of OPTN-based registry data, they come with their own limitations, which must be understood to avoid biased inference. This review discusses different registry-based data sources available in the United States, as well as the proper design and conduct of registry-based research.

Keywords

registry-based studies; retrospective studies; UNOS; SRTR; NIS

INTRODUCTION

Studies based on national registries and other administrative datasets have made enormous contributions to the field of organ transplantation. Registry-based studies offer a number of advantages over clinical trials or prospective cohort studies. They are relatively quick and inexpensive to conduct, and ethical approval is often straightforward. Typically, registries allow for many more subjects than would be feasible with primary data collection; the larger sample size and multicenter nature of many registries enhance study power and allow researchers to conduct sophisticated multivariate or multilevel analyses. Registries also draw

Contact Information: Dorry Segev, M.D., Ph.D., Associate Vice-Chair of Research, Department of Surgery, Johns Hopkins Medical Institutions, 720 Rutland Ave, Ross 34, Baltimore, MD 21205, 410-502-6115 (tel) 410-614-2079 (fax), dorry@jhmi.edu.

DISCLOSURE

The authors of this manuscript have no conflicts of interest to disclose as described by the American Journal of Transplantation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

from more transplant centers than would be feasible in most cohort studies (in particular at small or non-academic centers), meaning that inferences from registry-based studies are likely to generalize across the United States.

Historically, most registry-based studies in organ transplantation have used data collected by the Organ Procurement and Transplantation Network (OPTN). However, other large datasets exist that contain transplant-related data not available in OPTN-based datasets. Additionally, linkages of OPTN-based datasets to novel data sources can address questions which may not be answerable from OPTN data alone.

This review will explore various national datasets in the context of their applicability to transplant research. For each dataset, we will outline the data provided, identify key strengths and limitations, provide illustrative examples of research using the data, and discuss relevant analytical and study design considerations. Additionally, we will discuss the proper design and conduct of registry-based studies.

OPTN-BASED DATA SOURCES

Since 1987, OPTN has collected data on all transplant recipients and waitlist registrants for solid organ transplantation, as well as all live and deceased organ donors. Unique data collection forms exist for each organ, and separate forms exist for adult and pediatric patients (1). Most data are collected via one of three forms:

- The Transplant Candidate Registration (TCR) form includes information at the time of listing: demographic data (e.g. age at listing, race, gender); prior transplant history; basic clinical information (e.g. height, weight, ABO); comorbidities (e.g. diabetes, peptic ulcer, angina); and organ-specific information (e.g. exhausted access for kidney; portal vein thrombosis and TIPSS for liver). A TCR is completed for every waitlist registration; if a patient registers twice (for the same organ after organ failure, for the same organ at two different centers, or for a different organ), two TCRs are completed.
- The Transplant Recipient Registration (TRR) form includes information from the initial transplant admission: pre-transplant clinical data (e.g. height, weight, functional status); infectious disease status (e.g. HIV, CMV, EBV); data on the transplant procedure (e.g. cold ischemia time (CIT), procedure type); post-transplant clinical data (e.g. acute rejection during the initial hospitalization, creatinine at discharge for kidney, bilirubin and INR for liver); and information on immunosuppressive medications. A TRR is completed for every live-donor and deceased-donor transplant; if a patient receives several transplants over time, a TRR is completed for each transplant.
- The Transplant Recipient Follow-up (TRF) form includes information at each visit following a transplant: vital status, cause of death if applicable, graft status, patient education and employment status, and clinical information (e.g. height and weight, infectious disease detection). Theoretically, a TRF is completed for every surviving transplant recipient at six and twelve months post-transplant, and at twelve-month intervals thereafter until the organ fails or the patient dies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

- The Deceased Donor Registration (DDR) (submitted by the OPO) and Living Donor Registration (LDR) forms (submitted by the hospital performing the donor operation) include information at the time of organ donation: donor demographics, comorbidities, infectious disease status, and cause of death (for deceased donors) or post-operative clinical information (for live donors).

Additional forms include the Living Donor Follow-up form, the Donor Histocompatibility form, and the Post-Transplant Malignancy form. All forms are available at http://www.unos.org/.

In addition to the forms described above, the OPTN records waitlist status updates (e.g. MELD score changes for liver waitlist registrants, waitlist removals, status changes from active to/from inactive) and data generated through the organ allocation process (match runs including organ acceptance and/or decline).

The United Network for Organ Sharing (UNOS)

The OPTN data are linked by UNOS to the Social Security Death Master File to augment ascertainment of candidate and recipient death. The resulting data are available free of charge to researchers, and have been used in numerous important studies of transplantation (2–4).

