

# Price of Fairness in Kidney Exchange

John P. Dickerson, Ariel D. Procaccia, and Tuomas Sandholm  
Computer Science Department, Carnegie Mellon University  
Pittsburgh, PA, USA  
{dickerson,arielpro,sandholm}@cs.cmu.edu

## ABSTRACT

Kidney exchange provides a life-saving alternative to long waiting lists for patients in need of a new kidney. Fielded exchanges typically match under utilitarian or near-utilitarian rules; this approach marginalizes certain classes of patients. In this paper, we focus on improving access to kidneys for highly-sensitized, or hard-to-match, patients. Toward this end, we formally adapt a recently introduced measure of the tradeoff between fairness and efficiency—the *price of fairness*—to the standard kidney exchange model. We show that the price of fairness in the standard theoretical model is small. We then introduce two natural definitions of fairness and empirically explore the tradeoff between matching more hard-to-match patients and the overall utility of a utilitarian matching, on real data from the UNOS nationwide kidney exchange and simulated data from each of the standard kidney exchange distributions.

## Categories and Subject Descriptors

I.2.11 [Distributed Artificial Intelligence]: Multiagent systems;  
J.4 [Social and Behavioral Sciences]: Economics

## General Terms

Economics, Experimentation

## Keywords

Kidney exchange, price of fairness, equity-efficiency tradeoff

## 1. INTRODUCTION

The preferred treatment for kidney failure is transplantation. However, the demand for donor kidneys is far greater than supply. For example, 34,837 people were added to the US national waiting list in 2012, while only 10,851 left it due to receiving a kidney [26]. Demand is increasing worldwide.

Successful transplantation of a kidney relies on tissue-type compatibility between the donor organ and patient, among other factors. Compatibility is determined through a tissue-type *crossmatch* between a potential donor and patient’s blood; if the two types differ substantially, the patient’s body will reject the donor organ.

Some patients are *highly-sensitized*; there is a very low probability that their blood will pass a crossmatch test with a random organ. For these patients, finding a kidney is quite difficult (and

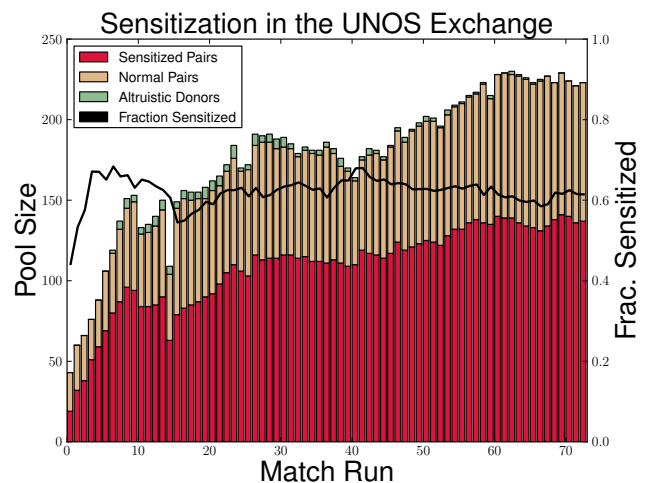
**Appears in:** *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

Copyright © 2014, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

median time on the waiting list jumps by a factor of three over less sensitized patients [26]).

Roughly 17% of the adult patients on the waiting list for *deceased* donor kidneys are highly-sensitized [18]. Recently, an allocation policy was designed for deceased donor kidneys that effectively balances fairness and efficiency while working within the currently fielded priority-based framework [10].

Complementing deceased donation is kidney exchange, which allows patients with a willing but medically incompatible *living* donor to swap their donor with other patients. The percentage of highly-sensitized patients in fielded kidney exchanges is quite high; roughly 60% of the UNOS nationwide kidney exchange is highly-sensitized, as shown in Figure 1.



**Figure 1: Composition of the UNOS national kidney exchange from inception. For each of 73 match runs (x-axis), the raw number of highly-sensitized patients, non-highly-sensitized patients, and altruists are plotted (left y-axis), as well as the percentage of patients who are highly-sensitized as a percentage of the pool size (right y-axis).**

In this paper, we focus on improving access to kidneys for highly-sensitized (and thus hard to match) patients. We explore the *price of fairness* in kidney exchange—the relative loss in total welfare from using a “fair” matching rule, instead of an overall utility-maximizing one [9]. Theoretically, we show that the price of fairness is small in the standard theoretical kidney exchange model. We then define two natural definitions of fairness in kidney exchange and empirically quantify the tradeoff between efficiency and fairness on real data from the UNOS nationwide kidney ex-

change (which runs our code), as well as on simulated data from the two most widely used kidney exchange distributions. We find that, on real data, prioritizing hard-to-match patients results in a price of fairness that is (often quite far from) zero. While our paper focuses on highly-sensitized patients, its techniques and results are easily adaptable to other notions of fairness in kidney exchange.

Kidney exchanges represent a truly fielded example of technology at the intersection of artificial intelligence and economics—an example that is both recent and as yet unsolved. Exchanges are saving lives on an ongoing basis and growing annually, both in the United States and worldwide. They provide a wealth of opportunities for theoretical and experimental research in both fields. For instance, key questions in practically tractable clearing algorithms [7, 13, 14, 15] and mechanism design [3, 5, 6, 11] remain unsolved. Advances in these areas will benefit the greater artificial intelligence community immensely, and provide valuable experience in fielding real-world artificial intelligence technology.

## 2. FAIRNESS IN KIDNEY EXCHANGE

In this section, we briefly overview the standard graph-based model of kidney exchange, expand the model to include a notion of sensitization, and formally define the price of fairness in the context of kidney exchange.

### 2.1 Kidney exchange model

The standard model for kidney exchange encodes an  $n$ -patient kidney exchange as a directed *compatibility graph*  $G = (V, E)$  by constructing one vertex for each patient-donor pair. An edge  $e$  from  $v_i$  to  $v_j$  is added if the patient in  $v_j$  wants and is compatible with the donor kidney of  $v_i$ . A 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.<sup>1</sup>

A cycle  $c$  in the graph  $G$  represents a possible kidney swap, where each vertex in  $c$  obtains the kidney of the previous vertex. We denote by  $k$ -cycle a cycle with  $k$  patient-donor pairs. In fielded kidney exchange, cycles of length at most only 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 UNOS kidney exchange,  $L = 3$  (i.e., only 2- and 3-cycles are allowed).

Fielded kidney exchanges also gain great utility through the use of *chains* (see, e.g., [4, 14, 16, 17, 21, 22, 23, 28]). Chains start with an altruistic donor donating his kidney to a patient, whose paired donor donates her kidney to another patient, and so on.

