Welfare and Equity Consequences of Transplant Organ Allocation Policies*

Tayfun Sönmez†
Department of Economics, Boston College
Distinguished Research Fellow, Koç University

M. Utku Ünver‡
Department of Economics, Boston College
Distinguished Research Fellow, Koç University

Incomplete and Preliminary

Abstract

Within the last decade kidney exchange became a mainstream paradigm to increase the number of kidneys transplants. However, compatible pairs do not participate in exchange and full benefit from exchange can only be realized if they participate. In this paper, we propose a new incentive scheme that relies on incentivizing participation of compatible pairs in exchange via an insurance for the patient for a second future renal failure. Welfare and equity analysis of this scheme is conducted and compared with welfare and equity outcomes of live donation and live donor organ exchange. The potential role of such an incentive scheme to strengthen the national kidney exchange system is also presented.

*M. Utku Ünver acknowledges the research support of NSF. We thank the participants at NBER Market Design Working Group Meeting, NSF/CEME Decentralization Conference, SMU, CIDE in Mexico City, Barcelona JOCS Seminar, CoED at Lund, Brown, and Cornell for their comments.

†Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467; sonmezt@bc.edu; www2.bc.edu/~sonmezt

‡Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467; unver@bc.edu; www2.bc.edu/~unver

1 Introduction

Several studies offered new policy suggestions and their welfare analyses in allocating transplant organs in a "partial equilibrium" setup: for example for deceased donation (cf. Zenios, Chertow, and Wein, 2000), or for live-donor exchanges (cf. Roth, Sönmez, and Ünver, 2004).

We present the first dynamic model that analyzes welfare and equity consequences of various deceased-donor allocation, live donation, and live-donor exchange policies across different patient groups, categorized based on the patients' blood types, the availability of paired live donors, and these donors' blood types when they are available, participating in different phases of transplantation process for different organs. The model gives us an explanation for some of the observed patterns in the data.

In particular, we propose a new exchange policy for live donors, which can substantially increase the number of pairs that can be matched through exchange while reducing inequality across blood type. We characterize the potential gains from this policy using our new model.

Currently compatible pairs generally do not participate in exchange, as the patient of the pair directly receives an organ from his donor. Only incompatible pairs participate in exchange. Incompatible pairs are either (a) blood-type incompatible (such as with an O blood-type patient and an A blood-type donor) or (b) blood-type compatible but tissue-type incompatible (such as the reciprocal of the above pair, with an A blood-type patient and O blood-type donor who is tissue-type incompatible with her patient). Because of this asymmetry, blood-type incompatible pairs are substantially more in number than blood-type compatible pairs participating in exchange. Moreover, for a blood-type incompatible pair to benefit from exchange with the exception of pairs with A and B blood types for the donor and the patient, a blood-type compatible pair is needed. However, the asymmetry in participation puts blood-type incompatible pairs at a high disadvantage and as a result not all pairs can benefit from exchange and the ones who can benefit have wait for their reciprocal blood-type compatible pairs to arrive at the pool. On the other hand, if compatible pairs can also participate in exchange, then the participation asymmetry will disappear, and exchange will benefit more than 90% of the pairs (cf. Roth, Sönmez, and Ünver, 2005a; Sönmez and Ünver, 2010).

However, it is not possible to force compatible pairs to participate in exchange. We propose to incentivize participation by linking deceased-donor queue with the exchange pool. It is a common practice to give priority to live donors on the deceased-donor queue in case they themselves get sick and need an organ transplant in the future. We propose giving similar incentives to the patients of compatible donors who give up their own compatible donor's organ in exchange for another pair's compatible organ. In this manner, the patient of a compatible donor receives a "guarantee" not to wait in the deceased-donor queue by getting a "priority" in case the organ he receives in exchange fails in the future.

Another benefit of this policy can be seen in creating unified national programs for exchange.

One of the biggest hurdles that needed to be overcome in kidney exchange in the US is unification of various kidney exchange programs. It is well established that running a single large program that encompasses all programs benefits more patients than running separate programs (cf. Roth, Sönmez, and Ünver, 2004, 2007). On the other hand, linking deceased-donor queue with the live-donor exchange queue can only be done through the national kidney exchange program governed by the federal contractor, United Network for Organ Sharing (UNOS), which also directs the deceased-donor allocation program in the US. We show that in an environment which has multiple exchange programs, if compatible pairs participate exchange, they would participate through the national program of UNOS, which has the jurisdiction over the deceased-donor queue, and in turn, this will attract other pairs to the national program rather than other programs. Hence, our proposed policy has the potential to unify various exchange programs to create a large exchange platform to exploit all benefits from exchange for the society.

1.1 Other Findings

We start our analysis by incorporating deceased-donor donation to the model, to predict the steady-state welfare consequences for different blood-type patients. Two types of deceased donation policies play an important role for many organs. The first commonly adopted policy is the *ABO-identical* allocation where a patient can only receive a transplant from a deceased donor with the same blood type. The second policy is the *ABO-compatible* allocation where a patient can receive a transplant from any compatible blood-type of deceased-donors. Both policies are governed through a priority allocation scheme which gives the greatest weight to waiting time for the ABO-identical allocation. We model the priority allocation policy through a *first-in-first-out* queue. The ABO-compatible allocation policy leads to a "pooling effect" by equalizing the waiting time of different blood types whose donors donate / patient receive to / from this group of patients / donors. On the other hand, ABO-identical allocation policy leads to separation of waiting times for different blood types with respect to the patient / donor inflow ratios of that blood type.[1]

Then, we consider live donation. Some patients have paired donors who would like to donate an organ, such as a kidney or part of the liver, to them. If they have blood- and tissue-type compatibility they donate to their patients and otherwise they are not utilized. Possibility of live donation helps unambiguously all patients, those with donors and without donors. We characterize the gains from live donation in our model. Patients without live paired donors benefit as patients with compatible

[1]For example, for minorities where B blood type could be a dominant blood type unlike the majority of the population, deceased donation rates do not differ from the rest of the population, yet people are more prone to need transplant due to life-style choices are other factors. Hence, the ABO-compatible and ABO-identical allocation policies are expected to lead to substantially different waiting times for B blood-type patients, who can receive organs from O blood-type deceased donors besides B blood-type deceased donors.

live paired donors drop from competition for deceased donors. Among different blood types, O blood-type deceased-donor queue benefits the least while AB queue benefits the most. O blood-type patients with a live donor have a lower chance to have a compatible live donor with respect to other types, as they can only receive from O blood-type donors. AB blood-type patients can receive from all types, and have the highest rate of compatible donors. On the other hand, A blood-type patients are more better off than B. As A is a more common blood type in the population, and hence for the paired donors of the patients with respect to B, and hence,a higher fraction of A blood-type patients benefit from live donation with respect to B. We quantify the amount of change in waiting time for a deceased donor in queues for different blood types when compatible live donation is feasible.

Next, we consider live-donor exchanges among incompatible pairs, for organs such as kidneys and livers. We characterize the welfare consequences of live-donor exchange on different blood type patients. A live-donor exchange involves the swap of paired live-donor organs of two pairs when the donors are incompatible with their own patients but compatible with the patient of the other pair. This causes patients with blood-type compatible donors to be matched immediately either through direct donation (if compatible) or through exchange (if compatible). On the other hand the patients who have blood-type incompatible donors need to wait in the pool, and conditional on survival, they get matched either through exchange with a live donor or with a deceased donor depending on the population characteristics.

2 A Dynamic Model of Transplant Patients

We consider a comprehensive dynamic organ transplantation model (for organs such as heart, kidney, liver, and pancreas) to which the deceased-donor queue, live donation for kidneys and livers, and live-donor kidney and liver exchange can be incorporated. We consider a continuum flow model in analysis where the number of patients are in Lebesgue measures at a steady state.

Consider patients who need a particular organ transplant. Each **patient** is represented by his blood type $X \in \mathcal{T} = \{A, B, AB, O\}$. Suppose p_X refers to the probability of having the probability of X blood type in the population distribution. We assume that there is an inflow π_X of blood-type X people getting sick per unit time. Suppose that in the population of new patients, the expected life time while living with the disease is distributed with a strictly increasing differentiable distribution function $F(\cdot)$[2] on the interval $[0, T]$.[3] Thus, among the inflow of π_X measure of blood-type X blood-type patients at a given time, the measure of patients who are alive after t years on is given by

[2]I.e., the probability density function $f(\cdot)$ is well defined and positive in $(0, T)$.

[3]This expectancy is different for different organs due to disease progression and techniques that can be used to substitute for the deficiency in the body because of the failing organ. For example, kidney patients, who can live on dialysis, have in general longer survival times.

$\pi_X[1 - F(t)]$.[4]

In Table 1, survival rates, $1 - F(t)$, for kidneys are listed. [5]

	Time					
	6 mo.	1 yr.	2 yr.	3 yr.	4 yr.	5 yr.
On dialysis (for kidneys)	84%	75%	61%	50%	42%	34%

Table 1: Survival rates $(1 - F(t))$ for kidney failures in the US.

At the steady state, when transplantation option is not present, the total mass of blood-type X blood-type patients is $\int_0^T \pi_X[1 - F(t)]dt$.[6] (See Figure 1.)

3 Organ Transplantation and Deceased-Donor Queue

The best remedy for organ failure is transplantation. A donor should be both blood- and tissue-type compatible with a patient, before her organ(s) can be transplanted.[7] O blood-type donors are blood-type compatible with all blood-type patients. A blood-type donors are blood-type compatible with A and AB blood-type patients and B blood-type donors are blood-type compatible with B and AB blood-type patients. On the other hand, AB blood-type donors are only blood-type compatible with AB blood-type patients. **Blood-type compatibility** is formally defined through a partial order $\triangleright$ over blood types, such that $X \triangleright Y$ means than X blood-type donors are blood-type compatible with Y blood-type patients (see Figure 2). Blood type distribution among the US ethnic groups are reported in Table 2.[8] In general O blood type is the most common one, while AB is the rarest; A is

[4]Hence, $\pi_X dt$ is the 2-dimensional *Lebesgue measure* of patients who enter in a small time interval dt. By a slight abuse of terminology, throughout the paper we will refer to 1-dimensional Lebesgue measures of agent sets, which are in general *inflow intensities* such as π_X, by **measure**. On the other hand 2-dimensional Lebesgue measures of agents sets such as $\pi_X dt$ will be referred to as **mass**.

[5]The kidney data includes 2005 estimates from National Kidney Organization 2012 Annual Report retrieved from http://www.usrds.org/2012/pdf/v2_ch5_12.pdf on 02/25/2102 while on dialysis.

[6]Although we assume that inflow of patients is constant over time, we could easily make it a function of time as well. For example, population growth is a reason for increase of inflow. Increase in longevity is another reason, which not only affects π_X but also F, as older people have a higher tendency to need organ transplantation. These can be incorporated in our model easily. In that case a steady state does not exist. However, we can carry all of our analysis in this paper and draw similar results in that model as a function of time. For simplicity and transparency of our analysis, we will use a model with constant inflows.

[7]Tissue-type incompatibility occurs for some organs such as kidneys, while it is not an issue for some other organs such as livers. Blood-type incompatibility is an issue for all organs.

[8]Retrieved from http://bloodbook.com on 03/18/2013. US general population is constructed using the ethnic proportions.

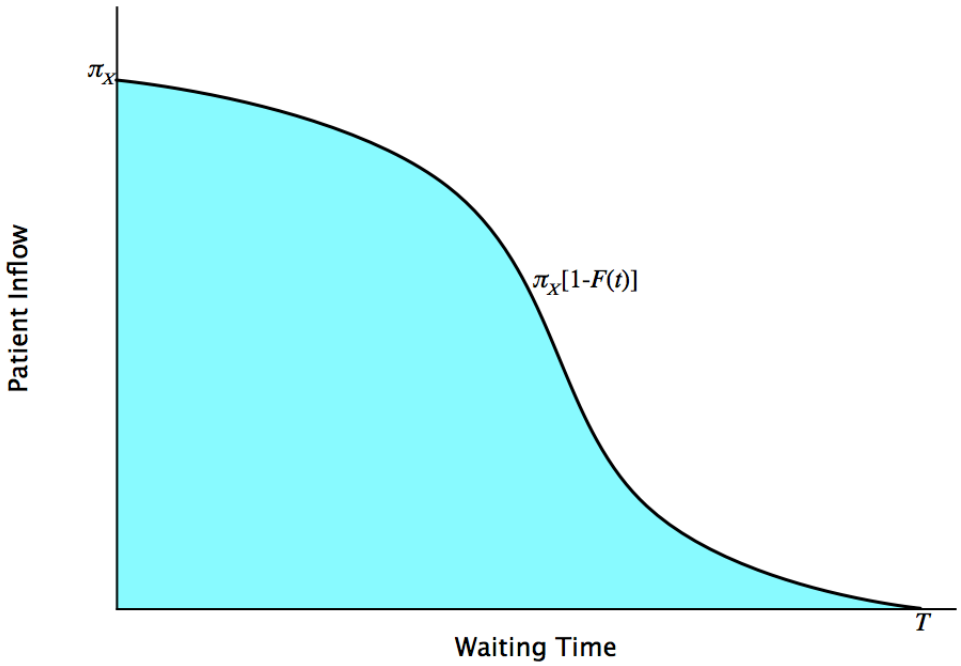

Figure 1: Steady-state X blood-type patient distribution over waiting time when organ transplantation is not possible. The shaded area is the mass of patients who are alive.

observed more commonly than B, while their rates are substantially different across different ethic groups: B has a big presence among Asian- and African-American groups while this is not the case for the white Americans. The world blood-type distribution more or less are similar geographically according to the origins of the US ethnic groups.

Once a donor is deemed compatible with a patient, she has to be also tissue-type compatible with the patient. Tissue-type compatibility requires that the patient's body does not form antibodies against a donor's DNA. Throughout the paper we assume that given a patient and a blood-type compatible donor, **tissue rejection** occurs with a probability $\theta < 1$.[9]. For some organs, such as livers, tissue rejection is not an important problem. In those cases, we can assume $\theta \approx 0$. On the the other hand, for other organs, such as kidneys, tissue rejection rate is significant, and hence, $\theta > 0$.

A common source of donation across organs is deceased donors. The deceased-donor queue is governed by a central entity. For example in the US, for all organ types, United Network for Organ

[9]In real life, tissue rejection probability may be different across the patient population. In those cases, we can instead assume, the rejection probability is a random variable $\hat{\theta}$ with a well-defined mean θ. As long as the probability density function of θ does not include in its support, full rejection probability 1, our analysis in this paper goes through using its mean θ. It is a reasonable assumption to assume that no patient ever rejects all blood-type compatible donors, as at least he will always be tissue-type compatible with his perfect HLA-matched donor, i.e., a donor who has the same 6 tissue-typing proteins on her DNA as him

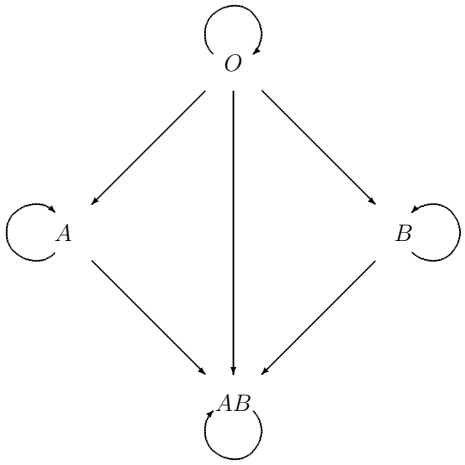

Figure 2: Blood-type compatibility partial order $\triangleright$

	Blood Types				Pop. %
	O	A	B	AB	— (1992)
African American	49%	27%	20%	4%	12.4%
Asian American	40%	28%	27%	5%	3.3%
Native American	79%	16%	4%	1%	0.8%
White American	45%	40%	11%	4%	83.4%
US all	45.6 %	37.8%	12.6%	4%	

Table 2: Blood Type Distribution in the US.

Sharing (UNOS) is the federal contractor that is in charge of the queue. We assume throughout the paper that any patient enrolled in the queue remains in the queue until he receives a transplant or he dies.

We denote the inflow of the X blood-type deceased donors, as $\delta_X < \pi_X$ per unit time. Across the blood types, the ratio δ_X/π_X need not be constant. For example, it is well known that among minority communities, organ failure is more prominent than the white American population while deceased donation rates are not that significantly different.As blood-type distribution of minorities are different from the white American population (especially B blood type is observed at a much higher frequency among Asian- and African-Americans, see Table 2), the ratio δ_X/π_X is not constant across blood types for the U.S. population: While a very high percentage of the donors, live or deceased, are white Americans, the patient rate of white Americans is much lower than their donation rate for kidneys and is only higher for lungs. On the other hand, for kidneys and hearts, patient rate of African-Americans is higher than their donation rates; while for kidneys and livers, patient rate

of Asian-Americans are higher from their donation rates. [10] Although these rates are distorted by many other factors such as live donation possibilities, we can conclude that especially for kidneys the ratio δ_B/π_B is lower than other blood types.

When a transplanted organ, i.e., graft, fails, the recipient reenters the deceased-donor queue as a new patient. Repeat patients survival function on the deceased-donor queue is "similar to" that of primary entrants (for example, that is the case for kidneys), so we assume $1-F$ is also their survival function while waiting in the deceased-donor queue. We assume that ϕ^d is the steady-state fraction of the previous recipients whose grafts fail and reenter the deceased queue per new deceased-donor transplant conducted. [11,12] Thus, if at steady state a ι measure of X blood-type patients receive a deceased-donor organ per unit time, then a $\phi^d\iota$ measure of previous recipients will reenter the queue per unit time. In 2005, $13.5\%, 7.9\%, 4.1\%, 5.5\%$ of all new kidney, liver, heart, and lung patients, respectively, were repeat entrants (Magee et al., 2007). In general, allocation policies do not differentiate primary transplant patients and repeat transplant patients.

3.1 The Deceased-Donor Queue Matching Protocols

The deceased-donor organs are allocated through the points system of UNOS, which is a priority mechanism. When a deceased donor arrives, the point total for each compatible patient for the donor is determined. The organ is offered to the patient with the highest point total. If it is rejected by the patient or his doctor for any reason, then the organ is offered to the next patient, so on. In general, different point schemes are used for different organs. Deceased donor allocation policies usually differ across organs and across geographic transplant regions, although usually a centralized mechanism is used in allocation. For example for kidneys, strict *ABO-identical* allocation policies are applied. That is, kidneys of blood-type X are only offered to blood-type X blood-type patients.[13] On the

[10]From the US Department of Health and Human Services - The Office of Minority Health web page for organ donation https://minorityhealth.hhs.gov/templates/browse.aspx?lvl=3&lvlid=12 retrieved on 02/25/2013.

[11] Fraction ϕ^d is formally calculated as follows: Suppose a measure ι of patients receive transplants per unit time at steady state. If the patient's life with a healthy graft ends, two things could be the reason: either the patient dies, or patient is alive but his graft fails. Of the patients leaving the status of living with a healthy graft, let $h_1(t)$ be the fraction that die after t years from the transplant and $h_2(t)$ be the fraction whose grafts fail after t years from the transplant. Thus, a random patient's expected lifetime with a healthy graft is distributed with a differentiable distribution function $H(\cdot)$ in some interval $[0,S]$ such that $\frac{dH(t)}{dt} \equiv h(t) \equiv h_1(t)+h_2(t)$ where t refers to years passed since the transplant. We assume that this distribution is independent of how long the patient waited initially in the queue to receive his previous transplant. Then the inflow of patients reentering the deceased-donor queue at each time is given by $\int_0^S \iota h_2(t)dt = \iota\int_0^S h_2(t)dt$. We set $\phi^d = \int_0^S h_2(t)d(t)$. Observe that $\phi^d < \int_0^S h(t)d(t) = 1$.

[12]For simplicity, we assume that it is constant, although it may possibly change as the age distribution of the patients receiving transplants changes in the deceased-donor queue, i.e., it may be a function of the waiting time.

