Market Failure in Kidney Exchange*†

Nikhil Agarwal, Itai Ashlagi, Eduardo Azevedo,
Clayton Featherstone and Ömer Karaduman‡

This version: 2018-02-11
First version: September 20, 2017
Preliminary: please do not cite without permission

Abstract

Kidney exchange facilitates over 800 transplants per year in the United States. We show that, despite this success, kidney exchange markets suffer from market failures that cause the loss of hundreds of transplants per year, but could be addressed with simple policies. Our argument has three parts. First, we document that the market is highly fragmented, with 65% of transactions happening in small platforms, often within hospitals, as opposed to in large, national platforms. Moreover, there is smoking-gun evidence these small platforms often perform inefficient matches. Second, we propose a simple model, showing that inefficiency arises for two reasons: hospitals do not fully internalize their patients' benefits from participation, and current mechanisms do not give hospitals adequate incentives. Third, we estimate the transplant production function, the key primitive of the model, to quantify the inefficiency and design practical mechanisms. Our estimates show that hospitals' production scale is too small to match patients efficiently. Consequently, market fragmentation leads to a deadweight loss of at least 200 transplants per year. Simple optimal mechanisms and policies that encourage participation are likely to have large positive effects, but eliminating this inefficiency requires a combined approach.

JEL: D47, D42, L11
Keywords: Kidney Exchange

*We thank Alex Garza, Sarah Taranto, and Jennifer Wainwright of UNOS; Jonathan Kopke and Michael Rees of APD; Cathi Murphey and Adam Bingaman of Methodist at San Antonio; and Garet Hill of NKR for their expertise and access to data. We thank Glenn Ellison, Parag Pathak, Al Roth, Jean Tirole, Utku Ünver, and seminar participants for helpful discussions. Abigail Ostriker provided excellent research assistance. Azevedo acknowledges support from Wharton's Dean's Research Fund, Agarwal from the NSF (SES-1729090), and Ashlagi from the NSF (SES-1254768).

†Disclaimer: The data reported here have been supplied by UNOS as the contractor for the Organ Procurement and Transplantation Network (OPTN). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the OPTN or the U.S. Government.

‡Agarwal: MIT and NBER, agarwaln@mit.edu. Ashlagi: Stanford University, iashlagi@stanford.edu. Azevedo: Wharton, eazevedo@wharton.upenn.edu. Featherstone: Wharton, claytonf@wharton.upenn.edu. Karaduman: MIT, omerk@mit.edu.

1 Introduction

Kidney exchange markets enable approximately 800 transplants per year. Patients in this market need a kidney transplant and most come with a living, but incompatible, donor. Transplants are organized using patient-donor swaps, and chains initiated by altruistic donors. The importance of this market has grown because of a shortage of organs from deceased donors, and because using monetary compensation to acquire organs is prohibited.[1] The economic value of a transplant is estimated at more than a million dollars; kidney transplants save lives, reduce healthcare costs, and saves taxpayer funds.[2]

We show that, despite significant success at organizing transplants, kidney exchange markets suffer from two serious and fixable market failures that together result in hundreds of lost transplants per year. Indeed, our descriptive evidence suggests that the market is fragmented, operates inefficiently and shows signs of agency problems. We use price-theoretic arguments to explain that this inefficiency arises from two standard market failures. Each failure corresponds to a specific problem with current institutions. Both problems yield suboptimal incentives for hospitals, the key decision makers (Roth et al., 2005; Ashlagi and Roth, 2014; Rees et al., 2009). First, large, national kidney exchange platforms use inefficient mechanisms. Due to biological compatibility, there is substantial variation in the additional transplants that result from different types patients and donors being submitted to the national platform. But, current mechanisms ignore these differences in social value. This limitation in the design pushes hospitals to match patients with high social value outside the platform, even when it is inefficient. Second, there is scope for agency problems: hospitals face most of the costs of participating in the national platforms, but receive only a fraction of the benefits. We show that fixing these two problems can result in hundreds of more transplants per year.

This argument is developed in three parts. First, we use administrative datasets and institutional detail to diagnose and document the causes and consequences of these market failures. Second, we argue that, although kidney exchange markets look different from other markets, these inefficiencies are caused by standard market failures. To analyze the market, it is useful to think about the role of a kidney exchange platform as a producer of transplants that takes patients and donors supplied by hospitals as inputs. This re-framing, based on neoclassical producer theory, allows us to use classic price theory arguments that have clear

[1] There are over 97,000 patients currently waiting for a kidney from a deceased donor, but less than a fifth are expected to be transplanted in the next year. Becker and Elias (2007) argue that this waitlist could be completely eliminated if there was monetary compensation for live donors, and advocate for the creation of this market. However, this type of transaction is widely panned by bioethicists, and almost all countries forbid monetary compensation for organs. The National Organ Transplantation Act (NOTA) prohibits the use of compensating donors in exchange for acquiring organs in the United States. The motivation for kidney exchange is to use donor swaps to help patients who need an organ in an ethically and legally acceptable way (Roth, 2007).

[2] Transplantation roughly doubles the life expectancy of patients with end-stage kidney disease, and is cheaper than the alternative treatment of dialysis. Medicare provides nearly universal coverage, irrespective of age, for patients with End-Stage Renal Disease (ESRD). The ESRD program comprises of about 7% of Medicare's annual budget (see USRDS, United States Renal Data System, 2016). The cost savings of transplantation relative to dialysis alone have been estimated to be over $270,000. See Section 2.

policy implications. Third, we estimate the key primitive of the model, a transplant production function, using data from the largest kidney exchange platform. The primitive allows us to estimate the magnitude of the inefficiencies, design practical alternative mechanisms, and suggest appropriate policy responses.

