Predicting kidney transplant outcomes with partial knowledge of HLA mismatch

Charles F. Manski[a,b,1], **Anat R. Tambur**[c], **and Michael Gmeiner**[a]

[a]Department of Economics, Northwestern University, Evanston, IL 60208; [b]Institute for Policy Research, Northwestern University, Evanston, IL 60208; and [c]Feinberg School of Medicine, Northwestern University, Chicago, IL 60611

We consider prediction of graft survival when a kidney from a deceased donor is transplanted into a recipient, with a focus on the variation of survival with degree of human leukocyte antigen (HLA) mismatch. Previous studies have used data from the Scientific Registry of Transplant Recipients (SRTR) to predict survival conditional on partial characterization of HLA mismatch. Whereas earlier studies assumed proportional hazards models, we used nonparametric regression methods. These do not make the unrealistic assumption that relative risks are invariant as a function of time since transplant, and hence should be more accurate. To refine the predictions possible with partial knowledge of HLA mismatch, it has been suggested that HaploStats statistics on the frequencies of haplotypes within specified ethnic/national populations be used to impute complete HLA types. We counsel against this, showing that it cannot improve predictions on average and sometimes yields suboptimal transplant decisions. We show that the HaploStats frequency statistics are nevertheless useful when combined appropriately with the SRTR data. Analysis of the ecological inference problem shows that informative bounds on graft survival probabilities conditional on refined HLA typing are achievable by combining SRTR and HaploStats data with immunological knowledge of the relative effects of mismatch at different HLA loci.

transplant risk assessment | HLA matching | nonparametric prediction | ecological inference

An important problem in organ transplantation is to predict the patient outcomes that would occur if an offered organ were to be transplanted into a specific recipient. Transplant clinicians regularly face this prediction problem. Some version of the problem arises with transplant of any organ.

We focus on the transplant of a kidney from a deceased donor. The clinically observed organ attributes commonly include information about the quality of the organ and about genetic features of the donor related to activation of the patient's immune system, measured by human leukocyte antigen (HLA) genotype. Observed recipient attributes include information about patient age, health, and HLA type.

Research in transplant immunology has shown that, other conditions remaining the same, patient outcomes vary with the degree of HLA mismatch between the donor and the recipient (e.g., refs. 1, 2). The greater the mismatch, the more likely it is that the patient's immune system will produce antibodies that attack the grafted organ, lowering its functionality and eventually destroying it. HLA mismatch is a potentially powerful predictor of outcomes, despite the development of immunosuppressive therapies that seek to decrease the response of a patient's immune system to the foreign tissue in a graft. Immunosuppression may reduce graft rejection, but it also may harm the patient by lowering the effectiveness of the immune system in protecting against a host of diseases.

When considering whether to accept an offered organ, clinicians and patients would benefit from accurate prediction of survival and functionality outcomes conditional on observed donor and recipient attributes. By "accurate prediction," we mean accurate probabilistic prediction of the frequency of favorable and unfavorable outcomes conditional on organ quality, patient age/health, and HLA mismatch. We do not mean perfect deterministic prediction of the outcomes that would be experienced by a patient receiving a specific organ. Perfect deterministic prediction is unrealistic to expect. In practice, one can at most seek accurate probabilistic predictions.

Accurate prediction of outcomes conditional on HLA mismatch would be especially beneficial because it would enable better genetic matching of donors and recipients. Better matching would reduce the need for immunosuppression, which increases the risks that recipients will develop infectious diseases and cancers. Available low-resolution data on donor and recipient HLA types do not suffice to realize the full promise of better matching, but these data have some predictive value. The analysis in this paper aims to improve prediction. We combine data in the Scientific Registry of Transplant Recipients (SRTR) on transplant outcomes conditional on low-resolution HLA types with HaploStats data on the distribution of refined HLA types within specific populations. The way we combine the data brings to bear knowledge from research on transplant immunology.

Accurate probabilistic prediction of treatment outcomes has long been an objective of evidence-based medical research, with more progress in some domains than others. No domain can claim complete accuracy. Nevertheless, simple online probabilistic risk assessment tools have become widely used by clinicians and patients to predict risk of breast cancer (3) and cardiovascular disease (4), among other diseases.

A similarly simple tool for prediction of kidney transplant outcomes is the Kidney Donor Profile Index (KDPI), accessible on the website of the Organ Procurement and Transplantation Network (OPTN) (5). However, this predictor does not use

Significance

An important problem in transplantation is to predict graft survival when an organ is transplanted into a recipient. We consider transplant of kidneys from deceased donors. Research in transplant immunology has shown that survival varies with the degree of genetic match between the donor and recipient, measured by human leukocyte antigen (HLA) genotypes. The greater the mismatch, the more likely it is that the patient's immune system will produce antibodies that attack the graft. When considering whether to accept an organ, clinicians would benefit from accurate prediction of survival conditional on HLA mismatch. The analysis in this paper aims to improve prediction using available data on transplant outcomes and the distribution of HLA types within specific populations.

information on HLA mismatch. The KDPI is determined only by demographic and health attributes of the deceased donor.

Clinicians who want to predict outcomes conditional on HLA mismatch can attempt to interpret and use findings reported in research articles that present evidence-based estimates of prediction models. Particularly visible has been Rao et al. (6), who developed various versions of the Kidney Donor Risk Index (KDRI), one of which was later transformed into the KDPI. Other estimated prediction models are reported in refs. 7–13 and elsewhere. To provide the background for our work, we briefly summarize the contribution in ref. 6, its transformation into the KDPI, and the current use of the KDPI.