[13]In the highly unlikely event that no X blood-type patient is available, then the organ is offered to a compatible patient.

other hand, livers are offered to *ABO-compatible* patients under bylaws of some regions. We inspect the welfare and distributional consequences of these two policies on different blood-type patients separately.

Given a fixed blood-type allocation policy, waiting time of a patient is often the most significant contributor to the points of a patient in deceased-donor allocation for many organs such as kidney, pancreas, or heart. Therefore, we will model deceased-donor allocation using **first-in-first-out** (**FIFO** from now on) queues for both the ABO-identical and ABO-compatible allocation schemes.[14]

We analyze these two FIFO matching protocols. We state the following lemma, which will help us model the steady state of the deceased-donor queue.[15]

Lemma 1 (FIFO Matching Protocol) *Consider the FIFO matching protocol. Suppose that there is an ordered ω measure of X blood-type patients available in the queue and a $\sigma \leq \omega$ measure of blood-type compatible donors arrive. Then*

- *if $\sigma = \omega$, then all donors, possibly except a finite (and thus of 0 measure) of them, are almost surely matched; and*
- *if $\sigma < \omega$, then all donors are almost surely matched.*

3.2 Steady State of The Deceased-Donor Queue

We are ready to characterize the steady state of the deceased-donor queue under the two FIFO matching protocols.

3.2.1 ABO-Identical Deceased Donation

Consider any blood type X. In the steady state, as $\delta_X < \pi_X$, there will always be a positive mass of X blood-type patients available in the deceased-donor queue. Moreover, as FIFO protocol is used, the organs of the δ_X measure will be transplanted to the longest waiting X blood-type patients who survived in the queue. Thus, by Lemma 1, these donors will be almost surely matched to the longest waiting cohort of δ_X measure of patients. We make the following observation regarding the reentries to the queue (see Figure 3).

[14]UNOS is preparing to switch to a new deceased-donor kidney allocation scheme that will use a quality-based allocation scheme for 20% of all allocation, while 80% of all allocation will continue to be done through its current FIFO type policy.

[15]This is in spirit similar to the Erdös and Rényi (1960) random graph convergence result. However, in substance it is different, as we are not using the maximal matching policy as in Erdös and Rényi (1960) but rather the FIFO matching policy.

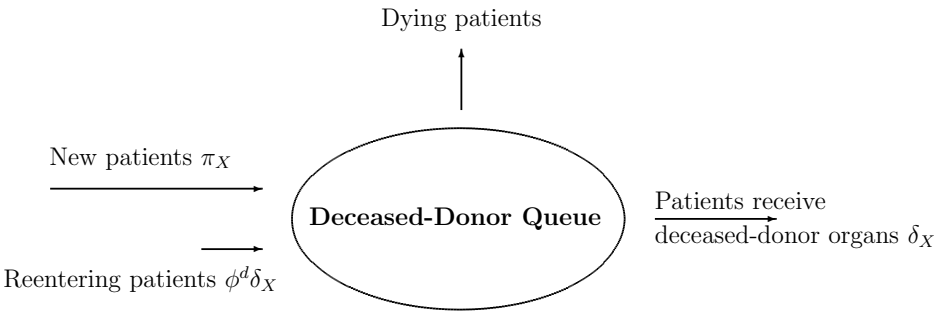

Figure 3: Inflow to and outflow from the X blood-type deceased-donor queue under the ABO-identical deceased donation policy at steady state.

Observation 1 *Under ABO-identical deceased-donor allocation policy, as a δ_X measure of X blood-type patients receive transplants per unit time, a $\phi^d\delta_X$ measure of previous recipients will reenter the deceased-donor queue per unit time due to graft failure.*

Let the receiving cohort have arrived t_X years before the current point in time. As there is a $[\pi_X + \phi^d\delta_X][1 - F(t_X)]$ measure of patients in this cohort including reentries and new arrivals, we should have

$$[\pi_X + \phi^d\delta_X][1 - F(t_X)] = \delta_X.$$

Hence, at steady state, the time spent on the X blood-type queue by the receiving cohort can be found through $t_X = F^{-1}(1 - \frac{\delta_X}{\pi_X+\phi^d\delta_X}) < T = F^{-1}(1)$. This is also the waiting time for X blood-type patients conditional on survival. Based on this analysis, we state the following characterization of the deceased-donor queue at steady state: (See also Figure 4.)

Theorem 1 (ABO-Identical Deceased Donation) *Under the ABO-identical FIFO deceased-donor allocation policy, at steady state, the (expected) waiting time for X blood-type patients in the deceased-donor queue is*[16]

$$t_X^{\mathbf{i}} \quad = F^{-1}\big(1 - \tfrac{\delta_X}{\pi_X+\phi^d\delta_X}\big), \tag{1}$$

and the mass of the patients in the deceased-donor queue is

$$\int_0^{t_X^{\mathbf{i}}} [\pi_X + \phi^d\delta_X][1 - F(t)]dt.$$

Proof. It follows from the analysis before the theorem. ■

[16]The superscript in deceased donation waiting time $t_X^{\mathbf{i}}$ or $t_X^{\mathbf{c}}$ refers to the type of deceased donation policy we explore, either ABO-**i**dentical or ABO-**c**ompatible, respectively.

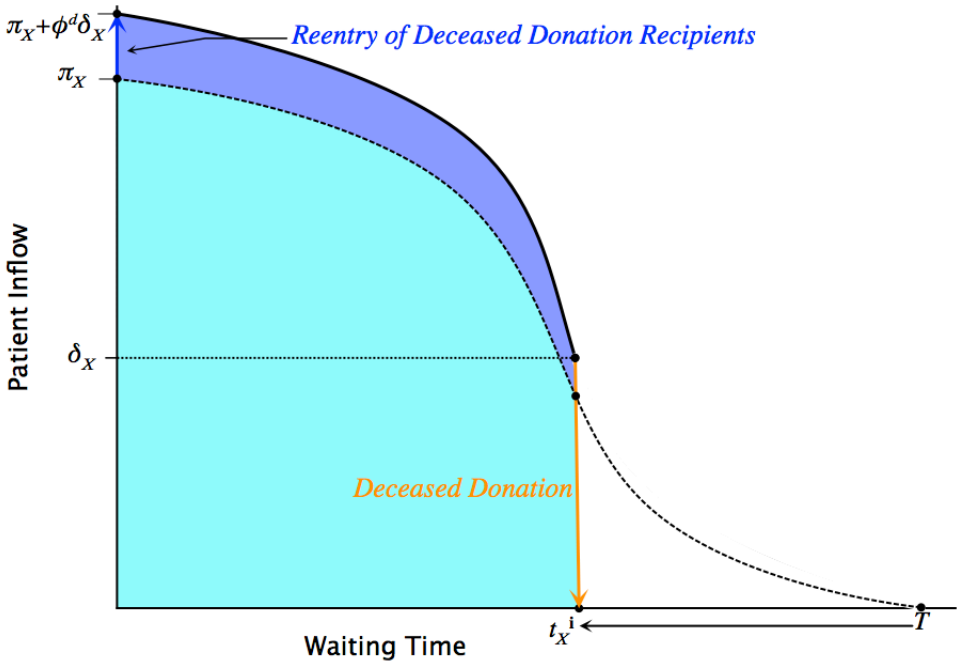

Figure 4: **X blood-type deceased-donor queue under the ABO-identical decease donation policy at steady state**: Incoming deceased donors, of a δ_X measure, are matched with a δ_X measure of the longest waiting patients at each time. New patients, who arrive at an inflow rate of π_X, and reentrants (whose previously received organs failed), who arrive with an inflow rate of $\phi^d\delta_X$, join the queue. Waiting time conditional on survival decreases from T to $t_X^i = F^{-1}(1 - \frac{\delta_X}{\pi_X+\phi^d\delta_X})$. Darker shaded region refers to the mass of reentrants in the queue.

3.2.2 ABO-Compatible Deceased Donation

The following lemmata relate the role of blood-type compatibility relationship to the waiting times of different blood types under the ABO-compatible deceased-donor allocation policy.

Lemma 2 *Let $X \neq Y$ be two blood types such that $X \triangleright Y$. Then under the ABO-compatible FIFO deceased-donor allocation, waiting times of X and Y blood-type patients at steady state satisfy $t^{\mathbf{c}}_Y \leq t^{\mathbf{c}}_X$.*

We make the following formal definition of *pooled* blood types:

Definition 1 *If blood types in some $\mathcal{S} \subseteq \mathcal{T}$ donate organs only to the blood types in $\mathcal{S}$ and they receive organs only from blood types in $\mathcal{S}$ at steady state, and there is no proper subset of $\mathcal{S}$ with this property, then we say that blood types in $\mathcal{S}$ are **pooled**.*

For example if O blood-type organs are transplanted to A and B blood-type patients besides O, and A and B blood-type organs are only transplanted to A and B blood-type patients, respectively, then $\{O, A, B\}$ is a pooled set. On the other hand neither $\{O, A\}$ is pooled (as O blood-type organs are also transplanted to B blood-type pairs) or $\{A, B\}$ is pooled (as both A and B blood-type patients also receive O blood-type organs). The whole blood type set $\mathcal{T} = \{O, A, B, AB\}$ is not pooled, either, as its proper subset $\{O, A, B\}$ is pooled. Lemma 3 characterizes the waiting times of pooled blood types:

Lemma 3 *For two distinct blood types X and Y, if Y blood-type patients receive X blood-type organs at steady state under the ABO-compatible FIFO deceased-donor allocation policy then $t^{\mathbf{c}}_X = t^{\mathbf{c}}_Y$.*

Moreover, if blood types in $\mathcal{S} \subseteq \mathcal{T}$ are pooled together then the waiting time of each $X \in \mathcal{S}$ is given by

$$t^{\mathbf{c}}_X \quad = t_{\mathcal{S}} \equiv F^{-1}\left(1 - \frac{\sum_{X\in\mathcal{S}} \delta_X}{\sum_{X\in\mathcal{S}} (\pi_X + \phi^d \delta_X)}\right) \tag{2}$$

Observe that $t^{\mathbf{i}}_X = t_{\{X\}}$ as defined in Equation 2 for all blood types X.

Using Lemmata 2 and 3 together with the FIFO feature of the deceased-donor queue policy and the partial order structure of the blood-type compatibility relationship, we can determine which types will be pooled together under the ABO-compatible allocation policy:

Theorem 2 (ABO-Compatible Deceased Donation) *At steady state, suppose Y blood type has the longest ABO-identical allocation waiting time and X blood type has the shortest ABO-identical allocation time among all blood types W with $W \triangleright Y$. Suppose further that $t^{\mathbf{i}}_X < t^{\mathbf{i}}_Y$. Then X and Y blood-type patients will be pooled together (possibly with other types) under ABO-compatible FIFO allocation. Moreover, we can treat X and Y together as a* composite blood type *$\{X, Y\}$ with deceased-donor inflow $\delta_{\{X,Y\}} = \delta_X + \delta_Y$ and patient inflow $\pi_{\{X,Y\}} = \pi_X + \pi_Y$ such that $W \triangleright \{X, Y\}$ for all blood types W with $W \triangleright Y$, and $\{X, Y\} \triangleright Z$ for all blood types Z with $X \triangleright Z$.*

Theorem 2 can be iteratively used to determine the ABO-compatible deceased-donor waiting times for all blood types with the simple mathematical fact that for all $a, b, c, d > 0$ whenever $\frac{a}{b} < \frac{c}{d}$ then $\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$:

Pooling procedure for blood types for ABO-compatible deceased donation:

1. Find all waiting times t_X as defined in Equation 2 for all $X \in \mathcal{T}$.[17]
2. Suppose X has the longest t_W among all $W \in \mathcal{T}$. Let Y have the shortest t_W among all $W \in \mathcal{T}$ with $W \triangleright X$.
   (a) If $Y = X$ then X is not pooled with any other blood type and $t^{\mathbf{c}}_X = t_X$. Repeat Step 1 for the remaining blood types $\mathcal{T} \setminus \{X\}$.
   (b) If $Y \neq X$ then X is pooled with Y (possibly together with other types). Replace the two blood types X and Y with the composite blood type $\mathcal{S} = X \cup Y$ and update the blood-type compatibility partial order $\triangleright$ as defined in Theorem 2. Repeat Step 1 for the new blood type set $\mathcal{T} := \big(\mathcal{T} \setminus \{X, Y\}\big) \cup \{\mathcal{S}\}$.

4 Live Donation

Organs such as kidney and liver have live donation possibilities. Especially live kidney donation is very common and $PP\%$ of all donation has been by live donors in 2011.[18]

We will refer to a live donor as a **paired donor**. We will assume that each patient has at most one paired donor. We assume λ fraction of incoming patients have a paired donor (such as a spouse). We also assume that the blood types of the patient and donor are independent and uncorrelated.[19] The patient and his paired donor are represented as a **pair**. The blood types of the pair, $X - Y \in \mathcal{T} \times \mathcal{T}$, X being the patient's and Y being the donor's blood type, determines the type of the pair.

If the paired-donor of a patient is both blood- and tissue-type compatible then we refer to the pair as a **compatible** pair, and otherwise as an **incompatible** pair. Recall that by assumption

[17] With a slight abuse of notation, even if X is not a set, it also refers to the set $\{X\}$.

[18] Each human has two kidneys and can have a perfectly healthy life with a single kidney. Also the risk associated with live donation surgery is very small. There is $PP\%$ chance that something will go wrong for the donor, and $PP\%$ chance that the donor will die complications due to surgery. On the other hand, live-donor liver donation is done through donation of part of a liver, and it is much riskier (there is $PP\%$ chance that he donor will die due to complications associated with donation). The ratio of live donation is much smaller, $PP\%$ for liver.

[19] In reality, if the paired donor is a blood relative of the patient, the blood types of the patient and donor are correlated through degree of relation and genetic laws. Hence, potentially figuring out the exact correlation can be complicated. For our purposes, we simply assume the blood types of the patient and his paired donor are uncorrelated to make our arguments.

there is a θ probability chance that a blood-type compatible donor being tissue-type incompatible with a patient. Given a patient with living donor, let p_Y be the probability of the donor being Y blood type.

Transplanted organs from live donors can also fail as in the case of transplants from deceased donors. As in the case of deceased donors, we assume that reentering patients have the same survival function $1-F$ as new patients. However, it is well known that the live-donor grafts survive longer than deceased-donor grafts. We assume that $\phi^l \leq \phi^d$ is the fraction of live-donation recipients reentering the deceased-donor queue per each live-donor organ transplant at steady state. We further assume that the reentrants (who received a organ previously from either a deceased donor or a live donor) do not have a paired live donor upon reentry.

Consistent with the donation rates throughout the world, we assume in the rest of the paper the following:

Assumption 1 *There is a shortage of deceased donor organs even in the absence of patients with living donors, i.e.,* $(1-\lambda)\pi_X + \phi^d\delta_X \geq \delta_X$ *for all* $X \in \mathcal{T}$.

We can calculate the inflow measures of different compatible and incompatible pair types:

- An O blood-type patient needs an O blood-type donor. Thus, $(1-\theta)p_O\lambda\pi_O$ is the inflow measure of O blood-type patients with a compatible live donor. On the other hand, $\theta p_O\lambda\pi_O$ is the measure of incompatible $O-O$ pairs, $\pi_O\lambda p_Y$ is the measure of $O-Y$ pairs with $Y \in \{A, B, AB\}$, who are all incompatible.

- An $X \in \{A, B\}$ blood-type patient can get an organ from O or X blood-type donor. Thus, given $Y \in \{X, O\}$, $(1-\theta)p_Y\lambda\pi_X$ is the inflow measure of X blood-type patients with a compatible Y blood-type live donor; on the other hand, $\theta p_Y\lambda\pi_X$ is the measure of incompatible $X-Y$ pairs. We have $p_Y\lambda\pi_X$ as the inflow measure of $X-Y$ pairs with $Y \in \{A, B, AB\} \setminus \{X\}$. The latter are incompatible pairs.

- An AB blood-type patient can get an organ from all blood-type donors. Thus, $(1-\theta)p_Y\lambda\pi_{AB}$ is the inflow measure of compatible $AB-Y$ pairs, and $\theta p_Y\lambda\pi_{AB}$ is the inflow measure of incompatible $AB-Y$ pairs for all $Y \in \mathcal{T} = \{O, A, B, AB\}$.

For a patient with a paired donor and of blood type X, let p_X^l denote the probability that the donor is compatible with the patient. Thus, $p_X^l\lambda\pi_X$ is the inflow measure of X blood-type patients with compatible live donors. These patients receive organs from their paired donors upon entry and they do not wait in the deceased-donor queue. We make the following observation regarding the allocation and reentry rates of deceased- and live-donor organ recipients:

Observation 2 *At steady state,*

- *a $p_X^l \lambda \pi_X$ measure of X blood-type patients receive live donation per unit time without waiting in the deceased-donor queue, and hence, a $\phi^l p_X^l \lambda \pi_X$ measure of previous live donation recipients reenter the deceased-donor queue per unit time; and*

- *a δ_X measure of X blood-type patients receive deceased-donor organs per unit time under ABO-identical FIFO allocation policy, and hence, a $\phi^d \delta_X \pi_X$ measure of previous deceased donation recipients reenter the queue per unit time.*

Hence, when the the total inflow measure of patients entering or reentering to the X blood-type deceased-donor queue under the ABO-identical FIFO allocation policy is given as

$$\pi_X^{\mathbf{L},\mathbf{i}} = \underbrace{\pi_X}_{\text{new patients}} + \underbrace{\phi^d \delta_X}_{\text{reentry / deceased}} + \underbrace{\phi^l p_X^l \lambda \pi_X}_{\text{reentry / live}} - \underbrace{p_X^l \lambda \pi_X}_{\text{compatible pairs}} . \tag{3}$$

Above, "reentry / deceased" and "reentry / live" refer to the reentering previous deceased- and live-donor organ recipients, respectively. Equation 3 and Observation 2 imply that the ABO-identical allocation waiting time conditional on survival in the X blood-type deceased-donor queue is given by

$$t_X^{\mathbf{L},\mathbf{i}} = F^{-1}\left(1 - \frac{\delta_X}{\pi_X^{\mathbf{L},\mathbf{i}}}\right) \tag{4}$$

conditional on survival (see Figure 5).

The analysis in Theorem 2 can be used to find which blood types are pooled together under ABO-compatible deceased-donor allocation policy by using $\pi_X + \phi^l p_X^l \lambda \pi_X - p_X^l \lambda \pi_X$ instead of π_X for all X. This analysis also helps us to pin down the waiting times for a deceased-donor organ under ABO-compatible allocation. In particular, we will make use of the following lemma:

Lemma 4 *Fix a blood type X. ABO-compatible deceased-donor allocation waiting time for every blood type Y, $t_Y^{\mathbf{c}}$, is continuous and weakly increasing in π_X and continuous and weakly decreasing in δ_X; moreover, for $t_X^{\mathbf{c}}$ is strictly increasing in π_X and strictly decreasing in δ_X.*

We are ready to make a more detailed analysis of how different blood types are affected by the availability of live donation. Due to the partial-order structure of blood-type compatibility across blood types, not all blood types will be affected equally when live donation is possible. For example, O blood type patients are at a disadvantage with respect to other types in finding a compatible paired donor. In general A blood type is more prominent in the population than B. Therefore, at random A blood-type patients will have a higher chance of finding a compatible paired donor then B types, given that they can all receive from O blood-type donors as well as their own types. Finally, AB blood-type patients have the highest chance of finding a compatible paired donor.