The Scientific Registry of Transplant Recipients (SRTR)

The SRTR supplements OPTN data with data from various secondary sources. Notably, the SRTR obtains additional ascertainment of graft failure and death from the Centers for Medicare and Medicaid Services (CMS), cancer ascertainment from the Surveillance, Epidemiology, and End Results program (SEER), and additional death ascertainment from the National Death Index (NDI). The SRTR dataset is used to compile SRTR program-specific reports (5, 6) and has provided data for many high-impact papers in organ transplantation (7–9). The SRTR provides standard analysis files (SAFs) to researchers by request, for a fee. The markedly improved ascertainment of kidney graft loss is a key difference between SRTR and UNOS data, and constitutes a strong argument for using SRTR (rather than UNOS) data for analysis of kidney transplant outcomes: a 2005 study found that, of 4040 graft failures reported by either the OPTN or CMS, 22% were reported only by the OPTN and 13% were reported only by CMS (1).

The United States Renal Data System (USRDS)

The United States Renal Data System (USRDS) includes data on all patients in the United States who developed end-stage renal disease requiring renal replacement therapy- either dialysis or a kidney transplant- since 1995. In contrast to the OPTN data, which includes only transplant recipients and waitlist registrants, the USRDS dataset contains data on patients irrespective of access to transplantation. Data from 1988–1994 are available but include only patients insured by Medicare. Data are drawn from a variety of sources, including the Centers for Medicare and Medicaid Services (CMS), OPTN, the ESRD Networks, and USRDS special studies (10).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Providers are required to file the CMS Medical Evidence Report (Form-2728) within 45 days of ESRD onset. This form captures demographics, insurance coverage, primary cause of renal failure, dialysis type, dialysis access type, laboratory values (e.g. Hba1c, creatinine), 20 comorbidities (e.g. COPD, diabetes, MI), functional status, and access to nephrology care (10). There are, however, several limitations. Unfortunately, there are no data on severity of comorbidities, and validation studies have shown low sensitivity for some comorbidities (11). Furthermore, for most patients this form is only filed at ESRD onset, so changes over time cannot be assessed.

Place, time, and cause of death are ascertained for all patients via the CMS ESRD Death Notification (Form-2746), required to be filed by the provider within 45 days of a patient death. Other outcomes are ascertained from OPTN data and include listing for transplantation, receipt of a transplant, and graft loss. OPTN data are included in the Standard Analysis Files; researchers can obtain additional transplant analytic files which include the UNOS kidney and kidney pancreas transplant follow-up datasets linked to USRDS (10).

For the subset of ESRD patients insured through Medicare, claims data are available and capture detailed longitudinal information on diagnoses, comorbidities (via ICD-9 codes), treatment modalities, and cost. Information on hospitalizations are derived from institutional claims, allowing for analyses of hospital readmissions (12, 13). Beginning in 2006, datasets including detailed prescription drug information are available from Medicare Part D. In addition to Medicare claims data for ESRD patients, the USRDS dataset also contains all claims from a randomly chosen 5% sample of Medicare participants irrespective of ESRD, allowing researchers to study relevant outcomes (e.g. CKD or progression to ESRD) in the general Medicare population.

A disadvantage of Medicare claims is that inferences may not be generalizable to the non-Medicare population. Importantly, while all patients with ESRD are eligible for Medicare, patients under 65 lose eligibility three years after receipt of a kidney transplant. Furthermore, although claims data can be highly informative and allow longitudinal assessment of comorbidities, they do not perfectly capture them. A 2010 single-center study comparing ascertainment of cardiovascular disease events via USRDS-derived Medicare claims data vs. electronic medical records found that Medicare claims captured 82–91% of events, depending on the algorithm used (14).

In addition to patient-level data, information about each dialysis facility is ascertained annually through a variety of sources including the CMS Facility Compare data, the Independent Renal Facility Cost Report (CMS 265-94), and the CDC National Surveillance of Dialysis Associated Diseases. This can be linked to patient data through Form-2728, which captures dialysis facility at ESRD onset. Examples of facility-level data include volume, number of deaths, facility ownership, chain affiliation, freestanding versus hospital-based, zip code, and staffing (e.g. number of NPs, social workers, etc.) (10)

Researchers have used the USRDS dataset to study a variety of topics including access to transplantation (15, 16), survival (17), complications (18), facility ownership and access to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

transplant (19), and treatment costs (20). USRDS SAFs are available for a fee from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK); in addition to the Core dataset, various SAFs (e.g. Medicare payment data, 5% Medicare sample, transplant dataset) are sold separately (10).

OTHER DATA SOURCES

Pharmacy and Private Claims Data

Some researchers have created novel linkages between OPTN data and external claims data. Due to missingness or inconsistency in linking variables, linkage may require a complex algorithm (21). Private claims provide the same advantages and limitations as discussed above for Medicare claims, albeit with different potential selection biases. Pharmacy claims potentially provide better ascertainment of immunosuppression and other post-transplant medication than OPTN data, but come with their own challenges and limitations. Linkages to private payer claims data have been used in studies of costs of liver transplantation (22) and post-donation morbidity in live kidney donors (23).