Rao et al. (6) used a proportional hazards model to interpret nationwide administrative data for the United States on the outcomes of deceased-donor kidney transplants, with the data being available in files provided by the SRTR (https://www.srtr.org.). The model predicts the probability of all-cause graft failure as a function of time since transplant and of variables that describe organ quality, recipient age/health, and HLA mismatch. All-cause graft failure is defined as patient return to dialysis, retransplant, or death. The KDRI is a summary statistic that measures the risk of graft failure for a transplant with specified characteristics relative to the risk for a specified reference transplant. The proportional hazards model assumes that this relative risk remains constant as a function of time since transplant.

In 2013, the OPTN approved a new national deceased-donor Kidney Allocation System (KAS) that makes official use of the KDPI, a truncated and renormalized version of the KDRI. Israni et al. (14) describe the KAS. The KDPI truncates the KDRI by considering only organ attributes. These include donor age, height, weight, ethnicity, history of diabetes and hypertension, cause of death, serum creatinine, hepatitis C virus status, and donation-after-circulatory-death status. It does not compute relative risk conditional on recipient attributes and HLA mismatch.

The KDPI renormalizes the truncated KDRI by measuring the risk of graft failure in percentile terms. The KDPI is calculated to the nearest integer percentage value and ranges from 0 to 100%. A donor with a KDPI of 0% has a KDRI less than all donors in a specified reference population. For integers x > 0, a donor with a KDPI of x% implies that the donor's KDRI exceeds more than (x − 1)%, but not more than x%, of all donors in the reference population.

Some in the kidney transplant research community have expressed a favorable view of the truncated, donor-only, version of the KDRI as a tool for transplant risk assessment. Lee and Abramowicz (15) comment as follows on the fact that the truncated version of the KDRI uses only organ attributes to predict outcomes:

Of note, several factors pertaining to the recipient and/or transplant procedure (cold ischaemic time, degree of HLA mismatching, single versus double versus en-bloc kidneys) can also be used to calculate a 'full' KDRI. Since these factors are generally not known at the time when offers are made, and are candidate-specific, the donor-only KDRI is the version that was implemented. In addition, virtually no predictive ability is lost with the 'donor factors only' version. Indeed, it has a concordance statistic of 0.6 compared with the full KDRI version which has a concordance statistic of 0.601 (p. 1287).

Reiterating the comment on predictive power, an OPTN document states (16):

Virtually no predictive ability is lost by using a donor-only version of the KDRI (c = 0.596) compared to a full version of the KDRI (c = 0.601) that includes the degree of HLA matching, cold ischemic time, and transplant procedure type (single vs. double vs. en-bloc) (p. 6).

We think it premature for the transplant community to endorse use of the donor-only version of the KDRI to predict patient outcomes. Contrary to the comment of Lee and Abramowicz (15) regarding the information available when organ offers are made, clinicians typically have some knowledge of HLA mismatch and other candidate-specific attributes that may affect evaluation of

offered organs. Neither Lee and Abramowicz (15) nor the OPTN (16) describes the specific C statistic that they cite, or why this statistic adequately summarizes predictive power. Even if their interpretation of the C statistic should be well grounded in the context of the model in ref. 6, this would not imply more broadly that outcome prediction using only donor attributes is adequate. Another finding may emerge when making predictions with a different model. Moreover, another finding may emerge when making predictions with richer information about HLA mismatch than was available to Rao et al. (6) in their analysis of SRTR data. These concerns motivate the present paper.

Our analysis is in 3 parts. We first consider outcome prediction using only the incomplete characterization of HLA provided in the SRTR data. Our work differs methodologically from that of Rao et al. (6) and other studies in that we compute nonparametric estimates of graft survival rather than estimates invoking a proportional hazards model. Considering 1-y and 5-y survival, we find that lower survival probabilities are associated with higher values of recipient age, organ KDPI, and the number of HLA mismatches. The estimated reductions in survival probability with increases in these risk factors are much larger when predicting 5-y survival than when predicting 1-y survival.

Not only are the reductions larger when predicting 5-y survival than 1-y survival, but they are differentially so. This finding provides evidence against the proportional hazards model. It is consistent with research in transplant immunology showing that generation of de novo donor-specific antibodies following transplant, a consequence of HLA mismatch, plays a cumulative rather than time-invariant role in graft loss. Moreover, it is known that late antibody-mediated rejection is less responsive to treatment, and thus more associated with graft loss (refs. 17–19).

We next critique an HLA imputation method suggested recently in the transplant literature. The idea is for clinicians who have partial knowledge of HLA mismatch to use available population data on the frequency distribution of distinct HLA genotypes to impute genotypes for specific donors and recipients. This done, the suggestion is for a clinician contemplating whether to accept an offered organ to act as if he or she has complete knowledge of mismatch. We show that imputation does not improve prediction of transplant outcomes. To the contrary, it may diminish the accuracy of predictions and thereby generate suboptimal transplant decisions.

The third part of our analysis considers the predicament of a clinician who possesses more complete knowledge of mismatch than is available in the SRTR data. This situation would be salutary if accurate probabilistic predictions of outcomes conditional on clinically observed mismatch were available. However, such predictors are not currently available. Bringing to bear econometric research on partial identification of conditional probability distributions, we combine SRTR and HaploStats data to yield credible partial predictions conditional on the clinically observed mismatch information.