In the first part, we document key facts about the market to motivate our model. We merge administrative data on the universe of transplants in the United States and proprietary data from the largest kidney exchange platform. The main dataset covers transplants from 2008 to 2014. The data reveal several signs of inefficiency. Rather than most transactions taking place at a few large platforms, the market is highly fragmented. Approximately 65% of kidney exchanges involve sets of patients and donors that belong to the same hospital. Crucially, we find smoking-gun evidence that many within-hospital exchanges are inefficient from a social perspective. An efficient market would use donors that are easy-to-match only to transplant patients that are hard-to-match. We show that small within-hospital exchanges often violate this simple rule, but large national platforms do not.

We also find evidence consistent with inefficient hospital incentives causing these problems. Many hospitals do not participate in the large national platform. Consistent with barriers due to fixed costs of participating in kidney exchange, smaller hospitals are less likely to participate. Previous surveys and a study of the payment structure in this market suggests that costs of participating can be a significant fraction of hospital revenue from kidney exchange. However, these costs are small relative to the social benefit of transplants. Even when hospitals do participate, the typical hospital does not conduct all kidney exchanges through the national platform. Instead, most hospitals continue to operate as a competing small kidney exchange platform. They tend to submit particularly hard-to-match patients and donors to the national platform and transplant their easy-to-match types separately.

These facts lead us to our model in the second part. The key insight underlying our approach is that although kidney exchange markets do not directly use monetary incentives, it is useful to analyze them using neoclassical producer theory. The platform produces a final good (transplants) from intermediate goods (submissions of patients and donors) supplied by a competitive fringe (hospitals), according to a production function.

This re-framing is based on three key institutional features. First, hospitals supply patients and donors to a platform voluntarily and are not forced to participate. Instead of submitting a patient-donor pair to a large platform, a hospital can transplant it with other pairs that it is treating. Thus, kidney exchange platforms must reward hospitals with transplants to procure necessary inputs (patients and donors).[3] Second, due to biological compatibility constraints, some types of patients and donors enable more transplants than others, i.e. some inputs are more productive than others. For example, blood type O donors can donate to patients of any blood type. This fact makes them both scarce and valuable when organizing kidney

[3]The importance of providing incentives to participate differentiates kidney exchange from some other market design applications with a central planning authority such as school choice and labor market clearinghouses (Roth and Peranson, 1999; Pathak, 2017). In contrast, providing socially efficient incentives for participation is a central issue in the economics of platforms (Rochet and Tirole, 2003; Armstrong, 2006; Weyl, 2010).

exchanges.[4] Because hospitals can perform exchanges within their set of patients and donors, a platform may have to increase the rewards for submitting valuable types in order to procure them. Third, due to biological compatibility constraints, some types of patients and donors are in abundance and the platform has to ration transplants within this set.[5] A transplant can therefore be transferred to a hospital by selecting its patient instead of another hospital's. This fact makes transplants a natural numeraire good, even though monetary transfers are not allowed.

The key primitive of the model is the transplant production function. Its shape determines important features of the economics of kidney exchange. The returns to scale determine how efficient it is to match patients in large exchanges. The marginal products of different types of submissions determine whether some submissions produce more transplants than others. We consider a steady-state version of this model, where all units are measured in flows. The steady-state interpretation is not essential, but appropriate for the empirical analysis in the third part.

The costs of supplying patients and donors to the platform are governed by two separate functions. Both are in units of the numeraire good, transplants per year. The first takes a revealed preference approach to hospital costs and yields hospital welfare. The second considers social costs, i.e. transplants within the hospitals that are lost because a patient or a donor is submitted to the platform. In the empirical analysis, we take social welfare to be the total flow of transplants nationwide because the social value of a transplant is much larger than the costs of participating in a platform. The two welfare measures coincide if hospitals maximize the number of their own patients that are transplanted. In that case, hospital welfare equals the total number of transplants performed, both in the national platform and within the hospitals. But we allow these two measure to be different, to allow for agency problems between hospitals and their patients, or behavioral wedges.[6]

Our main result, Theorem 1, formally shows that market failure can result from two sources of inefficiency. The first source is based on an inefficient mechanism. Most kidney exchange platforms run optimal matching algorithms to select which patients are matched. When a hospital submits a patient to the exchange, the hospital is rewarded according to the *probability* with which the patient is matched. In sharp contrast, Theorem 1 shows that, to maximize hospital welfare, hospitals should be rewarded based on the *marginal product* of their submissions. Because existing platforms do not reward hospitals for the social value of their submissions, even a hospital that maximizes the number of its own patients that are

[4] In a simple model, Roth et al. (2007) show that certain "over-demanded" types enable two additional transplants at a platform while others add no value to the pool. Our calculations will generalize this idea by computing the additional number of transplants generated by a rich classification of types using data from the largest exchange.

[5] A patient with blood type O can only receive a kidney from a blood type O donor. This feature creates an abundance of O patients with donors of other blood types. Similarly, patients that do not have a related living donor are also in abundance. Many patients and donors in such submissions will inevitably remain unmatched. See Roth et al. (2007) for arguments based on a limit economy.