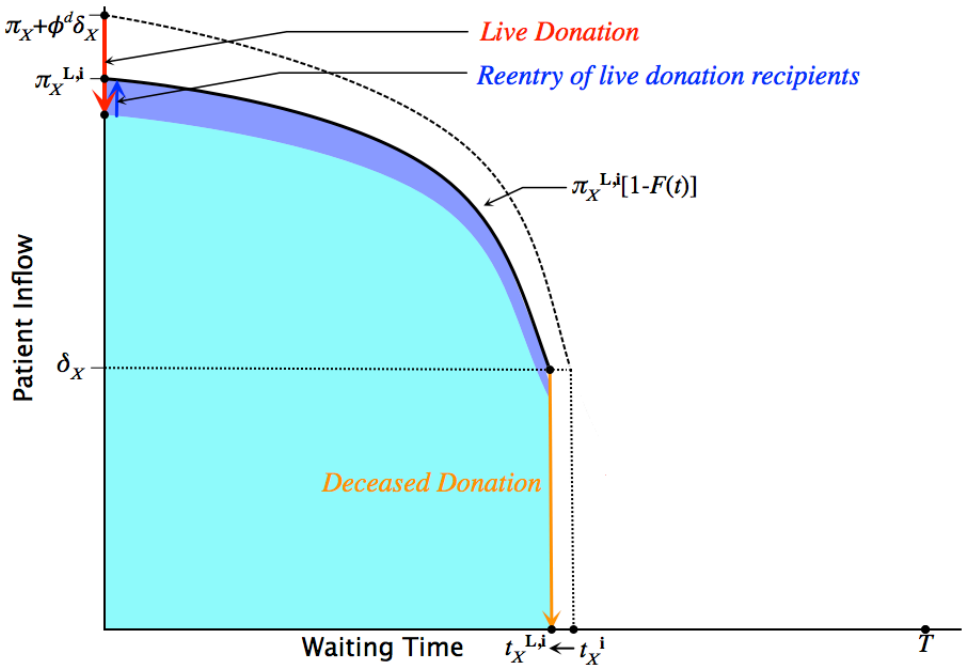

Figure 5: **X blood-type deceased-donor queue under the ABO-identical decease donation policy and live donation at steady state**: Inflow π_X of patients increases by the inflow of reentering previous deceased- and live-donor recipients, $\phi^d\delta_X$ and $\phi^l p_X^l \lambda\pi_X$, respectively; and decreases by $p_X^l\lambda\pi_X$, the outflow of patients who immediately receive an organ from their compatible paired donors. As a result waiting time for a deceased donor decreases from $t_X^{\mathbf{i}} = F^{-1}\left(1-\frac{\delta_X}{\pi_X+\phi^d\delta_X}\right)$ to $t_X^{\mathbf{L,i}} = F^{-1}\left(1-\frac{\delta_X}{\pi_X+\phi^d\delta_X+\phi^l p_X^l\lambda\pi_X-p_X^l\lambda\pi_X}\right) = F^{-1}\left(1-\frac{\delta_X}{\pi_X^{\mathbf{L,i}}}\right)$ conditional on survival for the ABO-identical deceased-donor allocation policy. Darker shaded region refers to the mass of reentering previous live-donor organ recipients in the queue.

However, depending on the exact shape of the survival function $1-F$ and the deceased donor to new patient inflow rates across blood types δ_X/π_X, O blood type does not necessarily experience the lowest decrease in the waiting time and AB blood type does not necessarily experience the greatest improvement.

On the other hand, for the benchmark case, where δ_X/π_X, the deceased donor to new patient inflow ratio, is the same for each blood type, we can make unambiguous predictions regarding the magnitude of the effects of live donation on deceased donor waiting times under both ABO-identical and ABO-compatible allocation policies. The following theorem states this result. [20]

Theorem 3 (Direct Live Donation and Deceased Donation) *Live donation will unambiguously decrease the steady state ABO-identical and ABO-compatible deceased-donor allocation waiting times for all blood types.*

Consider the benchmark case that the ratio δ_X/π_X is constant across blood types X. Then under live donation, no blood types pool under ABO-compatible deceased-donor allocation; and furthermore regardless of the deceased-donor allocation protocol, ABO-identical or ABO-compatible,

- *O blood-type patients have the shortest waiting time;*
- *AB blood-type patients have the longest waiting time, and*
- *provided that $p_A > p_B$, B blood-type patients have a longer waiting time than A blood-type patients.*

5 Live-Donor Exchange

In this section we analyze the effect of having a live-donor exchange program on waiting times of different patient groups for a donor organ. In practice, the donor of a compatible pair usually directly donates to the patient and the patient leaves the pool before he enters the deceased-donor queue. For the incompatible pairs, we assume that a live-donor kidney exchange program operates in parallel time to the deceased-donor queue. Incompatible pairs are listed in the exchange program. While waiting for a deceased-donor organ in the queue, they also wait for a paired-exchange to be conducted with another incompatible pair in the exchange program.

Formally, a two-way **exchange** matches two pairs where the patient of the first pair is compatible with the second pair and the patient of the second pair is compatible with the donor of the first pair.

[20] Although these conclusions seem to have been reached with the help of our assumption that blood types of patients are uncorrelated with their paired donors, a version of this result will also hold true even if there is positive correlation in a pair's blood types; however the magnitude of the difference in eventual waiting times will not be as extreme.

We refer to such pairs as **mutually compatible** pairs.[21] We refer to the queue of the pairs in the exchange program as the **exchange pool**. An **exchange matching** is a set of exchanges between mutually compatible pairs such that each pair is matched in at most one exchange.

There can be different policies determining which types of mutually compatible pairs are matched with each other, as a pair type can be mutually compatible with several other types. We will assume that the donor exchange is conducted in an *optimal manner* by matching the most possible measure of pairs at each point in time. However while selecting among a particular pair from a given type $X-Y$, organ exchange is also operated in a FIFO basis.

Not all incompatible pairs have similar features in terms of their abundance. In reality, for example although an incompatible $A-O$ blood-type patient-donor pair is harder to encounter than an $O-A$ pair, as $A-O$ pairs are incompatible only when there is tissue rejection between the A blood-type patient and O blood-type donor, while $O-A$ pairs are always incompatible.

Based on this observation, we make the following assumption: For a given patient - live donor pair type $X-Y$, we refer to type $Y-X$ as its **reciprocal** type.

Assumption 2 *We assume that any pair type $X-Y$ such that $X \neq Y$ and $X \triangleright Y$, its inflow rate to the exchange pool is not less than its reciprocal type $Y-X$'s inflow rate to the exchange pool, i.e. $\theta p_X \pi_Y \leq p_Y \pi_X$.*[22]

Another assumption we make is about the prevalence of $A-B$ and $B-A$ type pairs. This assumption is made out of notational convenience, and symmetric version of the results would hold if we did not make this assumption, without loss of generality.

Assumption 3 *We assume that $p_A \pi_B \leq p_B \pi_A$, i.e. $A-B$ type pairs do not inflow any slower than $B-A$ type pairs to the exchange pool.*[23]

To give an idea how easily this assumption is satisfied, recall that for kidneys, around $\theta = 0.1$ and for livers, $\theta = 0$. For all organs that can be used in exchange in real life, this inequality holds with a large slack for all populations.

[21] We can also think of exchanges that can match more than two pairs, such as three-way, four-way etc. For simplicity we will focus on two-way exchanges in our analysis, however, our results can easily be extended to cover three-way and four-way exchanges as in Roth, Sönmez, and Ünver (2007). Any sizes of exchanges greater than four will not change the results as reported in this paper.

[22] A simple requirement that would make the second condition of the assumption hold is that donor and patient inflow rates across blood types have a similar ratio i.e., $\pi_X/\pi_Y \approx p_X/p_Y$ for all blood types X, Y. This would be ensured if live donation and getting sick rates are not too different for different blood types.

[23] As a separate note, for kidneys Terasaki, Gjertson, and Cecka (1998) report that $A-B$ pairs make up of 5% of all pairs while $B-A$ pairs make up of 3%. However, our assumption has nothing to do with this observation and all our results would symmetrically hold if $B-A$'s were more than $A-B$'s.

Through this assumption, all incompatible $X-Y$ pairs with $Y \triangleright X$ can be matched immediately with $Y-X$ pairs, as $Y-X$ pairs will always be more in mass than $X-Y$ pairs in the exchange pool. Observe that the probability of mutual compatibility between an $X-Y$ pair and and an $Y-X$ pair is $(1-\theta)^2 > 0$. We state a slightly different version of Lemma 1 proving this case:

Lemma 5 (Live-Donor Exchange Matching Protocol) *Consider an ω measure of pairs denoted by the set M and a $\sigma \leq \omega$ measure of pairs denoted by set N (possibly intersecting with M), that are mutually blood-type compatible with the pairs in M. Suppose these sets are formed randomly using the governing population distributions. Then, almost surely there exists an exchange matching that matches all pairs in N with pairs in M.*

Proof. It follows from Erdös and Rényi (1960) random graph convergence theorem. ■

Using the terminology in Ünver (2010), we classify the pairs into several categories, based on their desirability in exchange.

Overdemanded pair types are the ones with a blood type donor which can donate to her patient's blood type yet they are not of the same blood type. There are $A-O, B-O, AB-A, AB-B, AB-O$ pair types. **Underdemanded pair types** are those with a blood type donor that cannot feasibly donate to her patient's blood type, excluding types A-B and B-A. That is, underdemanded types are reciprocals of overdemanded types, and they include $O-A, B-O, A-AB, B-AB, O-AB$. **Reciprocally demanded pair types** are $A-B$ and $B-A$, as they can be matched with each other in a donor exchange, when tissue incompatibility does not exist. Finally **Self-demanded pair types** are the ones with the same blood-type donor and patient: $O-O, A-A, B-B, AB-AB$.

The names associated with these classes will be more meaningful after our analysis. The following lemma shows the role of overdemanded types in exchange (similar results were also reported in Roth, Sönmez, and Ünver, 2007; Ünver, 2010):

Lemma 6 (Live-Donor Exchange Blood-Type Feasibility) *An underdemanded type pair can only be matched with an overdemanded type pair in an exchange, and overdemanded types can be used to match other overdemanded, underdemanded, reciprocally demanded, or self-demanded type pairs. Additionally, reciprocally demanded type pairs can only be used to match the other reciprocally demanded type pairs and self-demanded type pairs can only be used to match their own type pairs. In particular:*

- *An underdemanded $O-A$ ($O-B$, respectively) type pair can only be matched in an exchange with a pair from overdemanded types $A-O$ ($B-O$, respectively) or $AB-O$. An underdemanded $A-AB$ ($B-AB$, respectively) type pair can only be matched in an exchange with a pair from overdemanded types $AB-A$ ($AB-B$, respectively) or $AB-O$. An underdemanded $O-AB$ type pair can only be matched in an exchange with an overdemanded $AB-O$ type pair.*

- *A reciprocally demanded $A-B$ ($B-A$, respectively) type pair can only be matched in an exchange with a pair from the other reciprocally demanded $B-A$ ($A-B$, respectively) or from overdemanded types $AB-A$ ($AB-B$, respectively) or $AB-O$.*

- *A self-demanded $X-X$ type pair can be matched in an exchange with a pair from the same type. Additionally, an $O-O$ type pair can only be matched with a pair from overdemanded types $A-O, B-O$, or $AB-O$; an $A-A$ ($B-B$, respectively) type pair can only be matched with a pair from overdemanded types $AB-A$ ($AB-B$, respectively) and $AB-O$; and an $AB-AB$ type pair can only be matched with a pair from overdemanded types $AB-A, AB-B$, or $AB-O$.*

5.1 ABO-Identical Exchange & Deceased Donation

Next we model how the live-donor exchange pool and deceased-donor queue would evolve under live-donor donation and optimal exchange technologies. In this subsection we focus on ABO-identical decease donation. Live-donor exchange is mostly prevalent for kidneys and kidney deceased-donation is mostly ABO-identical. Recall that only incompatible patient - live-donor pairs participate in exchange. It turns out that we can make two-way exchange in an ABO-identical manner as well by matching $X-Y$ type pairs with $Y-X$ type pairs as they become available, and this would be optimal in the sense that the measure of transplants at each moment will be maximized. We will discuss ABO-compatible policies for deceased donation and exchange in the next subsection, which require substantially different tools in analysis.[24]

Theorem 4 (ABO-identical exchange is optimal) *A policy that dictates matching the longest-waiting pairs of a type with their longest-waiting reciprocal type pairs at each point in time is optimal in the sense that it matches the maximum measure of pairs possible at each point in time. Moreover, this policy maximizes the mass of pairs that can be matched within any close time interval, and in particular, matches a larger mass of pairs than waiting the pairs to arrive and running the exchange once at the end of the time interval.*

Thus, it is straightforward to calculate the total flow of patients that receive live-donor organs through exchange for each blood type under the optimal policy proposed in Theorem 4. We denote

[24]With the availability of live donor exchange, we can separate patients into different *groups* based on the existence or non-existence of a live donor and if she exists, compatibility or incompatibility of the live donor. We can measure the effect of each policy on each of these groups. There are 29 patient groups based on this criterion. As compatible and incompatible pairs of blood-type compatible pair types receive organs at time 0 under the optimal live-donor exchange policy, we do not distinguish among them. Therefore, we denote each patient group by the pair type $X-Y$ if it includes pairs and only by the blood type X of the patient, if it refers to patients without live donors (as subscript of the relevant welfare variables such as time of waiting for a transplant.

the measure of pairs matched X blood type patients at each point in time under the optimal policy that are patients of

- overdemanded type pairs as e_X^{od},
- self-demanded type pairs as e_X^{sd},
- reciprocally demanded type pairs as e_X^{rd}, and
- underdemanded type pairs as e_X^{ud}.

We denote the overall measure of X blood-type patients matched through the optimal exchange policy at each point in time as e_X:

$$\begin{aligned}
e_O &= \underbrace{\theta p_O \lambda \pi_O}_{e_O^{sd}:\, O-O \text{ pairs}} + \underbrace{\theta p_O \lambda (\pi_A + \pi_B + \pi_{AB})}_{e_O^{ud}:\, O-A,\, O-B,\, O-AB \text{ pairs}}, \\
e_A &= \underbrace{\theta p_A \lambda \pi_A}_{e_A^{sd}:\, A-A \text{ pairs}} + \underbrace{\theta p_O \lambda \pi_A}_{e_A^{od}:\, A-O \text{ pairs}} + \underbrace{p_A \lambda \pi_B}_{e_A^{rd}:\, A-B \text{ pairs}} + \underbrace{\theta p_A \lambda \pi_{AB}}_{e_A^{ud}:\, A-AB \text{ pairs}}, \\
e_B &= \underbrace{\theta p_B \lambda \pi_B}_{e_B^{sd}:\, B-B \text{ pairs}} + \underbrace{\theta p_O \lambda \pi_B}_{e_B^{od}:\, B-O \text{ pairs}} + \underbrace{p_A \lambda \pi_B}_{e_B^{rd}:\, B-A \text{ pairs}} + \underbrace{\theta p_B \lambda \pi_{AB}}_{e_A^{ud}:\, B-AB \text{ pairs}}, \text{ and} \qquad (5) \\
e_{AB} &= \underbrace{\theta p_{AB} \lambda \pi_{AB}}_{e_{AB}^{sd}:\, AB-AB \text{ pairs}} + \underbrace{\theta (p_O + p_A + p_B) \lambda \pi_{AB}}_{e_{AB}^{od}:\, AB-O,\, AB-A,\, AB-B \text{ pairs}}.
\end{aligned}$$

We use these inflow measures to analyze how the availability of live-donor exchange affects the waiting time in the deceased-donor queue. As more patients receive live donation in this new regime with respect to the case where only direct live donation was feasible, the waiting times of patients improve across all blood types. Some of these pairs will be matched immediately when they enter the pool: all overdemanded and self-demanded type pairs, and the scarcer reciprocal type $B-A$ pairs; and some are matched after waiting in the pool: underdemanded type pairs and the more abundant reciprocal type $A-B$ pairs. The latter pes of pairs are not as fortunate as overdemanded and self-demanded type pairs. Not all of them will be matched, either through exchange or deceased-donor donation. They will wait in the exchange pool and the deceased-donor queue simultaneously. In this case, they will either

- be "pooled" with patients of the same blood type in deceased-donor queue, and some of them will receive deceased-donor organs while the remaining ones (that are alive) will receive organs through exchange at the same time as their cohort of patients without live donors receive deceased-donor organs; or
- wait shorter than their cohort of patients without any live donors and receive live-donor organs through exchange before their cohort in the deceased-donor queue receive deceased donation.

First, we focus on the ABO-identical deceased-donor allocation policy. In order to determine the waiting times. For each blood type X, let

$$\pi^e_{X-Y} = \begin{cases} \theta p_Y \lambda \pi_X & \text{if } Y \triangleright X \\ p_Y \lambda \pi_X & \text{otherwise} \end{cases} \tag{6}$$

refer to the inflow measure of new $X-Y$ pairs to the exchange pool and

$$\pi^d_X = \underbrace{(1-\lambda)\pi_X}_{\text{new w/o live donors}} + \underbrace{\phi^d \delta_X}_{\text{reentry / deceased}} + \underbrace{\phi^l p^l_X \lambda \pi_X}_{\text{reentry / live}} + \underbrace{\phi^l e_X}_{\text{reentry / exchange}} \tag{7}$$

be the inflow measure of reentering and new X blood-type patients without live donors. We calculate the following ratios for each blood type X:

1. The deceased-donor inflow to the inflow of reentering and new patients without live donors

$$r^d_X = \frac{\delta_X}{\pi^d_X} = \frac{\delta_X}{(1-\lambda)\pi_X + \phi^d \delta_X + \phi^l p^l_X \lambda \pi_X + \phi^l e_X}.$$

2. For each underdemanded type $X-Y$ (i.e., $Y \neq X$ and $Y \triangleright X$), inflow of incompatible $Y-X$ to inflow of $X-Y$ pairs:

$$r_{X-Y} = \frac{\pi^e_{Y-X}}{\pi^e_{X-Y}} = \frac{\theta p_X \lambda \pi_Y}{p_Y \lambda \pi_X}.$$

3. If there is reciprocal $X-Y$ type then inflow of $Y-X$ to $X-Y$ ratio,

$$r_{X-Y} = \frac{\pi^e_{Y-X}}{\pi^e_{X-Y}} = \frac{p_X \lambda \pi_Y}{p_Y \lambda \pi_X}.$$

The ratio $r^d_X = \frac{\delta_X}{\pi^d_X}$ is relevant if we wanted to allocate all X blood type deceased donors to only X blood-type patients without live donors. For an pair type $X-Y$ with $Y \not\triangleright X$, i.e., underdemanded or reciprocally demanded, the ratio $r_{X-Y} = \frac{\pi^e_{Y-X}}{\pi^e_{X-Y}}$ is relevant if we did not want $X-Y$ pairs to receive any deceased donors but only live donors through exchange with their reciprocal type $Y-X$ pairs under the optimal exchange policy. In these cases, conditional on survival, the waiting time of X blood-type patients without live donors would be given as $t^d_X = F^{-1}\big(1 - \frac{\delta_X}{\pi^d_X}\big)$, and waiting time of $X-Y$ (when $X-Y$ is an underdemanded type or $X-Y = A-B$) would be given as $t_{X-Y} = F^{-1}\big(1 - \frac{\pi^e_{Y-X}}{\pi^e_{X-Y}}\big)$.[25]

However, underdemanded or reciprocally demanded $X-Y$ type pairs have another option besides waiting for their reciprocal type pairs to arrive. if deceased donors arrive earlier, they can receive an organ from a deceased donor. By assumption, we assume that patients choose whichever organ,

[25]The waiting time for $B-A$ type pairs is 0 as they are on the shorter side of the market when compared to $A-B$ type pairs by assumption.

deceased-donor or live-donor, becomes available first. [26] Hence, the patient of an $X-Y$ type pair will never wait for a $Y-X$ pair for exchange if a deceased organ comes first, i.e. if $t_{X-Y} < t_X^d$. As time is decreasing in r ratios, all we need to do is to compare these ratios in an iterative manner to decide whether any underdemanded type or $A-B$ type pairs will receive deceased-donor organs:

Pooling procedure for patient - live donor pairs and patients without live donors under ABO-identical deceased-donor allocation policy:

1. Let $X-Y_1, ..., X-Y_k$ be the ordered list of underdemanded or reciprocally demanded types with X blood-type patients ascending in r_{X-Y} ratio. Define for each $\ell = 0, ..., k$:

$$r^d_{X,X-Y_1,...,X-Y_\ell} = \frac{\delta_X + \pi^e_{Y_1-X} + ... + \pi^e_{Y_\ell-X}}{\pi^d_X + \pi^e_{X-Y_1} + ... + \pi^e_{X-Y_\ell}}. \quad (8)$$

2. For $\ell \in \{0, ..., k-1\}$, suppose pair types $X-Y_1, ..., X-Y_\ell$ have already been deemed to be receiving both deceased donors and live donors through exchange.