Nonparametric Prediction of Graft Failure with SRTR Data

The Proportional Hazards Model. The use by Rao et al. (6) of a proportional hazards model (20) to predict all-cause graft failure adheres to the widespread medical application of this model to predict survival following the occurrence of a specified event. In transplant contexts, the proportional hazards model assumes that the hazard of graft failure at any point in time following the date of transplant is an unrestricted function of time multiplied by a parametric function of observed covariates characterizing the organ and recipient. This function measures the risk of failure for transplants with a specified value of the covariates relative to that with a reference covariate value. Following standard practice, Rao et al. (6) suppose that the relative risk function is log-linear. Thus, the assumed hazard rate at date T for a transplant with covariates x is $h(T|x) = h_0(T) \cdot \exp(xb)$, where $h_0(T)$ is the T-varying baseline hazard, $\exp(xb)$ is the x-varying relative risk, and b is a parameter vector.

The popularity of the proportional hazards model stems from the simple way that it separates the dependence of survival on time and observed covariates. However, this simplicity comes with a potential cost in credibility. The assumption that relative risk is a time-invariant log-linear function of covariates is strong.

An unfortunate feature of application of the proportional hazards model in medical research has been that its realism is rarely questioned. This is evident in ref. 6 and other applications of the model to predict transplant outcomes. Transplant researchers have used the model to study patient outcomes without substantiating it. We are aware of no biological basis to assume that the relative risk of graft failure across organ and recipient covariates remains invariant as the length of time since transplant grows. We are aware of no basis to assume that a log-linear function expresses how relative risk varies across organ and recipient covariates.

Nonparametric Prediction. The literature on survival analysis has developed many models that weaken the assumptions of the proportional hazards model in various ways (21). When ample data are available, nonparametric prediction of transplant outcomes conditional on donor and recipient thought to have appreciable predictive power becomes feasible. This reduces the rationale to assume proportional hazards or other questionable assumptions about how risk of failure varies over time and covariates. The SRTR standard analysis files (SAFs) provide ample data. The SAFs are an updated version of the same data that were used by Rao et al. (6) to develop the KDRI. They provide (donor and recipient) data with personal identifiers removed, made available by the SRTR for research purposes (http://www.srtr.org/requesting-srtr-data/about-srtr-standard-analysis-files/).

The research reported in our paper uses data on the attributes and outcomes of kidney transplants in the United States contained in the SAFs of the SRTR. We are not permitted by the SRTR to share the SAF data with other parties. However, researchers who wish to perform their own research with the data may submit a request to the SRTR and obtain the SAF data, with SRTR approval.

To demonstrate the possibilities, we use SAF data for deceased-donor kidney transplants performed in the years 2009 to 2018 to estimate nonparametrically the probability of all-cause graft survival for 1 or 5 y, conditional on specified (organ, recipient) covariates. We use race and age to characterize the recipient, with age measured in years. We use the KDPI to characterize the quality of an organ, with the KDPI calculated using the formula developed by the OPTN. The SRTR dataset is sufficiently large that it may be feasible to obtain meaningfully precise nonparametric estimates that condition on further measured covariates of the organ and recipient. We choose not to go in this direction because we want to perform analysis that conditions on HLA mismatch in as general a manner as the SRTR data permit.

We use the available HLA data to characterize mismatch. For each organ and recipient, the SRTR SAF file records types for 3 HLA loci: A, B, and DR. The file codes HLA type for almost all donors and recipients with 2-digit resolution. We convert occasional 4-digit resolution codings to 2 digits (details are provided in *SI Appendix*).

Humans have 2 antigens at each HLA locus, from their paternal and maternal heritage. Hence, the number of mismatches at each locus takes one of the 3 values {0, 1, 2}. Thus, there are $3 \times 3 \times 3 = 27$ possible values for (A, B, DR) mismatch, with the value {0, 0, 0} indicating a perfect match and the value {2, 2, 2} indicating a complete mismatch. We condition on the specific 27-valued mismatch pattern, rather than simply counting the number of mismatches (0 to 6). There is biological reason to think that the effect of mismatch on organ failure varies with the specific loci and genotype of mismatches, not just with the number of mismatches (22, 23).

To formalize the prediction problem, let y denote the number of years following transplant when all-cause graft failure occurs.

Let x indicate the (A, B, DR) mismatch between donor and recipient. Let z denote the other organ and recipient attributes on which we condition prediction (with these being recipient race, recipient age, and organ KDPI). In principle, we can use the SAF data to estimate $P(y > T|z, x)$ for any number T of years. For specificity, we focus on survival for at least 1 or 5 y, that is, on $P(y > 1|z, x)$ and $P(y > 5|z, x)$. These conditional probabilities may be restated as conditional expectations $E\{1[y > 1]|z, x\}$ and $E\{1[y > 5]|z, x\}$, where $1[\cdot]$ is the indicator function. Hence, they may be estimated with standard nonparametric regression methods.

Data. The SRTR data extend back to 1987, but we restrict analysis to the period from 2009 onward. One reason is that it has been standard to use molecular rather than cruder serological methods for HLA typing from 2008 onward. Another is that immunosuppression therapy practices have been approximately stable since the mid-2000s. For illustrative purposes, we focus on transplants to the SRTR category of "white" recipients, which excludes Caucasians of Arab or Middle Eastern origin. Estimates may likewise be obtained for other races.

The SAF documents 48,945 kidney transplants to white recipients in the 2009 to 2018 period. Of these, we restrict attention to patients who were on the waiting list to receive transplant of a single kidney, excluding 699 en bloc patients, 446 double-transplant patients, and 201 patients who were initially waiting to receive a paired kidney/pancreas transplant. Among the 47,599 transplants eligible for analysis, we drop 2,382 cases with missing data on some covariate used in our analysis, yielding 45,217 cases with complete data. Given that the rate of missing data is only 0.05, dropping these cases can at most bias our findings by this small fraction overall. Dropping the cases with missing data implies no bias if data are missing at random.