Nationwide Inpatient Sample (NIS)

The Nationwide Inpatient Sample (NIS), maintained by the Agency for Healthcare Research and Quality (AHRQ), contains data on hospitalizations at about 1,000 hospitals across the United States, comprising a twenty-percent sample of hospitals (24, 25). The sample varies from year to year. Sampling is stratified across five criteria (geographic region, public vs. private, urban vs. rural, teaching vs. non-teaching, and bed size). Data available through the NIS include patient demographics; ICD-9-CM diagnoses and procedures; hospital charges and length of stay; discharge disposition; anonymized physician and hospital identifiers; and hospital characteristics (e.g. geographic region, teaching status, bed size). Much of the data in the NIS (notably charges, physician IDs, and many diagnoses/comorbidities) are unavailable in OPTN data. In contrast with Medicare claims data, the NIS includes patients with public and private payers, as well as uninsured patients. Also, since the NIS is a sampling of all hospitalizations, it includes patients who are not on a transplant waiting list but may nevertheless be of interest for transplant-related research questions.

However, NIS data have several important limitations. First, despite the stratified sampling mechanism used to create the NIS, a random sample of all hospitals is not necessarily an unbiased sample of transplant centers or transplant patients. For example, the total number of kidney transplants in the NIS sample rose from 2967 (at 47 distinct centers) in 2007 to 4119 (at 42 distinct centers) in 2008, even though the total number of transplants in the United States showed little increase (OPTN data). Also, NIS data lack longitudinal information (i.e. a single individual cannot be tracked across multiple hospitalizations). Therefore, NIS-based studies are limited to short-term, same-hospitalization outcomes such as cost, complications, or perioperative mortality. Finally, because identifiers cannot be released, NIS cannot be linked to OPTN data.

Despite these limitations, NIS data have been used to examine exposures unavailable in OPTN data (including hospital and surgeon characteristics (26, 27) and *C. difficile* exposure (28)); complication and cost outcomes (28, 29); potential deceased donors who are ineligible

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

for donation due to HIV (30); and insurance status of deceased donors (31). NIS files may be purchased from AHRQ.

State Inpatient Databases (SID)

The State Inpatient Databases (SID) contain data on hospital admissions from individual states. Forty-seven states (excluding Alabama, Delaware, and Idaho) participate in the SID (32). They contain many of the same elements as the NIS, and in fact the NIS sample is drawn from the SID. Unlike the NIS data, which represent a 20% sample of hospitals in the United States, the SID are comprehensive (100%) for the states and years for which they are available. Also, for some states and some years, “revisit files” allow researchers to link multiple hospital admissions for a single individual (24), and some states provide anonymous physician identifiers, allowing researchers to link patients in multiple hospitals treated by the same physician (33). While national studies using the SID would be possible in principle, they would be expensive and logistically complex, since data for each state and year must be purchased separately and data use agreements must be negotiated separately.

Data from the SIDs have been used to study perioperative complications of live liver donors in New York (34) and the relationship between hospital/surgeon volume and inpatient mortality in liver resection and transplantation in Maryland, Florida and New York (33). SID files may be purchased from the HCUP.

Additional data sources

The University Health Consortium (UHC) is an alliance of 120 academic medical centers and 290 affiliated hospitals in the United States. The UHC maintains a database, available to member institutions, of deidentified patient data, including patient demographics, ICD-9 diagnosis and procedure codes, and billing and cost data (35). UHC data have been used to study perioperative complications in live liver donation (34) and linked to OPTN data in studies of costs of liver transplantation (35, 36). Data are available from the University Health Consortium, http://www.uhc.edu/.

Even without linkage, other national datasets can be used as negative controls in comparison to transplant data. For example, we have previously compared long-term survival in live kidney donors (from OPTN data) to matched, healthy non-donor controls drawn from the third National Health and Nutrition Examination Survey (NHANES III) (4, 37).

A summary of the advantages and disadvantages of the datasets described above appears in Table 1.

THE CONDUCT OF REGISTRY-BASED STUDIES

By the time a clinical trial or other prospective study begins, the investigators have already had to design a protocol and justify it to an institutional review board (and, likely, to funders). This process helps researchers carefully consider the questions they wish to address and the appropriateness of their methods, reducing the likelihood that severe design flaws will derail an expensive study or lead to biased inference.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

With registry-based studies, barriers to the conduct of research are much lower. The data have already been gathered and are available at the start of the study. Even if investigators wrote a research protocol at the outset, they face no technical barriers in modifying analytical plans or investigating new questions on the fly. However, careful, *a priori* study design is as important for conducting proper analysis of registry data as it is for prospective studies or clinical trials. We will now consider general concepts and common pitfalls of study design and analytical approach in the context of registry-based studies.

