Estimating the Benefits of Incorporating Compatible Pairs in Kidney Exchange

ZHUOSHU LI, Washington University in St. Louis
SANMAY DAS, Washington University in St. Louis
SOFIA CARRILLO*, Cornell University
JASON WELLEN, Washington University School of Medicine

Kidney exchange has been studied extensively from the perspective of market design, and a significant focus has been on better algorithms for finding chains and cycles to increase the number of possible matches. A more dramatic benefit could come from incorporating compatible pairs into the mechanism, but this possibility has been relatively understudied. It is possible that the reason incorporation of compatible pairs in exchanges has not taken off has been the lack of quantification of the potential benefits to recipients in compatible pairs. The recent introduction of the Living Donor Kidney Profile Index (LKDPI), which can be transformed into an expected survival time for the graft, presents an opportunity to better estimate the potential benefits, and to present compatible pairs with a compelling medical reason to participate in an exchange rather than proceeding with a direct donation. Using data from a major transplant center, we develop a novel simulator for LKDPIs that generates realistic distributions of graft survival and combine this with a well-known compatibility simulator in a manner that is faithful to data on real arriving pairs. We use our simulator across different matching mechanisms to estimate both increased numbers of transplants of incompatible pairs (almost doubling the number transplanted) as well as improved match quality for recipients in compatible pairs (increasing expected graft survival by between 1 and 2 years). Our results are robust across several different exchange sizes in the static setting, dynamic settings where compatible pairs must be immediately matched, and across assumptions about incompatible to compatible pair ratios. The results are also promising for hard-to-match subpopulations, including blood group O recipients and highly sensitized patients.

1 INTRODUCTION

In the end stages of renal failure, when a patient needs a transplant, one excellent option, when available, is to receive a kidney from a living donor who is willing to donate to them. Close to a third of kidney transplants annually in the US are living donor transplants.¹ As of today, most living donor kidney transplants match recipients directly with a donor willing to donate a kidney to that specific recipient. However, in some cases the donor and the recipient may not be medically compatible (due to ABO blood-type incompatibility or a positive crossmatch), in which case they can enter a *kidney exchange*, a type of barter market where incompatible donors donate to others with the understanding that their recipient will receive a medically compatible kidney from someone else [Abraham et al., 2007, Roth et al., 2004, 2005b, 2007].

While paired kidney donation of this kind has had success in the United States, a raft of coordination problems and exchange fragmentation has prevented it from accounting for a truly significant fraction of transplants. One proposal to transplant more recipients from incompatible pairs has been to incorporate *compatible* pairs into exchanges [Gentry et al., 2007]. While prior work has estimated the possible benefits of such donation in terms of the number of additional patients that could be transplanted from incompatible pairs, the recent development of new metrics for the quality of a living donor transplant [Massie et al., 2016] present an opportunity to reassess the possible benefits in the context of realistic models of compatible pair behavior, while also

¹Work performed while an REU student at Washington University.


*Working paper.*
evaluating benefits in terms of both additional transplants made possible and improved outcomes from transplants. Further, it is reasonable to believe that compatible pairs may be more willing to enter exchanges if (1) their waiting times are kept low, and (2) they have a more precise idea of the potential benefit to doing so.

In this paper, we use detailed data on compatible pairs that have been transplanted at a major US transplant center in the last three years to estimate the benefits of incorporating compatible pairs into kidney exchanges. Doing so requires some measure of match quality when transplanting a particular donor’s kidney into a particular recipient. We estimate expected survival of a graft from the recently proposed Living Donor Kidney Profile Index (LKDPI) [Massie et al., 2016], and use this as our measure of quality. We impose the basic incentive compatibility constraint that, for compatible pairs to be transplanted through exchange instead of directly, each recipient must receive a graft with lower LKDPI, or increased expected survival time, compared with that of her original donor.

There is the potential for significant benefit from including directed donation pairs in kidney exchanges that also include incompatible pairs. The benefit can arise from two fronts: (1) a significant increase in the number of incompatible donors who find matches; (2) an increase in the quality of matches, since factors like HLA match [Massie et al., 2016, Saidman et al., 2006] etc. play a role in expected graft survival. The main goal of this paper is to estimate the potential benefits along both these fronts in a realistic manner. In doing so, we will also contribute to the literature on matching with cardinal utilities by providing a realistic data-generation mechanism for cardinal utilities.

During the years from 2014 to 2016, we were able to obtain data on 184 living donor kidney transplantations that took place at the major transplant center (henceforth “Center”). Of these 184, 171 were directed donations from a compatible donor to his/her paired recipient. We obtained complete information that enabled computation of the LKDPI on 166 of these pairs, which we use to estimate distributions of LKDPI scores (and hence expected graft survival) within compatible pairs and across pairs. We were able to obtain complete antibody and antigen data on 121 of these pairs, which enables donor-recipient compatibility checking.

The first question we can ask is about the heterogeneity of match qualities across pairs and the effects of this heterogeneity on the quality of the final matching. At one extreme, LKDPIs across pairs could be completely independent of the original LKDPIs within the pairs. This would correspond to maximally heterogeneous match qualities and offer the highest possible benefits to recipients in compatible pairs of participating in the exchange. At the other extreme, LKDPIs could be completely determined by the characteristics of the donor or the recipient in a pair, in which case there would be no social gains from trade [Anshelevich et al., 2013]. In reality, LKDPI does take into account various match characteristics (for example, HLA mismatches and body weight ratios), but where the gains from trade may fall in the spectrum is an empirical question. Our experiments confirm that the distribution of match quality (LKDPIs) from “external” donors is far from independent of the match quality within a compatible pair, and this has significant implications on the possible gains from trade to the compatible pairs. As a benchmark, we conduct counterfactual tests that assume no incompatibilities among any of the pairs, and that all 166 pairs participate in a pareto-improving kidney exchange with 2 and 3 cycle swaps. This improves the

---

2 LKDPI itself is a somewhat complex number to interpret. It is intended to be on the same scale as the KDPI for cadaveric kidneys, which is a percentile measure. Thus an LKDPI of 10 indicates that the kidney is comparable to the 10th percentile of cadaveric kidneys in terms of quality (with lower numbers being better). However, since some living donor kidneys can be better than any cadaveric kidney, LKDPI values can also be negative.