Among the 45,217 transplants with complete data, 40,316 were performed in the years 2009 to 2017, for which the 2018 SAF file records the 1-y graft survival outcome. A total of 21,310 were performed in 2009 to 2013, for which the 5-y graft survival outcome was recorded. These are the sample sizes for estimation of $P(y > 1|z, x)$ and $P(y > 5|z, x)$, respectively.

The SAF data record the exact date of the transplant as well as the exact dates of death and graft failure (if any). If both graft failure and death are recorded, we use the first date. These variables are not rounded in analysis. The 2 outcome variables created are indicator variables for (1) failure occurring strictly more than 365 d after the transplant, or no failure and more than 365 d have passed, and (2) failure occurring strictly more than 1,825 d after the transplant, or no failure and more than 1,825 d have passed.

We think it important to call attention to the fact that the SRTR data on HLA genotypes are incomplete in 2 respects, both implying undercounts of mismatches. First, the data provide types for 6 donor HLA loci (A, B, C, DP, DQ, and DR) but only 3 recipient ones (A, B, and DR). Hence, the data do not reveal mismatches at the C, DP, and DQ loci. Lack of knowledge of DQ mismatches may be particularly concerning, because research in transplant immunology suggests that these mismatches can have severe consequences for organ failure (24–29).

Second, the SRTR data code only low-resolution, 2-digit genotypes, each representing multiple alleles. Modern HLA typing codes higher resolution 4-digit types, with each identifying a unique allele. Hence, donor and recipient HLA types that appear matched with 2-digit coding may be mismatched with 4-digit coding.

Methods

To estimate $P(y > 1|z, x)$ and $P(y > 5|z, x)$, we form 27 subsamples of data, each with a distinct value of x, with the covariate indicating HLA mismatch. Estimation proceeds separately for all 27 values of A, B, and DR mismatch. For each of the 2 outcome variables, we use a bivariate kernel regression procedure, the npregress command in Stata, to estimate the conditional probability at specified values of patient age and KDPI. We use a local-constant

ECONOMIC SCIENCES

MEDICAL SCIENCES

Epanechnikov kernel, with 2-dimensional bandwidth following norms in the literature on variable-bandwidth kernel estimation. Details are provided in *SI Appendix*.

The SRTR data are a census listing all transplants performed in the United States, rather than a random sample of transplants. It is therefore not obvious how to measure the statistical precision of the kernel estimates. Without commenting on this question, previous research predicting transplant outcomes has used standard statistical methods to measure precision, essentially viewing the SRTR data as a random sample from a population of potential transplants. In the absence of a ready alternative, we follow this practice. We use the Stata percentile bootstrap procedure to compute an approximate 95% confidence interval for each estimate.

Although our analysis is nonparametric, it maintains the possibly unrealistic assumption that the ex post conditional distribution of outcomes for consummated transplants equals the ex ante distribution of outcomes for potential transplants. A clinician making a transplant decision wants to predict outcomes before choosing whether to accept an offered organ. The SRTR data reveal only the outcomes experienced with transplants that were in fact performed.

Equality of ex ante and ex post outcome distributions has been assumed, generally without comment, throughout the research literature that uses SRTR data or another source of observational data to assist clinicians making transplant decisions; yet, the accuracy of the assumption is unknown. The assumption would be easy to motivate if transplant decisions were made randomly, as in a trial. However, it need not hold when clinicians make purposeful transplant decisions, as has been the norm. The difference between ex ante and ex post outcome distributions is often called "selection bias."

The possibility of differences between ex ante and ex post outcome distributions may limit the predictive power of observational data sources such as the SRTR, but it does not imply that such data are useless as evidence supporting ex ante prediction. The econometric literature on partial identification of treatment response weakens the assumption of equal ex ante and ex post outcome distributions. It shows that observational data may still be informative to some degree, yielding bounds on treatment response (ref. 30, chap. 7). We do not perform this type of partial identification analysis here, but we consider it to be an important subject for future research. We do use a different form of partial identification analysis later in the present paper, when we examine another inferential problem in prediction.

Results

A kernel nonparametric regression estimate is easy to compute for any value of the conditioning covariates (z, x). Hence, nonparametric prediction is amenable to development of an online prediction tool. A clinician could input the covariate value for a (donor, recipient) case of interest and obtain survival estimates and confidence intervals.

Presentation in a research article of estimates for all values of (z, x) is not realistic due to space constraints, so we suffice with a simple summation. We have computed estimates for the 27 values of the HLA mismatch indicator and for a grid of 9 values of (recipient age, organ KDPI), namely, $\{30, 50, 70\} \times \{25, 50, 75\}$. This yields $27 \times 9 = 243$ estimates in all. Sparsity of SRTR data near certain values of (z, x) occasionally prevents computation of a precise kernel estimate. When computing estimates of 1-y and 5-y survival, we restrict attention to, respectively, 25 and 22 values of HLA mismatch for which there exist at least 50 observations. We moreover exclude (age, KDPI) combinations where there exist no observations within the relevant local bandwidth.

Table 1 gives a flavor of the nuanced findings, presenting a selection of 9 kernel estimates and their confidence intervals. Considering probability of 1-y survival, the table fixes values for (A mismatch, B mismatch, age) and shows the variation in estimates with the values of DR mismatch and KDPI.