A *matching*  $M$  is a collection of vertex-disjoint cycles and chains in the graph  $G$ . Note that the elements of the matching must be disjoint because no donor can give more than one of his kidneys. Then, 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  $u : \mathcal{M} \rightarrow \mathbb{R}$ . Formally:

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

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

<sup>1</sup>This model works for general  $n$ -participant barter exchanges, where vertices trade and receive utility from items instead of kidneys.

while marginalizing others. We formalize this notion in the next section.

### 2.2 The price of fairness

In practice, the sensitization level of a patient is presented as a Calculated Panel Reactive Antibody (CPRA) score that varies from 0 to 100. The CPRA score is an estimate of the percentage of donors that are tissue-type incompatible with the patient (i.e., the percentage of donors with whom a patient would have a positive, or failing, crossmatch).

Assume each non-altruistic vertex  $v$  has a sensitization level  $v_s \in [0, 100]$ , representing the CPRA level of  $v$ 's patient. Altruistic vertices need no CPRA specification because they have no associated patient. Let  $\tau \in [0, 100]$  represent a threshold delimiting low from high sensitization; in practice,  $\tau \geq 80$ . Partition  $V$  into  $\{V_L \cup V_H \cup A\}$ , such that:

- $V_L$  is not highly-sensitized:  $\{v \mid v \in V \setminus A \wedge v_s < \tau\}$
- $V_H$  is highly-sensitized:  $\{v \mid v \in V \setminus A \wedge v_s \geq \tau\}$
- $A$  are altruistic donors (with no patients)

Since the highly-sensitized patients will, by definition of CPRA, have fewer incoming edges on average than lowly-sensitized patients, one worries that a mechanism maximizing overall efficiency might favor easier-to-match vertices in  $V_L$  to the detriment of those in  $V_H$ . Similarly, if a mechanism prioritizes harder-to-match vertices in  $V_H$ , one worries that the overall efficiency of the matching might drop.

Bertsimas, Farias, and Trichakis recently defined the *price of fairness* to be the “relative system efficiency loss under a fair allocation assuming that a fully efficient allocation is one that maximizes the sum of [participant] utilities [9].” Caragiannis et al. defined an essentially identical concept in parallel [12]. We adopt that notion here.

Let  $u_f : \mathcal{M} \rightarrow \mathbb{R}$  be a *fair* utility function. Formally, a utility function is fair when its corresponding optimal match  $M_f^*$  is viewed as fair, where  $M_f^*$  is defined as:

$$M_f^* = \operatorname{argmax}_{M \in \mathcal{M}} u_f(M)$$

Given a fair utility function  $u_f$  and the utilitarian utility function  $u$ , the price of fairness is defined to be:

$$\text{POF}(\mathcal{M}, u_f) = \frac{u(M^*) - u(M_f^*)}{u(M^*)}$$

That is,  $\text{POF}(\mathcal{M}, u_f)$  is the relative loss in match efficiency (from the utilitarian point of view  $u$ ) due to the maximization of a fair utility function  $u_f$  over some family of matchings  $\mathcal{M}$ .

In the next section, we show that the price of fairness in the standard *theoretical* model for kidney exchange is quite small, for any reasonable fair utility function  $u_f$ . Then, in Sections 4 and 5, we present two families of fair utility functions motivated by our experiences with the UNOS national kidney exchange; on *real* data, the price of fairness is frequently far from zero.

## 3. THE (THEORETICAL) PRICE OF FAIRNESS IS LOW

In this section, we give bounds for the price of fairness under the standard model of kidney exchange. The price of fairness is upper-bounded by a small number of vertices (under reasonable assumptions, with high probability).

### 3.1 Upper bound over all fair utility functions

Different notions of fairness may result in more (or less) of an effect on overall system efficiency. As an extreme example, forcing a matching to include at least one highly-sensitized patient (if possible) intuitively restricts the solution space less than forcing a matching to include as many highly-sensitized patients as possible. We must consider this when stating theoretical bounds on the price of fairness.

We derive our bound under the fair utility function  $u_{H \succ L}$  that lexicographically ranks any highly-sensitized vertex over any lowly-sensitized vertex. For any matching  $M \in \mathcal{M}$ , let  $M_H = M \cap V_H$  be the subset of highly-sensitized vertices matched by  $M$ . Formally:

$$u_{H \succ L}(M) = \begin{cases} u(M) & \text{if } |M_H| = \max_{M' \in \mathcal{M}} |M'_H| \\ 0 & \text{otherwise} \end{cases}$$

This utility function gives nonzero weight only to those matches that include the maximum possible number of highly-sensitized patients. We informally argue that price of fairness guarantees on  $u_{H \succ L}$  are upper bounds to the price of fairness of any “reasonable” fair utility function. Indeed, any utility function that does not first maximize the number of highly-sensitized pairs matched will leave a thicker remaining market in which non-highly-sensitized pairs have more options for matching—and thus the resulting match will see less of an efficiency loss.

### 3.2 Model with ABO-blood types and two levels of sensitization

We work in a theoretical model that considers both *blood types* and sensitization. Like a positive crossmatch, the blood type of a donor can prevent the acceptance of a kidney by a patient. At a high level, the absence or presence of certain proteins splits human blood types into four groups: O, A, B, and AB. Barring other considerations, patients of type AB are compatible with any blood type, patients of type A (B) are compatible with blood types A (B) and O, and patients of type O can only accept from their own blood type. Thus, O-patients are the hardest blood type to match, while O-donors are the easiest.

Much of the theoretical market design work in kidney exchange uses the *ABO-model*, which labels vertices in a compatibility graph with their patient and donor blood types (see, e.g., [3, 6, 11, 14, 25, 27]). Work in this model has not directly addressed sensitization.

We draw random graphs in accordance with the canonical method (see, e.g., [6]). Partition the  $n$  incompatible patient-donor pairs of some large, directed compatibility graph  $G(n)$  into  $V^{X-Y}$  of type  $X-Y$ , for each combination of blood types  $X$  and  $Y$  of the patient and donor respectively. When required, we will further partition each set  $V^{X-Y}$  into  $V_L^{X-Y}$  and  $V_H^{X-Y}$ , the lowly- (highly-)sensitized pairs of type  $X-Y$ . The frequency of each blood type  $X$  is denoted by  $\mu_X$ . Note that vertices with blood type-compatible patient-donor pairs may still enter the pool due to tissue-type incompatibility. We assume that a donor and a lowly-sensitized (highly-sensitized) patient who are blood type compatible are tissue-type incompatible with constant probability  $p_L$  ( $p_H$ ). Let  $\bar{p}$  represent the average level of sensitization in the pool; that is,  $\bar{p} = \lambda p_L + (1 - \lambda)p_H$ , where  $\lambda$  is the fraction of the pool that is lowly-sensitized.

