Modern Artificial Intelligence for Health Analytics: Papers from the AAAI-14

Balancing Efficiency and Fairness in Dynamic Kidney Exchange

John P. Dickerson
Carnegie Mellon University
dickerson@cs.cmu.edu

Tuomas Sandholm
Carnegie Mellon University
sandholm@cs.cmu.edu

Abstract

The preferred treatment for kidney failure is a transplant; however, demand for donor kidneys far outstrips supply. *Kidney exchange*, an innovation where willing but incompatible donor-patient pairs can exchange organs—via cycles and altruist-initiated chains—provides a life-saving alternative to long waiting lists for deceased-donor organs. Typically, fielded exchanges act *myopically*, considering only the current pool of pairs when planning the cycles and chains. Yet kidney exchange is inherently dynamic, with participants arriving and departing. Also, many planned exchange transplants do not go to surgery due to various failures. Thus, it is important to consider the future when matching.

Motivated by our experience running the computational side of the US nationwide kidney exchange, we present FUTUREMATCH, a framework for *learning to match in a general dynamic model*. We validate FUTUREMATCH on the real nationwide exchange data. Not only does dynamic matching result in more expected transplants than myopic matching, but also even dynamic matching under economically inefficient objectives that enforce equity can result in statistically significant increases in social welfare over efficient myopic matching.

Introduction

Chronic kidney disease is a life-threatening health issue that affects millions of people worldwide; its societal burden is likened to that of diabetes (Neuen et al. 2013). Damage from kidney disease can cause irreparable loss of organ function and, eventually, complete kidney failure. Such failure requires either continual dialysis or an organ transplant to maintain life.

The preferred treatment for kidney failure is transplantation. However, the demand for donor kidneys is far greater than supply. In the US alone, the waiting list for a kidney transplant had 100,175 patients as of April 21, 2014 (UNOS). Demand is increasing: for example, 36,397 people were added to the US national waiting list in 2013, while only 16,415 left it due to receiving a kidney.

Patients can receive a transplant organ from either a deceased or living donor. Roughly two thirds of transplanted kidneys are sourced from cadavers, while one third come from living donors. Patients who are fortunate enough to find a willing living donor must still contend with *compatibility* issues like blood type, tissue type, and other medical or logistical factors. If a willing would-be donor is incompatible with a patient, the kidney cannot be transplanted.

Kidney exchange (Rapaport 1986) is a recent innovation that allows patients to swap willing but incompatible donors. Figure 1 shows a graphical view of a pool consisting of three patient-donor pairs, where an arrow from pair i to pair j means the patient at j is compatible with the donor at i. Also shown is a donor without a paired patient who is willing to donate a kidney altruistically. The basic kidney exchange problem is then to recommend a "good" set of organ swaps.

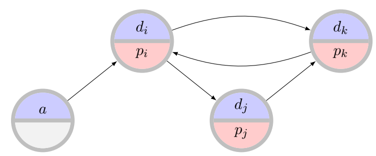

Figure 1: Example kidney exchange pool with three patient-donor pairs and one altruistic donor.

In this paper, we concentrate on balancing two objectives—efficiency and fairness—in *dynamic* kidney exchange, where the future composition of the kidney exchange pool is explicitly considered during the optimization process.[1] In terms of efficiency, we are interested in maximizing the total number of transplants performed over time. As in many healthcare applications, emphasizing overall efficiency can come at disproportional cost to certain classes of patients; thus, we also consider fairness, where "hard-to-match" patients are given varying levels of preference during the matching process.

Motivated by our experience running the computational side of the United Network for Organ Sharing (UNOS) US nationwide kidney exchange, which has grown to include 133 transplant centers since its inception in Oct. 2010, we present FUTUREMATCH, a general framework for learning how to match in dynamic environments. To our knowledge, FUTUREMATCH is the first data-driven learning framework for *complex* (i.e., where the goal is more complicated than just pairing up vertices) online matching. We validate the framework on real data drawn from the nationwide exchange. We find that using FUTUREMATCH *even with economically inefficient objectives*—like maximizing the match

[1] This workshop paper summarizes recent work (Dickerson, Procaccia, and Sandholm 2014; Dickerson and Sandholm 2014).

size subject to equity constraints—results in significantly higher social welfare than efficient but myopic matching.