3 We do not have data for the remaining 45 pairs because of a change in the software system, so there is no selection bias.
average LKDPI of transplanted kidneys from 37.15 to 25.5, corresponding to about 1.5 years of expected graft survival. We can estimate the hypothetical benefit if all donor-recipient pair LKDPIs were independent draws from the same distribution, and we find that the new average LKDPI achieved would be 2.67, corresponding to more than a 5 year benefit in terms of expected graft survival. Interestingly, we provide evidence that the variability is largely driven by characteristics of the donor rather than the recipient, so there could be benefits from increasing the pool of possible donors, as has recently been suggested [Farina et al., 2017].

Based on this observation, we argue for the importance of constructing a minimal simulator that produces realistic LKDPI / match quality values, and describe the construction of such a simulator, which closely matches the characteristics we observe. Having established the potential for gains from matching simply among already compatible pairs, we then turn to estimating impacts in differently-sized populations when both compatible and incompatible pairs are present using this simulator, paired with the standard Saidman simulator [2006] for generating recipient-donor pairs and compatibilities. These two simulation mechanisms together enable us to simulate realistic living donor kidney scenarios of any size with compatible and incompatible pairs. We use the simulator to estimate both the increase in the number of recipients in incompatible pairs who would be matched if compatible pairs participated in the exchange, as well as the increase in the expected graft survival for recipients in compatible pairs that participate in the exchange.

We find that with compatible pairs joining the kidney exchange, the percentage of matched incompatible pairs almost doubles. For example, with a small pool of 50 donor-recipient pairs, 64% of incompatible pairs are matched, compared with 33% when the two- & three-cycle swap is only run within the incompatible pairs. With a large pool of size 1000, the percentage of matched incompatible pairs reaches 95%, compared with 53% if we only run two- & three-cycle swap within incompatible pairs. These results are similar to those of Gentry [2007], who also estimate that the proportion of incompatible pairs matched can be doubled by participation of compatible pairs. They focus only on compatible recipients gaining a donor age benefit. Our methods, combined with the LKDPI, also allow us to estimate the benefits to recipients in compatible pairs. If the optimizer maximizes expected survival of grafts over the entire population, there is an increase of 1.21-2 years in expected graft survival among recipients from compatible pairs. If the optimizer instead maximizes number of transplants, this number is between 0.7 and 0.9 years.

An important practical consideration is likely to be that of waiting time. Compatible pairs may not be willing to wait even in order to find a potentially better match. Therefore, we consider a dynamic matching model where the incompatible pool matches either in a greedy or patient fashion (a la Akbarpour et al. [2017]), but the compatible pairs match greedily (from the incompatible pool if it improves the match for the compatible-pair recipient and directly from donor to recipient otherwise). Even with this pessimistic restriction, we estimate substantial benefits, going from matching 35% of incompatible pairs to 55% for the arrival and departure rates we examine.

In this dynamic setting, we also look at the effects on two hard-to-match subpopulations, namely blood group O recipients and highly sensitized patients. We estimate that the positive impacts on blood group O recipients are more substantial than in the general population (an increase from 18% to 46%), while those on the highly sensitized population are similar to the general population (an increase from 24% to 37%).

By bringing quantitative estimates of these benefits into the light, we can inform policy debates. For example, how much expected benefit would be needed to convince compatible pairs to enter an exchange? How long would they be willing to wait in a dynamic setting? These are all questions that can begin to be addressed from the foundation of the models and simulator we develop in this work.
2 KIDNEY EXCHANGE MODEL

In this section we describe the basics of kidney exchange. We largely follow Dickerson and Sandholm’s [2015] description. A kidney exchange can be represented as a directed compatibility graph \( G = (V, E) \) [Roth et al., 2004, 2005a,b]. Each vertex in the graph is a patient-donor pair in the pool. A directed edge \( e \) is constructed from vertex \( v_i \) to vertex \( v_j \) if the patient \( v_j \) is compatible with the donor kidney of \( v_i \). Edges exist or do not exist due to medical characteristics (most importantly blood type, tissue antibodies and antigens) of the patient and the donor. There may also be other logistical constraints, but those are not relevant for our work here. In this pool, the donor of vertex \( v_i \) is willing to give her kidney if and only if the patient of \( v_i \) receives a kidney. A weight \( w_e \) can be assigned to an edge \( e \). While this typically has been used in the literature to represent the priority of a transplantation (and therefore the utility to the system in some senses), we use it to represent the match quality when recipient \( v_j \) receives \( v_i \)’s donor kidney (we discuss the function we use to determine match quality in the next section). In this graph, a sequence of transplants occurs when several vertices form a cycle \( c \). A \( k \)-cycle refers to a cycle with exactly \( k \) pairs. In this paper, we only consider 2-cycles and 3-cycles, as is typical in fielded kidney exchange (incorporating cycles longer than 3 offers limited benefit given logistical constraints). Fielded exchanges also gain from chains, where an altruist donor without a paired patient enters the pool and start a directed path of transplants. We do not include chains in this work.