[6] The two welfare measures can be thought of as decision utility, which we infer from hospital behavior, and experience utility, which we observe by counting transplants. We allow the two measures to differ arbitrarily as in general theoretical results in behavioral public finance (Farhi and Gabaix, 2017).

transplanted has to perform socially inefficient matches. Hospital are in a difficult position; they must decide between helping its patients or performing socially efficient matches. This problem can be ameliorated using simple, dynamic mechanisms that reward hospitals according to marginal products. One example is a simple point systems that picks transplants based on point balances and awards points to hospitals that contribute more to the platform.[7]

The second source of inefficiency is when there are agency problems, in the sense that hospital welfare differs from social welfare. These problems arise if hospitals do not maximize the number of their own patients that are transplanted. It creates inefficiency, for example, if a hospital performs too many internal matches to avoid administrative costs of kidney exchange, even though these matches happen at an inefficiently small scale.

Theorem 1 shows that there is no inefficiency if neither of these two sources are present. This decomposition of the sources of market failure is consistent with long-standing concerns of surgeons, insurers, platforms, and researchers, and even with recent policy changes. Roth et al. (2005) and Ashlagi and Roth (2014) have long recognized that hospitals may have incentives to match patients internally in static models. Surgeons and insurers have noted that it may be in the interest of insurers to subsidize exchanges, and proposed that they do so (Rees et al., 2012). Our analysis pinpoints the sources of market failure to specific wedges, which has substantial implications. For example, Ashlagi and Roth (2014) propose a stylized mechanism that, in a static model, guarantees that hospitals can always match as many patients as they could match internally. Our model clarifies that this kind of mechanism is inefficient in an real-life, dynamic platform, where hospitals have incentives to withhold submissions to use them later. The problem is that this kind of mechanism does not set hospital rewards to be equal to marginal products, which is the relevant target.

The third part of our argument uses the data to quantify the importance of these issues and to design responses. We estimate the platform's transplant production function, which is the key primitive in the model. Our data and setting is particularly well-suited for estimating this production function for two reasons. First, we have detailed administrative data from the largest kidney exchange platform, the National Kidney Registry (NKR), including data on biological compatibility. Second, we have detailed information about the matching algorithms and operational procedures used in kidney exchange platforms. We build a detailed empirical simulation model of a kidney exchange platform using this information.[8] Most parameters are known, and only a handful of parameters need to be estimated or calibrated. The model fits the very data and allows us to estimate the flow of transplants that would be produced given a flow of submissions.

[7]Because our simplified steady-state model is not dynamic, it does not pin down the game form of an optimal dynamic mechanism. Existing theoretical work shows that optimal dynamic mechanisms in related settings are complicated. However, simple token mechanisms that keep track of point balances are both simple and highly efficient, making these mechanisms good candidates for practical implementation. We discuss these theoretical and practical issues in Section 6.

[8]There is a long tradition in economics of estimating production functions. Unfortunately, the standard methods are not suited for our empirical setting, because they rely on low-dimensional production functions, typically Cobb-Douglas with a few types of inputs, and on data on observed inputs and outputs for many firms (Marschak and Andrews, 1944; Olley and Pakes, 1996).

The estimated production function yields three sets of results. First, we measure the returns to scale of the production function, and estimate the inefficiency from market fragmentation. We find that the largest kidney exchange platform is well above the minimum efficient scale.[9] At the same time, almost all single-hospital platforms are far below the efficient scale. This difference in productivity suggests considerable inefficiency due to fragmentation because 65% of the transplants are organized within hospital. We then estimate the gains from moving all the production to the efficient scale,[10] assuming that single-hospital platforms have the same production function as the NKR. Because this exercise involves an extrapolation, we perform extensive robustness analyses to present a range of estimates. Our most conservative estimates suggest that at least 200 transplants are lost each year due to market fragmentation. More central estimates place this number at more than 400. Thus, consistent with the descriptive evidence and the shape of the production function, fragmentation has a large efficiency cost.

Second, we assess the current mechanism and design optimal mechanisms.[11] Recall that Theorem 1 shows that optimal mechanisms reward submissions approximately according to marginal products, while current mechanisms reward submissions according to probabilities of matching. Motivated by this result, we calculate the marginal products and the probabilities of matching for each type of submission using the estimated production function. There is substantial variation in marginal products. Moreover, marginal products are considerably different from the probabilities of matching, which implies that existing mechanisms are far from optimal. That is, we measure a large wedge between current and socially optimal rewards pinpointed by theorem 1. We also use regression tree analysis to show that marginal products are well predicted by a simple set of characteristics of submissions. This observation suggests there are simple mechanisms that give hospitals points based on submissions or transplantation that are approximately optimal.

Third, we study the importance of the two sources of market failure. In our price-theoretic approach, the loss in hospital welfare due to the inefficient mechanism is given by a standard Harberger triangle. The area of the triangle depends on the wedge between current and optimal rewards, on the marginal products of various submissions, and on the elasticity of supply from hospitals. We have estimated the wedges and the marginal products, but our data does not have enough information to credibly estimate the elasticity of supply. Therefore, we calculate this deadweight loss under a broad range of assumptions about elasticities. Except under extreme assumptions, the deadweight loss is significant, but considerably lower than the total inefficiency due to market fragmentation. Therefore, under most reasonable assumptions, both the current mechanism and agency problems cause significant inefficiency in the market. Taken together, these results suggest that improving the design of the mech-

[9]Because returns to scale are roughly constant at a large enough scale, gains due to scale economies alone from merging the major kidney exchange platforms are likely small. This observation on the scale economies can also explains why, instead of tipping to a single platform, multiple large national exchanges co-exist (Ellison and Fudenberg, 2003).