Kidney Exchange Model

We begin by overviewing the standard computational model for kidney exchange, which encodes a kidney exchange pool as a directed *compatibility graph* $G = (V, E)$ by constructing one vertex for each patient-donor pair in the pool. An edge e from v_i to v_j is added if the patient in v_j wants the donor kidney of v_i. A paired donor is willing to give her kidney if and only if the patient in her vertex v_i receives a kidney. The weight w_e of an edge e represents the utility to v_j of obtaining v_i's donor kidney.

A cycle c in the graph G represents a possible kidney swap, with each vertex in the cycle obtaining the kidney of the previous vertex. If c includes k patient-donor pairs, we refer to it as a *k-cycle*. For example, the compatibility graph in Figure 1 includes two possible cycles: a 2-cycle between vertex v_i and v_k, and a 3-cycle consisting of vertices v_i, v_j, and v_k. In kidney exchange, cycles of length at most some small constant L are allowed—all transplants in a cycle must be performed simultaneously so that no donor backs out after his patient has received a kidney but before he has donated his kidney. In most fielded kidney exchanges, including the United Network for Organ Sharing (UNOS) nationwide kidney exchange, $L = 3$ (i.e., only cycles of length at most 3 are allowed).

Fielded kidney exchanges gain great utility through the use of chains (Roth et al. 2006; Rees et al. 2009). Chains start with an altruistic donor donating his kidney to a patient, whose paired donor donates her kidney to another patient, and so on. The compatibility graph in Figure 1 includes four possible chains: $\langle a, v_i \rangle$, $\langle a, v_i, v_j \rangle$, $\langle a, v_i, v_k \rangle$, and $\langle a, v_i, v_j, v_k \rangle$. Chains can be (and typically are) longer than cycles in practice because it is not necessary to carry out all the transplants in a chain simultaneously. There is a chance that a bridge donor backs out of his/her commitment to donate—which has happened (albeit rarely) already in the United States. Cycles cannot be executed in parts because if someone backs out of a cycle, then some pair has lost a kidney (i.e., their "bargaining chip"). In contrast, if someone backs out of a chain, no pair has lost their bargaining chip (although it is unfortunate that the chain ends).

A *matching* M is a collection of disjoint cycles and chains in the graph G. The cycles and chains must be disjoint because no donor can give more than one of her kidneys. Given the set of all legal matchings $\mathcal{M}$, the *clearing problem* in kidney exchange is to find a matching M^* that maximizes some utility function $u : \mathcal{M} \to \mathbb{R}$. Formally:

$$M^* = \underset{M \in \mathcal{M}}{\operatorname{argmax}}\, u(M)$$

The standard clearing problem for finite cycle cap $L > 2$ is NP-hard (Abraham, Blum, and Sandholm 2007). Abraham, Blum, and Sandholm (2007) took the first serious computational step toward solving the kidney exchange problem by providing a specialized branch-and-price-based (Barnhart et al. 1998) integer program solver; subsequent work by Dickerson, Procaccia, and Sandholm (2013; 2014) has increased solver speed and generality. We use an adapted version of that clearing algorithm as the batch clearing algorithm module in this work.

Three Sample Utility Functions

In fielded kidney exchanges, one typically finds the maximum weighted cycle cover (i.e., $u(M) = \sum_{c \in M} \sum_{e \in c} w_e$). This *utilitarian* objective can favor certain classes of patient-donor pairs while marginalizing others, a phenomenon that we explore—and help alleviate—in this paper.

The medical and legal communities in kidney exchange are concerned about a wide variety of match characteristics. In our experience, the most frequently discussed include the number of overall matches, the number of overall transplants, the quality of transplants, and the preference applied, if any, to specific subgroups in the exchange (children, sensitized patients, underrepresented ethnicities). Other concerns might include some sort of equitable treatment between participating transplant centers, minimizing legal exposure, and fair compensation.

In this paper, we consider two different kidney exchange models—deterministic, where post-algorithmic match failures are not quantified in the optimization problem and failure-aware, where they are—and three formal matching objectives in each of the two models:

1. MAXCARD: Maximize the total number (i.e., cardinality) of patients who are algorithmically matched (in the deterministic model) or receive transplants in expectation (in the failure-aware model);
2. MAXCARD-FAIR: Maximize the total number of patients who are algorithmically matched (in the deterministic model) or receive transplants in expectation (in the failure-aware model), where "marginalized" patients are weighted in the objective by some constant factor β more than others; and
3. MAXLIFE: Maximize the total time algorithmically-matched (deterministic) or transplanted (failure-aware) donor organs will last in patients.