In this paper, a matching \( M \) is therefore a set of disjoint cycles in the compatibility graph \( G \). The cycles must be disjoint because no donor can give more than one of her kidneys (some recent work explores multi-donor donation [Ergin et al., 2017, Farina et al., 2017] but we do not consider this here). Given a pre-defined utility function \( u : M \rightarrow \mathbb{R} \) and the set of all legal matchings \( M \), we are trying to find a matching which maximizes \( u, M^* \in \arg \max_{M \in M} u(M) \). Kidney exchanges typically find the maximum weighted cycle cover, formally, \( u(M) = \sum_{c \in M} \sum_{e \in c} w_e \). In this paper, we consider two objectives, the number of matches (effectively \( w_e = 1 \forall e \)), and expected total graft survival (where \( w_e \) is defined as the expected graft survival for the recipient in edge \( e \)).

An integer programming (IP) solver is usually used to find the optimal solution [Abraham et al., 2007, Ashlagi et al., 2015, Constantino et al., 2013, Dickerson et al., 2016]. We use the position-indexed chain-edge formulation (PICEF) [Dickerson et al., 2016] method to find the optimal solution when doing two- & three-cycle swap.

3 MODELING MATCH QUALITY

Historically, much work on matching (and welfare economics broadly) has focused on ordinal preferences rather than cardinal utility [Anshelevich and Das, 2010]. This sidesteps the problem of having to make interpersonal comparisons of utility, and research has focused on outcomes in terms of objectives like stability and Pareto optimality [Nisan et al., 2007]. However, with the increasingly important social roles played by matching mechanisms [Diakopoulos, 2015, 2016, O’Neil, 2017, Roth, 2015], it is imperative to understand the outcomes of mechanisms in terms of overall social welfare (however this is defined for a given application) as well as distributional effects. Doing so necessitates considering specific models of utility [Ashlagi and Roth, 2014, Hajaj et al., 2015, Li et al., 2014, Mattei et al., 2017].

There is value in traditional parametric models that are used for utility, and these have been central to model development. Examples of such models include utilities that decay exponentially in waiting time [Akbarpour et al., 2017, Anderson et al., 2017], and random utility models for specific match pairs [Das and Kamenica, 2005]. However, a common criticism of such models is that it is unclear how general or valuable results are when the utility model itself is not grounded.
in reality. In our case, we are explicitly looking for a realistic model that can be used for decision-making. Further, in order to convince compatible pairs to enter kidney exchanges, we must be able to quantify the expected benefit to them in some meaningful manner, therefore, we need an individual model of match quality that can be reasoned about from the perspectives of agents in the market. One important consideration that we defer to future work is the waiting cost to agents in terms of cost and quality of life. For compatible pairs, this is a complex modeling problem from a practical standpoint, because the baseline waiting time is itself highly variable. The time from initial workup to transplantation for a compatible pair is at least several months long because of the barrage of necessary testing, and for part of this time the pair is not even sure that they will be judged compatible. Therefore, in this paper we focus on match quality, and subject our analyses to pessimistic assumptions (greedy dynamic matching), and various robustness checks (varying pool sizes can proxy for match frequency, for example).  

**Quantifying match quality.** Transplant surgeons often have to make decisions on whether a proposed transplant is worthwhile to proceed with. The Kidney Donor Profile Index (KDPI) was developed as a means of assessing the quality of a cadaveric (deceased donor) kidney [Rao et al., 2009]. Recently, the Living Kidney Donor Profile Index (LKDPI) has been proposed as an analog for living donations [Massie et al., 2016]. LKDPI takes into account characteristics of both the donor and the recipient.

KDPI is a percentile score. For example, a score of 4 implies that the kidney is in the “top 4%” of cadaveric kidneys. LKDPI is intentionally designed to be on the same scale (as mentioned in the Introduction, since living donor kidneys can be better than any cadaveric kidney, the LKDPI often takes negative values as well). Therefore, optimizing for LKDPI, while a useful proxy, is semantically ill-founded. However, since LKDPI is computed based on a survival model (Cox regression [Cox and Oakes, 1984]), one can translate the model to a model of expected graft survival (or graft half-life), the survival time of the transplanted organ in the donor [Hariharan et al., 2000].  

We have found that an exponential curve fits the graft half-life as a function of LKDPI almost perfectly (see Figure 1), and can thus estimate expected graft survival as $14.78e^{-0.01239x}$ where x is the LKDPI. We can use this measure in place of LKDPI where it is more appropriate. Thus the edge weight $w_e$ in each cycle is defined as the estimated expected graft survival of the recipient.

### 4 EXCHANGES BETWEEN COMPATIBLE PAIRS: A SINGLE CENTER ANALYSIS

#### 4.1 Data description

Massie et al [2016] come up with the LKDPI measure based on several important characteristics for determining graft survival. We gathered de-identified data on all donor and recipient characteristics that are used in computing LKDPI from all directed living-donor transplants performed at the center in a three year period (2014-2016). There were 166 such transplants with complete characteristics for calculating LKDPI and graft survival; 121 of them also include complete HLA antibody and antigen information. The distribution of each characteristic is shown in Table 1. We also analyze

---

4 The question of how to analyze waiting cost from the perspective of the matching market is also complex; however, one reasonable way to think about it is as costs to the healthcare system. For example, dialysis costs $70,000-$100,000 per year [Held et al., 2016, Krawiec and Rees, 2014, Liyanage et al., 2015], and this is a cost that must be borne by some agent (individuals, private insurance, hospitals, or the government). Incorporating this can be useful when the modeling task is to assess matching policies and how they change costs over the entire system, rather than from the perspective of individual agents, hospitals, and so on.