The small set of estimates in Table 1 do not show much pattern, a result that is not surprising given the great flexibility of nonparametric estimation. However, strong and interesting patterns emerge when we summarize all of the findings in a succinct manner. We first exclude kernel estimates for bandwidths that have no variation in the survival outcome (a confidence interval of [1, 1] as seen in Table 1) because this is a consequence of small

Table 1. Kernel estimates and confidence intervals for probability of 1-y survival by DR mismatch (mm) and KDPI (A mm = 1, B mm = 1, age = 30 y)

	KDPI = 25	KDPI = 50	KDPI = 75
DR mm = 0	0.97	0.96	0.89
	[0.91, 1]	[0.85, 1]	[0.80, 0.96]
DR mm = 1	0.97	1.00	1.00
	[0.90, 1]	[1, 1]	[1, 1]
DR mm = 2	0.92	0.96	0.94
	[0.84, 0.98]	[0.87, 1]	[0.90, 0.98]

Confidence intervals are shown in brackets.

sample size. With this additional exclusion, we restrict attention to 207 1-y estimates and 194 5-y estimates. Given this, we perform linear least-squares fits of the 207 and 194 kernel estimates of survival probability to recipient age, organ KDPI, the total number of HLA mismatches, and a constant. Table 2 presents the findings.

The results in Table 2 naturally cannot convey the full nuance of the kernel estimates, but they are nevertheless revealing. Considering both 1-y and 5-y survival, the summary shows that lower survival probabilities are associated with higher values of recipient age, organ KDPI, and the number of HLA mismatches. With the exception of a single parameter estimate, all are statistically significant by conventional criteria. The single exception is the estimate for number of mismatches when predicting 1-y survival.

We think it particularly interesting that all estimates are much larger in magnitude when predicting 5-y survival than when predicting 1-y survival. The summary suggests that a 10-y increase in patient age is associated with a 0.006 reduction in the probability of 1-y survival and a 0.03 reduction in the probability of 5-y survival. A 10-unit increase in organ KDPI is associated with a 0.006 reduction in the probability of 1-y survival and a 0.01 reduction in the probability of 5-y survival. A unit increase in the number of mismatches is associated with a 0.001 reduction in the probability of 1-y survival and a 0.009 reduction in the probability of 5-y survival.

Not only are estimates larger in magnitude when predicting 5-y survival than 1-y survival, but they are differentially so. The 5-y estimates for (age, KDPI, mismatches) are, respectively, approximately (4.4, 2.4, 7.2)-fold their 1-y values. This finding of differential temporal change across covariates provides evidence against the proportional hazards assumption maintained in previous research.

Transplant Decisions with Haplotype Imputations

We observed above that the SRTR data on (donor, recipient) HLA genotypes are incomplete. Similarly, clinicians often have incomplete data on HLA types. Hoping to overcome this problem, some transplant researchers have proposed imputation of complete high-resolution genotypes conditional on the partial typing information that clinicians possess.

Imputation studies use available data on the frequency distribution of distinct HLA genotypes within specific populations. Such a database is embedded in HaploStats (https://www.haplostats.org/haplostats?execution=e2s1), a web application provided by the National Marrow Donor Program Bioinformatics group (details are provided in ref. 31). HaploStats facilitates clinician access to HLA genotype frequency data collected for various ethnic/national populations. A clinician can input the HLA type data available for a donor or recipient. The HaploStats application accesses its genotype frequency data and outputs the frequency distribution conditional on the partial HLA data that the clinician provided.

The conditional frequency distribution output by HaploStats is formatted in order of the prevalence of specific haplotypes, that is, groups of HLA alleles that are inherited in conjunction. Geneugelijk et al. (32), Geneugelijk and Spierings (33), and others

Manski et al.

Table 2. Linear least-squares fits of kernel estimates of survival probability to covariates

Covariates	1-y survival	5-y survival
Age = (30, 50, 70 y)	−0.000622	−0.00276
	(0.000170)	(0.000304)
KDPI = (25, 50, 75)	−0.000634	−0.00149
	(0.000133)	(0.000274)
Number (A, B, DR) mm = (0 to 6)	−0.00124	−0.00894
	(0.00169)	(0.00309)
Constant	0.990	1.007
	(0.0113)	(0.0238)
Observations	207	194

Robust SEs are shown in parentheses.

suggest use of the most prevalent haplotype to impute an unobserved donor or recipient HLA genotype. An extension of the idea is to perform multiple imputation, which considers a small number of the most prevalent haplotypes.

Haplotype imputation may seem an appealing way to overcome incompleteness of clinically available HLA data. However, the appeal does not survive under scrutiny. We counsel against using imputations to make transplant decisions.

The obvious issue is that a donor or recipient may not actually have the imputed haplotype, even if the imputation is the most prevalent haplotype within his or her ethnic/national population. When an imputed haplotype is not accurate, computed HLA mismatch may be incorrect, with consequent misprediction of transplant outcomes. Accurate probabilistic prediction of outcomes requires consideration of the entire frequency distribution of haplotypes conditional on the partial HLA data that the clinician possesses.

Making transplant decisions with haplotype imputations on average yields worse patient survival outcomes than making decisions without imputation. To demonstrate this, some simple analysis is presented in *SI Appendix*.

Using More Complete Knowledge of HLA Mismatch: Partial Predictions Combining SRTR and HaploStats Data

We now consider transplant outcome prediction when a clinician has more complete knowledge of HLA mismatch than is available in the SRTR data. This knowledge would be advantageous if the research literature were to provide accurate probabilistic predictions of outcomes conditional on the available clinical knowledge of mismatch. However, current risk assessment tools analyzing SRTR data can condition their predictions at most on 2-digit (A, B, DR) mismatch. Hence, a clinician who possesses more complete knowledge of mismatch cannot draw on the research literature to use this knowledge.

While the SRTR data alone do not enable prediction conditional on more complete knowledge of mismatch, some degree of more refined conditional prediction becomes possible by combining SRTR and HaploStats data. The methodology for achieving this has been developed in research on the ecological inference problem, with application to risk assessment for personalized medicine (34). We first present the basic analysis, which uses only empirical evidence without invoking assumptions that restrict the conditional predictions. We then introduce assumptions that are credible when predicting transplant outcomes and that enable more informative predictions.