Proposition 1 gives a bound on the price of fairness in random graphs in the ABO-model parameterized in a realistic way (i.e.,  $\bar{p}$  mirrors that of a dense kidney exchange pool, and the blood type distribution mimics that of the US population). The proposition and proof sketch build on the efficiency result presented in §5 of Ashlagi and Roth [6]. Like their work, Proposition 1 considers a model without chains.

**PROPOSITION 1.** *Assume that  $\bar{p} < 2/5$ ,  $\mu_O < 3\mu_A/2$ , and  $\mu_O > \mu_A > \mu_B > \mu_{AB}$ , and  $\lambda \geq (1 - \bar{p})$ . Denote by  $\mathcal{M}$  the set of matchings in  $G(n)$ . Then, almost surely as  $n \rightarrow \infty$ ,*

$$POF(\mathcal{M}, u_{H \succ L}) \leq \frac{2}{33}.$$

(And this is achieved using only cycles of length at most 3.)

**PROOF SKETCH.** Let pairs of type A-B and B-A be called *reciprocal* pairs. Call any non-reciprocal pair  $X-Y$  whose donor is not ABO-compatible with its patient an *under-demanded* pair, and any pair  $X-Y$  such that  $X \neq Y$  ( $X = Y$ ) and whose donor is ABO-compatible with its patient an *over-demanded* (self-demanded) pair.

From the results of Ashlagi and Roth [6], with high probability there exists an efficient matching  $M^*$  in  $G(n)$  that matches all over- and self-demanded pairs, as well as all reciprocal pairs. Trivially, all highly-sensitized pairs are also matched in these specific subgroups. Thus, we need only consider highly-sensitized pairs in under-demanded pools. We do this exhaustively.

- In  $M^*$ , all pairs in  $V^{AB-B}$  are matched to as many pairs in  $V^{B-AB}$  as possible. If  $|V^{AB-B}| \geq |V^{B-AB}|$ , then all highly-sensitized under-demanded B-AB pairs will be matched. Note  $|V^{AB-B}| \propto \bar{p}\mu_B\mu_{AB}$  and  $|V^{B-AB}| \propto (1 - \lambda)\mu_B\mu_{AB}$ . Then by the assumption on  $\lambda$ ,  $\bar{p}\mu_B\mu_{AB} \geq (1 - \lambda)\mu_B\mu_{AB}$ , and the above cardinality inequality holds almost surely.<sup>2</sup>

- We now match all highly-sensitized pairs in  $V_H^{O-A}$  and  $V_H^{O-B}$  to pairs in  $V^{A-O}$  and  $V^{B-O}$ , respectively, using 2-cycles. This can be done by the same argument as above.

WLOG, assume  $|V^{B-A}| \leq |V^{A-B}|$ . Then, in  $M^*$ , all reciprocal pairs in  $V^{B-A}$  are matched to as many pairs as possible in  $V^{A-B}$ , leaving  $V_r^{A-B}$  leftover reciprocal pairs. In  $M^*$ , these leftover pairs in  $V_r^{A-B}$  are fully matched through  $V^{B-O}$  and  $V^{O-A}$  in 3-cycles. By executing the 2-cycles above, we may prematurely exhaust  $V^{B-O}$  and prevent up to  $|V_r^{A-B}|$  3-cycles from executing. (Note that if  $V^{O-A}$  is exhausted, the  $V^{A-B}$  through  $V^{B-O}$  cycle can be closed at length 2 for no efficiency loss.)

By Lemma 5.1 of Ashlagi and Roth [6], the absolute difference in reciprocal pairs is  $o(n)$ ; that is,  $|V_r^{A-B}|$  is less than any other subgroup in the pool (since the size of any other subgroup is linear in  $n$ ). Therefore, the *number* of 3-cycles lost is sub-linear in  $n$ ; thus, this term (relative to price of fairness) is insignificant as  $n$  grows.

- Next, we match all pairs in  $V_H^{A-AB}$  to over-demanded pairs in  $V^{B-A}$ ; this leaves  $|V_L^{A-AB}| \propto \lambda\mu_A\mu_{AB}$  remaining A-AB pairs.

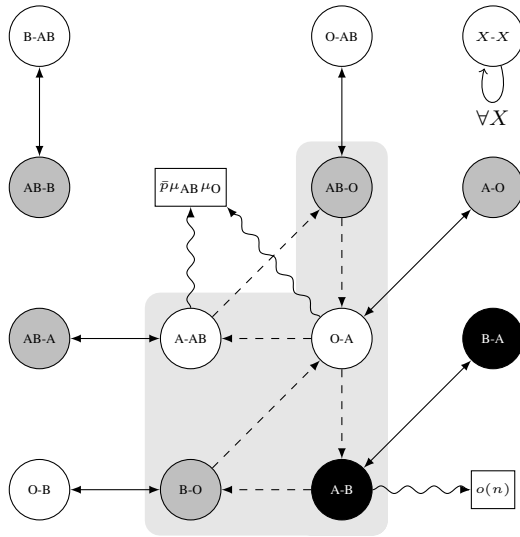
In the efficient matching  $M^*$ , all over-demanded pairs in  $V^{AB-O}$  are matched through 3-cycles with pairs in  $V^{A-AB}$  and  $V^{O-A}$ . If the remaining  $|V_L^{A-AB}|$  vertices in  $V^{A-AB}$  are exhausted before  $V^{AB-O}$ , then the remaining vertices in  $V^{AB-O}$  can still be exhausted through 2-cycles to, e.g.,  $V^{O-A}$  at no efficiency loss. (Note  $|V_L^{O-A}| \propto \lambda\mu_O\mu_A > \lambda\mu_O\mu_{AB} \geq (1 - \bar{p})\mu_O\mu_{AB} > \bar{p}\mu_O\mu_{AB} \propto |V^{AB-O}|$ , since  $\bar{p} < \frac{2}{5} < \frac{1}{2}$ .) This occurs unless the fair matching accrues efficiency loss through highly-sensitized, under-demanded O-AB pairs, as explained next.

- Finally, we are left with highly-sensitized pairs in  $V^{O-AB}$ , the hardest to match group. In the efficient allocation,  $V^{O-AB}$  is unmatched entirely. In a matching under  $u_{H \succ L}$ , at most

<sup>2</sup>As  $n \rightarrow \infty$ , the size of a set will be very close to its expectation.