[10]This is in the spirit of misallocation analyses in the productivity literature (Hsieh and Klenow, 2009).

[11]This is in the spirit of the sufficient statistics approach in the public finance literature (Dixit and Sandmo, 1977; Saez, 2001; Chetty, 2009). In particular, our analysis suffers from the standard caveat that marginal products depend on the composition of submissions, which is endogenous.

anism and policies for encouraging hospital participation in the national platforms are both likely to bring sizable benefits.

After presenting our main argument, we discuss important limitations of our analysis and some extensions. First, designing a mechanism raises theoretical issues when selecting a dynamic game form, and practical issues when using sufficient statistics in the real world. We explain the existing theory and open problems related to these issues. Second, while our main analysis considers mechanisms that maximize hospital welfare. Third, we consider imperfect competition, and how the abuse of market power by oligopolistic platforms can reduce efficiency. Fourth, we consider potential limitations due to the our assumptions. Fifth, we discuss recent changes in kidney exchange markets that have attempted to address the two market failures that we describe. Finally, we discuss the relationship with the literature.

Our paper is organized as follows. Section 2 describes institutional details and data sources. The three parts of our argument are developed in Sections 3, 4, and 5, respectively. Section 6 discusses the limitations and extensions, and Section 7 concludes.

2 Background and Data

2.1 Basics of Kidney Exchange

This section describes the basics of kidney transplantation and kidney exchange, and can be skipped by readers who are familiar with this literature.

End-Stage Renal Disease (ESRD) afflicts more than half a million Americans. The disease is almost universally covered by Medicare, including for patients under the age of 65. The Medicare ESRD program accounts for 7% of its budget, mostly spent on patients undergoing dialysis (USRDS, United States Renal Data System, 2016). The preferred treatment for ESRD patients is transplantation, which increases the quality and length of life by several years and is cheaper than dialysis. Transplantation saves several hundred thousand dollars per Medicare beneficiary, and saves more for privately insured patients (Wolfe et al., 1999; Held et al., 2016; Irwin et al., 2012). Moreover, the health risks to suitably screened living donors are small. Taken together, these facts point to living donor kidney transplants having a large economic value. Held et al. (2016) places the economic value of an average kidney transplant at $1.1 million.[12]

Unfortunately, there is a severe shortage of organs for transplantation. Each year, approximately 13,000 patients are transplanted using organs from deceased donors and another

[12] Held et al. (2016) conduct a comprehensive cost-benefit analysis with sensitivity to a range of assumptions. They place a value of a year of perfect health at $200,000 and adjust for differences between the quality of life between dialysis and a transplant. The costs include differences in expected medical costs incurred over the lifetime of a dialyzed patient and a transplanted patient, discounted at the rate of 3% per year. The cost savings on dialysis alone are significant. In 2014, Medicare paid $87,638 per year per dialysis patient, but only $32,586 in post-transplant costs per year per patient (USRDS, United States Renal Data System, 2016, Chapters 7 and 11).

5,500 from living donors. Demand far outstrips this supply with approximately 35,000 patients added to the deceased donor kidney waitlist in recent years. The shortage has resulted the kidney waitlist growing to almost 100,000 patients, and about 8,000 patients on the list dying or being categorized as too sick to transplant in each of the previous few years.[13] Monetary compensation cannot be used to address this shortage because of ethical and legal reasons, and compensation is forbidden in in almost every country (Becker and Elias, 2007), including the US.[14]

Kidney exchange is an innovative way to ameliorate this shortage.[15] This form of transplantation serves patients who have a willing live donor, but who is not biologically compatible. Such patients can swap donors with others in the same situation, enabling transplants for many patients. These swaps are organized by **kidney exchange platforms** that match patients and donors registered with them. The platforms receive three types of **submissions**. The most common type is a **pair**, consisting of a patient and her willing but incompatible donor. The second type is an **altruistic donor**, who is willing to donate a kidney to a stranger. Finally, there are some **unpaired patients**, who do not have a willing donor.

The platform organizes transplants using two types of kidney exchange transactions. The first, called a **cycle**, involves a set of pairs. The donor from one of the pairs is transplanted to the patient in the next pair until the cycle is closed. All transplants are carried out simultaneously to reduce the risk that a pair donates a kidney without also receiving one. Cycles are usually limited to at most three pairs due to logistical constraints.[16] The second type, called a **chain**, is initiated when an altruistic donor donates to a patient in an incompatible pair. The donor from this pair can then continue the chain by donating to the next pair and so on until the chain terminates with an unpaired patient. Chains can be very long in principle because the transplants do not have to be performed simultaneously.[17] However, our data from the National Kidney Registry indicates that most chains involve four to five transplants. Initially, cycles were the most common type of transaction. But chains became more important over time, and account for about 90% of the transplants in our data from a major platform.[18]

There are two types of biological compatibility constraints on kidney transplants: **ABO blood-type** and **tissue-type** compatibility (Danovitch, 2009). A donor is blood-type incompatible with a patient if the donor has a blood antigen that the patient lacks. There are two blood antigens, known as A and B. An individual's blood type is either A or B if she has

[13] Statistics taken from https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/ (accessed December 21, 2017).

[14] The National Organ Transplant Act (NOTA 1984) makes it illegal to obtain organs for transplantation by compensating donors.