The Ecological Inference Problem. Let each member of a population be characterized by an outcome y and by covariates (z, x, w). In the transplant setting of this paper, y is the length of all-cause graft survival and (z, x) are the (donor, recipient) attributes considered earlier, including knowledge of (A, B, and DR) 2-digit HLA mismatch. The symbol w expresses further information on HLA mismatch that is observed by a clinician but not coded in the SRTR data. In our application, w represents mismatches for HLA

antigens C and DQ (HaploStats does not provide information on the DP antigen).

Suppose that data are available from 2 sampling processes, each of which has a severe missing data problem. One sampling process yields observable realizations of (y, z, x), but the corresponding realizations of w are not recorded. This is the situation with the SRTR data. The other sampling process yields observable realizations of (w, x, z), but the realizations of y are not recorded. This is the situation with the HaploStats data, when z denotes ethnicity/nationality. The 2 sampling processes reveal the distributions $P(y|z, x)$ and $P(w|z, x)$. The term ecological inference, which originated in political science and sociology, describes the problem of inference on $P(y|z, x, w)$, given knowledge of $P(y|z, x)$ and $P(w|z, x)$.

The law of total probability relates these distributions, making the identification problem transparent. Hold (z, x) fixed at specified values. For any value of w, say w = j,

$$P(y|z,x) = P(w=j|z,x) \cdot P(y|z,x,w=j) + P(w \neq j|z,x)P(y|z,x,w \neq j).$$
[1]

Knowledge of $P(y|z, x)$ alone reveals nothing about $P(y|z, x, w = j)$, because any distribution $P(y|z, x, w = j)$ satisfies the equation if $P(w = j|z, x) = 0$. Partial conclusions may be drawn if one has evidence revealing $P(w = j|z, x)$, provided that it is positive. The joint identification region for $P(y|z, x, w = j)$ and $P(y|z, x, w \neq j)$, given knowledge of $P(y|z, x)$ and $P(w|z, x)$, is the set of pairs of distributions that satisfy [1].

The practical challenge is to characterize the feasible distributions. Analysis is simple when the objective is to learn the probability $P(y > T|z, x, w = j)$ that a graft will survive for at least T years and one makes no assumptions that restrict $P(y|z, x, w)$. The identification region for $P(y > T|z, x, w = j)$ is the interval

$$P(y > T|z,x,w=j) \in [0,1]$$
$$\cap \left[\frac{P(y > T|z,x) - P(w \neq j|z,x)}{P(w=j|z,x)}, \frac{P(y > T|z,x)}{P(w=j|z,x)} \right].$$
[2]

A version of this result was sketched by Duncan and Davis (35). A proof is given by Horowitz and Manski (36), who extend the analysis to derive identification regions for conditional means $E(y|x, z, w)$ and α-quantiles $Q_\alpha(y|z, x, w)$.

Bounded-Variation Assumptions. Bound [2] has a simple form, but it is often wide in practice. Indeed, the lower bound is positive, and hence informative, only if $P(y > T|z, x) > P(w \neq j|z, x)$. The upper bound is less than 1, and hence informative, only if $P(y > T|z, x) < P(w = j|z, x)$.

Tighter bounds may be achievable if one combines empirical evidence on $P(y|z, x)$ and $P(w|z, x)$ with credible assumptions. Ref. 34. characterizes the identifying power of bounded-variation assumptions. These are inequalities restricting how $P(y > T|z, x, w)$ varies with (z, x, w).

The state of knowledge in transplant immunology makes various bounded-variation assumptions credible when considering kidney transplants. We caution that the term "credible" is necessarily subjective, reflecting the current judgement of the field. Hence, we cannot guarantee that any assumption is accurate. Nevertheless, we feel comfortable using 2 types of bounded-variation assumptions.

First, recall that the number of (donor, recipient) mismatches at any HLA locus takes one of the values {0, 1, 2}. Study of transplant immunology strongly suggests that, considering any locus and holding all else equal, survival probability decreases as the number of mismatches at this locus grows. This implies a set of bounded-variation assumptions. Let (x, w) and (x′, w′) be alternative mismatch vectors, with (x′, w′) > (x, w) in the vector sense that (x′, w′) has at least as many mismatches as (x, w) in every HLA locus and has more mismatches in at least 1 locus.

Then, it is credible to assume that for any value of covariates z and survival length T,

$$P(y > T|z, x, w) \geq P(y > T|z, x', w'). \qquad [3]$$

Second, consider mismatches at different HLA loci. There exist 2 main classes of loci, which operate biologically in different ways. Class I includes the (A, B, C) loci. Class II includes the (DP, DQ, DR) loci. Study of transplant immunology suggests that DQ mismatches have the most severe consequences for graft rejection, followed by DR mismatches. Next in severity of their consequences are mismatches at the A and B loci. It is thought that C mismatches have the least severe consequences. The implications of DP mismatches are not clear. These ideas about relative severity imply a set of bounded-variation assumptions.

To illustrate, we show the inequalities implied by comparison of DQ and A mismatches. Consider $P(y > T|A = a, B = b, C = c, DR = r, DQ = q)$, where (a, b, c, r, q) are numbers of mismatches. Now, hold 3 of the 4 values (a, b, c, r) fixed and permute the remaining value with q. For example, hold (b, c, r) fixed and permute a and q. The inequalities are as follows:

$$\begin{aligned} a < q &\Rightarrow P(y > T|A = a, B = b, C = c, DR = r, DQ = q) \\ &\leq P(y > T|A = q, B = b, C = c, DR = r, DQ = a) \\ a > q &\Rightarrow P(y > T|A = a, B = b, C = c, DR = r, DQ = q) \\ &\geq P(y > T|A = q, B = b, C = c, DR = r, DQ = a) \end{aligned}. \qquad [4]$$

Analogous inequalities relate the severity of other HLA mismatches, as previously mentioned.