5 After graft failure, the donor typically needs another transplant.

6 One could also use expected graft survival as input to an expected “Quality Adjusted Life Year” (QALY) [Braithwaite et al., 2008, Torrance and Feeny, 1989, Vergel and Sculpher, 2008] computation over the lifetime of the recipient.
the correlation of every pair of characteristics, shown in Figure 2, which serves as a fundamental building block for designing the simulator in Section 5.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor Age</td>
<td>48.22</td>
<td>12.68</td>
</tr>
<tr>
<td>Donor eGFR</td>
<td>98.11</td>
<td>15.08</td>
</tr>
<tr>
<td>Donor Systolic BP</td>
<td>124.14</td>
<td>13.11</td>
</tr>
<tr>
<td>Donor BMI</td>
<td>27.78</td>
<td>4.46</td>
</tr>
<tr>
<td>Recipient Weight (Female)</td>
<td>180.7</td>
<td>42.26</td>
</tr>
<tr>
<td>Recipient Weight (Male)</td>
<td>190.34</td>
<td>39.9</td>
</tr>
<tr>
<td>Donor Weight (Female)</td>
<td>160.75</td>
<td>30.06</td>
</tr>
<tr>
<td>Donor Weight (Male)</td>
<td>200.8</td>
<td>32.8</td>
</tr>
<tr>
<td>Donor Sex</td>
<td>F: 0.7</td>
<td>M: 0.3</td>
</tr>
<tr>
<td>Rec Sex</td>
<td>F: 0.35</td>
<td>M: 0.65</td>
</tr>
<tr>
<td>Donor African-American</td>
<td>Y: 0.05</td>
<td>N: 0.95</td>
</tr>
<tr>
<td>Donor Cigarette Use</td>
<td>Y: 0.32</td>
<td>N: 0.68</td>
</tr>
<tr>
<td>Donor/Rec Related</td>
<td>Y: 0.50</td>
<td>N: 0.50</td>
</tr>
<tr>
<td>Donor Blood Type</td>
<td>O: 0.6, A: 0.3, B: 0.07, AB: 0.03</td>
<td></td>
</tr>
<tr>
<td>Rec Blood Type</td>
<td>O: 0.46 A: 0.39 B: 0.12 AB: 0.03</td>
<td></td>
</tr>
<tr>
<td>Donor/Rec ABO compatible</td>
<td>Y: 0.88</td>
<td>N: 0.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Donor/Rec related</th>
<th>Donor/Rec unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor/Rec HLA-B Mismatches</td>
<td>0: 0.18, 1: 0.32, 2: 0.5</td>
<td>0: 0.01, 1: 0.1, 2: 0.89</td>
</tr>
<tr>
<td>Donor/Rec HLA-DR Mismatches</td>
<td>0: 0.13, 1: 0.06, 2: 0.81</td>
<td>0: 0.01, 1: 0.06, 2: 0.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Counterfactual Matrix: all unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donor/Rec HLA-B Mismatches</td>
<td>0: 0.009, 1: 0.091, 2: 0.9</td>
</tr>
<tr>
<td>Donor/Rec HLA-DR Mismatches</td>
<td>0: 0.02, 1: 0.04, 2: 0.94</td>
</tr>
</tbody>
</table>

Table 1. Distribution of each characteristic of the center’s data. F/M means Female/Male, Y/N represents Yes/No, and Rec is a shortening of Recipient.
4.2 Counterfactual analysis within the center

Typically, if a donor and recipient are deemed medically compatible, a directed transplant is performed, with the donor’s kidney going to the recipient. However, there may be cases where the match quality is low even if they are compatible, and perhaps the recipient could receive a better kidney through an exchange; for example, they may be able to receive a kidney from a younger donor, or avoid an immunologically risky donor/recipient combination, like child to mother or husband to wife [Gentry et al., 2007]. Such scenarios are hypothetical, and may seem unlikely at first glance. To validate our conjecture, for these donor-recipient pairs, we computed the expected graft survival (EGS) of each pair, and then performed counterfactual simulations to assess the potential to improve outcomes. The two counterfactual simulations share a Pareto improvement restriction—no recipient may receive a kidney with a shorter (i.e., worse) EGS for them than the EGS for them of the kidney from their original paired donor.

**Optimal.** In the first simulation, we find the best matching, allowing arbitrary length cycles; this can also be treated as a bipartite matching problem (with the restriction that the matching must be perfect) between donors on one side and recipients on the other.

**Two and Three-cycle swap.** In the second simulation, we only allow either a direct donation from the donor to the recipient or through a two- and three-cycle kidney exchange, to more closely approximate realistic logistical constraints.

We consider two subsets of the data. The “complete” 166-pair subset, assuming no HLA incompatibilities, and the “restricted” subset of 121 pairs for which we have complete antibody/antigen information and can determine all incompatibilities and rule out such transplants. The distribution of EGS and corresponding LKDPI among the real pairs and in the results of our counterfactual simulations are shown in Figure 3. The mean and median EGS and LKDPIs are given in Tables 2 and 3 below.

We can see there is a median improvement of 1.93 years of expected graft survival for the Optimal and 1.38 years for the two- & three-cycle swap (over a median half-life of 10.84 years). We also see that including compatibility constraints itself does not have a huge effect on the results.