[15] The first kidney exchange was in Korea (Kwak et al., 1999).

[16] In the early days kidney exchange was limited to pairwise exchanges (cycles with only two pairs). Cycles of length three were later introduced to realize the gains from this expansion (Saidman et al., 2006).

[17] See Rees et al. (2009) for the first long non-simultaneous chain

[18] Chains were initially controversial. While some researchers and practicioners pointed out the value of non-simultaneous chains (Roth et al., 2006), others were opposed (Gentry et al., 2009). Eventually, it became clear that chains considerably increase the efficiency with which patients can be matched (Ashlagi et al., 2012), and they became widely adopted.

one of these antigens, AB if she has both, and O if she has neither. A donor is tissue-type incompatible with a patient if the donor has certain cell-surface proteins (antigens) that the patient has an immune response to.[19] Patients differ considerably in their level of sensitization. The most common measure of sensitization is the **Panel Reactive Antibody (PRA)** score. A patient's PRA is between 0 and 100, and denotes the percentage of a representative population of donors that a patient is tissue-type incompatible with.

2.2 Key Institutional Features and the Economics of Kidney Exchange

There are three key institutional features that drive outcomes, and will play a key role in our analysis.

First, kidney exchange takes place both in large, national platforms, and within indvidual hospitals. There are three major national platforms currently operating in the United States: the National Kidney Registry (NKR), which is the largest, the Alliance for Paired Kidney Donation (APD), and the United Network for Organ Sharing (UNOS) KDP Pilot Program.[20] These large platforms match patients using optimization software that maximizes a weighted number of transplants. They differ on the exact algorithm and on operational details.[21] Besides these major platforms, there are small regional platforms, and there are single hospitals that also organize kidney exchanges. As we will see in the next section, most of the hospitals that participate in large national platforms also match patients outside the platform.

Moreover, hospitals are not forced to participate in platforms. Platforms effectively reward hospitals with transplants in order to receive submissions (Roth et al. 2005; Ashlagi and Roth 2014) as hospitals perform the transplants on the patients that they submit to a platform. Rewards can also be explicit. For example, most platforms reward hospitals that submit altruistic donors by matching one of their unpaired patients.

Second, there is substantial variation in the social value of different submissions due to biological compatibility. One reason for this variation is blood type compatibility, as articulated

[19] The immune system recognizes foreign cells based on certain cell-surface proteins, known as antigens. The organism has antibodies that bind to these antigens, tagging foreing cells which are then attacked. Hence, if we put a cell with an antigen in the body of a person who has antibodies for that antigen, the immune system will attack it. Each donor has up to 6 possible human leukoctye antigen (HLA) proteins out of a list of hundreds. Similarly, a recipient has a list of antibodies to some, possibly large, subset of the HLA antigens. If the recipient has an antibody to one of the donor kidney's antigens, the recipient's immune system will attack the kidney, leading to rejection. A recipient is tissue-type compatible with a donor kidney if she has no antibodies corresponding the antigens of the donor kidney *(Danovitch, 2009)*. Note that a transplant between certain incompatible patient-donor has become possible due to development of desensitization technologies (Orandi et al., 2014).

[20] Historically, live donation mostly involved single hospitals and small platforms, and donors with an immediate relationship to the patient. Over time, kidney exchange grew, culminating in the passage of the Charlie W. Norwood Living Organ Donation Act in 2007. The Act explicitly states that paired kidney donation is not forbidden by NOTA's provision outlawing the transfer of an organ in exchange for "valuable consideration." This legal clarification allowed for the expansion of kidney exchange platforms.

[21] See Abraham et al. (2007); Ashlagi et al. (2016); Anderson et al. (2014); Dickerson et al. (2012).

in an important result by Roth et al. (2007). They calculated how many patients can be transplanted in a simplified model with only pairs, and a large-market approximation that effectively focuses on only blood type compatibility. To simplify the explanation of their results, assume that the only blood types are O and A.[22] Denote a pair with patient blood type X and donor blood type Y as X-Y, and let $q_{\text{X-Y}}$ be the number of such pairs in a pool. Assume that $q_{\text{A-O}} < q_{\text{O-A}}$, which is the empirically relevant case. Roth et al. (2007) showed that the number $f(\boldsymbol{q})$ of transplants that can be performed is approximately

$$f(\boldsymbol{q}) = 2 \cdot q_{\text{A-O}} + 1 \cdot (q_{\text{A-A}} + q_{\text{O-O}}) + 0 \cdot q_{\text{O-A}}.$$

This result follows because A-A and O-O pairs can be matched with pairs of the same type. Roth et al. (2007) call these pairs **self-demanded**. Self-demanded pairs have a marginal product of 1, in the sense that they generate 1 additional transplant when they join the pool. However, an O-A pair can only be transplanted using a cycle with one of the valuable A-O pairs. Thus, there will be many leftover O-A pairs, and they can only be transplanted if more A-O pairs join the pool. A-O pairs are called **over-demanded**, and have a marginal product of 2. O-A pairs are called **under-demanded**, and have a marginal product of 0. An under-demanded pairs competes with another under-demanded pairs and adds no value to the pool. Roth et al. (2007) showed that this qualitative pattern holds even in a model with all possible blood types. For example, with two-way exchanges and more A-B than B-A pairs, the total number of transplants is

$$\begin{aligned} f(\boldsymbol{q}) =& 2 \cdot (q_{\text{A-O}} + q_{\text{B-O}} + q_{\text{AB-O}} + q_{\text{AB-A}} + q_{\text{AB-B}} + q_{\text{B-A}}) \\ &+ 1 \cdot (q_{\text{A-A}} + q_{\text{B-B}} + q_{\text{O-O}} + q_{\text{AB-AB}}) \\ &+ 0 \cdot (q_{\text{O-A}} + q_{\text{O-B}} + q_{\text{O-AB}} + q_{\text{A-AB}} + q_{\text{B-AB}} + q_{A-B}). \end{aligned} \tag{1}$$