   - If $r_{X-Y_{\ell+1}} < r^d_{X,X-Y_1,...,X-Y_\ell}$ then $X-Y_{\ell+1}$ pairs receive both live donors through exchange with $Y_{\ell+1}-X$ pairs and deceased donors with the rest of the X blood-type patients without live donors and $X-Y_1, ..., X-Y_\ell$ pairs. We continue with Step 2 with $\ell := \ell+1$.
   - If $r_{X-Y_{\ell+1}} \geq r^d_{X,X-Y_1,...,X-Y_\ell}$ then all types $X-Y_{\ell+1}, ..., X-Y_k$ only receive donors through exchange, but no deceased donors. We terminate the procedure.[27]

Based on this procedure, we state the following theorem:

Theorem 5 (Direct Live Donation, ABO-Identical Exchange and Deceased Donation) *Consider the ABO-identical deceased-donor allocation and live-donor exchange policies. Consider a blood type X. Conditional on survival, the waiting time for a donor and the measure receiving donation are given as follows for different X blood-type patient groups:*

[26] This assumption can be rationalized by the risk associated with dying while waiting for an organ and high risk aversion. To model this choice explicitly under a wider class of preferences, we can introduce additional structure regarding the cardinal preferences of the patients and the shape of the survival distribution $1-F(t)$. The patients could be willing to wait more for a live donor than a deceased donor as the former kind of graft survives longer, while waiting for an organ is riskier and could result with death, and is usually inferior in life quality to living with a functioning graft. The patients will be willing to wait as long as the second disutility does not outweigh the first utility marginally. When $1-F(t)$ is concave (i.e., for $t < t'$ dying at time t' is more likely than at time t), an endogenous incentive driven waiting time gap can be explicitly derived: at steady state when patients can receive a deceased-donor organ t years after entry, each patient will be willing to wait at most $\tau(t)$ years additionally for a live-donor organ. All our calculations can be modified to include this time gap function without much change.

[27] When some $X-Y$ pairs receive deceased donation and they reenter the pool, whether the patient of such a pair reenters as part of a pair or he reenters without a live donor does not have any impact on waiting times. As $X-Y$ pairs will be pooled with X blood-type patients without live donors, what matters is the total inflow of new and reentering $X-Y$ pairs and X blood-type patients without live donors, which is the same under either assumption.

- *X blood-type patients who have compatible live donors immediately receive their live donor's organ upon entry.*

- *X blood-type patients who are part of incompatible overdemanded, self-demanded, and if $X = B$ then, $B - A$ type pairs, immediately receive a live donor organ through exchange upon entry.*

- *Suppose patients of underdemanded and reciprocally demanded types $X - Y_1, ..., X - Y_\ell$ receive both live donation through exchange and deceased donation while patients of underdemanded and reciprocally demanded types $X - Y_{\ell+1}, ..., X - Y_k$ only receive donation through live-donor exchange. Then:*

  - *Conditional on survival, X blood-type patients without live donors and patients of $X - Y_1, ..., X - Y_\ell$ type pairs wait for a donor for*

$$t_X^{\mathbf{E},\mathbf{i}} = F^{-1}\left(1 - \frac{\delta_X + \pi^e_{Y_1-X} + ... + \pi^e_{Y_\ell-X}}{\pi^d_X + \pi^e_{X-Y_1} + ... + \pi^e_{X-Y_\ell}}\right). \tag{9}$$

  - *Conditional on survival, for all $m \in \{\ell + 1, ..., k\}$, patients of $X - Y_m$type pairs wait for live-donor exchange for*

$$t^{\mathbf{E},\mathbf{i}}_{X-Y_m} = F^{-1}\left(1 - \frac{\pi^e_{Y_m-X}}{\pi^e_{X-Y_m}}\right). \tag{10}$$

Proof. It follows from the procedure discussed before the statement of the theorem. ■

5.2 ABO-Compatible Exchange & Deceased Donation

For some organs such as livers, a deceased-donor queue patient can get precedence in receiving any ABO-compatible deceased-donor organ. If an egalitarian concern is in place, a similar practice can be adopted for exchange: $AB - O$ type pairs can be used to match $A - AB$ $B - AB$ or $O - AB$ pairs, not just $O - AB$ pairs and if two-way exchange is the only available exchange policy, saving any of them would be efficient in Pareto sense. However, a FIFO policy can also be adopted and an $AB - O$ type pair can be used in exchange with the longest waiting of these two types. However these overdemanded types can also receive organs from the deceased donor queue and they will determine which source to go, either exchange or deceased donor, according to their waiting time. Sorting out what patient group gets from what source leads to a seemingly complex graph-theory problem. However, thanks to techniques from combinatorial optimization theory, we can solve this cumbersome problem quite easily.[28]

[28]The same technique can be adopted to determine which blood types will be pooled when exchange is not possible.

Consider the following two-sided graph, with sides labeled as $\mathbf{O}$ and $\mathbf{U}$. Side $\mathbf{O}$ consists of 4 nodes O, A, B, AB representing the deceased donor blood types, 5 nodes representing overdemanded pair types $A-O, B-O, AB-O, AB-A, AB-B$ and type $B-A$ which is on the short side among the two reciprocally demanded types $A-B$ and $B-A$:

$$\mathbf{O} \ = \{O, A, B, AB, A-O, B-O, AB-O, AB-A, AB-B, B-A\}. \tag{11}$$

The other side consists of also 10 nodes, 4 representing the blood types of patients without live donors, 5 for the underdemanded pair types, and 1 for the $A-B$ pair type:

$$\mathbf{U} = \{O, A, B, AB, O-A, O-B, O-AB, A-AB, B-AB, A-B\}. \tag{12}$$

The nodes in both sides rate connected with a blood-type feasibility link and these links are represented by a matrix of 0's and 1's, $C = [c_{i,j}]_{i\in\mathbf{O}, j\in\mathbf{U}}$. Two types $i \in \mathbf{O}$ and $j \in \mathbf{U}$ are linked, i.e. $c_{i,j} = 1$, when (1) if i is a blood type, the patient of from a j type (i.e., if $j = X$ is a blood type, the X blood-type patient of an $X-Y$ pair and if $i = X-Y$ is a pair type then an X blood-type patient without live donor) can receive an organ from an i blood-type deceased donor, and (2) if i and j are pair types, then i and j pairs are mutually blood-type compatible to be matched in a two-way exchange. The graph induced by C is depicted in Figure 6.

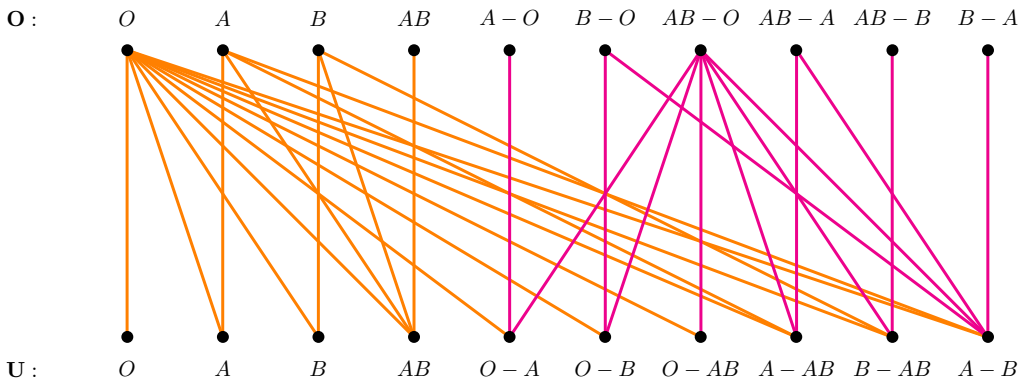

Figure 6: ABO-compatible exchange and deceased donation feasibility graph $(\mathbf{O}, \mathbf{U}, C)$. Lighter links correspond to deceased donation possibilities and darker links correspond to exchange possibilities.

Each node $h \in \mathbf{O} \cup \mathbf{U}$ is assigned a weight w_h such that w_h is the inflow measure of the type of patient without live donor / deceased donor / pair group h in question:

$$w_h \ = \ \begin{cases} \delta_h & \text{if } h \in \mathbf{O} \cap \mathcal{T} \\ \pi^d_h & \text{if } h \in \mathbf{U} \cap \mathcal{T} \\ \pi^e_h & \text{if } h \in (\mathbf{O} \cup \mathbf{U}) \cap \mathcal{T} \times \mathcal{T} \end{cases} \tag{13}$$

where π^e_h is defined in Equation 6 and π^d_h is defined in Equation 7.

Now, we determine the *least privileged* node subset of $\mathbf{U}$ as follows: For any $\mathbf{V} \subseteq \mathbf{U}$ and $\mathbf{P} \subseteq \mathbf{O}$ define

$$\mathbf{C_V}(\mathbf{P}) = \{i \in \mathbf{P} \mid c_{i,j} = 1 \text{ for some } j \in \mathbf{V}\}, \tag{14}$$

and[29]

$$r^d_{\mathbf{V}}(\mathbf{P}) = \frac{\sum_{i \in \mathbf{C_V(P)}} w_i}{\sum_{j \in \mathbf{V}} w_j}. \tag{15}$$

Here, $\mathbf{C_V}(\mathbf{P})$ is the set of deceased donor blood-types and overdemanded pair types in $\mathbf{P} \subseteq \mathbf{U}$ that can feasibly matched through deceased donation or exchange with some type in set $\mathbf{V} \subseteq \mathbf{U}$; and $r^d_{\mathbf{V}}(\mathbf{O})$ is the supply-to-demand ratio for $\mathbf{V}$ i.e., the ratio of inflow measures of deceased donor and pairs on short supply that can be matched with patients without live donors and pairs on long supply within $\mathbf{V}$. This ratio is the generalization of the r^d ratio defined in Equation 8. Now we can find the subset of $\mathbf{U}$ which minimizes r^d:[30]

$$\mathbf{V}_1 = \arg\min_{\mathbf{V} \subseteq \mathbf{U}} r^d_{\mathbf{V}}(\mathbf{O}); \quad \text{and} \tag{16}$$

$$\mathbf{P}_1 = \mathbf{C}_{\mathbf{V}_1}(\mathbf{O}). \tag{17}$$

Then we iteratively construct the partition $\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_k$ of $\mathbf{U}$ such that

$$\mathbf{V}_\ell = \arg\min_{\mathbf{V} \subseteq \mathbf{U} \setminus \cup_{m=1}^{\ell-1} \mathbf{V}_m} r^d_{\mathbf{V}}(\mathbf{O} \setminus \cup_{m=1}^{\ell-1} \mathbf{P}_m); \quad \text{and} \tag{18}$$

$$\mathbf{P}_\ell = \mathbf{C}_{\mathbf{V}_\ell}(\mathbf{O} \setminus \cup_{m=1}^{\ell-1} \mathbf{P}_m) \tag{19}$$

This means that the patient and pair groups belonging to $\mathbf{V}_\ell$ are the least fortunate, i.e. **bottleneck**, after serving the groups in $\mathbf{V}_1, ..., \mathbf{V}_{\ell-1}$ and we can assign all deceased donors and pairs belonging to $\mathbf{P}_\ell$ exclusively to patients and pairs of groups in $\mathbf{V}_\ell$ that are deemed blood-type feasible by matrix C. This result follows from the minimum cut - maximum flow theorem of Ford and Fulkerson (1956) in combinatorial optimization theory.[31] Even when we do that their waiting time in the exchange and deceased donor queues will not be lower than the other groups in $\mathbf{U} \setminus \cup_{m=1}^{\ell} \mathbf{V}_m$, as r^d ratio is lowest for $\mathbf{V}_\ell$ once $\mathbf{V}_1, ..., \mathbf{V}_{\ell-1}$ are fixed. Moreover, we can make this assignment feasibly, i.e. matching all deceased donors and incompatible pairs of groups in $\mathbf{V}_\ell$ with compatible patients and mutually compatible pairs of groups in $\mathbf{P}_\ell$ at the same waiting time:

$$t^{\mathbf{E,c}}_{\mathbf{V}_\ell} = F^{-1}(1 - r^d_{\mathbf{V}_\ell}(\mathbf{P}_\ell)). \tag{20}$$

[29]Similarly define $\mathbf{C_P}(\mathbf{V}) = \{j \in \mathbf{V} \mid c_{i,j} = 1 \text{ for some } i \in \mathbf{P}\}$. We will use these notations later in the proof of Proposition 1 as well.

[30]If there are more than one such sets then take largest of them, which is uniquely defined.

[31]For example, see Katta and Sethuraman (2006), Yılmaz (2009), and **?** for uses of the minimum cut - maximum flow theorem in the probabilistic matching framework.

Theorem 6 (Direct Live Donation, ABO-Compatible Exchange and Deceased Donation) *Under direct live donation and ABO-compatible exchange and deceased donation, the waiting time for X blood-type patients without live donors for each $X \in \mathcal{T}$, and for all underdemanded type pairs and pair type $A-B$, the waiting time is characterized by $t^{\mathbf{E},\mathbf{c}}_{\mathbf{V}_\ell}$ in Equation 20 where $\mathbf{V}_\ell$ and $\mathbf{P}_\ell$ are defined as in Equations 13-19.*

One can wonder how likely pooling of waiting times would occur between live-donor exchange and deceased donation. For kidneys, deceased-donor queue additions and removals for each blood type are summarized in Table 3 using US OPTN data in 2011 in Appendix B. Rates r^d_X are extrapolated from the data as $46.5-48.5\%$ for O, $45.7-46.9\%$ for A, $35.9-37.0\%$ for B and $37.5-38.2$ for AB. On the other hand, it is more difficult to access data on both recipient and donor blood types to determine the arrival rates of pair types. However, assuming roughly similar arrival rates for reciprocal types $X-Y$ and $Y-X$, we can conclude that $r_{X-Y} = \pi^e_{Y-X}/\pi^e_{X-Y} = \theta p_X \pi_Y / p_Y \pi_X = \theta = 0.11 << r^d_X$ for all $Y \not\trianglerighteq X$, i.e. underdemanded types $X-Y$. There is a lot of slack in this inequality and even if reciprocal types do arrive at different rates, as long as the arrival ratio is not too unbalanced in the favor of overdemanded types, we would still expect this inequality to hold. Hence, we expect that all underdemanded type pairs will receive both deceased donation and live donation through exchange even if full benefits from kidney exchanges are fulfilled in the US. On the other hand, it would be also good to estimate r_{A-B}. Terasaki, Gjertson, and Cecka (1998) report that $A-B$ pairs make up of 5% and $B-A$ pairs make up of 3% of all pairs. Hence, $r_{A-B} = \pi^e_{B-A}/\pi^e_{A-B} = 0.60$ is a good lower bound for this ratio. Observe that $r_{A-B} > r^d_A$. Hence, our expectation is that even using only two-way exchanges no $A-B$ pair would end up receiving deceased donation if all benefits from exchange are fulfilled in the US.[32] In the light of these estimations, Figure 7 depicts A blood-type deceased-donor queue at the steady state as an example of our predictions.

6 A New Proposal: Incentivizing Compatible Pairs to Participate in Exchange

One shortcoming of the current live-donor exchange practices is that they only utilize *incompatible* pairs. However, if *compatible* pairs can be incentivized to participate in exchange, then some sort of a supply balance will be satisfied between reciprocal type pairs in exchange, and hence the exchange will bring more benefit for all groups of patients. One sensible way of incentivizing compatible pairs to participate in exchange is to give their patients priority in the deceased donor queue regardless

[32] There is only one slight caveat in this prediction. Some patients may have a very high tissue-rejection rate, i.e. $\theta = 0.11$ may not be uniform for all patients. Such patients' only hope could be deceased donation, and in many instances they cannot even find compatible deceased donors. However, the number of such patients are relatively lower.

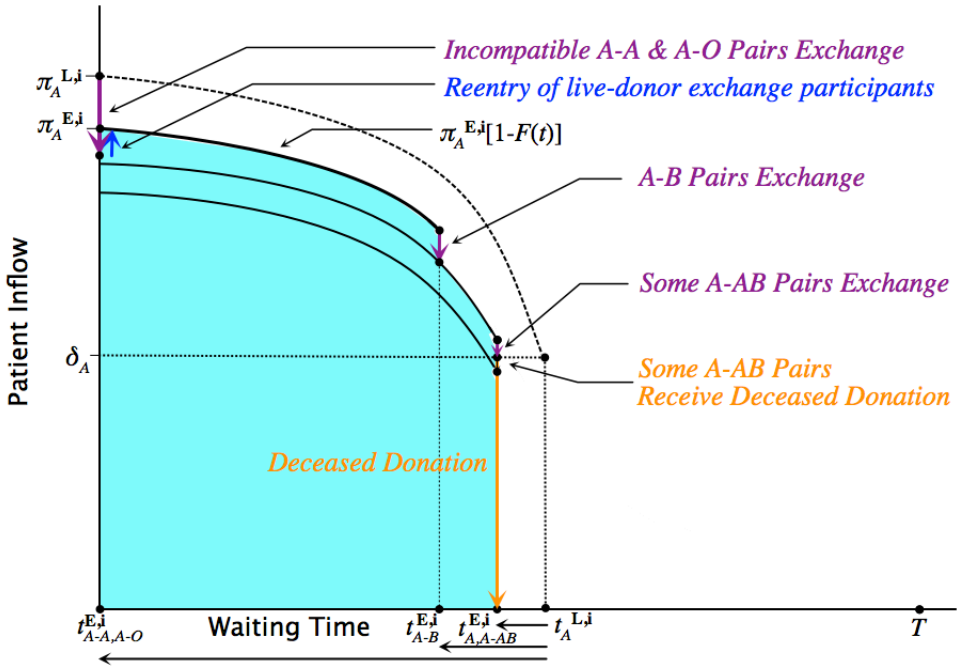

Figure 7: **A blood-type deceased-donor queue under the ABO-identical deceased donation policy, live donation, and exchange of incompatible live donors at steady state**: Inflow π_A of patients decreases by $p_A^l \lambda \pi_A = (1-\theta)(p_A+p_O)\lambda\pi_A$ as a result of immediate compatible direct live donation and a further $e_A^{od} + e_A^{sd} = \theta(p_A+p_O)\lambda\pi_A$ as a result of incompatible live-donor exchange for $A-O$ and $A-A$ types who do not wait in the deceased-donor queue. Assuming that $p_A\pi_B \geq p_A\pi_B$ and $A-B$ type pairs do not end up receiving deceased donation at time $t^{\mathbf{E},\mathbf{i}}_{A-B}$ (as found in Equation 10), a measure of $e_A^{rd} = p_A\pi_B$ of $A-B$ pairs are matched through exchange. Assuming that $A-AB$ type pairs both receive deceased donation and participate in exchange, at time $t_A^{\mathbf{E},\mathbf{i}}$ (as found in Equation 9 for $A-AB$ type pairs and A blood-type patients without live donors) a measure of $e_A^{ud} = \theta p_A \lambda \pi_{AB}$ of $A-AB$ pairs are matched through exchange. As a result waiting time in the deceased-donor queue decreases from $t_A^{\mathbf{L},\mathbf{i}}$ to $t_A^{\mathbf{E},\mathbf{i}}$, as well. The mass of reentering previous live-donor exchange participants are not shaded because of the complexity of the figure.

of their waiting time if their graft transplanted as a result of exchange fails. As noted earlier in the Introduction, live donors are already incentivized in a similar manner. if a live-donor's organ fails in the future she gets a priority in the decease-donor allocation. A similar practice of prioritizing not only the donor but also patient of a compatible pair may face little resistance in the medical community. In this section, using the tools we developed in the earlier sections, we analyze the welfare an equity effects of such a prioritization policy. Thus, when the transplanted graft of a patient who was part of a compatible pair and acquired this graft through exchange fails, we assume that the FIFO structure of deceased-allocation policy is altered, and such patients are placed to the front of the queue. In this case, we can analyze the welfare effects of this policy with respect to the alternative, regular exchange without compatible pairs.[33]

We will focus on ABO-identical FIFO deceased-donation policy in this section and the next one, as this is the primary policy adopted for kidney allocation, which has the most prominent exchange programs in the world.