Posing a research question

Proper study design begins with a well-articulated hypothesis. A researcher may hypothesize, for example, that prolonged cold ischemia time is associated with increased risk of graft loss in deceased donor transplants. The hypothesis may come from existing knowledge of biological processes, from clinical observation, or from a research finding in another field. In a prospective study, the hypothesis cannot come from the data itself.

Large registries can place thousands of variables at an investigator's fingertips. A few hours of clever coding could automatically examine pairwise correlations among all the variables in the OPTN database. However, many seemingly high correlations are inevitably spurious, a result of either confounding or statistical chance. Investigators should resist the temptation to fish for statistical associations in the absence of a biologically plausible hypothesis. Modern statistical packages generally include commands to perform stepwise regression, a set of techniques which essentially tests for statistical association at random from a list of possible exposures; in most situations, stepwise regression is best avoided.

Population selection

A well-defined research question inherently addresses a specific population. For example, an investigation of the association between CIT and graft loss in deceased donor kidney transplant (DDKT) recipients addresses the population of DDKT recipients. The study design should reflect the population of interest. There are three populations for the researcher to consider: the *target population*, the *source population*, and the *study population* (38). The target population refers to a category of patients about whom the researcher hopes the study will give valid inference: for example, adult DDKT recipients, as above. The individual membership of the target population inherently cannot be specified; generally, the goal of clinical research is to provide insight into disease processes or treatments of *future* patients.

By contrast, the *source population* is a specific, enumerable set of individuals with the characteristics that define the target population. For example, a researcher might choose all adult, first-time, deceased donor kidney-only recipients from 2005–2012 appearing in the SRTR registry. The *study population* consists of individuals who are actually included in a study. In a prospective study, individuals eligible for inclusion may not be contacted, or may refuse consent; such individuals fall in the source population, but not the study population. In registry studies, the source population and study population are generally the same, unless the study design calls for using data from only a subset of eligible individuals (e.g. in a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

matched design, the matching algorithm may select only a subset of study patients from the source population) (4, 39).

In selecting a source population, the researcher must strike a balance between including a broad range of patients representative of the target population, and excluding atypical individuals whose outcomes might bias the results. For example, analyses of the general waitlist or transplant population sometimes exclude pediatric patients, patients with a prior history of transplant, and/or multiorgan registrants/recipients in order to describe the experience of a "typical" adult patient (8, 40); consequently, analyses of these populations may not generalize to the excluded groups. Exact inclusion/exclusion criteria will depend on the nature of the research question, but typically, criteria will include at least an age range (e.g. adult patients), a date range (e.g. transplants from 2005–2010), and an organ/procedure type (e.g. kidney-only deceased donor transplant recipients).

A study will suffer from *selection bias* if individuals in the study population are not representative of the target population, and this difference affects inference. In a comprehensive registry of transplant recipients, selection bias will not be a problem unless investigator-specified inclusion criteria are flawed. However, in registries that include only a subset of patients (e.g. the NIS 20% sample, or Medicare claims, which contain data only on Medicare patients), some selection bias may be inherent in the dataset.

Data quality

In prospective studies, investigators work to ensure standardization of measurement, data collection, and data entry. Investigators of registry studies do not have this luxury. Registry data may be collected at hundreds of different transplant centers, by thousands of individuals. Often, data are not gathered primarily for research purposes, but rather in the course of clinical care, billing, or regulation. As a result, missing data and mismeasured data are realities of most registries. Careful exploratory data analyses of key variables are necessary to identify potential threats to data quality, and proper analytical techniques are required to avoid bias.

Data missingness may arise by design (e.g. in the case of MELD at transplant in liver recipients, which is missing from all OPTN data prior to the introduction of MELD-based allocation in 2002) or because it was not recorded or entered by treatment providers (e.g. CIT, which is missing for 30.3% of live donor transplants between 1990–2005 (41)). In general, when considering OPTN data, those data used for organ allocation and recipient priority determination are generally the least missing, and those used in the PSRs are a close second; all other elements in the OPTN data require careful exploration before trusting them for a research study. Strategies for dealing with missingness include (in order of robustness): casewise deletion (complete-case analyses), missing indicator variables, and imputation (42). Researchers need to be aware that, by default, most statistical software packages address missingness with casewise deletion without warning the user. For example, a study of liver transplant recipients between 1990–2005 that "adjusts for MELD" becomes effectively a study of only those recipients between 2002–2005 (because of the missingness pattern described above), without warning the user that 4/5 of the study population was

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

dropped. Published examples of misled naïveté in the face of missing data are, disappointingly, not uncommon.