Each of these objectives amounts to setting weights on edges in the input graph (e.g., Figure 1). We give edge weighting algorithms for the MAXCARD and MAXCARD-FAIR objectives.[2] The MAXCARD-FAIR objective can viewed as a generalized form of MAXCARD (that is, MAXCARD is just MAXCARD-FAIR with an empty set of vertices who are preferred by the objective).

Deciding which class of vertices are preferred is a complex ethical and medical decision. We use two common preference criteria in this paper: pediatric status and sensitization. Children (in the US, those who are under age 18) are typically treated preferentially in medical systems; we follow that rule here. For kidney exchange it has explicitly been articulated that such pediatric patients should be preferred not only because they have a lot of life left (barring their kidney disease) but also because having poor kidney function stunts growth. Some patients are *highly sensitized*, which means they are extremely unlikely to be medically compatible with a random organ. For these patients, finding a kidney is difficult (UNOS). The percentage of highly-sensitized patients in fielded kidney exchanges is high; over 60% of the patients in the UNOS kidney exchange are highly sensitized (Kidney Paired Donation Work Group 2013).

[2] Due to space constraints, we guide the reader to §3.2.2 of Dickerson and Sandholm (2014) for a derivation of the MAXLIFE edge weighting function, which learns a match quality metric from historical US data between 1987 and mid-2013.

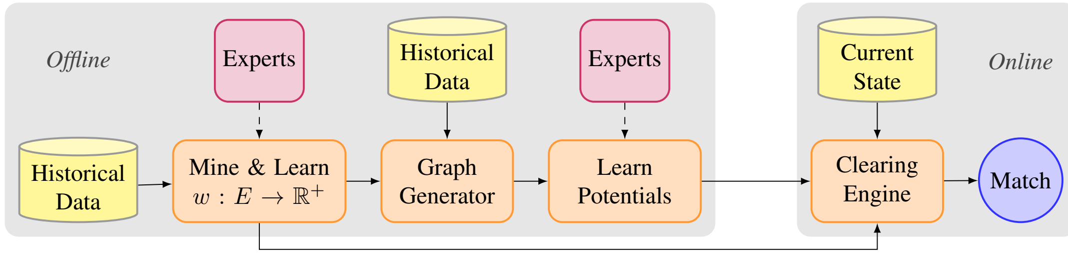

Figure 2: The FUTUREMATCH framework.

	$\lvert V\rvert = 300$		$\lvert V\rvert = 400$		$\lvert V\rvert = 500$		$\lvert V\rvert = 600$		$\lvert V\rvert = 700$		$\lvert V\rvert = 800$		$\lvert V\rvert = 900$	
	Gain	p	*Gain*	p	*Gain*	p	*Gain*	p	*Gain*	p	*Gain*	p	*Gain*	p
MAXCARD	+2	✓	+4	✓	+5	✓	+6	✓	+10	✓	+11	✓	+13	✓
MAXCARD-FAIR, $\beta = 1$	+1	✓	+4	✓	+6	✓	+8	✓	+9	✓	+11	✓	+12	✓
MAXCARD-FAIR, $\beta = 2$	+1		+2	✓	+3	✓	+3	✓	+5	✓	+6	✓	+10	✓
MAXCARD-FAIR, $\beta = 3$	+1		+0		+3	✓	+1		+1	✓	+3	✓	+2	
MAXCARD-FAIR, $\beta = 4$	-1		+1		+1		+1		+3	✓	+3		+2	
MAXCARD-FAIR, $\beta = 5$	+0		+0		+1		+1		+1		+2		+3	
MAXLIFE	+2	✓	+3	✓	+6	✓	+8	✓	+7	✓	+11	✓	+9	✓

Table 1: Median gains in expected total number of transplants under FUTUREMATCH. A ✓ represents statistical significance (Wilcoxon signed-rank test, $p \ll 0.01$).