Suppose an endogenous proportion ρ of all compatible pairs take up of this option and participate in exchange. We will maintain the following assumption in this and next sections.

Assumption 4 *For any underdemanded type $X-Y$ (i.e., $X \triangleright Y$ and $X \neq Y$), suppose $[\rho(1-\theta)+\theta]p_X\pi_Y \leq p_Y\pi_X$.*

This assumption ensures that the measure of arriving underdemanded pairs is greater than the measure of arriving reciprocal overdemanded pairs who are either incompatible or compatible and willing to participate in exchange for each underdemanded pair type $X-Y$. This is a simplification. If this is not the case, any excess of the compatible pairs will not participate in exchange but the patients will directly receive transplant from their own donors upon reentry. As a result compatible pairs never wait in the deceased-donor queue.

Moreover, although we assumed ρ is exogenously determined, we would expect, in equilibrium in the long run, participation percentage ρ is maximized to match the maximum possible number of incompatible pairs through exchange so that if a non-participating compatible pair were to switch to participate in exchange, it will not be able to help an additional incompatible pair to receive donation through exchange, hence a version of the above assumption will hold endogenously.

We assume that we use incompatible pairs in exchange as much as possible and if they are not feasible to be used anymore, then we use compatible pairs in exchange. We first show that this approach does not decrease the number of possible incompatible pairs matched in exchange:

Patients of compatible pairs are unambiguously weakly better off under this policy for any ρ. What about other groups of patients?

[33]We will abbreviate the new policy with superscript **I** and add to our variables, referring to **I**ncentivized exchange with compatible pairs, while regular exchange will continued to be abbreviated by **E**.

Hence an optimal incentivized exchange policy dictates matching all incompatible $X-Y$ pairs with its incompatible reciprocal-type pairs, and if there is excess incompatible $X-Y$ left using compatible $Y-X$ pairs (if they exist) to save them in exchange. We state this result formally through the following theorem:

Theorem 7 (ABO-identical optimal exchange with incentivized compatible pairs) *When some compatible pairs participate in exchange with the condition that their reentrant patients are prioritized in the deceased-donor queue, a policy that dictates matching incompatible self-demanded pairs of one type with each other, and the longest-waiting underdemanded and $B-A$ type pairs of with their reciprocal-type incompatible and willing compatible pairs at each point in time is optimal in the sense that it matches the maximum measure of pairs possible at each point in time.*

Moreover, this policy maximizes the mass of pairs that can be matched within any closed time interval, and in particular, matches a larger mass of pairs than waiting the pairs to arrive and running the exchange once at the end of the time interval.

The following theorem outlines the predictable differences of the outcomes under exchange with incentivized compatible pairs with respect to regular exchange.

Theorem 8 (Incentivizing compatible pairs to participate in exchange) *Under the ABO-identical deceased-donation and exchange policies with incentivized compatible pairs, with respect to regular exchange*

1. *weakly more patients are matched for each patient group at each point in time, that is, for each blood type X and Y, X blood-type patients without live donors, incompatible $X-Y$ pairs, and if they exist, compatible $X-Y$ pairs; underdemanded type pairs are matched at a strictly higher rate.*

2. *no compatible pairs of type $X-X$ participate in exchange (incompatible $X-X$ blood types save each other through exchange);*

3. *no O blood-type patients are prioritized upon reentry, however reentrant A, B and AB blood-type patients of previously compatible pairs that participated in exchange will get prioritized;*

4. *waiting time for underdemanded-type pairs strictly decreases; waiting time for O, A, and B blood-type deceased donors may increase or decrease; waiting time for AB blood-type deceased donors increases; waiting time for other patient groups is 0 and does not change.*

The proof of this theorem, especially of Statement 4, is also of independent interest, as it quantifies the sources of changes to the rates of being matched for different patient groups when we switch from regular exchange to exchange with incentivized compatible pairs. Additionally, Figure 8 provides an example for A blood-type patients illustrating these effects.

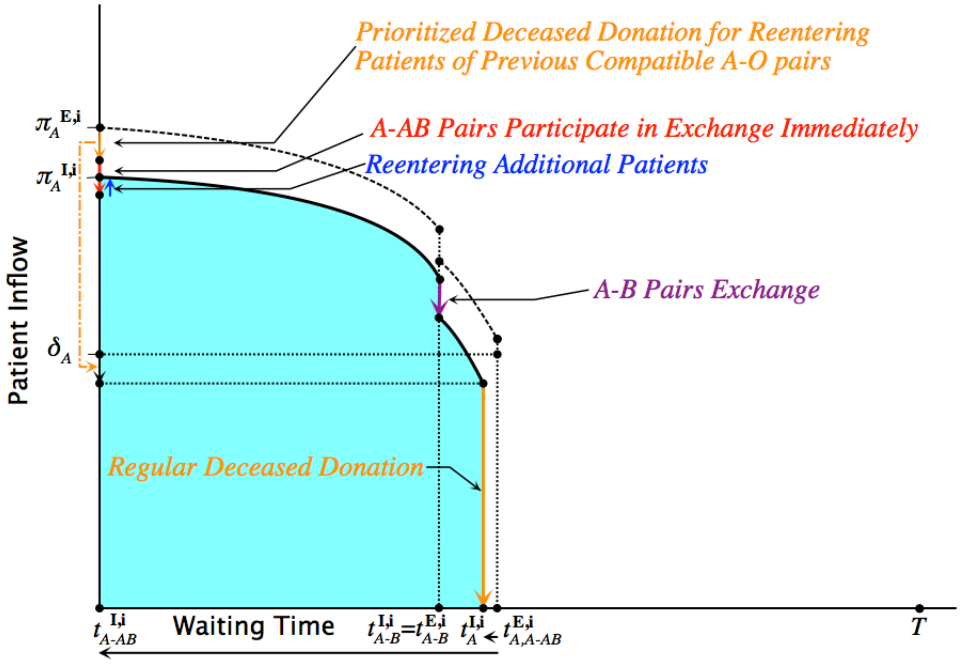

Figure 8: **A blood-type deceased-donor queue under the ABO-identical deceased donation policy, live donation, and incentivized exchange of incompatible live donors at steady state** when $\rho = 1$, i.e., all compatible pairs participate in exchange, and $p_X\lambda\pi_Y = p_Y\lambda\pi_X$ for all underdemanded $X-Y$ type pairs: Inflow π_A of patients decreases by $p_A^l\lambda\pi_A = (1-\theta)(p_A+p_O)\lambda\pi_A$ as a result of immediate compatible direct live donation and a further $e_A^{od} + e_A^{sd} = \theta(p_A+p_O)\lambda\pi_A$ as a result of incompatible live-donor exchange for $A-O$ and $A-A$ types who do not wait in the deceased-donor queue. Assuming that $p_A\pi_B \geq p_A\pi_B$ and $A-B$ type pairs do not end up receiving deceased donation at time $t_{A-B}^{\mathbf{E,i}}$ (as found in Equation 10), a measure of $e_A^{rd} = p_A\pi_B$ of $A-B$ pairs are matched through exchange. Assuming that $A-AB$ type pairs both receive deceased donation and participate in exchange, at time $t_A^{\mathbf{E,i}}$ (as found in Equation 9 for $A-AB$ type pairs and A blood-type patients without live donors) a measure of $e_A^{ud} = \theta p_A\lambda\pi_{AB}$ of $A-AB$ pairs are matched through exchange. As a result waiting time in the deceased-donor queue decreases from $t_A^{\mathbf{L,i}}$ to $t_A^{\mathbf{E,i}}$, as well. The mass of reentering previous live-donor exchange participants are not shaded because of the complexity of the figure.

7 Multiple Exchange Platforms and Exchange with Incentivized Compatible Pairs

Although in our model, we assumed that there is a unique central live-donor organ exchange authority, in reality many parallel platforms compete with each other in the case of kidney exchange in the US. Due to vagueness of original National Organ Transplant Act of 1984 regarding legality of exchanges, it had to be amended in 2007 and the US national kidney exchange program started under the provision of UNOS only in 2010. UNOS is also the federal contractor that oversees the deceased donor allocation in the US. On the other hand, regional kidney exchange programs had started in early 2000s. For example, New England Program for Kidney Exchange was founded in 2004, while Ohio Solid Organ Consortium has been conducted ad-hoc kidney exchanges since early 2000s. Currently most number of kidney exchange operations are done in smaller non-profit programs rather than the UNOS national program. The downside of this is that the pairs with difficult-to-match patients dues to severe tissue sensitivity have a much higher chance to be matched in a large pool of pairs rather than in a small pool. What happens is that smaller programs match internally easier-to-match pairs and left-over difficult-to-match pairs form the majority of the national program pair pool. Therefore, such pairs have a very small chance to be matched under the current realm of the market formation. The advantage of a large kidney exchange program is that it will provide a more efficient system than several smaller programs (for example, see the simulations reported in RSÜ 2005a; 2007).

The consolidation of multiple programs in a single large kidney exchange program is difficult. RSÜ 2005b showed that there is no incentive compatible exchange mechanism that would make all smaller programs reveal their all pairs to the centralized national program, when smaller programs only care about maximizing the number of their registered pairs matched (Ashlagi and Roth, 2014, also see). Hence, it is an often debated challenge how to create a single exchange pool with voluntary participation.

It turns out that our proposal of incentivizing compatible pairs to participate in exchange can also help us to create a single large exchange pool. Although there are multiple programs for exchange, and only one of them is also in charge of the administration of the deceased donor queue, as in the case of UNOS. Hence, we can give the right of incentivizing compatible pairs through priorities upon reentry *only* to the UNOS program. We show that such a policy design will cause compatible pairs to register *only* at the UNOS national exchange program, which in turn will attract all other pairs to the UNOS program. Therefore, at equilibrium there will be a unique exchange pool with actual pairs in it – namely the UNOS national program – driving all other exchange programs out of business.

7.1 The Exchange Participation Game for Pairs

Consider the following dynamic game. Suppose there are $n+1$ live-donor kidney exchange platforms $P_0, P_1, ..., P_n$. Platform P_0 is the UNOS national exchange program.

Exchange with incentivized compatible pairs is available only in the UNOS program, P_0, which also oversees the deceased donation program. Hence, only the UNOS program gives a priority to the reentering patient of a compatible pair that previously participated in an exchange conducted through its program.

Each platform uses an *ABO-compatible* FIFO optimal exchange policy to maximize the measure of pairs matched at every instant, while the national program uses the optimal policy by incentivizing compatible pair participation with deceased donation. In the ABO-compatible FIFO policy, ties among pairs who arrive at the same time are broken through an even lottery as long as it does not affect efficiency as explained in Subsection 5.2. Hence, an B-O pair can be matched with an excess A-B pair (i.e., one remains unmatched after all arriving B-A pairs are matched) or an O-B pair with equal probability if they have waited the longest and either matching would result with the same efficiency outcome in terms of maximizing the pairs matched.

We assume that an exogenously determined ρ–fraction of compatible pairs from overdemanded types automatically participate in exchange at platform P_0, and they are not strategic agents such that Assumption 4 is satisfied. It is straightforward to extend our results to the case when compatible pairs are strategic agents and ρ is endogenously determined through their own risk attitudes etc.

We assume that compatible overdemanded pairs are always immediately matched, whether they participate in exchange or not. If there is no available pair in the exchange platform for a compatible pair to be matched, the compatible pair's donor donates to her patient immediately and the compatible pair leaves the pool. Patients without live donors are not strategic agents, either. All patients simultaneously wait at the deceased donation queue.

On the other hand, each patient with an incompatible donor is a strategic agent and would like to maximize his lifetime expected utility while listing at an exchange platform. As each patient can die while waiting for a transplant, we assume that receiving an earlier transplant is preferable to receiving a later transplant. For simplicity of the analysis we assume that a pair chooses whichever donor arrives earlier from a deceased donor or a live donor. Our results would not have changed if we explicitly modeled the utility functions of patients over time and live versus deceased donors (such as, using a measure of expected survival of the transplant).

An incompatible pair can opt in or out of the exchange pool at any point in time after it arrives. An incompatible pair that registers in an exchange platform waits for a live donor's kidney or a deceased donor kidney and has the option to choose to wait for which one.

A patient without a compatible live donor and has not registered in any exchange platform waits to receive a deceased donor under the ABO-identical FIFO allocation policy.

We inspect the Nash equilibria of this game. The first lemma is obvious to prove, and we just state it:

Lemma 7 *At any Nash equilibrium of the participation game, if $X-Y$ type pairs registers for exchange at two distinct exchange platforms with positive probability then their expected waiting times are the same at these platforms.*

In this game any strategy that tells pairs not to participate in exchange at any platform is weakly dominated. For different ρ, there exists an equilibrium in dominated strategies. For example, there exists an equilibrium in which no pair participates in exchange when $\rho = 0$. When $\rho > 0$, there are equilibria in which no self-demanded or reciprocally demanded pair participates in exchange. Hence, we focus on equilibria in undominated strategies:

Proposition 1 *In the participation game, there are pure strategy Nash equilibria in undominated strategies. The total measure of patients matched through exchange or deceased donation is the same and maximal across all such equilibria for the given ρ; moreover, this total measure strictly increases in ρ.*

There are indeed multiple pure strategy equilibria where different measures of pairs register at different programs. Some of these equilibria can be constructed in a straightforward manner: Denote one equilibrium by σ^* where all pairs register at P_0. Consider another strategy profile σ' where a sufficiently small fraction ϵ of all pair types register at platform P_1, while the rest of the pairs register at P_0, including all of the participating compatible pairs. Now, P_1 works as a mini version of P_0 with the same ratio of different pair types registering. Hence, all pairs are matched at the same time at both P_0 and P_1 through exchange and (if needed) deceased donation. Thus, σ' is also an equilibrium, as no pair has any incentive to deviate.

On the other hand, this kind of an equilibrium allows only a *sufficiently small* fraction of pairs to register at platforms other than P_0. As otherwise, there will be excess compatible pairs registering at P_0 and underdemanded pairs registering at other platforms will have unilateral incentives to deviate and register at P_0. Thus, this maximum fraction ϵ is inversely related to ρ: as ρ, the participation rate of compatible pairs, increases, the size of other platforms decrease at equilibria.

Our model does not consider explicitly "difficult to match" pairs, and assumes that each pair has the same tissue type incompatibility probability. In reality, there exist positive measures of highly sensitized pairs and their chances of being matched are much smaller when the size of the exchange pool is small. Hence, from a practical point, a large exchange platform will be more desirable than several small platforms, although all equilibria in undominated strategies are efficient.

Moreover, it does not matter where some pair type participate in exchange for efficiency purposes. For example any positive measure of $O-O$ type incompatible pairs can participate in exchange at any

platform and they could all be matched with each other without affecting the efficiency of exchange. On the other hand, if a positive measure of incompatible $A-O$ pairs participated at a platform where there are no underdemanded pairs, this would decrease the efficiency of the exchange. Hence, we will refer to all pair types that are not self-demanded as **efficiency critical pair types**.

Hence, it is important to create a large exchange platform with efficiency critical pair types. Our main result of this section states under what conditions with compatible pair participation such a large program can be created:

Theorem 9 *In the participation game,*

- *the maximum total equilibrium measure of registrants at platforms other than the national exchange program, P_0, decreases with increasing ρ; and*
- *if*

$$\rho > \frac{\sum_{X-Y\in\mathbf{O}\setminus\{B-A\}} \theta p_Y \pi_X + p_A \pi_B}{\sum_{X-Y\in\mathbf{O}\setminus\{B-A\}} (1-\theta) p_Y \pi_X},$$

then the total measure of efficiency critical pairs participating at P_0 is more than the sum of their respective participation rates in other platforms in every undominated pure strategy equilibrium.

Assumption $\rho > \frac{\sum_{X-Y\in\mathbf{O}\setminus\{B-A\}} \theta p_Y \pi_X + p_A \pi_B}{\sum_{X-Y\in\mathbf{O}\setminus\{B-A\}} (1-\theta) p_Y \pi_X}$ in the theorem makes sure that the measure of compatible pairs participating in exchange are relatively high. In particular with the independent pairing assumptions across blood types and identical donation and patient blood type distributions reported on in the United States for kidney exchange, $\rho > 32\%$ with $\theta = 0.11$.

References

Ashlagi, Itai and Alvin E. Roth (2014), "Free riding and participation in large scale, multi-hospital kidney exchange." *Theoretical Economics*, forthcoming.

Athanassoglou, Stergios and Jay Sethuraman (2011), "House allocation with fractional endowments." *International Journal of Game Theory*, 40, 481–513.

Erdös, Paul and Alfréd Rényi (1960), "The evolution of random graphs." *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5, 17–61.

Ford, L. R. and D. R. Fulkerson (1956), "Maximal flow through a network." *Canadian Journal of Mathematics*, 8, 399–404.

Katta, A-K. and J. Sethuraman (2006), "A solution to the random assignment problem on the full preference domain." *Journal of Economic Theory*, 131, 231–250.

Korte, Bernhard and Jens Vygen (2002), *Combinatorial Optimization: Theory and Algorithms*. Springer.

Magee, J. C., M. L. Barr, G. P. Basadonna, M. R. Johnson, S. Mahadevan, M. A. McBride, D. E. Schaubel, and A. B. Leichtman (2007), "Repeat organ transplantation in the United States, 1996–2005." *American Journal of Transplantation*, 7, 1424–1433.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2004), "Kidney exchange." *Quarterly Journal of Economics*, 119, 457–488.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2005a), "A kidney exchange clearinghouse in New England." *American Economic Review Papers and Proceedings*, 95, 376–380.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2005b), "Transplant center incentives in kidney exchange." Harvard University and Koç University, unpublished mimeo.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2007), "Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences." *American Economic Review*, 97, 828–851.

Sönmez, Tayfun and M. Utku Ünver (2010), "Altruistically unbalanced kidney exchange." Working paper.

Terasaki, Paul I., David W. Gjertson, and J. Michael Cecka (1998), "Paired kidney exchange is not a solution to ABO incompatibility." *Transplantation*, 65.

Ünver, M. Utku (2010), "Dynamic kidney exchange." *Review of Economic Studies*, 77, 372–414.

Yılmaz, Özgür (2009), "Random assignment under weak preferences." *Games and Economic Behavior*, 66, 546–558.

Zenios, S.A., G.M. Chertow, and L.M. Wein (2000), "Dynamic allocation of kidneys to candidates on the transplant waiting list." *Operations Research*, 48, 549–569.

A Appendix: Remaining Proofs

Proof of Lemma 1. We prove it by contradiction: If $\sigma = \omega$ then suppose an infinite or uncountable number of donors are unmatched, and if $\sigma < \omega$ then suppose a donor is unmatched with a positive probability under the FIFO policy. Then, in either case, an infinite or uncountable number of patients are unmatched as well. But then, take a donor who is unmatched, then there exists almost surely a compatible unmatched patient, as the probability of finding no tissue-type compatible patient is $\lim_{n\to\infty} \theta^n = 0$. ■

Proof of Lemma 2. Since $X \neq Y$ and $X \triangleright Y$, we have $Y \not\triangleright X$. Moreover, $W \triangleright Y$ for all blood types W such that $W \triangleright X$.