Measurement error (and the risk for misclassification bias) can occur for a variety of reasons: patients may not remember their health history, or may not report their history honestly; lab results may be flawed; providers may make mistakes in recording or entering data. In most statistical techniques, some data points are more *influential* than others, meaning that they have a greater effect on the overall summary statistic (43). This is particularly likely to be true of *outlier* points. For example, if a weight of 250 pounds is mistakenly recorded as 250 kilograms, the erroneous measurement may lead to an artificially high estimate of mean weight among a group of patients. Measurement error can be difficult to detect, but researchers should examine the distribution of key variables, particularly over time. While not all outliers are influential in a statistical model, researchers should carefully consider values which fall outside of the normal range, and how influential these observations are (a determination for which statistical methods exist).

Statistical models

Modern statistical software largely automates the process of performing hundreds of statistical analyses, from simple analysis such as a t-test to complicated multilevel regression models. However, these tools make it easy for researchers to ignore the mathematical assumptions underlying statistical models. For example, linear regression assumes the expected value of an outcome variable Y varies linearly with each exposure variable x; that variance in the outcome is constant across all Y; and that residuals are normally and independently distributed (43). If these assumptions are not checked, linear regression may easily lead to mistaken inference. The next section of this review will consider specific statistical models of particular relevance to transplantation.

Survival models

The most common form of outcome in transplantation is *time-to-event* or *survival* outcome, in which patients at risk for an event (e.g. graft failure or death) are followed starting at a defined point (e.g. date of transplant) until either the time of event or end-of-followup (*censorship*). *Non-parametric* models (e.g. Kaplan-Meier curves) make no assumption about the distribution of times to event; they can be fit to any survival dataset (44). *Semi-parametric* models make some assumptions about the distribution of events. For example, Cox proportional hazards models allow the risk of an event over time (the *hazard*) to vary according to the data. However, when two groups of patients are compared, Cox models assume that the relative risk of the event in one group compared to the other group (the *hazard ratio*) stays constant over time. For example, a hazard ratio of 2.0 comparing kidney recipients who had received a prior kidney transplant to first-time kidney recipients implies that the risk of graft failure is twofold higher for retransplant recipients at all times, from the day of transplant through the duration of the graft. Cox proportional hazards models have found wide use in transplantation (7, 8, 17). However, caution is warranted. Some exposures (such as surgery) may increase risk in the short term while decreasing risk in the long term, in which case the assumptions of a Cox model are violated. *Fully parametric* models assume that hazard fits a specific distribution, such as a Weibull distribution (45) or generalized

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

gamma distribution (46). The proportional hazards assumption is not a requirement for some parametric models. All of the models described above assume that the risk of censorship is independent of the risk of the outcome of interest (the assumption of *non-informative censorship*).

Prediction models

Statistical models intended to predict individual outcomes are common in transplantation; examples include the MELD (Model of End-stage Liver Disease) score (47) and the Kidney Donor Risk Index (KDRI) (48). Prediction models require assessment of the model's predictive accuracy. Predictive accuracy can be partitioned into *calibration* (agreement between predicted and observed data) and *discrimination* (chance that an observation with a higher observed value will have a higher predicted value). Calibration can be assessed with Hosmer-Lemeshow tests (49) or calibration plots (50). Discrimination can be assessed with the area under the receiver operating curve (ROC) (49), or with the net reclassification index (NRI) or the integrated discrimination index (IDI) (51). Prediction models should be *validated* by using them to predict data separate from the data used to create the model. This can take the form of *internal validation*, in which a model is validated using additional records from the same dataset used for model building (e.g. cross-validation (52) or a bootstrap (53)), or *external validation*, in which a model is validated using a separate data source (54). An example of using several different methods to validate a prediction model in transplantation is our work on the Probability of Discard/Delay (PODD) model for deceased donor kidneys (55).

Statistical software

A variety of statistical packages are commonly used for statistical analysis of transplant data. SAS (SAS Institute, Cary, NC) has been commercially available since the 1970s and has wide use in health research. SAFs from UNOS and the SRTR are made available in SAS format, although software exists to convert them to other formats. Stata (Statacorp, College Station, TX) is a more recent alternative which has become common in academia. An advantage of Stata is that it is available for a relatively low one-time cost, whereas a SAS license requires users to pay a yearly fee. R (R foundation, Vienna, Austria) is a free, open-source statistical package. All three of the above packages have scripting languages which allow researchers to write complex custom programs. R and Stata both benefit from large libraries of user-written functions which supplement the official programs.