![Fig. 2. Correlation matrix of each pair of characteristics.](image)
Table 2. Mean and median EGS for two counterfactual simulations, compared to reality over the last three years at the center. Figure (a) shows the 121-subset of data with HLA antigens and antibodies, and Figure (b) shows the 166-subset assuming no incompatibilities.

Table 3. Mean and median LKDPI for two counterfactual simulations, compared to the reality over the last three years at the center. Figure (a) shows the 121-subset of data with HLA antigens and antibodies, and Figure (b) shows the 166-subset assuming no incompatibilities.

Fig. 3. Distribution of the expected graft survival (left) and LKDPI (right) of the original matched pairs and matched pairs in the two counterfactual simulations, using 121 subset of real data with HLA antigens and antibodies from the center over the last three years.

(some of the improvement in the larger set is simply due to having a thicker market). Beyond the specific results, it is surprising to see the high number of transplants that were performed with LKDPIs above 50, since these indicate that the average cadaveric kidney would have been better for the recipient, in contrast to the conventional wisdom that living donor kidneys are always better. The optimized matches from the counterfactual “exchange” are much better, with many fewer “bad” matches and many more with LKDPI of 20 or lower, predictive of excellent outcomes.
Fig. 4. Distribution of the expected graft survival (left) and LKDPI (right) of the original matched pairs and matched pairs in the two counterfactual simulations, using the 166 full dataset of real data from the center over the last three years. We can see the distribution is similar to Figure 3.

4.3 Discussion
This is a proof-of-concept for the potential of improving quality of matching. One immediate question arises from the fact that we are using three years worth of data on recipients and donors in a static setting; this is obviously unrealistic. However, the main point is to estimate realistic distributions from data; we can use projections to then analyze differently-sized static markets (from smaller ones to larger ones that could be realized through regional pooling or already-functioning national exchanges). We turn to these questions and beyond in the next section.

5 INCLUDING COMPATIBLE PAIRS IN KIDNEY EXCHANGES
In addition to improving match quality, we may also be able to improve the number of matches by including compatible pairs to thicken the exchange with incompatible pairs. This could also lower costs for transplant centers by allowing for more internal matches where the transplant center does not need to go to a regional or national exchange to find a match for an incompatible pair. In order to estimate the possible benefits more systematically over different possible population sizes, we need to efficiently and correctly simulate LKDPIs over donor and recipient populations.

5.1 Determinants of match quality, and design of a Compatibility+LKDPI simulator
This would be simple if LKDPIs were distributed in a manner that was easy to correctly estimate, for example, independently, or independently conditional on the LKDPI of the original compatible pair. Unfortunately, this turns out not to be the case. To get a simple benchmark of how much this may affect the results, we can simulate different distributions based on data from the center.

We first build a counterfactual matrix of estimated graft survival based on the original (166-pair) data by calculating LKDPI values for each of these 166 pairs. We then investigate the expected graft survival of donor-patient pairs under the Optimal and two- & three-cycle swapmatching algorithms when resampling the matrix in different ways. To simulate independent LKDPIs, we resample individual LKDPIs from the whole matrix. To simulate donor-dependent LKDPIs, we shuffle all donors for a given recipient, and to simulate recipient-dependent LKDPIs, we shuffle all recipients for a given donor. The results are shown in Table 4. The first row shows statistics.

---

7This could have positive and negative effects overall, by perhaps increasing fragmentation, but lowering costs. However, many centers choose not to participate in broader exchanges much of the time in practice, for a variety of reasons.
from the original compatible matching. As we see, most of the methods for generating LKDPIs vastly overestimate the possible gains, and the evidence is consistent with the observation that the determination of LKDPI/expected graft survival is largely based on the donor’s characteristics [Massie et al., 2016]. These results demonstrate the need for a good simulator.

The central empirical facts that allow us to construct an efficient simulator are analyses of the joint distributions of variables involved in determining compatibility (PRA and ABO compatibility, based on the simulator of Saidman et al. [2006]) and computing LKDPI (See Table 1), and analysis of the possible underlying mechanisms of dependence. In particular, compatibility is solely a function of blood type and antibodies, while LKDPI considers many other factors, most of which have limited relationship to those (Figure 2). Since the state of practice in kidney transplantation has been to always assume that any living donor is excellent (a practice called into question by our results above), it is unlikely that there is any selection bias in the characteristics we sample for typical compatible pair arrivals. We first generate a donor-recipient pair, with all LKDPI-related characteristics generated sequentially in a manner that respects the data distributions in Table 1 and the correlation structure shown in Figure 2. We then generate the PRA (percentage reactive antibodies) and compatibility based on the Saidman model. Details of our simulator are in Appendix A, Algorithm 1. The last line of Table 4 shows that the simulator produces results very close to the real data.

<table>
<thead>
<tr>
<th>Method</th>
<th>EGS original</th>
<th>EGS 2&amp;3 swap</th>
<th>EGS Optimal</th>
<th>LKDPI original</th>
<th>LKDPI 2&amp;3 swap</th>
<th>LKDPI Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original 166 dataset</td>
<td>9.67</td>
<td>11.14</td>
<td>11.58</td>
<td>37.15</td>
<td>25.50</td>
<td>22.46</td>
</tr>
<tr>
<td>Sample from the whole matrix</td>
<td>9.23</td>
<td>14.40</td>
<td>15.30</td>
<td>40.51</td>
<td>2.67</td>
<td>-2.5</td>
</tr>
<tr>
<td>Shuffle all donors per recipient</td>
<td>9.19</td>
<td>14.16</td>
<td>14.94</td>
<td>40.92</td>
<td>4.11</td>
<td>-0.47</td>
</tr>
<tr>
<td>Shuffle all recipients per donor</td>
<td>9.21</td>
<td>11.74</td>
<td>12.50</td>
<td>40.70</td>
<td>20.6</td>
<td>15.49</td>
</tr>
<tr>
<td>Sample from the simulator</td>
<td>9.38</td>
<td>11.40</td>
<td>11.80</td>
<td>39.21</td>
<td>24.50</td>
<td>20.09</td>
</tr>
</tbody>
</table>

Table 4. The EGS and LKDPI comparison of different sampling methods and different market clearing algorithm.