Suppose to the contrary of the claim, $t^{\mathbf{c}}_Y > t^{\mathbf{c}}_X$. Then the longest-waiting Y blood-type patients would receive the maximum number of organs that would otherwise go to X blood-type patients under the FIFO policy, as they are waiting longer than the longest-waiting X blood-type patients. Hence, either Y blood-type patients do not wait at all, i.e. $t^{\mathbf{c}}_Y = 0$ or X blood-type patients never receive transplant $t^{\mathbf{c}}_X = T$. Either case contradicts the assumption. ■

Proof of Lemma 3. Suppose Y blood-type patients receive X blood-type organs at steady state under the ABO-compatible FIFO allocation policy. By Lemma 2, $t^{\mathbf{c}}_Y \leq t^{\mathbf{c}}_X$. Suppose the inequality is strict. Then either all X blood-type organs would go to longest-waiting X blood-type patients, which would contradict the fact that X blood-type organs are transplanted to Y blood-type patients, or X blood-type patients would not be waiting at all in the deceased-donor queue, which would contradict the assumption that $t^{\mathbf{c}}_Y < t^{\mathbf{c}}_X$. Hence, $t^{\mathbf{c}}_Y = t^{\mathbf{c}}_X$.

Next, suppose that blood types in some $\mathcal{S} \subseteq \mathcal{T}$ are pooled together. Then there is a chain of blood types $\{X_1, ..., X_k\} = \mathcal{S}$ such that X_1 receives from X_1 and X_2, ..., X_{k-1} receives from X_{k-1} and X_k. By the previous paragraph, all types in $\mathcal{S}$ have the same waiting time under the ABO-compatible allocation scheme. Moreover, the supply-demand equations for these types are given as, for all $X \in \mathcal{S}$,

$$\sigma_X = [\pi_X + \phi^d \sigma_X][1 - F(t_{\mathcal{S}})]$$

where $t_{\mathcal{S}}$ is the common waiting time and σ_X is the measure of organs supplied to X blood-type patients. At steady state, we observe an inflow $\phi^d \sigma_X$ of reentrants to the queue. Moreover, $\sum_{X\in\mathcal{S}} \sigma_X = \sum_{X\in\mathcal{S}} \delta_X$. Hence summing up left-hand sides and right-hand sides of these equations, respectively, we get $\sum_{X\in\mathcal{S}} \delta_X = [\sum_{X\in\mathcal{S}}(\pi_X + \phi^d \delta_X)][1 - F(t_{\mathcal{S}})]$. Solution for $t_{\mathcal{S}}$ is given as in Equation 2. ■

Proof of Theorem 2. By Lemma 2, $t^{\mathbf{c}}_Y \leq t^{\mathbf{c}}_X$. As $t^{\mathbf{i}}_X$ is the shortest among $t^{\mathbf{i}}$ for types that Y can receive from, the only way $t^{\mathbf{c}}_Y \leq t^{\mathbf{c}}_X$ can happen is that Y blood-type patients receive X blood-type organs at steady state or X pools with another type which has a higher $t^{\mathbf{i}}$ than Y. However, the latter

is not correct by assumption. Therefore, Y and X blood-type patients are pooled (possibly together with other types). By Lemma 3, $t^{\mathbf{c}}_Y = t^{\mathbf{c}}_X$. Moreover, by transferring some of the X blood-type organs Y and X blood-type patients receive to other compatible-type patients, the waiting time of Y and X blood-type patients can be adjusted above $t_{\{X,Y\}}$ but no higher than $t^{\mathbf{i}}_Y$. Similarly, by transferring some of the X blood-type organs that Y blood-type patients are receiving to X blood-type patients, and substituting those with other compatible organs for Y, the waiting time of Y and X blood-type patients can be adjusted below $t_{\{X,AB\}}$ but no lower than $t^{\mathbf{i}}_X$. Observe that the waiting time of no donor blood type that is compatible with Y blood-type patients can be made shorter than $t^{\mathbf{i}}_X$ or longer than $t^{\mathbf{i}}_Y$, at steady state, under the constraint of Lemma 3, which says that all donating blood types to Y blood-type patients will have the same waiting time. Hence, the composite type of X and Y behaves like Y when it is receiving organs and behaves like X when it is donating organs with deceased-donor inflow $\delta_X + \delta_Y$ and patient inflow $\pi_X + \pi_Y$, by Lemma 3. ■

Proof of Lemma 4. Suppose that for a given X, the non-negative real line for π_X can be divided into a sequence of open intervals marked by $0 = \epsilon_0 < \epsilon_1 < \epsilon_2 < ...$ such that for any k, for any $\pi_X \in (\epsilon_k, \epsilon_{k+1})$ the sets of pool types remain constant; and the sets of pooled types do change in transition from ϵ_k^- to ϵ_k^+ for each k.

For any $\pi_X \in (\epsilon_k, \epsilon_{k+1})$, Equation 2 gives the waiting time of any pooled set $\mathcal{S}$. Moreover, $t^{\mathbf{c}}_{\mathcal{S}}$ strictly increases in π_X for the pooled set $\mathcal{S}$ that includes X and the waiting times of other types do not change.

Moreover, waiting times are continuous in π_X and bounded in this open interval. Hence, left- and right-hand limits exist at each ϵ_k. Next, for some k suppose at $\pi_X = \epsilon_k$ for some blood type left-hand limit is higher than its value at $\pi_X = \epsilon_k$ for the waiting time, i.e. $\lim_{\pi_X \to \epsilon_k^-} t^{\mathbf{c}}_Z > t^{\mathbf{c}}_Z \mid_{\pi_X = \epsilon_k}$ for some Z. Suppose at ϵ_k, Z is pooled in $\mathcal{S}_1 \in 2^{\mathcal{T}}$. However, as the total inflow of patients, $\sum_{Y \in \mathcal{T}} \pi_Y$ at $\pi_X \to \epsilon_k^-$ can be made arbitrarily close to its value at $\pi_X = \epsilon_k$, for some types of a pooled set $\mathcal{S}_2 \in 2^{\mathcal{T}} \setminus \{\mathcal{S}_1\}$ at $\pi_X \to \epsilon_k^-$, we necessarily have $\lim_{\pi_X \to \epsilon_k^-} t_{\mathcal{S}_2} < t_{\mathcal{S}_2} \mid_{\pi_X = \epsilon_k}$. This can only happen if some $Y \in \mathcal{S}_1 \cap \mathcal{S}_2$ that donates to a blood type in $\mathcal{S}_2$ at $\pi_X \to \epsilon_k^-$, which is no longer pooled within $\mathcal{S}_2$ but within $\mathcal{S}_1$ at $\pi_X = \epsilon_k$. But then, this contradicts the definition of ABO-compatible FIFO policy as some deceased donors of Y blood type, which is no longer pooled in $\mathcal{S}_2$ at $\pi_X = \epsilon_k$, could be given to the patients of one or more blood types in $\mathcal{S}_2$ and their waiting time can be decreased without making it smaller than the waiting time for $\mathcal{S}_1$ at $\pi_X = \epsilon_k$.

The cases where $\lim_{\pi_X \to \epsilon_k^-} t^{\mathbf{c}}_Z < t^{\mathbf{c}}_Z \mid_{\pi_X = \epsilon_k}$, $\lim_{\pi_X \to \epsilon_k^+} t^{\mathbf{c}}_Z > t^{\mathbf{c}}_Z \mid_{\pi_X = \epsilon_k}$, and $\lim_{\pi_X \to \epsilon_k^+} t^{\mathbf{c}}_Z < t^{\mathbf{c}}_Z \mid_{\pi_X = \epsilon_k}$ are handled in the symmetric manner, leading to a contradiction. Hence, this shows that all blood types' ABO-compatible waiting times are continuous in π_X.

Since, each waiting time $t^{\mathbf{c}}_Y$ is continuous at each $\pi_X = \epsilon_k$ for all $Y \in \mathcal{T}$ and it is weakly (and strictly for $Y = X$) decreasing at each open interval $\pi_X \in (\epsilon_k, \epsilon_{k+1})$, then it is weakly (and strictly for $Y = X$) decreasing in π_X.

The proof for "decreasing and continuous in δ_X" is analogous to the proof for "increasing and continuous in π_X" and follows the above proof. ■

Proof of Theorem 3. Observe that we have $p_O^l = p_O(1-\theta)$, $p_A^l = (p_O + p_A)(1-\theta)$, $p_B^l = (p_O + p_B)(1-\theta)$, and $p_{AB}^l = 1-\theta$. Hence, $p_O^l < p_A^l, p_B^l < p_{AB}^l$. First, consider the ABO-identical deceased-donor allocation policy. By Equation 4, for any X,

$$t_X^{\mathbf{L,i}} \;= F^{-1}\big(1 - \tfrac{\delta_X}{(\pi_X - c_X) + \phi^d \delta_X}\big), \tag{21}$$

where $c_X = (1-\phi^l)p_X^l \lambda \pi_X \in (0, \pi_X)$. As $t_X^{\mathbf{L,i}}$ is increasing in net patient inflow, comparing Equation 1 with Equation 21 we conclude for all X, $t_X^{\mathbf{L,i}} < t_X^{\mathbf{i}}$.

Next, consider the ABO-compatible deceased-donor allocation policy. Assume that we introduce patient - live donor pairs for each blood type one at a time. The net effect of having patients with live donors is a decrease in the new patient inflow π_X by c_X for each X (as in the case of ABO-identical allocation policy). Hence, using Lemma 4 for all four blood types consecutively, we conclude that $t_X^{\mathbf{L,c}} < t_X^{\mathbf{c}}$ for all X.

In the rest of the proof, we analyze the benchmark case where δ_X/π_X is constant across all blood types X. Then $c_O \le c_X$ for all X and $c_{AB} \ge c_X$ for all X. These in turn imply that $t_O^{\mathbf{L,i}} \ge t_X^{\mathbf{L,i}}$ for all X and $t_{AB}^{\mathbf{L,i}} \le t_X^{\mathbf{L,i}}$ for all X, respectively, since $t_X^{\mathbf{L,i}}$ is decreasing in c_X We also have

$$\frac{\delta_O}{\pi_O - c_O} \le \frac{\delta_A}{\pi_A - c_A}\,,\; \frac{\delta_B}{\pi_B - c_B} \le \frac{\delta_{AB}}{\pi_{AB} - c_{AB}\delta_{AB}}.$$

Then by Theorem 2 and the procedure following this theorem, using $\pi_X - c_X\delta_X$ instead of π_X for all X, we observe that none of the blood types are pooled together when live donation is possible under the ABO-compatible deceased donation policy. Thus, we also have $t_X^{\mathbf{L,c}} = t_X^{\mathbf{L,i}}$ for all X. Further assume that $p_A > p_B$. Then $p_A^l > p_B^l$. Therefore, $c_A > c_B$, which in turn implies $\frac{\delta_B}{\pi_B - c_B} < \frac{\delta_A}{\pi_A - c_A}$, and hence, $t_A^{\mathbf{L,i}} < t_B^{\mathbf{L,i}}$. ■

Proof of Theorem 4. Under the proposed policy, by Lemma 6 all self-demanded pairs can be matched with their own type pairs as soon as they arrive, and all pairs of type $B-A$ that has the lower inflow rate by assumption than $A-B$ pairs, will be matched under as soon as they arrive with their reciprocal type pairs. Hence, under this policy only $A-B$ pairs will remain in the exchange pool at any point in time. These pairs can only be matched with overdemanded pairs by Lemma 5, as $B-A$ pairs are already committed to other $A-B$ pairs.

Next consider underdemanded type pairs. These are $Y-X$ type pairs such that $Y \neq X$ and $Y \triangleright X$. By Assumption 2, we have $\theta p_Y \pi_X \le p_X \pi_Y$. By Lemma 5, they can only be matched with overdemanded types. Recall that the inflow of each $Y-X$ type pair to the exchange pool is $p_Y \lambda \pi_X$. Their reciprocal types $X-Y$, which is overdemanded, has the inflow measure $\theta p_X \lambda \pi_Y < p_Y \lambda \pi_X$. Hence, we can match all such overdemanded pairs $X-Y$ (by Lemma 6) as soon as they enter

the pool under the proposed policy with their reciprocal types pairs. As all overdemanded, self-demanded, and type $W-Z$ reciprocally demanded pairs are matched as soon as they arrive, by Lemma 6, the proposed policy achieves the maximum measure of pairs matched. At steady state, as no incompatible overdemanded-type, self-demanded and $B-A$ types each does wait in the pool, gets immediately matched, and saves one additional pair, the maximum mass of possible exchanges is also conducted in this manner in any closed time interval.

On the other hand, if we do not conduct the exchanges immediately whenever they become available, but after a closed time interval, then some of the patients of overdemanded, self-demanded, and $B-A$ type pairs who have arrived earlier will not survive. Hence, when we conduct the exchanges at the end of the time interval, we will match a strictly smaller mass of possible pairs than we would have matched under the proposed policy. ■

Proof of Theorem 6. We prove the theorem using the concept of *flow networks* developed in the combinatorial optimization and graph theory literature (see for example Korte and Vygen (2002) for an excellent survey).

This tool will be used to show that, for each $\ell \in \{1,...,k\}$, for each patient group $i \in \mathbf{V}_\ell$ (as defined in Equations 16 and 18), we can feasibly serve deceased donors / pairs belonging to groups in $\mathbf{P}_\ell$ (as defined in Equations 17 and 19) to patients of group i at a rate $w_i r^d_{\mathbf{V}_\ell}(\mathbf{P}_\ell)$ (as defined in Equations 13, 15, and 20).

A *flow network* in our context is the directed graph with *nodes* $\mathbf{N} = \{\sigma, \tau\} \cup \mathbf{U} \cup \mathbf{V}$ such that σ is referred to as the *source* and τ is referred to as the *sink*. An *edge* of the flow network originating from node i and pointing at node j is denoted by (i,j). In particular, each $\mathbf{U}$ node is pointed at by σ. Hence, for each $h \in \mathbf{U}$ (σ, h) is in the network. Also each node in $\mathbf{V}$ points at t. Hence, for each $h \in \mathbf{V}$ $(h, \tau) \in \mathbf{V}$. Moreover, there are edges starting from each node in $\mathbf{U}$ and ending at some nodes in $\mathbf{V}$: for each $i \in \mathbf{U}$ and $j \in \mathbf{V}$, (i,j) is a directed edge if and only if $c_{i,j} = 1$. Let E be the set of edges of the network.

We will send flows from the source σ through the edges of the graph and these flows will reach the sink. For this purpose, each edge $(i,j) \in E$ has also a *capacity* $q(i,j) > 0$ denoting the maximum flow it can carry. For all other pairs of nodes $(i,j) \notin E$, let $q(i,j) = 0$. Let $q = (q(i,j))_{i,j \in \mathbf{N}}$ denote the capacity vector for all the edges. A *flow network* is denoted by the pair $(\mathbf{N}, q)$. Fix a flow network $(\mathbf{N}, q)$.

A *flow function* $f : \mathbf{N} \times \mathbf{N} \to \mathbb{R}$ is a mapping such that for each $i, j \in \mathbf{N}$ we have (i) if $q(i,j) > 0$ then $0 \leq f(i,j) \leq q(i,j)$ and if $q(i,j) = 0$ then $f(i,j) \leq 0$, (ii) $f(j,i) = -f(i,j)$, and (iii) if $i \notin \{\sigma, \tau\}$ then $\sum_{h \in \mathbf{N}} f(i,h) = 0$. Property (i) says that an edge cannot carry a flow higher than its capacity, and in particular, for positive capacity edges the flow cannot be negative and for zero capacity edges the flow cannot be positive. Property (ii) is a technical one and used for ease of notation making sure that the flow is a directed quantity but not a scalar: the flow of the reverse of an edge is the negative

of the flow of the edge. Property (iii) says that for any node other than the source and the sink, the flows from it and flows into it cancel out, i.e., all flows entering it also leave the node. Let $\mathcal{F}$ be the set of flow functions. We refer to $f(i,j)$ as the *flow from node i to j under f*. For a subset of nodes $\{\sigma\} \subseteq \mathbf{S} \subseteq \mathbf{N} \setminus \{\tau\}$, the *flow from* $\mathbf{S}$ *(to* $\mathbf{N} \setminus \mathbf{S}$*) under* f is denoted by $f(\mathbf{S}) = \sum_{i \in \mathbf{S}, j \in \mathbf{N} \setminus \mathbf{S}} f(i,j)$. Such a subset of nodes $\mathbf{S}$ is denoted as a *cut*.

The *total capacity of a cut* $\mathbf{S}$ is defined as $q(\mathbf{S}) = \sum_{i \in \mathbf{S}, j \in \mathbf{N} \setminus \mathbf{S}} q(i,j)$, i.e., it is the sum of the capacities of edges originating from a node in $\mathbf{S}$ and ending at a node in $\mathbf{N} \setminus \mathbf{S}$. A *minimum cut* $\mathbf{S}$ is a cut such that $q(\mathbf{S}) = \min_{\{\sigma\} \subseteq \mathbf{S}' \subseteq \mathbf{N} \setminus \{\tau\}} q(\mathbf{S}')$, i.e. a cut with the minimum total capacity.

The flow of f is its flow from cut $\mathbf{N} \setminus \{\tau\}$ to cut $\{\tau\}$, which is also equal to its flow from cut $\{\sigma\}$ to cut $\mathbf{N} \setminus \{\sigma\}$. The *maximum flow* over the flow network $(\mathbf{N}, q)$ is defined as $\max_{f \in \mathcal{F}} f(\mathbf{N} \setminus \{\tau\})$.

The following is the fundamental theorem that relates the capacities of the edges to the maximum flow that can be carried over a flow network:

Minimum Cut - Maximum Flow Theorem (Ford and Fulkerson (1956)): The maximum flow over a flow network is equal to the total capacity of one of its minimum cuts.

One direction of the theorem's statement, i.e., the maximum flow cannot exceed the total capacity of a minimum cut is obvious by the definition of a flow function. The other direction is proven through this theorem.

For our flow network used in the proof of our theorem, we define the capacities as follows (see Figure 9, where the edges are denoted by lines with arrows and their capacities are written on the lines; it defines a flow network using the feasible exchange and deceased donation graph given in Figure 6): For an edge (i,j) such that $i \in \mathbf{U}$ and $j \in \mathbf{V}$, we set its capacity to $q(i,j) = +\infty$. Hence, it can carry any load. On the other hand, for edge (j,τ) for each $j \in \mathbf{O}$, we set its capacity $q(j,\tau) = w_j$, the arrival rate of the deceased donor / pair type j to the pool, as defined in Equation 13. For edge (σ,i) for each $i \in \mathbf{U}$, we set its capacity $q^\chi(\sigma,i) = \chi w_i$, where w_i is the arrival rate of the patient without live donor / pair type i to the pool, as defined in Equation 13, and $\chi \in \mathbb{R}_+$ is a parameter that will be changed in our construction. We refer to such a flow network as a χ–parametric flow network.

The idea behind this construction is as follows: as we increase χ continuously starting from 0, the flows carried from the source to the rest of the network are set to be equal to the capacities of the edges from the source for an appropriately defined flow function $f^\chi \in \mathcal{F}$. As χ is close to zero, all the flows can be carried over the network and hence, $\{\sigma\}$ is a minimum cut. We will be able to increase these continuously until a break point occurs $\chi_1 < 1$, i.e. the minimum cut becomes a proper superset of $\{\sigma\}$. To see that, to the contrary of the claim suppose $\chi_1 \geq 1$. We have the total capacity of cut $\mathbf{N} \setminus \{\tau\}$ equal to $q^{\chi_1}(\mathbf{N} \setminus \{\tau\}) = \sum_{j \in \mathbf{O}} w_j$, which should be greater than or equal to maximum flow over the network. On the other hand, the total capacity of cut $\{\sigma\}$ is equal to

$q^{\chi_1}(\{\sigma\}) = \chi_1 \sum_{i\in\mathbf{U}} w_i$. We increase χ to χ_1 so that the flow of f^{χ_1} is equal to $q^{\chi_1}(\{\sigma\})$. However, this is a contradiction by Assumptions 1, 2, and 3, as the flow of f^{χ_1}, $q^{\chi_1}(\{\sigma\}) > q^{\chi_1}(\mathbf{N}\setminus\{\tau\})$, the maximum flow over the network at χ_1.

Hence, at $\chi = \chi_1 < 1$ there will be a minimum cut larger than $\{\sigma\}$, such that we will not be able to carry all the flows if we exceed χ above χ_1. Let $\{\sigma\} \subsetneq \mathbf{N}_1$ be this minimum cut. If there are multiple such cuts, let $\mathbf{N}_1$ be the largest of them. It is straightforward to see that there is a minimum cut, which includes all minimum cuts as subsets.