Current platform rules largely ignore this variation in social value of submissions, inducing hospitals to perform socially inefficient internal matches. To illustrate this, consider a hospital with two over-demanded pairs. The hospital could match these two in a pairwise exchange. However, if the hospital submits the pairs to the platform, the hospital receives on average twice the probability that a pair is matched. In the Roth et al. (2007) stylized model, the probability is 1 while our estimates suggest that the probability is closer to 80% because the platform may not necessary use a pair. Therefore, the hospital expects only 1.6 transplants from submitting the two pairs to the platform. This fact pushes the hospital to match its patients outside the platform. In constrast, if the hospital submits the pairs to the platform, the two pairs together generate twice their marginal products. The Roth et al. (2007) stylized model suggests that the platform could generate 4 additional transplants using these pairs. Our estimates using a more realistic empirical model places this number at 3. Either way, matching these two pairs within the hospital is socially inefficient. This inefficiency occurs even though the hospital wants to help its patients. Our theoretical analysis will summarize all the ways in which such inefficiencies can happen, and clarify how platform rules can be redesigned to eliminate this problem.

[22] This restriction is purely for simplicity, as the results in Roth et al. (2007) hold with a richer set of blood types.

An important corollary of Roth et al. (2007)'s results is that transplants are a natural numeraire in a kidney exchange platform. Because hospitals have a large number of under-demanded pairs, it is easy for a platform to transfer transplants from one hospital to another by choosing which under-demanded pairs to match, without compromising efficiency.

The final feature is that hospitals do not necessarily maximize a reasonable utilitarian measure of patient welfare. We refer to any such divergence as an agency problem between hospitals and patients. One major source of agency problems is that hospitals incur most of the costs of kidney exchange, but receive a small fraction of the social benefits. While the social value from a transplant is over a million dollars, hospital revenues are between \$100,000 to \$160,000 per transplant.[23] Variable profits are likely much smaller. Thus, small costs of performing kidney exchange through a platform are insignificant from a social perspective, but can be important for hospitals. Conversations with hospital staff indicate that participation in kidney exchange platforms involves significant logistical and administrative hassle. There are also direct cost arising from biological testing and platform fees.[24] Indeed, previous surveys and interviews have found that both logistical and financial costs are important barriers to participation (Ellison, 2014; American Society of Transplant Surgeons, 2016). Besides this channel, it is also possible that hospitals do not perfectly maximize patient welfare due to behavioral reasons. For example, there is anecdotal evidence of considerable heterogeneity across hospitals in the level of sophistication. Some hospitals report using optimization software to match patients and others manually search for ad hoc matches.

The upshot of this discussion is that it is plausible that hospitals do not perfectly maximize patient welfare. Such agency problems may exist even if hospitals have complex objectives, that take into account social objectives that are different from profits. Our analysis will incorporate this by allowing for the possibility of agency problems, and by using the data to determine whether agency problems are important.

2.3 Data

We draw upon two main, de-identified, data sources: administrative data on the universe of transplants and waitlist registrations in the United States, and proprietary data from the largest kidney exchange platform, the National Kidney Registry (NKR). The data from the NKR contain the information used in the algorithm and to organize kidney exchanges. The data contains rich demographic and biological information on patients and donors. We also observe registration dates, registration hospitals, and which patients are related to which donors. Given the biological information, we can accurately determine the set of biologically feasible transplants. We also have detailed data on transplants performed by NKR. The

[23]See Held et al. (2016); USRDS, United States Renal Data System (2013). This amount includes payments for hiring surgery teams, drugs, equipment and capital.

[24]Platforms require extensive biological testing, which is particularly complicated because donors and patients are in different hospitals. Platforms also charge fees, which are paid by hospitals. NKR charges annual fees of about \$10,000 plus about \$4,000 per transplant. See (National Kidney Registry, 2016) for NKR's fees, and Rees et al. (2012) and Wall et al. (2017) for a broader discussion of the costs of kidney exchange borne by hospitals.

data includes transplant date, the donors and patients involved, and the chain or cycle configuration. We obtained records on all transplants and associated patients and donors in the NKR between 2008 and November 2014. Complete data on all NKR registrations, including those that were not transplanted, is only available from April 2012 to November 2014.

In addition, we use an administratively sourced dataset on organ transplants in the United States. This study used data from the Organ Procurement and Transplantation Network (OPTN). The OPTN data system includes data on all donor, wait-listed candidates, and transplant recipients in the US, submitted by the members of the Organ Procurement and Transplantation Network (OPTN). The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN contractor. The OPTN members include essentially all hospitals, and the data contains a record for each transplant performed in the US, as well as records for a patient's registration status on the deceased donor waitlist. These records contain most of the fields that are in the data from the NKR.[25] For transplants, the OPTN data contains fields that allow us to determine with high accuracy whether a transplant is part of kidney exchange. We restrict attention to all kidney transplants from 2008 to November 2014, as well as any waitlist registration that was active during that window.