CONCLUSION

Registry-based studies have made substantial contributions to the field of transplantation, and will no doubt continue to do so in the future. In addition to the commonly used OPTN-based datasets, there are a number of other registries which can be used to address a wide variety of research questions beyond the scope of the OPTN data. However, the availability of rich, comprehensive datasets does not obviate the need for careful study design. Properly conducted, registry-based research can provide novel insights to improve patient care and advance our understanding of the field of organ transplantation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Acknowledgments

This work was supported by grant number 1R01DK0960008 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The analyses described here are the responsibility of the authors alone and do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Abbreviations

OPTN	Organ Procurement and Transplantation Network
UNOS	United Network for Organ Sharing
SRTR	Scientific Registry for Transplant Recipients
OPO	Organ Procurement Organization
STAR	Standard Transplant And Analysis
CMS	Centers for Medicare and Medicaid Services
SER	Surveillance, Epidemiology, and End Results
NDI	National Death Index
SAF	Standard Analysis File
USRDS	United States Renal Data System
ESRD	End-Stage Renal Disease
NIDDK	National Institute of Diabetes and Digestive Kidney Diseases
NIS	Nationwide Inpatient Sample
AHRQ	Agency for Healthcare Research and Quality
SID	State Inpatient Database

References

1. Dickinson DM, Dykstra DM, Levine GN, Li S, Welch JC, Webb RL. Transplant data: sources, collection and research considerations, 2004. American Journal of Transplantation. 2005; 5(4p2): 850–861. [PubMed: 15760413]
2. Young BY, Gill J, Huang E, Takemoto SK, Anastasi B, Shah T, et al. Living donor kidney versus simultaneous pancreas-kidney transplant in type I diabetics: an analysis of the OPTN/UNOS database. Clinical journal of the American Society of Nephrology : CJASN. 2009; 4(4):845–852. [PubMed: 19201918]
3. Massie AB, Zeger SL, Montgomery RA, Segev DL. The effects of DonorNet 2007 on kidney distribution equity and efficiency. Am J Transplant. 2009; 9(7):1550–1557. [PubMed: 19459793]
4. Segev DL, Muzaale AD, Caffo BS, Mehta SH, Singer AL, Taranto SE, et al. Perioperative mortality and long-term survival following live kidney donation. JAMA: the journal of the American Medical Association. 2010; 303(10):959–966.
5. Dickinson D, Arrington C, Fant G, Levine G, Schaubel D, Pruett T, et al. SRTR Program-Specific Reports on Outcomes: A Guide for the New Reader. American Journal of Transplantation. 2008; 8(4p2):1012–1026. [PubMed: 18336703]
6. Massie AB, Segev DL. Rates of false flagging due to statistical artifact in CMS evaluations of transplant programs: results of a stochastic simulation. American journal of transplantation : official

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2013; 13(8):2044–2051.

7. Port FK, Bragg-Gresham JL, Metzger RA, Dykstra DM, Gillespie BW, Young EW, et al. Donor characteristics associated with reduced graft survival: an approach to expanding the pool of kidney donors. Transplantation. 2002; 74(9):1281–1286. [PubMed: 12451266]
8. Merion RM, Ashby VB, Wolfe RA, Distant DA, Hulbert-Shearon TE, Metzger RA, et al. Deceased-donor characteristics and the survival benefit of kidney transplantation. JAMA. 2005; 294(21): 2726–2733. [PubMed: 16333008]
9. Schold JD, Kaplan B, Baliga RS, Meier-Kriesche HU. The broad spectrum of quality in deceased donor kidneys. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2005; 5(4 Pt 1):757–765.
10. US Renal Data System. Researcher's Guide to the USRDS Database: 2011 ADR Edition. Bethesda, MD: 2011.
11. Merkin SS, Cavanaugh K, Longenecker JC, Fink NE, Levey AS, Powe NR. Agreement of self-reported comorbid conditions with medical and physician reports varied by disease among end-stage renal disease patients. Journal of clinical epidemiology. 2007; 60(6):634–642. [PubMed: 17493523]
12. McAdams-Demarco MA, Grams ME, Hall EC, Coresh J, Segev DL. Early hospital readmission after kidney transplantation: patient and center-level associations. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2012; 12(12):3283–3288.
13. McAdams-Demarco MA, Grams ME, King E, Desai NM, Segev DL. Sequelae of early hospital readmission after kidney transplantation. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2014; 14(2):397–403.
14. Lentine KL, Schnitzler MA, Abbott KC, Bramesfeld K, Buchanan PM, Brennan DC. Sensitivity of billing claims for cardiovascular disease events among kidney transplant recipients. Clinical journal of the American Society of Nephrology : CJASN. 2009; 4(7):1213–1221. [PubMed: 19541817]
15. Segev DL, Kucirka LM, Oberai PC, Parekh RS, Boulware LE, Powe NR, et al. Age and comorbidities are effect modifiers of gender disparities in renal transplantation. Journal of the American Society of Nephrology : JASN. 2009; 20(3):621–628. [PubMed: 19129311]
16. Abbott KC, Glanton CW, Agodoa LY. Body mass index and enrollment on the renal transplant waiting list in the United States. J Nephrol. 2003; 16(1):40–48. [PubMed: 12649534]
17. Kucirka LM, Grams ME, Lessler J, Hall EC, James N, Massie AB, et al. Association of race and age with survival among patients undergoing dialysis. JAMA : the journal of the American Medical Association. 2011; 306(6):620–626.
18. Abbott KC, Swanson SJ, Richter ER, Bohen EM, Agodoa LY, Peters TG, et al. Late urinary tract infection after renal transplantation in the United States. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2004; 44(2):353–362. [PubMed: 15264195]
19. Garg PP, Frick KD, Diener-West M, Powe NR. Effect of the ownership of dialysis facilities on patients' survival and referral for transplantation. N Engl J Med. 1999; 341(22):1653–1660. [PubMed: 10572154]
20. Schnitzler MA, Gheorghian A, Axelrod D, L'Italien G, Lentine KL. The cost implications of first anniversary renal function after living, standard criteria deceased and expanded criteria deceased donor kidney transplantation. J Med Econ. 2013; 16(1):75–84. [PubMed: 22905738]
21. Gilmore AS, Helderman JH, Ricci JF, Ryskina KL, Feng S, Kang N, et al. Linking the US transplant registry to administrative claims data: expanding the potential of transplant research. Med Care. 2007; 45(6):529–536. [PubMed: 17515780]
22. Buchanan P, Dzebisashvili N, Lentine KL, Axelrod DA, Schnitzler MA, Salvalaggio PR. Liver transplantation cost in the model for end-stage liver disease era: looking beyond the transplant admission. Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society. 2009; 15(10):1270–1277.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