$|V_H^{O-AB}| \propto (1 - \lambda)\mu_O\mu_{AB}$  could find a matching elsewhere, possibly cannibalizing a 3-cycle through  $V^{AB-O}$  to form a 2-cycle, or through  $V^{AB-B}$  and  $V^{B-O}$  or  $V^{AB-A}$  and  $V^{A-O}$  to form a 3-cycle at a cost of two 2-cycles. In any of these three (exhaustive) cases, every match of a highly-sensitized O-AB pair results in an efficiency loss of one pair, or an overall absolute efficiency loss proportional to  $(1 - \lambda)\mu_{AB}\mu_O$  pairs.

This exhausts all highly-sensitized pairs in the under-demanded subgroups not fully matched by the efficient allocation  $M^*$ . All highly-sensitized pairs in self- and over-demanded subgroups are still matched, as in the efficient allocation. Any newly unmatched reciprocal pairs are lowly-sensitized. Thus, we have matched all highly-sensitized pairs in the pool. This came at an absolute loss proportional to at most  $(1 - \lambda)\mu_{AB}\mu_O \leq \bar{p}\mu_{AB}\mu_O$  pairs. Figure 2 visualizes this new matching (and shows where losses occur with respect to the efficient matching  $M^*$ ).



**Figure 2: An example matching used in Proposition 1. Patient-donor pairs are ovals: under- and self-demanded pairs are white, over-demanded pairs are gray, and reciprocal pairs are black. Regular edges appear in the efficient matching, while dashed edges represent 3-cycles from the efficient matching that may be disturbed via fair matching. Efficiency loss is denoted with rectangular nodes.**

To determine a bound on the price of fairness, we must first determine the actual expected loss in number of vertices (through normalization), and then determine the relative loss with respect to the expected size of the efficient matching  $M^*$ .

The expected size of  $M^*$  is at least  $E$ , such that:

$$E \propto \bar{p} [2\mu_{AB}\mu_B + 2\mu_{AB}\mu_A + 3\mu_{AB}\mu_O + 2\mu_A\mu_O + 2\mu_B\mu_O + \mu_O^2 + \mu_A^2 + \mu_B^2 + \mu_{AB}^2] + 2\mu_A\mu_B$$

This value is computed by counting only 2- and 3-cycles that are almost surely guaranteed to execute in the efficient matching. Then, the expected size of a fair matching under  $u_{H>L}$  is at least  $F \propto E - \bar{p}\mu_{AB}\mu_O$ , as calculated above.

An upper bound on the expected price of fairness is then  $\frac{E-F}{E}$ :

$$\text{POF}(\mathcal{M}, u_{H>L}) \leq \frac{(\bar{p}\mu_{AB}\mu_O)}{(\bar{p} [2\mu_{AB}\mu_B + 2\mu_{AB}\mu_A + 3\mu_{AB}\mu_O + 2\mu_A\mu_O + 2\mu_B\mu_O + \mu_O^2 + \mu_A^2 + \mu_B^2 + \mu_{AB}^2] + 2\mu_A\mu_B)}$$

Note that the notion of proportionality can be dropped, since both the numerator (the size of the fair matching, written in terms of the size of the efficient matching) and the denominator (the size of the efficient matching) are normalized by the same constant.

We prove the statement by giving an upper bound on the right-hand side of the inequality above. Considering the distribution constraint  $\sum_{X \in \{O,A,B,AB\}} \mu_X = 1$ , and the ordering constraints on each blood type, this is upper-bounded when  $\bar{p} = \frac{2}{5}$ ,  $\mu_O = \frac{1}{3}$ , and  $\mu_{AB} = \frac{2}{9}$  (which forces  $\mu_A = \mu_B = \frac{2}{9}$ ). Using these values yields an upper-bound of  $\text{POF}(\mathcal{M}, u_{H>L}) \leq \frac{2}{33}$  (with high probability).  $\square$

Different countries and regions within countries have different blood type distributions. From Proposition 1, a more realistic bound for this model can be drawn from the United States distribution of blood types ( $\mu_O \approx 0.44$ ,  $\mu_A \approx 0.42$ ,  $\mu_B \approx 0.10$ ,  $\mu_{AB} \approx 0.04$ ); this yields a bound of  $\text{POF}(\mathcal{M}, u_{H>L}) \lesssim 1.5\%$ .

A symmetric result follows if  $\mu_B > \mu_A$ , instead of  $\mu_A > \mu_B$  as assumed. We note that, from the reasoning in [6], similar results can be derived for  $\bar{p} > 2/5$ —so long as  $p_L$ ,  $p_H$ , and  $\bar{p}$  remain constant.

## 4. HOW SHOULD WE DEFINE FAIRNESS?

In this section, we present two definitions of fairness in kidney exchange—one using strict lexicographic preferences and the other using a sliding scale weighting function. Critically, we feel both definitions fit within the scope and practice of current policy in fielded exchange, a necessary consideration when fielding new technology in medicine (as noted by those who designed the recent deceased donor allocation scheme [10] and supported by our experience with the UNOS exchange). We briefly discuss challenges in implementing either fairness rule, then experimentally validate both rules in Section 5 on simulated and real exchange data.

### 4.1 Lexicographic fairness

First, we consider a fairness criterion that assumes lexicographic preferences over classes of vertices. This is a generalization of the  $u_{H>L}$  utility function used in Section 3. Informally, this fairness rule allows policymakers to state a hard preference for matchings containing a baseline percentage of highly-sensitized patients and, if and only if this constraint is fulfilled, express a secondary utilitarian preference. We present this rule below in the context of both deterministic and probabilistic kidney exchange.

#### 4.1.1 Deterministic model

Let  $\alpha^*$  be the maximum fraction of vertices in  $V_H$  that could be matched over  $\mathcal{M}$ , the set of possible matchings on some compatibility graph  $G$ . That is,

$$\alpha^* = \left( \max_{M' \in \mathcal{M}} |M'_H| \right) / |V_H|$$

For a given graph  $G$ ,  $\alpha^*$  can be computed using a simple modification of the standard kidney exchange integer program (known as the *cycle formulation* [1]).

$$\alpha^* = 1/|V_H| \cdot \max_{\text{s.t.}} \sum_c h(c)x_c \quad \forall v \sum_{c|v \in c} x_c \leq 1 \quad \forall c \quad x_c \in \{0, 1\}$$

Unless otherwise specified, vertices  $v$  range over the set  $V$ , while cycles and chains  $c$  range over the set of all legal cycles and chains  $C$  (possibly capped at some length  $L$ ). Here,  $x_c$  is a binary variable that is set to 1 if cycle  $c$  is included in the final matching; otherwise,

it is 0. The constant  $h(c)$  is the number of highly-sensitized patients in a cycle or chain  $c$ . Formally,

$$h(c) = |\{v \mid v \in c \wedge v \in V_H\}|$$

Note that  $h(c) = 0$  if a cycle or chain does not contain any highly-sensitized patients, so only those binary variables  $x_c$  corresponding to cycles or chains with at least one highly-sensitized patient need be included in the objective. However, all *vertices* belonging to at least one cycle or chain that contains at least one highly-sensitized vertex must be included in the vertex-disjointness constraints (to maintain the feasibility of the final disjoint cycle cover).