	$\lvert V\rvert = 300$		$\lvert V\rvert = 400$		$\lvert V\rvert = 500$		$\lvert V\rvert = 600$		$\lvert V\rvert = 700$		$\lvert V\rvert = 800$		$\lvert V\rvert = 900$	
	Gain	p	*Gain*	p	*Gain*	p	*Gain*	p	*Gain*	p	*Gain*	p	*Gain*	p
MAXCARD	-2	✗	-2	✗	-3	✗	-4	✗	-6	✗	-7	✗	-9	✗
MAXCARD-FAIR, $\beta = 1$	-1	✗	-1	✗	-1	✗	-2	✗	-3	✗	-3	✗	-5	✗
MAXCARD-FAIR, $\beta = 2$	+0		+0		+1	✓	+1	✓	+2	✓	+1		+1	
MAXCARD-FAIR, $\beta = 3$	+1	✓	+1	✓	+3	✓	+3	✓	+3	✓	+5	✓	+4	✓
MAXCARD-FAIR, $\beta = 4$	+1	✓	+2	✓	+3	✓	+4	✓	+4	✓	+5	✓	+5	✓
MAXCARD-FAIR, $\beta = 5$	+1	✓	+2	✓	+3	✓	+4	✓	+5	✓	+7	✓	+5	✓
MAXLIFE	-1	✗	-3	✗	-3	✗	-5	✗	-6	✗	-6	✗	-9	✗

Table 2: Median gains in expected total number of marginalized transplants under FUTUREMATCH. A ✓ or ✗ represents statistically significant gains or losses, respectively (Wilcoxon signed-rank test, $p \ll 0.01$).

While defining fairness is a contentious issue in social science, a recent paper by Dickerson, Procaccia, and Sandholm (2014) formalizes two natural "fair" utility functions and shows how to optimize either of these functions in the deterministic or failure-aware *static* models of kidney exchange. We adapt the *weighted* fairness rule from that paper to FUTUREMATCH. The weighted fairness rule adjusts edge weights by some re-weighting function $\Delta : E \to \mathbb{R}^+$. A simple example re-weighting function is multiplicative:

$$\Delta^{\beta}(e) = \begin{cases} (1+\beta)w_e & \text{if } e \text{ ends in } V_M \\ w_e & \text{otherwise} \end{cases}$$

Here, $V_M \subseteq V$ is the set of vertices with marginalized patients. Intuitively, for some $\beta > 0$, this function scales the weight of edges ending in marginalized vertices by $(1+\beta)$. For example, if $\beta = 1.5$, then the optimizer will value edges that result in a marginalized patient receiving a transplant at 250% of their initial weight (possibly scaled by factors like edge failure probability or chain position).

For any $M \in \mathcal{M}$, let M' be the matching such that every edge $e \in E$ has augmented weight $\Delta^{\beta}(e)$. Then the MAXCARD-FAIR utility function u_{Δ} is defined in terms of the utilitarian MAXCARD utility function u applied to the augmented matching M', such that $u_{\Delta}(M) = u(M')$. In our experiments, we vary the parameter β to empirically quantify its effects on each of the three objective functions.

The FUTUREMATCH Framework

We are interested in learning from demographically accurate data how to match *in the present* such that some overarching objective function is maximized over time. Scalability is important: heavy offline statistics can be computed and periodically updated, but the fielded clearing algorithm must run quickly (within minutes or at most hours).

Figure 2 depicts the FUTUREMATCH framework. A domain expert (e.g., a committee of medical and legal professionals) begins by describing an overall objective function for the exchange. Even *measuring* this objective can be difficult: for example, if the goal is to maximize the number of days added to patients' lives via kidney transplantation, then calculating the relative quality of a proposed match requires knowing some notion of utility for each edge—representing a potential transplant—in the compatibility graph. We learn this edge weight function $w : E \to \mathbb{R}^+$ from data, and gave examples objective functions in the previous section.

The learned weight function w is then fed into a parameterized instance generator that mimics the underlying distribution. This generator in turn feeds training and test sets into a system for learning the *potentials* of various element classes in the compatibility graph. Intuitively, given an element θ (e.g., vertex, edge, cycle, or chain type), a potential $P_\theta \in \mathbb{R}$ quantifies the expected utility to the exchange of that element in the future (Dickerson, Procaccia, and Sandholm 2012). Potentials are combined with w to quantify an edge-specific quality rating. In our experimental results, we learned potentials[3] for the combinations of different blood types for patient-donor pairs under each of the three objectives defined earlier.