Estimation of Bounds on $P(y > T|z, x, w)$. Consider any specified values of (T, z, x, j). Given knowledge of $P(y > T|z, x)$ and $P(w = j|z, x)$, the feasible values of $P(y > T|z, x, w = j)$ satisfy bound [2] and the inequalities of forms [3] and [4]. We presume that HaploStats evaluates $P(w = j|x, z)$ accurately, via a derivation described in *SI Appendix*. We acknowledge sampling imprecision in our nonparametric estimates of $P(y > T|z, x)$ by using our bootstrapped 95% confidence intervals rather than the kernel-regression point estimates. Thus, letting $P_L(y > T|z, x)$ and $P_U(y > T|z, x)]$ denote the lower and upper bounds on the confidence interval, we replace bound [2] by the wider bound

$$\begin{aligned} P(y > T|z, x, w = j) &\in [0, 1] \\ &\cap \left[\frac{P_L(y > T|z, x) - P(w \neq j|z, x)}{P(w = j|z, x)}, \frac{P_U(y > T|z, x)}{P(w = j|z, x)} \right]. \end{aligned} \qquad [2']$$

Illustrative Findings. In principle, a bound on $P(y > T|z, x, w = j)$ can be estimated for any specified value of (T, z, x) and for whatever further mismatch information, w, that a clinician observes. For illustrative specificity, we have computed bounds when w indicates the numbers of 2-digit (C, DQ) mismatches.

Holding fixed (age, KDPI), there are $3^5 = 243$ feasible values for (A, B, C, DQ, DR) mismatch. For example, consider a (donor, recipient) case with (age = 50 y, KDPI = 50). In the polar case with no mismatch at any locus, the estimated bounds on 1-y and 5-y survival are $P[y > 1|(50, 50), x = w = 0] \in [0.924, 1]$ and $P[y > 5|(50, 50), x = w = 0] \in [0.823, 1]$. In the other polar case with 2 mismatches at each of the 5 loci, the estimated bounds on 1-y and 5-y survival are $P[y > 1|(50, 50), x = w = 2] \in [0.496, 0.927]$ and $P[y > 5|(50, 50), x = w = 2] \in [0.544, 0.775]$.

These findings are reasonably representative in terms of the width of the bounds. They demonstrate that informative bounds on survival probabilities conditional on (z, x, w) can be obtained by combining SRTR data, HaploStats data, and credible assumptions. Nevertheless, a clinician would naturally like to obtain tighter bounds. If SRTR and HaploStats are the only data available, the only way to obtain tighter bounds is to make stronger assumptions than the bounded-variation ones used in our analysis.

Conclusion

Several previous studies have used SRTR data to predict kidney transplant outcomes conditional on donor and recipient covariates, including partial characterization of HLA mismatch. Whereas previous studies assumed proportional hazards models, we used nonparametric regression methods. These do not make the unrealistic assumption that relative risks are invariant as a function of time since transplant. To the contrary, we found that relative risks vary with time. Consistent with research on transplant immunology, we found that HLA mismatch plays a larger role in graft loss 5 y after transplant than 1 y after transplant.

Clinicians and patients would naturally like to refine the predictions possible with the SRTR data. It has been suggested that HaploStats statistics on the frequencies of haplotypes within specified ethnic/national populations might be used to impute complete HLA types, and thereby obtain an accurate assessment of mismatch. We have counseled against this, showing that imputation cannot improve predictions on average and sometimes yields suboptimal transplant decisions.

Nevertheless, the HaploStats frequency statistics are useful when combined appropriately with the SRTR data. We explained the ecological inference problem and showed how to combine the 2 data sources, generating partial predictions of transplant outcomes conditional on refined HLA typing. The data alone typically yield wide bounds on survival probabilities. However, tighter and more informative bounds are achieved when one brings to bear immunological knowledge of the relative effects of mismatch at different HLA loci.

While careful analysis of the ecological inference problem enables one to make the most of the available data, it would be better to collect more refined HLA data. The reason the SRTR data do not include HLA C and DQ types for organ recipients is that the OPTN has not required transplant centers to report these types. The OPTN could do so, thereby enriching the data available for prediction studies. Moreover, the OPTN could encourage reporting of 4-digit, rather than 2-digit, HLA types.