Given this setup, a match could be considered equitable if it satisfies some nonnegative, user-defined parameter  $\alpha \leq \alpha^*$ , such that the matching algorithm includes  $\alpha \cdot |V_H|$  highly-sensitized patients in the “optimal” match. E.g., if  $\alpha = 0.5$ , any returned match would include at least 50% of the number of highly-sensitized patients available. Formally, define the *deterministic lexicographic* fairness rule  $u_{H>L}^\alpha$  over any  $M \in \mathcal{M}$  as follows:

$$u_{H>L}^\alpha(M) = \begin{cases} u(M) & \text{if } |M_H| \geq \alpha \cdot |V_H| \\ 0 & \text{otherwise} \end{cases}$$

We implement this utility function by adding a single constraint to the standard integer program, yielding:

$$\begin{aligned} \max \quad & \sum_c |c| x_c \\ \text{s.t.} \quad & \sum_{c|v \in c} x_c \leq 1 \quad \forall v \\ & \sum_c h(c) x_c \geq \alpha \cdot |V_H| \\ & x_c \in \{0, 1\} \quad \forall c \end{aligned}$$

In the next section, we generalize this rule for probabilistic kidney exchange.

#### 4.1.2 Probabilistic (failure-aware) model

Most algorithmic matches in fielded kidney exchanges do not result in an actual transplant, even if a patient and donor are seen as ABO- and tissue-type compatible by the optimization software. A variety of issues arise, including sudden illness or death, improper medical testing, a patient finding a deceased or alternate living donor outside of the kidney exchange, or match failure due to more intensive (and expensive) post-match medical testing.

Recent work introduced the idea of *failure-aware* kidney exchange, where the optimizer endogenously balances the likelihood of match success with the raw size and composition of cycles and chains in the match [15]. Intuitively, instead of maximizing the number of pairs matched, the failure-aware clearing problem is to maximize the expected number of transplants, subject to each edge  $e$  in the graph having some probability of success  $q_e$ . In the lexicographic fairness model we consider in this paper, this is further constrained by ensuring some fraction of these expected transplants are to highly-sensitized patients.

Toward this end, define  $v^h(c)$ , the expected number of transplants to highly-sensitized patients of a cycle or chain  $c$ , as follows: **Cycles.** For cycles, the discounted utility is the same as in [15], but counting only those vertices in the cycle that are highly-sensitized. Formally,

$$v^h(c) = h(c) \prod_{e \in c} q_e$$

**Chains.** While cycles necessarily execute atomically, chains can execute partially and then fail—thus making the calculation of  $v^h$  a bit trickier. For a chain of length  $k$ , let  $c^i$  represent the initial  $i$

vertices in the chain (including the altruistic donor). Formally,

$$v^h(c) = \left[ \sum_{i=1}^{k-1} (1 - q_i) h(c^{i-1}) \prod_{j=0}^{i-1} q_j \right] + \left[ h(c) \prod_{i=0}^{k-1} q_i \right]$$

We now let  $N^*$  be the maximum *expected* number of vertices in  $V_H$  that could receive transplants over all possible matchings  $\mathcal{M}$  on  $G$ . That is,

$$N^* = \max_{M \in \mathcal{M}} v^h(M)$$

where for any matching  $M$ ,  $v^h(M)$  is defined as  $\sum_{c \in M} v^h(c)$ .

Then for a given graph  $G$ , this failure-aware  $N^*$  can be computed using a similar modification of the failure-aware kidney exchange integer program as follows:

$$\begin{aligned} N^* = \max \quad & \sum_c v^h(c) x_c \\ \text{s.t.} \quad & \sum_{c|v \in c} x_c \leq 1 \quad \forall v \\ & x_c \in \{0, 1\} \quad \forall c \end{aligned}$$

Now, as above, the user can supply some parameter  $\alpha \in [0, 1]$  that guarantees  $\alpha$  fraction of the maximum possible expected number of transplants for highly sensitized-patients. This preference is enforced by adding a single constraint to the standard failure-aware kidney exchange IP as follows:

$$\begin{aligned} \max \quad & \sum_c v(c) x_c \\ \text{s.t.} \quad & \sum_{c|v \in c} x_c \leq 1 \quad \forall v \\ & \sum_c v^h(c) x_c \geq \alpha \cdot N^* \\ & x_c \in \{0, 1\} \quad \forall c \end{aligned}$$

Here,  $v(c)$  is defined similarly to  $v^h(c)$ , only including lowly-sensitized transplants in the expectation calculation as well.

#### 4.1.3 A note on implementing this rule

Abraham, Blum, and Sandholm [1] presented the algorithm that, to our knowledge, is the most scalable optimal kidney exchange clearing algorithm created to date. An implementation of this algorithm is currently used by the UNOS exchange. Their solver uses *branch-and-price*, a technique that proves optimality by incrementally generating only a small part of the model during tree search [8]. This method is necessary, as the number of variables in the full integer program is too large to write down in memory (for all but small compatibility graphs).

Unfortunately, adding lexicographic preferences into the optimization model breaks the specific branch-and-price structure on which this solver relies (both in the deterministic model [1] and in the probabilistic model [15]). While a solver could still use column generation or general branch-and-price to solve this new problem, the addition of a matching-wide constraint—that a matching must contain cycles containing some fraction of a set of marked vertices—makes solving the pricing problem (see [8] for details) much more difficult than in the utilitarian formulation, where determining the price of a cycle not included in the current subproblem is relatively simple. Indeed, with such an allocation-wide constraint, finding a positive price cycle at a node in the search tree requires solving an integer program, whereas current solvers can use a (typically quite fast) depth-first search to find a positive price cycle in the standard kidney exchange model.

With these computational constraints in mind, in Section 4.2 we define a different fair utility function that respects the constraints of current solvers. This utility function circumvents a matching-wide fairness constraint. We will then compare both fair utility functions against the utilitarian one on real and simulated data in Section 5.