Finally, the fielded clearing algorithm incorporates the combined weight function w and set of potentials $\mathcal{P}_\Theta$ into its myopic weighted matching algorithm. For example, to combine patient-donor blood type potentials with the learned weight function w, we could use a function $f_w : E \to \mathbb{R}$ such that $f_w(e) = w(e) \cdot (1 - P_X - P_Y)$, with X the donor blood type at e's source and Y the patient blood type at e's sink. This incorporation of potentials into the myopic algorithm comes at very low or no cost to the runtime of the clearing algorithm; indeed, the final "potential-aware" input graph is simply a re-weighted version of the original compatibility graph, using the weights that encode the future.

A Realistic Simulator. When learning potentials offline, it is important to mimic closely the behavior of the fielded exchange online. If the distribution of incoming potential types is significantly different than expected, so too will be the estimates of potentials. We built a dynamic simulator of kidney exchange using data from the UNOS exchange (and APD (Ashlagi et al. 2011)). This work significantly extends that of Dickerson, Procaccia, and Sandholm (2013), which defined and experimentally evaluated a model of the evolution of dynamic kidney exchange. Critically, they did not perform dynamic optimization in that model—just myopic optimization applied sequentially in a dynamic model. They also sampled from a basic generator that is no longer accepted in the kidney exchange community (Saidman et al. 2006), while we sample from an accurate distribution—the historical UNOS exchange pool.

Our simulator works as follows. New pairs and altruistic donors enter the pool at each time period, while some leave the pool due to a variety of non-exchange-related reasons (e.g., becoming too ill to transplant). A matching is performed at each time period, which results in a set of matched pairs leaving the pool for $t > 0$ time periods. This reflects the length of time required to medically and logistically verify the implementability of the planned match. Matched patients then either leave the pool permanently after successfully receiving a kidney, or return to the pool after failing to receive a kidney. We set the relevant entrance and exit probabilities based on the real UNOS kidney exchange data (Kidney Paired Donation Work Group 2013).

A matching is determined at each time period based on either a *deterministic* or *failure-aware* clearing algorithm, which we briefly describe here. Both models compute an optimal matching $M^* = \text{argmax}_{M \in \mathcal{M}}\, u(M)$, where $u(M) = \sum_{c \in M} u(c)$. Here, $u(c)$ represents the utility of a cycle or chain c. In the deterministic model, $u(c) = \sum_{e \in c} f_w(e)$: that is, the sum of the weights of the constituent edges in a cycle or chain subject to the weight function w and potential mapping f_w. The deterministic model is susceptible to edge failures. For example, if a single edge in a 3-cycle fails, that entire cycle fails to execute. Similarly, if the third edge in a long chain fails, then the tail of that chain (after the failed edge) is cut off. Let $q_e \in [0, 1]$ be the probability that an edge succeeds. Then the failure-aware model defines the *discounted* utility for a cycle c as $u(c) = \left[\sum_{e \in c} f_w(e)\right] \cdot \left[\prod_{e \in c} q_e\right]$, and the discounted utility for a chain $c = \langle e_0, e_1, \ldots, e_{k-1}\rangle$ as

$$u(c) = \left[\sum_{i=1}^{k-1}(1 - q_i)\sum_{j=0}^{i-1} f_w(e_j) \prod_{j=0}^{i-1} q_j\right] + \left[\sum_{i=0}^{k-1} f_w(e_i) \prod_{i=0}^{k-1} q_i\right].$$

Results

We compare FUTUREMATCH against a baseline of myopic deterministic matching under each of the objectives. Conservatively, statistical significance was determined using the Wilcoxon signed-rank test, which is a nonparametric alternative to the paired t-test. Table 1 shows the median expected gain in the overall number of transplants from using FUTUREMATCH under each of the objectives. Each column labeled $|V| = k$ corresponds to a simulation over k patient-donor pairs and altruists sampled as described earlier.

Table 1 shows that the two objectives that do not regard fairness—MAXCARD and MAXLIFE—significantly beat myopic deterministic matching under the same objective. Interestingly, so too does MAXCARD-FAIR for low values of β. As β increases, the gain in *overall* number of transplants decreases (although it never drops below the deterministic matching algorithm with significance). This decrease in overall gain is incurred because marginalized patients, who (i) generally have lower in-degree, and (ii) have a higher probability of match failure, are being weighted more than easier to match pairs.