What are the properties of this minimum cut? Suppose $i \in \mathbf{N}_1 \cap \mathbf{U}$. Then observe that all $j \in \mathbf{O}$ such that $c_{i,j} = 1$ is also in $\mathbf{N}_1$. As otherwise the edge (i,j) with capacity $q(i,j) = +\infty$ would make the total capacity of the minimum cut equal to $+\infty$. However, this is a contradiction to $\mathbf{N}_1$ being a minimum cut, as the cut $\{\sigma\}$ has always a finite total capacity (see Figure 10 for an example of a possible minimum cut at some χ_1).

Hence, whenever $i \in \mathbf{N}_1 \cap \mathbf{U}$ then all $j \in \mathbf{O}$ with $c_{i,j} = 1$ also satisfy $j \in \mathbf{N}_1$. Let $\mathbf{V}_1 = \mathbf{N}_1 \cap \mathbf{U}$, and $\mathbf{P}_1 = \mathbf{N}_1 \cap \mathbf{O}$. By the above construction $\mathbf{P}_1 = \mathbf{C}_{\mathbf{V}_1}(\mathbf{O})$ (as defined in Equation 7).

The total capacity of $\mathbf{N}_1$ is equal to

$$q^{\chi_1}(\mathbf{N}_1) = \sum_{i\in\mathbf{N}\setminus\mathbf{V}_1} q^{\chi_1}(\sigma, i) + \sum_{j\in\mathbf{P}_1} q(j,\tau) = \chi_1 \sum_{i\in\mathbf{N}\setminus\mathbf{V}_1} w_i + \sum_{j\in\mathbf{P}_1} w_j.$$

On the other hand, the flow of f^{χ_1} over the network at χ_1 is given as

$$f^{\chi_1}(\{\sigma\}) = \sum_{i\in\mathbf{U}} f^{\chi_1}(\sigma, i) = \chi_1 \sum_{i\in\mathbf{U}} w_i.$$

This is maximum as all the capacity of the edges from σ are used i.e., $f^{\chi_1}(\sigma,i) = q^{\chi_1}(\sigma,i)$ for all $i \in \mathbf{U}$.

As $\mathbf{N}_1$ is a minimum cut, by the Minimum Cut-Maximum Flow Theorem,

$$q^{\chi_1}(\mathbf{N}_1) = f^{\chi_1}(\{\sigma\}).$$

Hence,

$$\chi_1 \sum_{i\in\mathbf{U}\setminus\mathbf{V}_1} w_i + \sum_{j\in\mathbf{P}_1} w_j = \chi_1 \sum_{i\in\mathbf{U}} w_i,$$

leading to

$$\chi_1 = \frac{\sum_{j\in\mathbf{P}_1} w_j}{\sum_{i\in\mathbf{V}_1} w_i} = r^d_{\mathbf{V}_1}(\mathbf{P}_1)$$

where r^d was defined in Equation 15.

Observe that even if we increase χ beyond χ_1, the flow over the edges $((\sigma,i))_{i\in\mathbf{V}_1}$ will not increase and no additional flow through the increased χ will flow through the nodes $j \in \mathbf{P}_1$. Therefore, we can remove the nodes in $\mathbf{V}_1$ and $\mathbf{P}_1$ from the network and repeat the above exercise iteratively. As result,

we determine a number of minimum cuts $\mathbf{N}_1, ..., \mathbf{N}_k$ with corresponding node sets in $\mathbf{U}$ as $\mathbf{V}_1, ..., \mathbf{V}_k$ and node sets in $\mathbf{O}$ as $\mathbf{P}_1, ..., \mathbf{P}_k$ with breakpoints $\chi_1 < ... < \chi_k < 1$ such that $\mathbf{P}_\ell = \mathbf{C}_{\mathbf{V}_\ell}(\mathbf{O} \setminus \cup_{\ell'=1}^{\ell-1} \mathbf{P}_{\ell'})$ and $\chi_\ell = \frac{\sum_{j \in \mathbf{P}_\ell} w_j}{\sum_{i \in \mathbf{V}_\ell}} = r^d_{\mathbf{V}_\ell}(\mathbf{P}_\ell)$ for each $\ell \in \{1, ..., k\}$.

This proves that for each patient group $i \in \mathbf{V}_\ell$, we can feasibly serve deceased donors / pairs belonging to groups in $\mathbf{P}_\ell$ to group i at a rate $f^{\chi_\ell}(\sigma, i) = \chi_\ell w_i = w_i r^d_{\mathbf{V}_\ell}(\mathbf{P}_\ell)$. Hence, we can feasibly match a measure $f^{\chi_\ell}(\sigma, i)$ of patients belonging to group i with arriving deceased donors (through deceased donation) and pairs (through echange) in $\mathbf{P}_\ell$ by Lemmas 1 and 5, respectively.

Define $t^{\mathbf{E},\mathbf{c}}_{\mathbf{V}_\ell} = F^{-1}(1 - r^d_{\mathbf{V}_\ell}(\mathbf{P}_\ell))$ for each ℓ. Observe that $t^{\mathbf{E},\mathbf{c}}_{\mathbf{V}_k} < t^{\mathbf{E},\mathbf{c}}_{\mathbf{V}_{k-1}} < ... < t^{\mathbf{E},\mathbf{c}}_{\mathbf{V}_1}$.

After $t^{\mathbf{E},\mathbf{c}}_{\mathbf{V}_k}$ years of entry, the measure of live patients belonging to the groups in $\mathbf{V}_k$ is exactly equal to $\sum_{j \in \mathbf{P}_k} w_j$, the measure of arriving deceased donors / pairs belonging to groups in $\mathbf{P}_k$. None of the other patients belonging to groups in $\mathbf{V}_1, ..., \mathbf{V}_{k-1}$ can be matched through deceased donation or exchange with deceased donors / pairs belonging to $\mathbf{P}_k$. Hence, they have to wait longer than $t^{\mathbf{E},\mathbf{c}}_{\mathbf{V}_k}$. Moreover, none of the patients of groups in $\mathbf{V}_k$ will be matched with deceased donors / pairs of groups in $\mathbf{P}_1, ..., \mathbf{P}_{k-1}$, as this will decrease their waiting time at the cost of increasing the waiting time of other groups, contradicting the FIFO protocol. We also proved above that all remaining live patients/pairs in V_k after $t^{\mathbf{E},\mathbf{c}}_{\mathbf{V}_k}$ years of entry (that is $r^d_{\mathbf{V}_k}(\mathbf{P}_k)$ fraction of the arriving rate) can be matched with all arriving deceased donors / pairs belonging to groups in $\mathbf{P}_k$. Hence, remaining live patients belonging to groups in $\mathbf{V}_k$ will be matched after $t^{\mathbf{E},\mathbf{c}}_{\mathbf{V}_k}$ years of entry. We repeat the above argument for each of the remaining sets $\ell = k-1, ..., 1$, concluding the proof of the theorem. ■

Proof of Theorem 7. Under the proposed policy, using Assumption 4 (similar to the role of Assumption 2 in the proof of Theorem 4), as soon as they arrive all incompatible and willing compatible pairs can be matched with incompatible pairs of their reciprocal type by Lemma 5. $A-B$ and $B-A$ type pairs will be matched with each other, and by Assumption 3, $A-B$ pairs will remain in the list while $B-A$ pairs will be matched as soon as they enter the exchange pool by Lemma 5. As no overdemanded pairs are left, more of $A-B$ type pairs or underdemanded-type pairs cannot be matched. For any blood type X, all incompatible $X-X$, self-demanded-type, pairs can already be matched without the use of compatible pairs as they arrive by Lemma 5. Therefore, none of the compatible $X-X$ type pairs is needed under the incentivized compatible-pair exchange scheme. Hence, the maximum measure of pairs possible are matched at each point in time under the proposed policy. At steady state, as no willing overdemanded-type, self-demanded and $B-A$ types each does wait in the pool, gets immediately matched, and saves one additional pair, the maximum mass of possible exchanges is also conducted in this manner in any closed time interval.

Similar to the proof of Theorem 4, if we do not conduct the exchanges immediately whenever they become available, but after a closed time interval, then some of the patients of overdemanded, self-demanded, and $B-A$ type pairs who have arrived earlier will not survive and some of the compatible pairs may withdraw themselves from exchange. Hence, when we conduct the exchanges

at the end of the time interval, we will match a strictly smaller mass of possible pairs than we would have matched under the proposed policy. ■

Proof of Theorem 8. Let $\psi^{\mathbf{i,c}}$ be the ABO-identical optimal policy explained in Theorem 7 under exchange with incompatible pairs and incentivized compatible pairs, and $\varphi^{\mathbf{i}}$ be the ABO-identical optimal policy explained in Theorem 4 under exchange with only incompatible pairs. Any reentrant patient is classified as a patient without live donor. Under $\psi^{\mathbf{i,c}}$, no unwilling compatible pairs and compatible self-demanded pairs, and under $\varphi^{\mathbf{i}}$, no compatible pairs participate in exchange; however, their patients immediately receive a live donation from their own donors. All willing overdemanded pairs are matched through exchange with their reciprocal types under both $\psi^{\mathbf{i,c}}$ and $\varphi^{\mathbf{i}}$ as soon as they enter the pool (by Assumption 4). We first prove Statement 2 and then the rest.

Proof of Statement 1: First consider underdemanded pairs. Suppose that an underdemanded $X-Y$ pair type is not pooled with X blood-type patients without live donors for deceased donation under $\varphi^{\mathbf{i}}$. Under $\psi^{\mathbf{i,c}}$, that type of pairs is matched at the rate

$$m_2^{X-Y} = [\rho(1-\theta)+\theta]p_X\lambda\pi_Y, \tag{22}$$

at each point in time while under $\varphi^{\mathbf{i}}$, they are matched at the rate

$$m_1^{X-Y} = \theta p_X\lambda\pi_Y, \tag{23}$$

which is strictly smaller.

Next suppose pair types $X_1-Y_1,...,X_\ell-Y_\ell$ are pooled altogether for deceased donation, and suppose among these pair types, $X_{\ell^*}-Y_{\ell^*}$ is underdemanded. Note that all of these pair types are either underdemanded or $A-B$. Each X_k-Y_k is matched at the rate $m_1^{X_k-Y_k}+\epsilon_1^{X_k-Y_k}$ under $\varphi^{\mathbf{i}}$, where the rate $\epsilon_1^{X_k-Y_k}>0$ is the measure of $X-Y$ pairs whose patients receive deceased donation and $m_1^{X_k-Y_k}$ is defined as in Equation 22. Under $\psi^{\mathbf{i,c}}$, $m_2^{X_k-Y_k}$ is the measure of the $Y-X$ pairs willing to participate in exchange, which is strictly larger than $m_1^{X_k-Y_k}$, while the rate of deceased donation does not change. Hence, while $m_2^{X_k-Y_k}-m_1^{X_k-Y_k}$ more of X_k-Y_k pairs participate in exchange under $\psi^{\mathbf{i,c}}$, fewer of such pairs may receive deceased donation. Suppose that $e_2^{X_k-Y_k}$ is the rate of X_k-Y_k pairs receiving deceased donation under $\psi^{\mathbf{i,c}}$. We will show that $\iota_k=[m_2^{X_k-Y_k}+\epsilon_2^{X_k-Y_k}]-[m_1^{X_k-Y_k}+\epsilon_1^{X_k-Y_k}]>0$ for all k. Suppose not for some k. In particular if there are multiple such k, let k be chosen with the smallest $\iota_k\leq 0$. Hence, as waiting times of all pairs $X_1-Y_1,...,X_\ell-Y_\ell$ is the same under $\varphi^{\mathbf{i}}$, X_k-Y_k's waiting time increases the most among all pairs or stays the same and no other pair's waiting time increases under $\psi^{\mathbf{i,c}}$. Hence, X_k-Y_k continues to be pooled with X blood-type patients without live donors under $\psi^{\mathbf{i,c}}$. As $m_2^{X_{\ell^*}-Y_{\ell^*}}-m_1^{X_{\ell^*}-Y_{\ell^*}}>0$, and for all $k^*\neq\ell^*$ we have, $m_2^{X_{k^*}-Y_{k^*}}-m_1^{X_{k^*}-Y_{k^*}}\geq 0$, then a higher share of deceased donors should go to X_k-Y_k pairs under $\psi^{\mathbf{i,c}}$ with respect to $\varphi^{\mathbf{i}}$. Hence, $\epsilon_2^{X_k-Y_k}-\epsilon_1^{X_k-Y_k}>0$ implying that $\iota_k>0$, a contradiction.

Hence unless $A-B$ is pooled by itself with A blood-type patients without live donors under $\varphi^{\mathbf{i}}$, any pooled paired group with X blood-type patients without live donors has strictly higher rate being matched at each point in time under $\psi^{\mathbf{i,c}}$.

We continue with other patient groups. All overdemanded pairs and self-demanded pairs receive live donation under both $\psi^{\mathbf{i,c}}$ and $\varphi^{\mathbf{i}}$ immediately after their arrive. We already showed that underdemanded pairs strictly benefit from $\psi^{\mathbf{i,c}}$. Moreover, by Assumption 3, Theorems 4 and 7, all $B-A$ pairs are matched with $A-B$ pairs through exchange as soon as they enter the exchange pool. This and the proof of for underdemanded pairs imply that $A-B$ pairs either benefit under $\psi^{\mathbf{i,c}}$ (if they are pooled with an underdemanded type for deceased donation under $\varphi^{\mathbf{i}}$) or they remain indifferent between the two policies (otherwise). Next consider X blood-type patients without live donors. As more of underdemanded-type pairs are matched through exchange and the same rate of $A-B$ pairs pairs participate in exchange under $\psi^{\mathbf{i,c}}$, overall fewer of underdemanded-type and $A-B$ type pairs will be left from the same cohort for deceased donation. Hence, weakly more X blood-type patients without live donors receive donation under $\psi^{\mathbf{i,c}}$.

Proof of Statement 2: Under $\psi^{\mathbf{i,c}}$, by Theorem 8 $X-X$ pairs are only matched in exchange with $X-X$ pairs. Moreover, all incompatible $X-X$ pairs are almost surely matched through exchange as soon as they arrive with each other. Hence no compatible $X-X$ pairs are used to match them.

Proof of Statement 3: Patient blood type O can form 4 types of pairs: $O-O$, $O-A$, $O-B$, and $O-AB$. None of them can form compatible pairs except $O-O$. By Statement 3, no compatible $O-O$ pairs participate in exchange. Hence, upon possible reentry under $\psi^{\mathbf{i,c}}$, no O blood-type patients are prioritized. On the other hand, positive measures of compatible overdemanded pairs with A, B, AB blood-type patients participate in exchange. Therefore, a positive measure of these patient reenter at steady state and they get prioritized.

Proof of Statement 4: First observe that the waiting time of underdemanded types strictly decreases by Statement 1. The waiting times of reciprocally demanded $B-A$ type pairs and $A-B$ type pairs do not increase by Statement 1. Moreover, self-demanded and overdemanded type pairs do not wait and get immediately matched under both policies. Finally we consider patients without live donors. To see how their waiting times are affected we consider he change of rates of exchange for compatible and incompatible pairs first. We do this analysis for all blood types separately.

- O blood-type patients:
  - *Compatible pairs:* $O-O$ is the only compatible type with O blood-type patients. However, incompatible $O-O$ pairs are already matched immediately with each other in exchange. Hence,
    $$\kappa_O = 0$$
    measure of compatible pairs with O blood-type patients participates in exchange.

– *Incompatible pairs:* A measure of $[\rho(1-\theta)+\theta]p_O\lambda[\pi_A+\pi_B+\pi_{AB}]$ incompatible pairs with O blood-type patients are matched through exchange with their reciprocal type pairs at each point in time. This is a net increase of

$$\iota_O = \rho(1-\theta)p_O\lambda[\pi_A+\pi_B+\pi_{AB}]$$

with respect to regular exchange. If some of these pair types are pooled for deceased donation under exchange with incentivized compatible pairs, then they are also pooled for deceased donation under regular exchange.

– *Patients without live donors:*

** Prioritized patients without live donors*: As no O blood-type reentrant patients are prioritized, all O blood-type deceased donors are still given to O blood-type patients without live donors and there is a

$$\phi^l\kappa_O = 0$$

measure of prioritized O blood-type reentrants per unit time.

** Regular patients without live donors*: On the other hand, some additional O blood-type patients are saved through exchange, an additional rate of

$$\phi^l\iota_O = \phi^l[\rho(1-\theta)]p_O\lambda[\pi_A+\pi_B+\pi_{AB}]$$

of O blood-type patients reenter with respect to regular exchange. These reentrant patients join the regular deceased donor queue. However, if some underdemanded pairs with O blood-type patients receive deceased donation under regular exchange regime then some of these fall from competition for deceased donation under exchange with incentivized compatible pairs. Depending the size of this fallout, the net effect on the net inflow of O blood-type patients without live donors can be negative or positive, but this additional inflow to the regular deceased donation queue will be no more than

$$\phi^l\iota_O.$$

Depending on which of the above effects dominates, the waiting time for regular O blood-type patients without live donors can slightly increase or decrease under exchange with incentivized compatible pairs.

- A blood-type patients:

– *Compatible pairs:* A measure of

$$\kappa_A = \rho(1-\theta)p_O\lambda\pi_A,$$

$A-O$ type compatible pairs participate in exchange to save $O-A$ type pairs. Self-demanded $A-A$ type compatible pairs do not participate in exchange.

– *Incompatible pairs:* A measure of $[\rho(1-\theta)+\theta]p_A\lambda\pi_{AB}$ of underdemanded type pairs $A-AB$ are matched through exchange in every point in time. This is a net increase of

$$\iota_A = \rho(1-\theta)p_A\lambda\pi_{AB}$$

with respect to regular exchange. If some of these pair types are pooled for deceased donation under exchange with incentivized compatible pairs, then they are also pooled for deceased donation under regular exchange.

The reciprocally demanded pair type $A-B$ continues to run a deficit as $B-A$ inflow is – by Assumption 3 – lower than $A-B$ inflow. If $A-B$ type pairs wait both for $B-A$ type pairs and decease donors under incentivized exchange, see the case for patients without live donors to understand the effect of incentivized exchange on their waiting times below. On the other hand if they are waiting exclusively for $B-A$s under incentivized exchange policies, then $A-B$ types wait for the same time under both regular and incentivized exchange, and exactly the same measure of them gets matched in every moment.

– *Patients without live donors:*

* *Prioritized patients without live donors*: Patients of some of the $A-O$ type compatible pairs that previously participated in exchange reenter as their grafts fail. Their inflow measure is

$$\phi^l\kappa_A = \phi^l\rho(1-\theta)p_O\lambda\pi_A.$$

These reentering A blood-type patients, who no longer have live donors, directly go to the top of the A blood-type deceased-donor queue instead of going to the bottom as under regular exchange. We will refer to this as *incentivized exchange burden*. This is also the rate of the decease donors reserved for these patients.

* *Regular patients without live donors*: An additional ι_A measure of $A-AB$ pairs are saved by $AB-A$ types through exchange, a measure of

$$\phi^l\iota_A = \phi^l\rho(1-\theta)p_A\lambda\pi_{AB}.$$

A blood-type patients reenter and join in the regular queue to the A blood-type patients without live donors. However, if some $A-AB$ pairs receive deceased donation under regular exchange regime then some of these fall from competition for deceased donation under exchange with incentivized compatible pairs. Depending the size of this fallout, the net effect on the net inflow of A blood-type patients without live donors for the regular queue can be negative or positive, but this additional inflow will be no more than

$$\phi^l\iota_A - \phi^l\kappa_A.$$

As a result, the waiting time for regular A blood-type patients without live donors, can slightly increase or decrease under exchange with incentivized compatible pairs (see Figure 8 for an example of the overall impact of this new exchange policy on A blood-type patients).