## 4.2 Weighted fairness

We now present a different formalization of fairness that relaxes the strict lexicographic preferences from the previous section. This definition generalizes the policy UNOS currently applies to highly-sensitized patients in their fielded kidney exchange.

Building on the standard (deterministic or probabilistic) kidney exchange integer programming formulation, we rewrite the objective as follows:

$$\max \sum_c v_\Delta(c)x_c$$

Here,  $v_\Delta(c)$  is the value of a cycle or chain  $c$  (either in the deterministic or probabilistic model) such that the weight of each edge  $e \in c$  is adjusted by some re-weighting function  $\Delta : E \rightarrow \mathbb{R}$ .

A simple example re-weighting function is multiplicative:

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

Intuitively, for some  $\beta > 0$ , this function scales the weight of edges ending in highly-sensitized vertices by  $(1 + \beta)$ . For example, if  $\beta = 0.5$ , then the optimization algorithm will value edges that result in a highly-sensitized patient receiving a transplant at 50% above their initial weight (possibly scaled by other factors like failure probability and chain position, as in the probabilistic model). We will use this multiplicative re-weighting in our experiments in Section 5.

For any  $M \in \mathcal{M}$ , let  $M'$  be the matching such that every cycle  $c \in M$  has augmented weight  $v_\Delta(c)$ . Then define the *weighted* fairness rule  $u_\Delta$  in terms of the utilitarian rule  $u$  applied to the augmented matching  $M'$ , such that  $u_\Delta(M) = u(M')$ .

### 4.2.1 A note on implementing this rule

Note that, unlike implementing the lexicographic fairness rule  $u_{H>L}^\alpha$ , this definition of fairness does not break the branch-and-price structure on which current scalable kidney exchange solvers rely. Indeed, the  $u_\Delta$  rule, for simple re-weighting functions like the multiplicative example above, can be implemented by first preprocessing a compatibility graph using  $\Delta$  to determine edge weights, and then solving the maximization problem using a standard kidney exchange solver.

## 5. EXPERIMENTAL VALIDATION

In this section, we compare the behavior of both the lexicographic fair rule  $u_{H>L}^\alpha$  and weighted fair rule  $u_\Delta$  (relative to the utilitarian rule  $u$ ) on (a) real data from 73 UNOS national kidney exchange match runs, from October 2010 to August 2013, and (b) simulated data from the standard kidney exchange compatibility graph models.

### 5.1 Results from the fielded UNOS exchange

The UNOS national kidney exchange began matching patient-donor pairs in October 2010. Initially matching monthly with approximately 70 hospitals, it has since grown to weekly matching with over 130 hospitals across the United States. We have been heavily involved with this exchange since its inception, and present fairness experiments on the first 73 match runs (through August of 2013) in this section.

The UNOS algorithm currently performs a utilitarian maximum weighted matching, where edge weights are set through a point-based system determined by a committee of doctors and policy-makers. Highly-sensitized patients in the UNOS exchange are those with CPRA of at least 80.

Results are presented for both the deterministic (denoted NO-FAIL) and probabilistic versions of the kidney exchange clearing

problem. Motivated by Dickerson, Procaccia, and Sandholm [15], we test the probabilistic model on two different edge failure probability distributions: one where each edge has a constant 70% rate of failure (denoted CONSTANT), and the other with a bimodal failure rate such that 25% of edge failures are drawn from  $U[0.0, 0.2]$  and the rest from  $U[0.8, 1.0]$  (denoted BIMODAL). These distributions are derived from real data (see §4.1 of [15] for details). We use a cycle cap of  $L = 3$ , as does UNOS, and include chains.

**Lexicographic rule.** We begin with the strictest version of the lexicographic fair rule:  $u_{H>L}^{1.0}$ . With  $\alpha = 1.0$ , this rule maximizes the number of highly-sensitized pairs in a match (which aligns with the theory of Section 3).

Metric	Minimum	Average	Maximum	St. Dev.
Loss (Objective) %	0.00%	2.76%	19.04%	4.84%
Loss (Cardinality) %	0.00%	4.09%	33.33%	8.18%
Loss (Cardinality)	0	0.55	4	1.10

**Table 1: Minimum, average, and maximum loss in objective value and match size due to  $u_{H>L}^{1.0}$ , across all 73 UNOS match runs, in the deterministic model.**

Table 1 presents results in the deterministic model. Under  $u_{H>L}^{1.0}$ , the price of fairness in this deterministic model is, on average, quite small; however, there are outlier cases in which large relative losses in the objective (of 19%) and overall match size (of 33%) are observed. We now explore this phenomenon in depth, in both the deterministic and probabilistic model.

Figure 3 presents cumulative distribution functions in efficiency loss with respect to the UNOS-weighted objective value, for the NO-FAIL, CONSTANT, and BIMODAL models.<sup>3</sup> In each of the models, roughly half of all UNOS match runs see no efficiency loss when prioritizing highly-sensitized candidates. However, (a) in each model, there exist a nontrivial number of matches with a nontrivial loss in efficiency, and (b) increasing the variability in failure rates increases the price of fairness. Indeed, in the BIMODAL model, some runs have a nearly 100% loss in efficiency! Intuitively, this is due to the optimizer being “forced” into including edges—possibly with a very low chance of successful execution—that result in a potential highly-sensitized transplant.

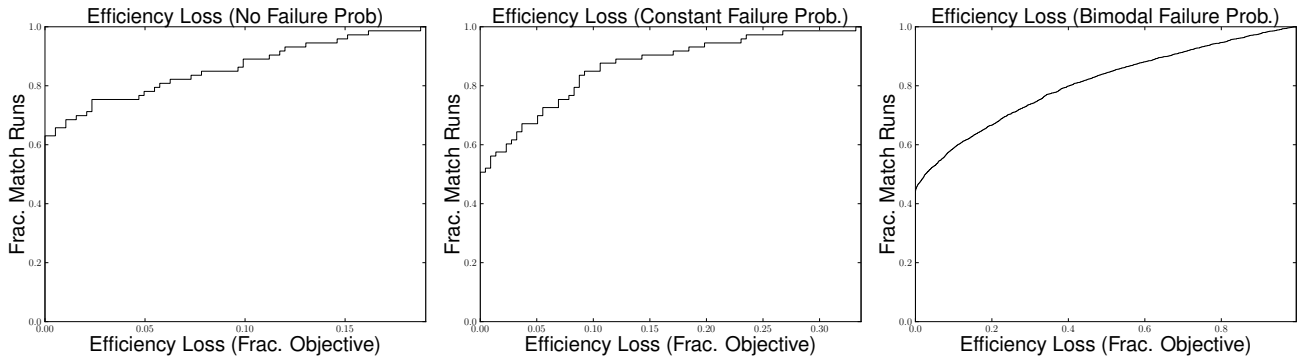
**Weighted rule.** We now give results for the weighted fair rule  $u_\Delta$ , where edges are re-weighted under the multiplicative rule  $\Delta^\beta$  defined in Section 4.2. Recall that  $\Delta^{0.0}$  values highly-sensitized transplants at the same rate as lowly-sensitized ones,  $\Delta^{0.5}$  values them at 50% over their base value,  $\Delta^{1.0}$  at 100% over their base value, etc.