1. G. Opelz, B. Döhler, Effect of human leukocyte antigen compatibility on kidney graft survival: Comparative analysis of two decades. *Transplantation* **84**, 137–143 (2007).
2. C. Süsal, G. Opelz, Current role of human leukocyte antigen matching in kidney transplantation. *Curr. Opin. Organ Transplant.* **18**, 438–444 (2013).
3. National Cancer Institute, The Breast Cancer Risk Assessment Tool. https://bcrisktool.cancer.gov/. Accessed 28 May 2019.
4. American College of Cardiology, ASCVD Risk Estimator Plus. http://tools.acc.org/ASCVD-Risk-Estimator-Plus/#!/calculate/estimate/. Accessed 28 May 2019.
5. Organ Procurement and Transplantation Network, Allocation Calculators. https://optn.transplant.hrsa.gov/resources/allocation-calculators/. Accessed 28 May 2019.
6. P. S. Rao et al., A comprehensive risk quantification score for deceased donor kidneys: The kidney donor risk index. *Transplantation* **88**, 231–236 (2009).
7. A. O. Ojo et al., Survival in recipients of marginal cadaveric donor kidneys compared with other recipients and wait-listed transplant candidates. *J. Am. Soc. Nephrol.* **12**, 589–597 (2001).
8. R. A. Metzger et al., Expanded criteria donors for kidney transplantation. *Am. J. Transplant.* **3** (suppl. 4), 114–125 (2003).
9. J. J. Wynn et al., Kidney and pancreas transplantation. *Am. J. Transplant.* **4** (suppl. 9), 72–80 (2004).
10. R. M. Merion et al., Deceased-donor characteristics and the survival benefit of kidney transplantation. *JAMA* **294**, 2726–2733 (2005).
11. J. D. Schold, H. U. Meier-Kriesche, Which renal transplant candidates should accept marginal kidneys in exchange for a shorter waiting time on dialysis? *Clin. J. Am. Soc. Nephrol.* **1**, 532–538 (2006).
12. B. L. Kasiske et al., A simple tool to predict outcomes after kidney transplant. *Am. J. Kidney Dis.* **56**, 947–960 (2010).
13. V. B. Ashby et al., A kidney graft survival calculator that accounts for mismatches in age, sex, HLA, and body size. *Clin. J. Am. Soc. Nephrol.* **12**, 1148–1160 (2017).
14. A. K. Israni et al., New national allocation policy for deceased donor kidneys in the United States and possible effect on patient outcomes. *J. Am. Soc. Nephrol.* **25**, 1842–1848 (2014).
15. A. P. Lee, D. Abramowicz, Is the Kidney Donor Risk Index a step forward in the assessment of deceased donor kidney quality? *Nephrol. Dial. Transplant.* **30**, 1285–1290 (2015).
16. Organ Procurement and Transplantation Network, A Guide to Calculating and Interpreting the Kidney Donor Profile Index (KDPI). https://optn.transplant.hrsa.gov/media/1512/guide_to_calculating_interpreting_kdpi.pdf. Accessed 29 May 2019.

17. R. C. Walsh *et al.*, Early and late acute antibody-mediated rejection differ immunologically and in response to proteasome inhibition. *Transplantation* **91**, 1218–1226 (2011).

18. C. Wiebe, P. Nickerson, Posttransplant monitoring of de novo human leukocyte antigen donor-specific antibodies in kidney transplantation. *Curr. Opin. Organ Transplant.* **18**, 470–477 (2013).

19. E. Velidedeoglu *et al.*, Summary of 2017 FDA public workshop: Antibody-mediated rejection in kidney transplantation. *Transplantation* **102**, e257–e264 (2018).

20. D. Cox, Regression models and life-tables, *J. R. Statist. Soc. Ser. B Stat. Methodol.* **34**, 187–220 (1972).

21. J. Kalbfleisch, R. Prentice, *The Statistical Analysis of Failure Time Data* (Wiley, Hoboken, NJ, 2002).

22. I. I. Doxiadis *et al.*, Association between specific HLA combinations and probability of kidney allograft loss: The taboo concept. *Lancet* **348**, 850–853 (1996).

23. A. R. Tambur, HLA-epitope matching or eplet risk stratification: The devil is in the details. *Front. Immunol.* **9**, 2010 (2018).

24. A. Tagliamacco *et al.*, DQ molecules are the principal stimulators of de novo donor-specific antibodies in nonsensitized pediatric recipients receiving a first kidney transplant. *Transpl. Int.* **27**, 667–673 (2014).

25. J. M. DeVos *et al.*, Donor-specific HLA-DQ antibodies may contribute to poor graft outcome after renal transplantation. *Kidney Int.* **82**, 598–604 (2012).

26. M. Willicombe *et al.*, De novo DQ donor-specific antibodies are associated with a significant risk of antibody-mediated rejection and transplant glomerulopathy. *Transplantation* **94**, 172–177 (2012).

27. A. R. Tambur, J. R. Leventhal, J. R. Zitzner, R. C. Walsh, J. J. Friedewald, The DQ barrier: Improving organ allocation equity using HLA-DQ information. *Transplantation* **95**, 635–640 (2013).

28. M. C. Freitas *et al.*, The role of immunoglobulin-G subclasses and C1q in de novo HLA-DQ donor-specific antibody kidney transplantation outcomes. *Transplantation* **95**, 1113–1119 (2013).

29. M. J. Everly *et al.*, Incidence and impact of de novo donor-specific alloantibody in primary renal allografts. *Transplantation* **95**, 410–417 (2013).

30. C. Manski, *Identification for Prediction and Decision* (Harvard University Press, Cambridge, MA, 2007).

31. M. Maiers, L. Gragert, W. Klitz, High-resolution HLA alleles and haplotypes in the United States population. *Hum. Immunol.* **68**, 779–788 (2007).

32. K. Geneugelijk, J. Wissing, D. Koppenaal, M. Niemann, E. Spierings, Computational approaches to facilitate epitope-based HLA matching in solid organ transplantation. *J. Immunol. Res.* **2017**, 9130879 (2017).

33. K. Geneugelijk, E. Spierings, Matching donor and recipient based on predicted indirectly recognizable human leucocyte antigen epitopes. *Int. J. Immunogenet.* **45**, 41–53 (2018).

34. C. Manski, Credible ecological inference for medical decisions with personalized risk assessment. *Quant. Econom.* **9**, 541–569 (2018).

35. O. Duncan, B. Davis, An alternative to ecological correlation. *Am. Sociol. Rev.* **18**, 665–666 (1953).

36. J. Horowitz, C. Manski, Identification and robustness with contaminated and corrupted data. *Econometrica* **63**, 281–302 (1995).

ECONOMIC SCIENCES

MEDICAL SCIENCES