23. Lentine KL, Schnitzler MA, Xiao H, Saab G, Salvalaggio PR, Axelrod D, et al. Racial variation in medical outcomes among living kidney donors. N Engl J Med. 2010; 363(8):724–732. [PubMed: 20818874]
24. Introduction to the HCUP Nationwide Inpatient Sample (NIS) 2011. Rockville, MD: Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project (HCUP); 2013.
25. Steiner C, Elixhauser A, Schnaier J. The healthcare cost and utilization project: an overview. Eff Clin Pract. 2002; 5(3):143–151. [PubMed: 12088294]
26. Hollingsworth JM, Hollenbeck BK, Englesbe MJ, DeMonner S, Krein SL. Operative mortality after renal transplantation--does surgeon type matter? J Urol. 2007; 177(6):2255–2259. discussion 2259. [PubMed: 17509333]
27. Scarborough JE, Pietrobon R, Tuttle-Newhall JE, Marroquin CE, Collins BH, Desai DM, et al. Relationship between provider volume and outcomes for orthotopic liver transplantation. J Gastrointest Surg. 2008; 12(9):1527–1533. [PubMed: 18612704]
28. Pant C, Anderson MP, O'Connor JA, Marshall CM, Deshpande A, Sferra TJ. Association of Clostridium difficile infection with outcomes of hospitalized solid organ transplant recipients: results from the 2009 Nationwide Inpatient Sample database. Transpl Infect Dis. 2012; 14(5):540–547. [PubMed: 22726461]
29. Friedman AL, Cheung K, Roman SA, Sosa JA. Early clinical and economic outcomes of patients undergoing living donor nephrectomy in the United States. Arch Surg. 2010; 145(4):356–362. discussion 362. [PubMed: 20404286]
30. Boyarsky BJ, Hall EC, Singer AL, Montgomery RA, Gebo KA, Segev DL. Estimating the potential pool of HIV-infected deceased organ donors in the United States. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2011; 11(6):1209–1217.
31. Herring AA, Woolhandler S, Himmelstein DU. Insurance status of U.S. organ donors and transplant recipients: the uninsured give, but rarely receive. Int J Health Serv. 2008; 38(4):641–652. [PubMed: 19069285]
32. Introduction to the HCUP State Inpatient Databases (SID). Rockville, MD: Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project (HCUP); 2013.
33. Nathan H, Cameron JL, Choti MA, Schulick RD, Pawlik TM. The volume-outcomes effect in hepato-pancreato-biliary surgery: hospital versus surgeon contributions and specificity of the relationship. J Am Coll Surg. 2009; 208(4):528–538. [PubMed: 19476786]
34. Patel S, Orloff M, Tsoulfas G, Kashyap R, Jain A, Bozorgzadeh A, et al. Living-donor liver transplantation in the United States: identifying donors at risk for perioperative complications. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2007; 7(10):2344–2349.
35. Axelrod DA, Gheorghian A, Schnitzler MA, Dzebisashvili N, Salvalaggio PR, Tuttle-Newhall J, et al. The economic implications of broader sharing of liver allografts. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2011; 11(4):798–807.
36. Salvalaggio PR, Dzebisashvili N, MacLeod KE, Lentine KL, Gheorghian A, Schnitzler MA, et al. The interaction among donor characteristics, severity of liver disease, and the cost of liver transplantation. Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society. 2011; 17(3):233–242.
37. Muzaale AD, Massie AB, Wang MC, Montgomery RA, McBride MA, Wainright JL, et al. Risk of end-stage renal disease following live kidney donation. JAMA : the journal of the American Medical Association. 2014; 311(6):579–586.
38. Dekkers OM, von Elm E, Algra A, Romijn JA, Vandenbroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. Int J Epidemiol. 2010; 39(1):89–94. [PubMed: 19376882]
39. Montgomery RA, Lonze BE, King KE, Kraus ES, Kucirka LM, Locke JE, et al. Desensitization in HLA-incompatible kidney recipients and survival. New England Journal of Medicine. 2011; 365(4):318–326. [PubMed: 21793744]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