The Pareto frontiers shown in Figure 4 represent the set of Pareto efficient matchings constrained by  $u_\Delta$  as  $\beta$  increases from 0, for each of the NO-FAIL, CONSTANT, and BIMODAL models. Intuitively, these plots visualize the relationship between favoring a subset of vertices versus the overall match efficiency. As with the lexicographic rule, higher diversity in the underlying failure probability distribution begets a greater price of fairness.

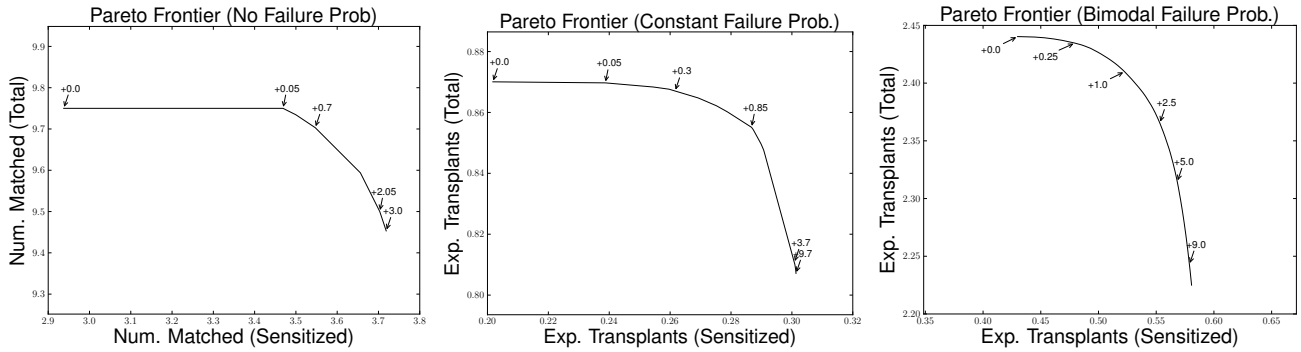
### 5.2 Simulated results from random graphs

Fielded kidney exchanges are still young and have relatively small pools, containing at most a couple of hundred pairs at a time. To explore fairness-aware matching behavior in larger pool sizes, and to validate the theory developed in Section 3, we now turn to generated data. We look at two models:

<sup>3</sup>Cumulative distribution functions of efficiency loss as a function of match size exhibit similar behavior and are omitted due to space constraints.



**Figure 3: Cumulative distribution functions of the price of fairness under the lexicographic fairness rule  $u_{H>L}^{1.0}$  according to UNOS' weighting policy, on 73 UNOS match runs since the inception of the exchange.**



**Figure 4: Pareto frontiers for  $u_{\Delta}$  under different failure probability distributions, for  $\beta \in \{0.0, 0.05, \dots, 10.0\}$ .**

1. The most well-known model of kidney exchange is due to Saidman et al. [24]. It generalizes the theoretical model in Section 3, and includes blood types, three tiers of CPRA (low, medium, and high), and various other medical aspects that affect pool composition. We refer to the base Saidman model, which is parameterized with US national data, as Saidman (US). We also parameterize the model using blood type and sensitization statistics from the 73 UNOS match runs (aggregate data is available in [19]), and denote this Saidman (UNOS).
2. Some recent theoretical work looks at kidney exchange graphs without blood types but with increasing sparsity (in the size of the pool) for highly-sensitized candidates [4, 5]. For these experiments, the probability of an incoming edge to a highly-sensitized pair is  $O(1/n)$ . The probability of an incoming edge to lowly-sensitized pairs is held constant, as before. We denote by Heterogeneous this family of random graphs.

Table 2 gives the average loss in efficiency for each of these models over multiple generated pool sizes, with 40 runs per pool size per model, under the strict lexicographic rule  $u_{H>L}^{1.0}$ . For all but the smallest pools, Heterogeneous graphs see no loss in efficiency at all. Efficiency loss in the Saidman (US) and Saidman (UNOS) families of graphs are low (and statistically indistinguishable), aligning with our earlier theoretical results (applied to the distribution of blood types in the US).<sup>4</sup>

<sup>4</sup>Losses under the weighted fair utility rule  $u_{\Delta}$  were similarly small and not included due to space constraints.

Size	Saidman (US)	Saidman (UNOS)	Heterogeneous
10	0.24% (1.98%)	0.00% (0.00%)	0.98% (5.27%)
25	0.58% (1.90%)	0.19% (1.75%)	0.00% (0.00%)
50	1.18% (2.34%)	1.96% (6.69%)	0.00% (0.00%)
100	1.46% (1.80%)	1.66% (3.64%)	0.00% (0.00%)
150	1.20% (1.86%)	2.04% (2.51%)	0.00% (0.00%)
200	1.43% (2.08%)	1.55% (1.79%)	0.00% (0.00%)
250	0.80% (1.24%)	1.86% (1.63%)	0.00% (0.00%)
500	0.72% (0.74%)	1.67% (0.82%)	0.00% (0.00%)

**Table 2: Average (St. Dev.) percentage loss in efficiency for three families of random graphs, under the strict  $u_{H>L}^{1.0}$  rule.**

## 6. CONCLUSIONS & FUTURE RESEARCH

Fielded kidney exchanges use utilitarian or near-utilitarian matching rules, at the cost of marginalizing certain classes of patient-donor pairs. In this paper, we focused on balancing overall exchange efficiency while improving access to kidneys for highly-sensitized patients. We defined the price of fairness in the standard kidney exchange model, and provided theoretical bounds in the major kidney exchange model. We introduced two natural definitions of fairness—lexicographic and weighted—and empirically explored the tradeoff between prioritizing hard-to-match patients and the overall efficiency of a utilitarian system, on real data from the UNOS nationwide kidney exchange and on simulated data from each of the standard kidney exchange distributions.