5.2 Experimental results using the LKDPI simulator

We can now use the LKDPI simulator to estimate the benefits in terms of both quality and quantity of transplants. We study the impact of different optimization objectives (survival and number of matches) on outcomes for both compatible and incompatible pairs. We are most interested in the improvement of (1) expected graft survival of compatible pairs compared with their original donation, since the incentive for compatible pairs to enter is to seek a better organ for the recipient; (2) the number of matched incompatible pairs compared with the number when running two- &
three-cycle swap only on incompatible pairs. We find the maximum weighted cycle cover, where
the weight can be (1) \( w_e = \text{expected graft survival of recipient} \), (2) \( w_e = 1 \), (maximizing
the number of matched pairs). Table 5 summarizes the possible objectives and the metrics that we measure.

In our first set of experiments, we fix the size of the pool and generate donor-recipient pairs
using the simulator. We find that the sizes of the compatible and incompatible pool are roughly
even. This matches the statistics of the center we have data from. In 2017, 217 compatible pairs and
181 incompatible pairs registered for initial transplant workups (though only 1/3 of them ended up
having a transplantation procedure in the center).

We then run two- & three-cycle swap under the Pareto improvement restriction, where compatible
pairs only swap if the expected graft survival is longer than their original ones. We find that (Figure
5), with participation of compatible pairs, the percentage of matched incompatible pairs doubles to
64% compared with only running two- & three-cycle swap within the incompatible pairs (33%).
This result is from the pool size of 50 donor-recipient pairs, and still holds no matter whether
we maximize the expected graft survival of the whole graph \( G \) or the number of matched pairs.
When the pool size increases to 1000, the percentage of matched incompatible pairs reaches 95%,
compared with 53% if we only run two- & three-cycle swap within incompatible pairs. These
results are similar to the results of Gentry [2007], where they also estimate that the proportion
of incompatible pairs matched could be doubled by participation of compatible pairs that would
gain a donor age benefit. From the perspective of compatible pairs, there is 2.04-2.36 years graft
survival improvement (for those who have improvement) if we focus on maximizing the expected
survival of the whole population, and 1.20-1.59 years graft survival improvement (for those who
have improvement) if we focus on simply maximizing the number of matched incompatible pairs.

### 5.2.1 Varying the size of the incompatible pair pool

While the rate of entry of compatible and incompatible pairs may be similar, it is possible that one or the other population is less likely to go through with a transplant. This could result in different ratios between the sizes of the two pools.

In order to study how our results would vary with different assumptions about this, we hold the
number of compatible pairs fixed and vary the number of incompatible pairs. Both compatible and
incompatible pairs are randomly generated using the population characteristics from Table 1 and
following Algorithm 1, where the compatibility is decided by Saidman’s simulator. The number
of compatible pairs we consider are 50, 100, 200 and 300, while the number of incompatible pairs
ranges from 10 to 200.

The performance of incompatible pairs – The number of matched pairs. For incompatible pairs,
we are primarily interested in the increase in the number of matches when compatible pairs join

<table>
<thead>
<tr>
<th>C-Or</th>
<th>ENM</th>
<th>P-I</th>
<th>P-CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compatible pairs original donation</td>
<td>Expected number of matched pairs</td>
<td>Pool with only incompatible pairs</td>
<td>Pool with both incompatible and compatible pairs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maximize EGS</th>
<th>Maximize ENM</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-maxSur</td>
<td>I-MaxNum</td>
</tr>
<tr>
<td>I-O-MaxSur</td>
<td>I-O-MaxNum</td>
</tr>
<tr>
<td>CI-MaxSur</td>
<td>CI-MaxNum</td>
</tr>
<tr>
<td>CI-O-MaxSur</td>
<td>CI-O-MaxNum</td>
</tr>
</tbody>
</table>

Table 5. Table indexing abbreviations we use corresponding to different optimization objectives, matching methods, and different subpopulation measurements.
Zhuoshu Li, Sanmay Das, Sofia Carrillo, and Jason Wellen

Fig. 5. The comparison between expected graft survival of compatible pairs by participating two- & three-cycle swap (blue solid line) and their original matching (blue dash line), expected graft survival of incompatible pairs when compatible pairs participate two- & three-cycle swap (red solid line) and only within incompatible pairs (red dashed line), and proportion of matched incompatible pairs when compatible pairs participate two- & three-cycle swap (black solid line) and only within incompatible pairs (black dashed line), where Figure (a) shows the results of maximizing the expected graft survival across the whole graph $G$, and Figure (b) shows the results of maximizing the number of matched pairs.

-the pool. Figure 6a shows the expected number of matched incompatible pairs/recipients when maximizing expected graft survival of all cycles (*-MaxSur) and maximizing the number of matched pairs ($w_c = 1$, *-MaxNum). For both objective functions, the optimal matching will match all the pairs (I/CI-O-MaxSur/MaxNum). In two- & three-cycle swap, both objective functions achieve similar performance (though *-MaxNum are slightly better then *-MaxSur). When the market is thick enough (compatible size is 300, CS=300), the number of matched incompatible pairs is very close to the optimal solution. In general, for two- & three-cycle swap, the pool with compatible pairs (CI-*) matches far more incompatible recipients than only running two- & three-cycle swap within the incompatible pairs (I-*).