We merged the OPTN and NKR data, so that we can identify almost all transplants, and most patients and donors who appear in both datasets. This merge is not trivial, because both datasets are de-identified. We summarize our linking procedure here. Because the OPTN dataset contains the universe of transplants, it is straightforward to find the transplant in the OPTN data that is closest on dimensions of transplant date, transplant center, and patient and donor age, gender, blood type, and HLA type. Ultimately, we were able to match around 95% of transplants conducted through the NKR to a very high degree of certainty.[26] For donors and patients who were never transplanted by the NKR, the matching process is more difficult because it is possible that they won't match to any OPTN data record at all.[27] We deal with this problem by implementing a naïve Bayes classifier, which outputs a probability that a match is genuine, and setting a cutoff (Russell and Norvig, 2003). These procedures create match the closest observations of patients and donors across datasets based on patient and donor age, blood-types, tissue-types and transplant center and date information. More details on this procedure are included in Appendix B.

The main limitation of these datasets is that OPTN provides little information on the patient-donor pairs that are available to a hospital for kidney exchange. We only observe patients and donors that were either transplanted or were submitted to the NKR. For patients and donors that were transplanted through a kidney exchange, we have information on which patient was transplanted with an organ from which donor. However, we do not know the form of exchange, whether it was a two-way or three-way cycle, or a chain, and the related

[25] The standard analysis files from the STAR data do not contain information on acceptable antigens, home centers, and transplant centers. These fields are available from UNOS on request.

[26] In fact, 85% of the matches were a perfect match along the dimensions just described.

[27] Donors only appear in the OPTN data if they ultimately donate an organ, while patients only show up if they receive an organ or if they register for the deceased donor waiting list.

Table 1: Summary Statistic for Kidney Exchange Transplants

	NKR		Non-NKR Across Hospital		Within Hospital	
N	1118		480		2781	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Patient Blood Type						
A	34.7%	(0.48)	37.1%	(0.48)	37.0%	(0.48)
B	19.0%	(0.39)	18.5%	(0.39)	17.1%	(0.38)
AB	5.7%	(0.23)	6.7%	(0.25)	5.4%	(0.23)
O	40.6%	(0.49)	37.7%	(0.49)	40.5%	(0.49)
Donor Blood Type						
A	36.8%	(0.48)	37.5%	(0.48)	33.2%	(0.47)
B	18.2%	(0.39)	16.5%	(0.37)	13.8%	(0.35)
AB	3.9%	(0.19)	5.6%	(0.23)	2.8%	(0.16)
O	41.1%	(0.49)	40.4%	(0.49)	50.2%	(0.50)
Panel Reactive Antibody (PRA) (Sensitization)						
Mean	34.95%	(0.40)	34.77%	(0.39)	17.80%	(0.31)
Fraction >90%	16.4%	(0.37)	15.3%	(0.36)	5.3%	(0.23)
Transplant Ourcomes and Quality Measures						
Donor Age	44.1	(11.8)	44.0	(11.4)	43.2	(11.8)
Donor Body Mass Index (BMI)	26.5	(4.0)	26.7	(4.0)	26.5	(4.3)
Donor Height (cm)	169.4	(9.8)	169.1	(10.1)	169.3	(9.9)
Donor Weight (kg)	76.3	(15.1)	76.8	(14.9)	76.3	(15.2)
Tissue Type Mismatch (0-6)	4.2	(1.3)	4.2	(1.3)	4.4	(1.2)
Mean Days on Dialysis	1026.6	(1088.1)	1061.1	(1134.8)	960.8	(984.7)

Note: Sample of all PKE transplants between January 1, 2008 and December 1, 2014.

patients and donors. In contrast, the NKR data is much richer. Tables 1 and 2 provide basic summary statistics, which we discuss in detail in the next section.

3 Descriptive Evidence

We now use our detailed administrative dataset to describe three key facts: the kidney exchange market is highly fragmented; this fragmentation leads to inefficiency; and there is evidence of broadly defined agency problems between hospitals and patients.

3.1 Fragmentation

We document market fragmentation by showing that a large fraction of transplants are facilitated by individual hospitals, instead of large national platforms. A market coordinated on a single platform would have all transplants facilitated through that platform. In contrast, a completely fragmented market would have no transplants facilitated through a multi-hospital

Table 2: Summary Statistic for Submissions to the NKR

	Altruistic Donors		Pairs		Unpaired Patients	
N	164		1265		501	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Patient Blood Type						
A	---	---	23.8%	(0.43)	51.1%	(0.50)
B	---	---	15.0%	(0.36)	16.0%	(0.37)
AB	---	---	2.6%	(0.16)	19.0%	(0.39)
O	---	---	58.6%	(0.49)	14.0%	(0.35)
Donor Blood Type						
A	44.5%	(0.50)	44.8%	(0.50)	---	---
B	14.0%	(0.35)	18.5%	(0.39)	---	---
AB	3.7%	(0.19)	5.1%	(0.22)	---	---
O	37.8%	(0.49)	31.5%	(0.46)	---	---
Match Power						
Recipient/Pair	---	---	21.6%	(0.21)	43.0%	(0.39)
Donor	27.6%	(0.16)	25.4%	(0.16)	---	---
Panel Reactive Antibody (PRA)	---	---	48.8%	(0.41)	44.4%	(0.45)
Pair Type						
Overdemanded	---	---	13.8%	(0.35)	---	---
Underdemanded	---	---	42.2%	(0.49)	---	---

Note: A pair is overdemanded if the patient is blood-type compatible with the related donor. Underdemanded pairs either are O-patients without O-donors or are AB-donors without AB-patients. Sample of all patients and donors registered in the NKR between April 4, 2012 and December 1, 2014.

platform. In this case, all transplants would be **within hospital** in which each transplanted patient and donor belong to the same hospital.