How to prioritize highly-sensitized patients—if they should be prioritized at all—is currently the most contentious issue regarding fairness in kidney exchange, which motivates this paper. We note that the price of fairness concept applies to valuing *any* subset of vertices—not just highly-sensitized ones—in the compatibility graph, possibly under different prioritization rules. A clear next step would be developing analogous theoretical results and empirical techniques applicable to fielded kidney exchange that generalize the equity versus efficiency tradeoff presented here to other notions of fairness (while mimicking present-day parameters of the compatibility pool, legal climate, and limits of medical knowledge). Recent work in finding Lorenz-dominant matchings is promising [20], but not yet applicable to fielded kidney exchange due to its simple theoretical model (e.g., only 2-cycles, no chains).

Kidney exchange is naturally dynamic, where patients and donors arrive to and depart from the pool over time. Both the theoretical and experimental sides of utilitarian dynamic kidney exchange are active and important areas of research [2, 5, 7, 13, 27]. Developing accurate models and scalable algorithms that consider the price of fairness in the dynamic setting will be of increasing importance as fielded kidney exchanges move from static to dynamic matching.

**Acknowledgments.** This material was funded by NSF grants IIS-1320620, CCF-1101668, CCF-1215883, and IIS-0964579, by an NDSEG fellowship, and used the Pittsburgh Supercomputing Center in partnership with the XSEDE, which is supported by NSF grant OCI-1053575. We thank Intel Corporation for machine gifts.

## 7. REFERENCES

- [1] D. Abraham, A. Blum, and T. Sandholm. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *ACM Conference on Electronic Commerce (EC)*, pages 295–304, 2007.
- [2] E. Anshelevich, M. Chhabra, S. Das, and M. Gerrior. On the social welfare of mechanisms for repeated batch matching. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 60–66, 2013.
- [3] I. Ashlagi, F. Fischer, I. Kash, and A. D. Procaccia. Mix and match. In *ACM Conference on Electronic Commerce (EC)*, pages 305–314, 2010.
- [4] I. Ashlagi, D. Gamarnik, M. Rees, and A. E. Roth. The need for (long) chains in kidney exchange. NBER Working Paper No. 18202, July 2012.
- [5] I. Ashlagi, P. Jaillet, and V. H. Manshadi. Kidney exchange in dynamic sparse heterogeneous pools. In *ACM Conference on Electronic Commerce (EC)*, pages 25–26, 2013.
- [6] I. Ashlagi and A. E. Roth. Individual rationality and participation in large scale, multi-hospital kidney exchange. In *ACM Conference on Electronic Commerce (EC)*, pages 321–322, 2011.
- [7] P. Awasthi and T. Sandholm. Online stochastic optimization in the large: Application to kidney exchange. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, pages 405–411, 2009.
- [8] C. Barnhart, E. L. Johnson, G. L. Nemhauser, M. W. P. Savelsbergh, and P. H. Vance. Branch-and-price: Column generation for solving huge integer programs. *Operations Research*, 46(3):316–329, 1998.
- [9] D. Bertsimas, V. F. Farias, and N. Trichakis. The price of fairness. *Operations Research*, 59(1):17–31, 2011.
- [10] D. Bertsimas, V. F. Farias, and N. Trichakis. Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. *Operations Research*, 61(1):73–87, 2013.
- [11] I. Caragiannis, A. Filos-Ratsikas, and A. Procaccia. An improved 2-agent kidney exchange mechanism. International Workshop on Internet and Network Economics (WINE), 2011.
- [12] I. Caragiannis, C. Kaklamanis, P. Kanellopoulos, and M. Kyropoulou. The efficiency of fair division. International Workshop on Internet and Network Economics (WINE), 2009.
- [13] J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Dynamic matching via weighted myopia with application to kidney exchange. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 1340–1346, 2012.
- [14] J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Optimizing kidney exchange with transplant chains: Theory and reality. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 711–718, 2012.
- [15] J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Failure-aware kidney exchange. In *ACM Conference on Electronic Commerce (EC)*, pages 323–340, 2013.
- [16] S. E. Gentry, R. A. Montgomery, B. J. Swihart, and D. L. Segev. The roles of dominos and nonsimultaneous chains in kidney paired donation. *American Journal of Transplantation*, 9(6):1330–1336, 2009.
- [17] S. E. Gentry and D. L. Segev. The honeymoon phase and studies of nonsimultaneous chains in kidney-paired donation. *American Journal of Transplantation*, 11(12):2778–2779, 2011.
- [18] HHS/HRSA/HSB/DOT. OPTN/SRTR annual data report, 2011.
- [19] Kidney Paired Donation Work Group. OPTN KPD pilot program cumulative match report (CMR) for KPD match runs: Oct 27, 2010 – Apr 15, 2013, 2013.
- [20] J. Li, Y. Liu, L. Huang, and P. Tang. Egalitarian pairwise kidney exchange: Fast algorithms via linear programming and parametric flow. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2014.
- [21] R. Montgomery, S. Gentry, W. H. Marks, D. S. Warren, J. Hiller, J. Houpp, A. A. Zachary, J. K. Melancon, W. R. Maley, H. Rabb, C. Simpkins, and D. L. Segev. Domino paired kidney donation: a strategy to make best use of live non-directed donation. *The Lancet*, 368(9533):419–421, 2006.
- [22] M. Rees, J. Kopke, R. Pelletier, D. Segev, M. Rutter, A. Fabrega, J. Rogers, O. Pankewycz, J. Hiller, A. Roth, T. Sandholm, U. Ünver, and R. Montgomery. A nonsimultaneous, extended, altruistic-donor chain. *New England Journal of Medicine*, 360(11):1096–1101, 2009.
- [23] A. Roth, T. Sönmez, U. Ünver, F. Delmonico, and S. L. Saidman. Utilizing list exchange and nondirected donation through ‘chain’ paired kidney donations. *American Journal of Transplantation*, 6:2694–2705, 2006.
- [24] S. L. Saidman, A. Roth, T. Sönmez, U. Ünver, and F. Delmonico. Increasing the opportunity of live kidney donation by matching for two and three way exchanges. *Transplantation*, 81(5):773–782, 2006.
- [25] P. Toulis and D. C. Parkes. A random graph model of kidney exchanges: efficiency, individual-rationality and incentives. In *ACM Conference on Electronic Commerce (EC)*, pages 323–332. ACM, 2011.
- [26] United Network for Organ Sharing (UNOS). <http://www.unos.org/>.
- [27] U. Ünver. Dynamic kidney exchange. *Review of Economic Studies*, 77(1):372–414, 2010.
- [28] S. Woodle, J. Daller, M. Aeder, R. Shapiro, T. Sandholm, V. Casingal, D. Goldfarb, R. Lewis, J. Goebel, and M. Siegler. Ethical considerations for participation of nondirected living donors in kidney exchange programs. *American Journal of Transplantation*, 10:1460–1467, 2010.