-Expected graft survival. We now investigate how expected graft survival of incompatible pairs changes when compatible pairs join the pool. The results of comparing *-MaxSur and *-MaxNum can be found in Figure 6b. Overall, *-MaxSur (solid lines) has longer expected graft survival than *-MaxNum (dash lines) as we expect. When compatible pairs participate, expected graft survival of incompatible pairs is lower than when running two- & three-cycle swap within incompatible pairs (I-*). Another interesting observation is that the expected graft survival of incompatible recipients decreases as the number of compatible pairs increases for both *-MaxSur and *-MaxNum.

The performance of compatible pairs – Expected graft survival. Under the Pareto improvement restriction, the compatible pairs are guaranteed to match with their original donor at least and they only swap if they can find a better organ for both the Optimal and two- & three-cycle swap. From Figure 7a we can see that for both objective functions (MaxSur and MaxNum), compatible pairs have a substantially longer graft survival for participating two- & three-cycle swap (CI-*) than if matched with their original donor (C-or). The size of the compatible pool does not have major influence on the performance. It is also obvious that the compatible pairs benefit more when the market clearing algorithm maximizes the expected graft survival rather than the number of matched pairs. The number of incompatible pairs who are not matched when maximizing graft
Fig. 6. (a) Expected number of matched incompatible pairs under maximizing expected graft survival (solid lines) and expected number of matched recipients (dash lines) when holding the number of compatible pairs (CS) as 50, 100, 200, 300; (b) Expected graft survival of incompatible recipients under maximizing expected graft survival (solid lines) and expected number of matched recipients (dash lines) when holding the number of compatible pairs (CS) as 50, 100, 200, 300 and varying the size of incompatible pairs from 10 to 210. Each point in the graph is an average of 500 simulations.

survival, but who would have been matched when maximizing the number of matches, is shown in Figure 7b.

Fig. 7. two- & three-cycle swap: Expected graft survival of compatible recipients under maximizing expected graft survival (solid lines) and expected number of matched recipients (dash lines) when holding the number of compatible pairs (CS) as 50, 100, 200, 300 and varying the size of incompatible pairs from 10 to 210.

Figure 8 provides an upper bound on expected graft survival of compatible pairs for maximizing graft survival (CI-O-MaxSur) and maximizing the number of matches (CI-O-MaxNum) by running the optimal matching algorithm. When the number of compatible pairs (CS=50) is 50, there is a roughly 1 year improvement for running optimal matching comparing to running two- & three-cycle swap if maximizing EGS, and a 0.5 year improvement if maximizing the number of matches.
6 MODELING DYNAMIC MARKETS

While different sizes of pools can proxy for different match frequencies, and patience in matching can improve outcomes [Akbarpour et al., 2017], there are also good arguments and practical concerns that favor greedy or frequent matching [Anderson et al., 2017, Das et al., 2015]. In particular, it is a reasonable, if pessimistic, assumption, that compatible pairs would be completely unwilling to wait, and would therefore insist on an immediate exchange, or else they would want to go ahead with the direct donation.

In this section, we build a dynamic model where patient-donor pairs arrive gradually over time. Incompatible pairs stay in the market until they find an acceptable swap or they perish (they may leave the market if the patient’s condition deteriorates to the point where kidney transplants become infeasible, for example). Compatible pairs must either be matched with an incompatible pair at the moment of arrival, or else the donor gives directly to the recipient immediately.

Arriving patient-donor pairs are still generated from our simulation model described above. Pairs arrive at the market according to a Poisson process, with rate parameter $m \geq 1$. The sojourn of an agent is drawn from an exponential distribution, with rate parameter $\lambda = 1$. If a compatible pair arrives, all feasible swaps for that pair are considered (where feasibility means that both compatibility requirements are satisfied and the recipient in the compatible pair receives a higher-quality kidney match). If there is more than one acceptable swap the newly entered pair chooses the one with the longest expected graft survival for its own recipient; ties are broken uniformly at random. Incompatible pairs can be matched either greedily, in the same manner as above, or using a *patient* algorithm [Akbarpour et al., 2017] which waits to match until the moment an agent is about to perish (with the caveat that incompatible pairs with possible matches in the incompatible pool are not considered as matches for compatible pairs).
We again find a substantial benefit in terms of the number of incompatible pairs matched under either mechanism (from approximately 35% to approximately 55%). Figure 9 also shows that compatible pairs for whom the new mechanism changes the match improve their expected graft survival by almost two and a half years. There is some effect of “competition” – since match quality is largely a function of the donor, and compatible pairs need to receive good donors in order to participate, the average expected graft survival of those who are transplanted in the incompatible pool actually goes down; however, the huge increase in the number of matches more than compensates in terms of the sum total of years of graft life (where pairs that don’t receive a transplant are assigned an EGS of 0). Therefore, our results demonstrate that the potential value of incorporating compatible pairs is high even under pessimistic assumptions about what wait times they would be willing to tolerate.

Fig. 9. Comparison of matched proportion of incompatible patients (left) and change in EGS (right) when running different matching algorithms for the incompatible pool in the dynamic setting.