Table 2 explores this tradeoff between fairness and efficiency explicitly. For the fairness-agnostic and lightly fairness-preferring objectives, a relative loss of a few marginalized transplants is realized—although this loss of marginalized transplants is always less (typically much less) than the overall gain in transplants. Increasing the optimizer's preference for marginalized patients results in statistically significant gains in the number of marginalized transplants at *no* statistically significant loss in the overall expected number of transplants. In fact, for a middle ground around $\beta = 2$, FUTUREMATCH often shows statistically significant gains in *both* overall transplant and marginalized transplant counts—a clear win over myopia.

Our experiments support the following conclusions:

- FUTUREMATCH under MAXCARD and MAXCARD-FAIR with low $\beta = 1$ results in a significant increase in the *overall* number of transplants compared to myopic, at the cost of a smaller decrease in the number of *marginalized* transplants.
- FUTUREMATCH under MAXCARD-FAIR with high β results in a significant increase in *marginalized* transplants, at *no cost* to the overall number of transplants under myopic matching.
- For a middle ground around $\beta = 2$, FUTUREMATCH can result in both more overall expected transplants and more marginalized transplants.

We note that we are *not* making policy recommendations; rather, we are giving a proof of concept that the FU-

[3] Again, due to space constraints, we guide the reader to §3.3 of Dickerson and Sandholm (2014) for a complete derivation of potentials and an explanation of the potential learning system.

TUREMATCH framework can effectively balance conflicting wants in an exchange. Indeed, the exact fairness quantification β that most effectively balances efficiency and fairness is a function of the underlying graph dynamics, which vertices are considered marginalized, and the ethical and legal wants of an exchange. All of these dimensions can be effectively encoded and validated through FUTUREMATCH.

References

Abraham, D.; Blum, A.; and Sandholm, T. 2007. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *ACM Conference on Electronic Commerce (EC)*, 295–304.

Ashlagi, I.; Gilchrist, D. S.; Roth, A. E.; and Rees, M. 2011. Nonsimultaneous chains and dominos in kidney-paired donation—revisited. *American Journal of Transplantation* 11(5):984–994.

Barnhart, C.; Johnson, E. L.; Nemhauser, G. L.; Savelsbergh, M. W. P.; and Vance, P. H. 1998. Branch-and-price: Column generation for solving huge integer programs. *Operations Research* 46(3):316–329.

Dickerson, J. P., and Sandholm, T. 2014. FutureMatch: Learning to match in dynamic environments. Draft.

Dickerson, J. P.; Procaccia, A. D.; and Sandholm, T. 2012. Dynamic matching via weighted myopia with application to kidney exchange. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 1340–1346.

Dickerson, J. P.; Procaccia, A. D.; and Sandholm, T. 2013. Failure-aware kidney exchange. In *ACM Conference on Electronic Commerce (EC)*, 323–340.

Dickerson, J. P.; Procaccia, A. D.; and Sandholm, T. 2014. Price of fairness in kidney exchange. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Kidney Paired Donation Work Group. 2013. OPTN KPD pilot program cumulative match report (CMR) for KPD match runs: Oct 27, 2010 – Apr 15, 2013.

Neuen, B. L.; Taylor, G. E.; Demaio, A. R.; and Perkovic, V. 2013. Global kidney disease. *The Lancet* 382(9900):1243.

Rapaport, F. T. 1986. The case for a living emotionally related international kidney donor exchange registry. *Transplantation Proceedings* 18:5–9.

Rees, M.; Kopke, J.; Pelletier, R.; Segev, D.; Rutter, M.; Fabrega, A.; Rogers, J.; Pankewycz, O.; Hiller, J.; Roth, A.; Sandholm, T.; Ünver, U.; and Montgomery, R. 2009. A nonsimultaneous, extended, altruistic-donor chain. *New England Journal of Medicine* 360(11):1096–1101.

Roth, A.; Sönmez, T.; Ünver, U.; Delmonico, F.; and Saidman, S. L. 2006. Utilizing list exchange and nondirected donation through ‘chain’ paired kidney donations. *American Journal of Transplantation* 6:2694–2705.

Saidman, S. L.; Roth, A.; Sönmez, T.; Ünver, U.; and Delmonico, F. 2006. Increasing the opportunity of live kidney donation by matching for two and three way exchanges. *Transplantation* 81(5):773–782.

United Network for Organ Sharing (UNOS). http://www.unos.org/.