- B blood-type patients: Symmetric version of A blood-type patients, except that $B-A$'s are immediately matched with $A-B$'s when they enter the pool by the assumption that $B-A$'s are on the short side.

- AB blood-type patients:

  - *Compatible pairs:* A total measure of

    $$\kappa_{AB} = \rho(1-\theta)[p_O + p_A + p_B]\lambda\pi_{AB}$$

    compatible $AB-O$, $AB-A$, and $AB-B$ type pairs participate in exchange to save their reciprocals at each point in time. Self-demanded compatible $AB-AB$ type pairs do not participate in exchange.

  - *Incompatible pairs:* All incompatible pairs with AB blood-type patients are either self-demanded or overdemanded. Hence, they are matched immediately when they arrive through exchange with their reciprocal types under both regular exchange and exchange with incentivized compatible pairs. Hence additionally a

    $$\iota_{AB} = 0$$

    measure of incompatible pairs with AB blood-type patients are matched under the new regime.

  - *Patients without live donors:*

    ** Prioritized patients without live donors*: The reentry burden of AB blood-type patients from previous compatible pairs that participated in exchange is

    $$\phi^l\kappa_{AB} = \phi^l\rho(1-\theta)[p_O + p_A + p_B]\lambda\pi_{AB},$$

    which is the rate of prioritization for AB blood-type reentrants to the deceased donor queue. This is also the rate of the decease donors reserved for these patients.

    ** Regular patients without live donors*: On the other hand, the same measure of AB blood-type patients reenter at each point in time under both regular exchange and exchange with incentivized compatible pairs. No pairs with AB blood-type patients are pooled for deceased donation under either regular exchange or exchange with incentivized compatible pairs. Hence, a

    $$\phi^l\iota_{AB} = 0$$

measure of additional AB blood-type reentrants from previous incompatible pairs reenter the deceased donor queue. Net increase of rate of entry to the regular AB blood-type deceased donor queue is negative and equal to

$$-\phi^l \kappa_{AB}.$$

As a result, the waiting time for regular AB blood-type patients without live donors unambiguously slightly increases under exchange with incentivized compatible pairs. This holds as all of the prioritized AB blood-type patients receive deceased donation under exchange with incentivized compatible pairs, while some patients from the same population would have died and not received deceased donation under the alternative regime, regular exchange.

■

Proof of Proposition 1. Fix $\rho \in [0,1]$ such that Assumption 4 holds. Consider the following strategy profile σ^*: all pairs register at P_0, the national program, with probability 1. As an optimal exchange mechanism is used, then under this profile the maximal measure of pairs are matched at each point in time as explained in Theorem 7 and 8. Moreover, σ^* is a pure strategy equilibrium in undominated strategies: as no pairs register in any other platform then it is a best response to register at P_0.

Consider an arbitrary pure strategy equilibrium profile σ in undominated strategies. Each pair registers at a unique exchange platform with probability one as soon as it arrives.

We prove that all pairs belonging to overdemanded pair types and pair type $B-A$ are matched with pairs belonging to underdemanded types or pair type $A-B$ immediately when they arrive under σ. To the contrary, suppose there is a platform P_a where a positive measure of pairs of a type $X-Y \in \mathbf{O} \cap \mathcal{T} \times \mathcal{T}$, i.e., overdemanded or type $B-A$, are not matched with pairs of types in $\mathbf{U} \cap \mathcal{T} \times \mathcal{T}$, i.e., either underdemanded or type $A-B$, at σ when they arrive with a positive probability (using the notation in Subsection 5.2).

Consider any pair type $W_1 - Z_1$ in set $\mathbf{U}$ that has $c_{X-Y,W_1-Z_1} = 1$ (i.e., that is mutually blood-type compatible with an $X-Y$ type pair using the same notation). All pairs of type $W_1 - Z_1$ should be matched immediately at σ, as otherwise such a pair x can register at P_a and can be immediately matched with probability 1 with one of the $X-Y$ pairs at σ. The reason for this is as follows: As pair x is of measure 0 and a positive measure of $X-Y$ pairs are either being matched with other overdemanded pairs or not being matched at all, the platform P_a, which is using an optimal exchange policy with randomization when there are multiple possible pairs to match, will match pair x immediately with probability 1. This implies that all $W_1 - Z_1$ type pairs are matched with probability 1 through exchange when they arrive at σ by Lemma 7.

Suppose $\mathbf{P}_1 \subseteq \mathbf{O} \cap \mathcal{T} \times \mathcal{T}$ is the set of overdemanded pair types or type $B-A$ with which $W_1 - Z_1$ type pairs are mutually blood-type compatible: that is, $\mathbf{P}_1 = \mathbf{C}_{\{W_1-Z_1\}}(\mathbf{O} \cap \mathcal{T} \times \mathcal{T})$. Observe that $AB-O \in \mathbf{P}_1$. Let $\mathbf{V}_1 \subseteq \mathbf{U} \cap \mathcal{T} \times \mathcal{T}$ be the set of underdemanded pair types or $A-B$ that are mutually blood-type compatible with the types in $\mathbf{P}_1$: that is, $\mathbf{V}_1 = \mathbf{C}_{\mathbf{P}_1}(\mathbf{U} \cap \mathcal{T} \times \mathcal{T})$. As $AB - O \in \mathbf{P}_1$, we have $V_1 = \mathbf{U} \cap \mathcal{T} \times \mathcal{T} = \{O-A, O-B, O-AB, A-AB, A-AB, B-AB, A-B\}$ (see Figure 6).

All pairs belonging to types in $\mathbf{V}_1$ should be matched immediately with probability 1 at σ, as otherwise, one pair that does not get matched immediately with positive probability can register at a platform where a positive measure of $W_1 - Z_1$ type pairs register at σ. As all $W_1 - Z_1$ pairs are matched immediately with pairs of types in $\mathbf{P}_1$ and this one pair is of measure 0, it would guarantee to be matched immediately as well.

Pairs of types in $\mathbf{U} \cap \mathcal{T} \times \mathcal{T}$ can only be matched with pairs of types in $\mathbf{O} \cap \mathcal{T} \times \mathcal{T}$. We have a measure of $e_1 = \sum_{X-Y \in \mathbf{U} \cap \mathcal{T} \times \mathcal{T} \setminus \{A-B\}} [\theta + \rho(1-\theta)] p_Y \lambda \pi_X + p_B \lambda \pi_A$ underdemanded and $A - B$ pairs being matched through exchange at every moment in time at σ. However, the total measure of overdemanded and $B - A$ pairs arriving at each moment is only $e_2 = \sum_{Y-X:\mathbf{O} \cap \mathcal{T} \times \mathcal{T} \setminus \{B-A\}} [\theta + \rho(1-\theta)] p_X \lambda \pi_Y + p_A \lambda \pi_B$. By Assumptions 2 and 4, $e_2 > e_1$. Hence, a positive measure of underdemanded pairs should wait under any feasible exchange scheme, contradicting the fact that all pairs of types in $\mathbf{U} \cap \mathcal{T} \times \mathcal{T}$ are matched immediately.

Thus, we showed that all pairs of types in $\mathbf{O} \cap \mathcal{T} \times \mathcal{T}$ matched to pairs of types in $\mathbf{U} \cap \mathcal{T} \times \mathcal{T}$ under equilibrium. As any positive measure of self-demanded types can be matched with each other at any platform, equilibrium σ maximizes the total measure of pairs being matched through exchange, and hence, through deceased donation, as well.

As ρ goes up, the measure of pairs of types in $\mathbf{O} \cap \mathcal{T} \times \mathcal{T}$ goes up. Hence, more underdemanded and overdemanded pairs are matched at any equilibrium, while the measure of reciprocally demanded pairs matched in exchange stays constant. On the other hand, if patients without live donors are pooled with some types in $\mathbf{U} \cap \mathcal{T} \times \mathcal{T}$ before ρ goes up, the measure of such patients being matched also increases. ■

Proof of Theorem 9. By Proposition 1, as the maximal measure of pairs are matched at pure Nash equilibria in undominated strategies, the worst equilibrium in undominated strategies for P_0 is the best equilibrium for other platforms. The measure of pairs matched in every moment in time through exchange in any pure strategy equilibrium in undominated strategies is given as

$$\sum_{X-X \in \mathcal{T} \times \mathcal{T}} \theta p_X \lambda \pi_X + 2 \sum_{X-Y \in \mathbf{O} \setminus \{B-A\}} [\theta + \rho(1-\theta)] p_Y \lambda \pi_X + 2 p_A \lambda \pi_B$$

where $[\rho(1-\theta)] p_Y \lambda \pi_X$ is the measure of compatible pairs participating in exchange, which also save the same amount of the underdemanded or $A - B$ pairs through exchange. In the worst equilibrium for P_0 only the compatible pairs participate in exchange at P_0 among all overdemanded and $B - A$

pairs. Hence no $B-A$ pair participates at P_0. Among the underdemanded pairs and $A-B$ type pairs, the participation at P_0 is such that exactly $\rho(1-\theta)]p_Y\lambda\pi_X$ survive and get matched with the compatible pairs. Contrary to the claim, suppose that as ρ increases, the participation of overall pairs decreases or stays the same at P_0 under the worst equilibrium. Then, more compatible pairs are available of each (feasible) type, the waiting time for the underdemanded pairs registered at P_0 decreases while the waiting time at other programs for the same types stay the same of increases. This leads to a contradiction.

Now, if

$$\rho \;>\; \frac{\sum_{X-Y\in\mathbf{O}\setminus\{B-A\}}\theta p_Y\pi_X + p_A\pi_B}{\sum_{X-Y\in\mathbf{O}\setminus\{B-A\}}(1-\theta)p_Y\pi_X},$$

then

$$2\sum_{X-Y\in\mathbf{O}\setminus\{B-A\}}\rho(1-\theta)p_Y\lambda\pi_X \;>\; 2\sum_{X-Y\in\mathbf{O}\setminus\{B-A\}}\theta p_Y\lambda\pi_X + 2p_A\lambda\pi_B,$$

where the left-hand side denotes the least measure of pairs matched at P_0 at each point in time at any equilibrium with undominated pure strategies and the right-hand side denotes the maximum total measure of efficiency critical pairs matched outside of P_0 at an equilibrium.

Therefore, more pairs of types in $\mathbf{O}$ register at P_0 at any pure undominated equilibrium, as half of the above measures belong to pairs of types in $\mathbf{O}$ registering at P_0 (left-hand side) and other platforms in total (right-hand side), respectively. To the contrary to the claim suppose less or equal measure of pairs of types in U register at P_0 than as more pairs of types in $\mathbf{U}$ are matched within P_0 then at all other platforms combined, some pair type $X-Y\in\mathbf{U}$ will have a lower waiting time at P_0 than some other platform outside, leading to a contradiction to Lemma 7. ■

B Appendix: ABO-Identical Exchange and Extrapolations Using US Kidney Transplant Data

Inflow rates of patients without live donors are not given in the data. Live donation rates include some direct live-donor transplants as well as some exchanges, which are not much wide spread, yet. A number of them is missing as some patients receive their compatible live donors kidney without ever being listed in the deceased-donor queue. We use blood-type distribution for deceased donors reported in Table 3 for both live donors and $\theta = 11\%$ as the probability of tissue rejection. For an O blood type patient with a live donor, the probability of having his donor compatible is $p_O^l = (1-\theta)p_O = 0.89\times 0.48 = 0.427$. If all exchanges involved patients arriving in 2011, then there were $2,272-442 = 1830$ direct live-donor transplants for O patients. On the other hand, if all

	Blood Types				Total
	O	A	B	AB	
Additions to the Queue	16,240	11,237	4,832	1,260	33,568
Live-Donor Recipients at the Queue	2,272	1,998	674	209	5,153
- through exchange	199	167	58	18	442
Deceased-Donor Organs	5,290	4,026	1,319	392	11,026
	Estimates				
p_X^l	42.7%	75.2%	53.3%	89%	
$r_X^d = \delta_X/\pi_X^d$	46.5-48.5%	45.7-46.9%	35.9-37.0%	37.5-38.2%	

Table 3: Arrivals to and transplants from the kidney deceased-donor queue in 2011 in the US. Data obtained from national data at http://www.http://optn.transplant.hrsa.gov on 02/25/2013. Deceased donor numbers reported in data for each blood type and the empirical fact that 1.48 kidneys are harvested from each deceased donor are used to find the number of deceased-donor kidneys available.

exchanges involved patients who arrived before 2011 or never listed in the deceased-donor queue , then there were 2,272 direct live-donor transplants for O patients. These two boundary numbers respond to 42.7% of all O patients with O donors. These in turn lead to between 4,855 and 5,321 O patients with live donors being added to the deceased-donor queue in 2011. Hence, the number of patients without live donors including reentries arriving in 2011 is given between $16,260 - 5,321 = 10,919$ and $16,260 - 4,855 = 11,385$. On the other hand, only 5,290 O deceased-donors arrived. These give us a lower bound on the rate $r_O^d = \delta_O/\pi_O^d = 0.465 - 0.485$. Similar calculations yield the r^d rates reported in the Table 3 for other blood types.

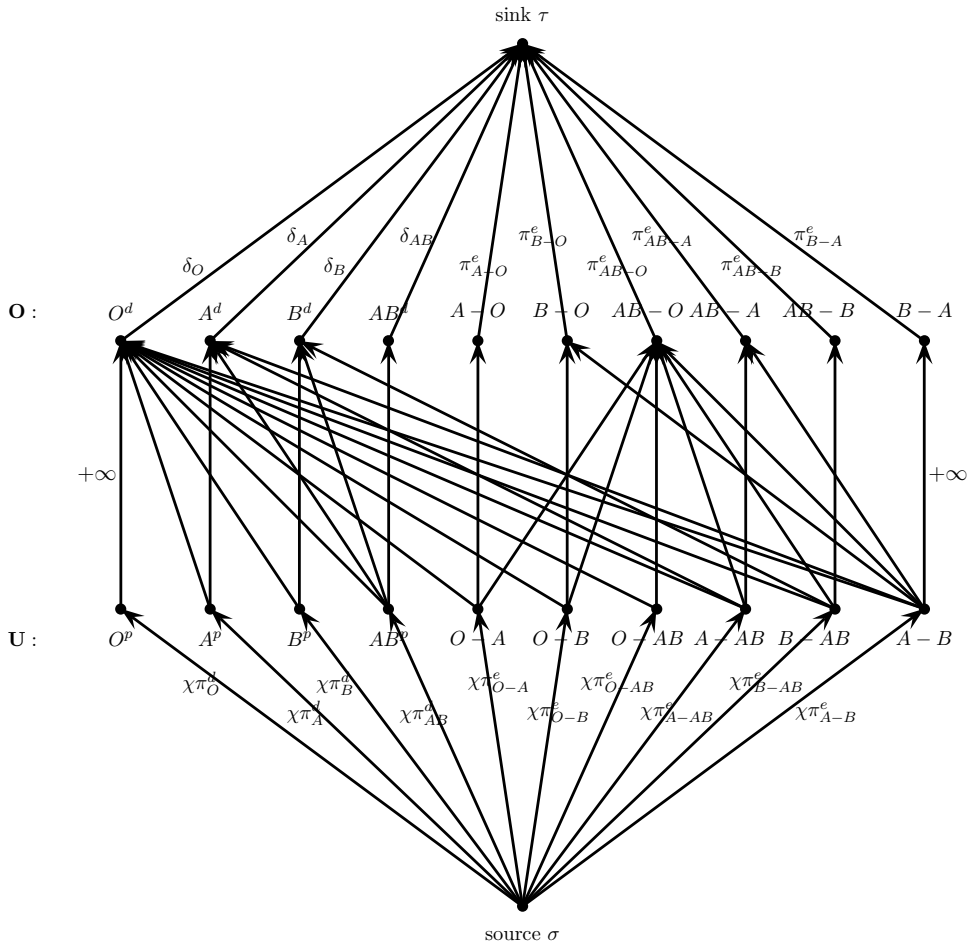

Figure 9: The χ–parametric flow network for the proof of Theorem 6, using the exchange and deceased donation feasibility graph in Figure 6. In order to prevent confusion, the nodes representing patients without live donors (i.e., blood types in **U** as defined in Equation 12) are superscripted by p and the nodes representing deceased donors (i.e., blood types in **O** as defined in Equation 11) are superscripted by d.

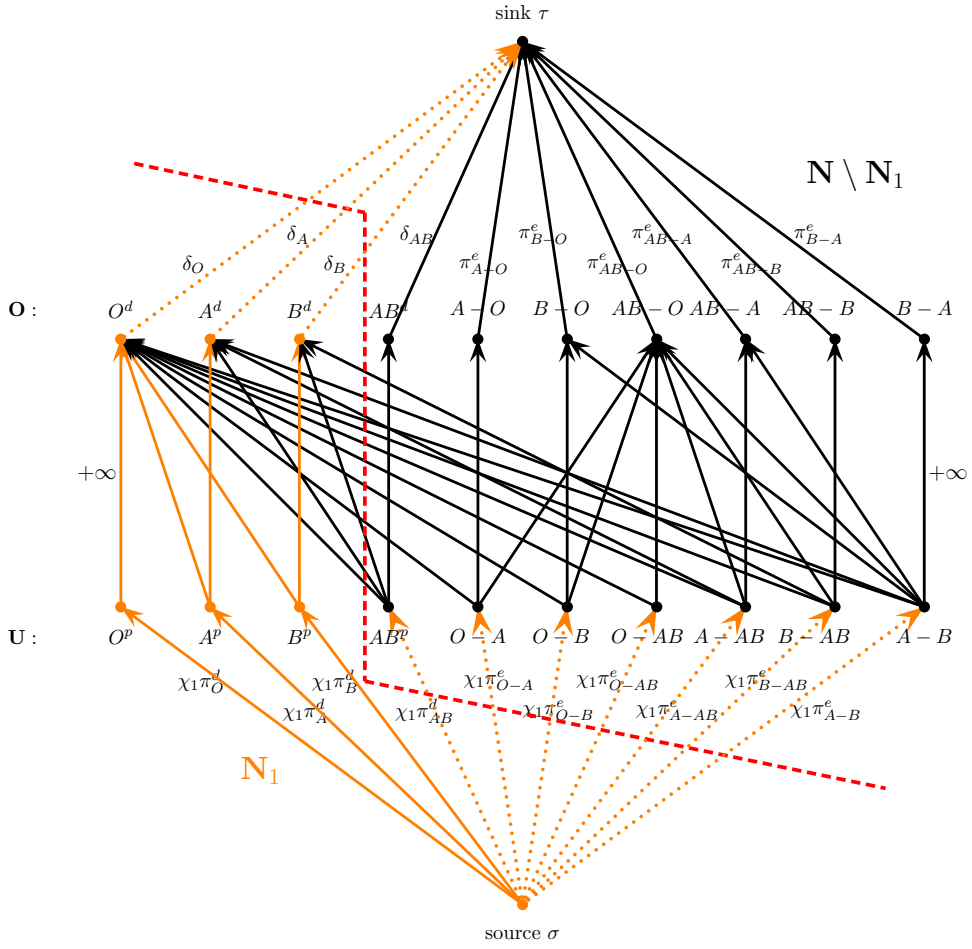

Figure 10: Example of a possible minimum cut $\mathbf{N}_1$ at some $\chi = \chi_1$ with $\mathbf{N}_1 \cap \mathbf{U} = \mathbf{V}_1 = \{O^p, A^p, B^p\}$. Hence, $\mathbf{N}_1 \cap \mathbf{O} = \mathbf{P}_1 = \mathbf{C}_{\mathbf{V}_1}(\mathbf{O}) = \{O^d, A^d, B^d\}$. The edges from $\mathbf{N}_1$ to $\mathbf{N} \setminus \mathbf{N}_1$ are denoted by dotted pointed lines. This cut's total capacity is $q^{\chi_1}(\mathbf{N}_1) = \delta_A + \delta_B + \delta_{AB} + \chi_1(\pi^d_{AB} + \pi^e_{O-A} + \pi^e_{O-B} + \pi^e_{O-AB} + \pi^e_{A-AB} + \pi^e_{B-AB} + \pi^e_{A-B})$.