6.1 Fairness considerations: Hard to match types

An important consideration in kidney exchanges is how they may differentially affect different populations. The populations one often worries about are those who are harder to match. Therefore, we consider the effects on two groups of hard-to-match patients, those with blood group O (patients with blood group O have fewer ABO-compatible living donors [Glander et al., 2010]) and highly sensitized patients, who are likely to have antibodies to a significant fraction of the population. We define highly sensitized patients as those whose PRA is greater than 80%, constituting approximately 30% of the patient population.

Figure 10-left shows that there is a significant improvement for the matched proportion of incompatible blood type O patients (from 0.18 to 0.46) when incorporating compatible pairs. Therefore the relative benefit to this group is actually higher than to the rest of the population. The matched proportion of incompatible highly sensitized patients improves to 37% from 24% when compatible pairs are included, a rate of increase roughly similar to that in the overall population.

7 CONCLUSION AND FUTURE RESEARCH

Living donor kidney transplantation has proven to be an important domain for the development of matching theory and algorithms. It is becoming increasingly important to study cardinal utilities in kidney exchange, and we believe this could open up more fertile avenues for research. Our main goal in this paper is to develop the framework and a robust framework for analyzing match quality in models of kidney exchange. Our framework is based on real donor and recipient data from a
major transplant center. We have also used the model to estimate the benefits, in terms of both quantity and quality of transplants, of including compatible pairs in kidney exchange. We find that if we were able to induce compatible pairs to join kidney exchanges, the percentage of matched incompatible pairs would increase dramatically, and there would also be a substantial increase in expected graft survival for recipients in compatible pairs. Quantifying the potential quantitative benefits of participating through LKDPI may also make compatible pairs more likely to join.

While our work here is largely in a static setting and a simple dynamic setting, the development of our realistic LKDPI simulator allows for the investigation of many different matching models. Of particular interest will be questions related to matching policy in the dynamic setting with both compatible and incompatible pairs, incorporating wait times, and possible systemic effects of including compatible pairs in exchanges, for example, changes in incentives for centers.

REFERENCES


Algorithm 1: Details of Generating Living Donor Pair

**Input:** Pair ID  
**Output:** A living donor pair

1. **Sample following characteristics based on the distribution of Table 1:**
2. Donor Age $\sim \mathcal{N}(48.22, 12.68)$;
3. Donor Sex: $P(F) = 0.7$, $P(M) = 0.3$;
4. Rec Sex: $P(F) = 0.35$, $P(M) = 0.65$;
5. Donor eGFR: Table-6 based on Donor Age;
6. Donor SBP Table-6 based on Donor eGFR;
7. Donor Weight: Sample based on Donor Sex;
8. Rec Weight: Sample based on Recipient Sex;
9. Donor BMI: $0.0948 \times \text{Donor Weight} + 11.387$;
10. Donor/Rec Weight Ratio: Donor Weight/Rec Weight;
11. Donor Blood Type: Based on Saidman’s simulator;
12. Recipient Blood Type: Based on Saidman’s simulator;
13. Donor is African American: Based on Donor Blood Type;
14. Donor cigarette use: $P(Y) = 0.32$, $P(N) = 0.68$;
15. Donor&Rec Related: Based on Table 1;
16. Check Donor&Rec ABO compatibility;
17. Donor&Rec HLA-B mismatches and Donor&Rec HLA-DR mismatches: Jointly sample from Table 1 based on whether the pair is related or not;
18. Donor&Rec isWifePatient: Based on Saidman’s simulator; if the recipient is female and the donor-recipient pair is unrelated, the probability that the donor is the recipient’s spouse is $0.4897$;
19. Recipient PRA: Based on Saidman’s simulator;
20. Generate crossmatch incompatibility: Based on PRA and isWifePatient;
21. Determine compatibility: The pair is compatible if and only if both ABO compatible and a negative crossmatch.

<table>
<thead>
<tr>
<th>Age(Years)</th>
<th>Average Measured GFR (ML/min/1.73m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-29</td>
<td>116</td>
</tr>
<tr>
<td>30-39</td>
<td>107</td>
</tr>
<tr>
<td>40-49</td>
<td>99</td>
</tr>
<tr>
<td>50-59</td>
<td>93</td>
</tr>
<tr>
<td>60-69</td>
<td>85</td>
</tr>
<tr>
<td>70+</td>
<td>75</td>
</tr>
</tbody>
</table>

Table 6. Average measured GFR by age in people.

### A SIMULATOR DETAILS

Our basic simulation model is based on the distribution of all relevant recipient and donor characteristics from the data of the center. The characteristics of each donor-recipient pair are generated from the distribution of the center’s data (See Table 1). We determine compatibility based on the simulator from Saidman et al [2006], which utilizes PRA and ABO compatibility. More specifically,
we first generate a donor-recipient pair, with all LKDPI-related characteristics generated sequentially in a manner that respects the data distributions in Table 1 and the correlation structure shown in Figure 2. We then generate the PRA (percentage reactive antibodies) and compatibility based on the Saidman model. The exact details of how we generate the characteristics can be found in Algorithm 1.

To note, in this simulator, (1) the estimated GFR (line 5) is generated from Table 6 which depends on age instead of using the distribution from Table 1\(^8\); (2) The BMI (line 9) is generated based on a regression on data from the transplant center; (3) When we consider a counterfactual pair, we always assume they are unrelated. (4) HLA-B and HLA-DR mismatches of a donor-recipient pair are generated based on whether the donor and recipient are related or not. When we need to decide the HLA-B and HLA-DR mismatches of a counterfactual pair, we use the distribution from the counterfactual matrix instead of the distribution from the original dataset.

\(^8\)See https://www.kidney.org/sites/default/files/docs/12-10-4004_abefaq5gtrrev1b_singleb.pdf.