40. Massie AB, Caffo B, Gentry SE, Hall EC, Axelrod DA, Lentine KL, et al. MELD Exceptions and Rates of Waiting List Outcomes. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2011; 11(11):2362–2371.
41. Simpkins CE, Montgomery RA, Hawxby AM, Locke JE, Gentry SE, Warren DS, et al. Cold ischemia time and allograft outcomes in live donor renal transplantation: is live donor organ transport feasible? American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2007; 7(1):99–107.
42. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. American journal of epidemiology. 1995; 142(12):1255–1264. [PubMed: 7503045]
43. Vittinghoff, EGDV.; Shiboski, SC.; McCulloch, CE. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. Springer; 2005.
44. Meier-Kriesche HU, Kaplan B. Waiting time on dialysis as the strongest modifiable risk factor for renal transplant outcomes: a paired donor kidney analysis. Transplantation. 2002; 74(10):1377–1381. [PubMed: 12451234]
45. Thompson D, Waisanen L, Wolfe R, Merion RM, McCullough K, Rodgers A. Simulating the allocation of organs for transplantation. Health Care Manag Sci. 2004; 7(4):331–338. [PubMed: 15717817]
46. Gleisner AL, Munoz A, Brandao A, Marroni C, Zanotelli ML, Cantisani GG, et al. Survival benefit of liver transplantation and the effect of underlying liver disease. Surgery. 2010; 147(3):392–404. [PubMed: 19962165]
47. Kamath PS, Wiesner RH, Malinchoc M, Kremers W, Therneau TM, Kosberg CL, et al. A model to predict survival in patients with end-stage liver disease. Hepatology. 2001; 33(2):464–470. [PubMed: 11172350]
48. Rao PS, Schaubel DE, Guidinger MK, Andreoni KA, Wolfe RA, Merion RM, et al. A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. Transplantation. 2009; 88(2):231–236. [PubMed: 19623019]
49. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA : the journal of the American Medical Association. 2001; 286(2):180–187.
50. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009; 338:b605. [PubMed: 19477892]
51. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008; 27(2):157–172. discussion 207–112. [PubMed: 17569110]
52. Altman DG, Royston P. What do we mean by validating a prognostic model? Statistics in medicine. 2000; 19(4):453–473. [PubMed: 10694730]
53. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001; 54(8):774–781. [PubMed: 11470385]
54. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999; 130(6):515–524. [PubMed: 10075620]
55. Massie AB, Desai NM, Montgomery RA, Singer AL, Segev DL. Improving distribution efficiency of hard-to-place deceased donor kidneys: Predicting probability of discard or delay. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2010; 10(7):1613–1620.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Summary of selected data sources available for transplant research.

Data source	Population	Strengths	Weaknesses
UNOS	Live and deceased donors, transplant candidates, transplant recipients	Represents entire U.S. transplant population; longitudinal followup; available free of charge	Lacks many comorbidities; Poor graft loss ascertainment
SRTR	Live and deceased donors, transplant candidates, transplant recipients	Represents entire U.S. transplant population; longitudinal followup; good graft loss ascertainment	Lacks many comorbidities
USRDS	ESRD patients, 5% sample of Medicare	Longitudinal followup; ESRD incidence on entire U.S. population; rich claims data	Claims data limited to Medicare participants
NIS	Inpatients at 20% sample of U.S. hospitals	Contains diagnoses and procedures unavailable from OPTN sources	No longitudinal followup; no long-term outcomes; 20% sample may not be representative of transplant population
SID	Inpatients at hospitals in most U.S. states	Contains diagnoses and procedures unavailable from OPTN sources; link multiple records of one patient in some cases	No long-term outcomes; each state/year must be purchased separately
External linkages	Depends on linked dataset	Access novel data beyond the scope of standard datasets	Linkage is challenging