Therefore, to measure fragmentation, we first identify within hospital transplants by checking whether the hospital associated with a patient transplanted through kidney exchange is the same as the hospital that recovered the organ.[28] From this set, we use our matching procedure to exclude all transplants facilitated by the largest exchange, the NKR. Unforunately, we cannot do the same for the two other national platforms, APD and UNOS KPD. However, these two platforms together accounted for just over 80 transplants during our sample period (Agarwal et al., 2018). Their relatively small size will allow us to obtain an informative lower bound on the extent of market fragmentation.

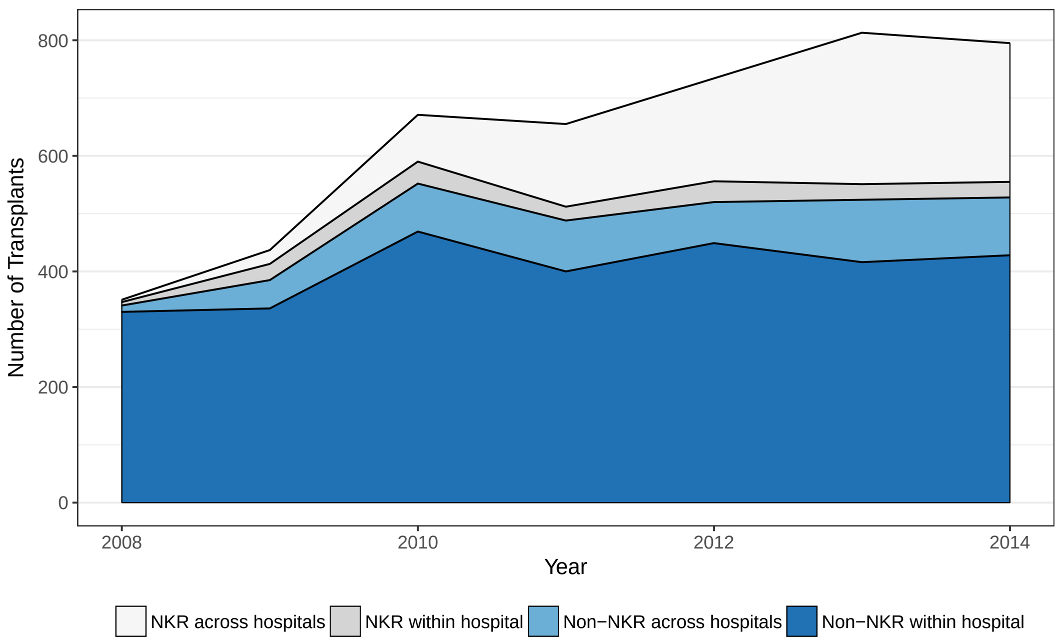


Figure 1: Market fragmentation and trends in kidney exchange

Notes: The figure displays the number of kidney exchange transplants in different categories. NKR or non-NKR classifies a transplant into whether it was facilitated by NKR. Within hospital and across hospital classify a transplant into whether the donor hospital was the same as the patient hospital. Large national platforms perform most of their matches across hospitals. Thus, the number of non-NKR across hospital transplants approximately corresponds to transplants in other exchanges, while the number of non-NKR within hospital approximately corresponds to internal matches performed by hospitals.

Figure 1 shows that the market is highly fragmented. 63% of kidney exchange transplants are non-NKR, within hospital transplants. Figure 1 shows that the vast majority of the transplants facilitated by the NKR are **across hospitals**, involving a living donor and a

[28]The common practice is to transport the organ after recovery instead of transporting the donor and recovering the organ elsewhere. A primary motivation for this practice is to safeguard the donor's interests and because, by the time of the transplant, the donor has built a relationship with her hospital and surgeon. The surgery performed on the donor requires extensive pre-planning and follow-up care. Conversations with surgeons suggest that these factors severely limit the willingness to transport the donor and conduct a surgery in another hospital.

patient from different hospitals. Our conversations and anecdotal evidence suggest that a similar pattern also holds true for other national platforms. This observation suggests that essentially all of the non-NKR within hospital transplants are facilitated by individual hospitals.[29] Moreover, there are over one hundred hospitals performing kidney exchanges during this period, and only 50 participated in the NKR in 2014. Taken together, these facts suggest a highly fragmented market structure.

The contrast between the fraction of within hospital exchanges in the overall market and the NKR is striking because the NKR does not use across hospital exchanges as a rule. Rather, the predominance of across hospital exchanges in the NKR is a by-product of maximizing the total number of transplants. This fact points to potential gains from co-ordinating across hospitals. We will formally analyze and quantify these gains in the subsequent sections.

Although the market is fragmented, figure 1 also shows that the total number of kidney exchange transplants grew from about 400 transplants in 2008 to about 800 in 2014.[30] The NKR grew rapidly during this period, becoming the largest kidney exchange platform and accounting for about one-third of all kidney exchange transplants. However, growth of the overall market seems to have slowed in recent years. The total number remains at around 800,[31] well below some estimates of the potential size of the kidney exchange market (Bingaman et al., 2012; Massie et al., 2013).

3.2 Smoking-Gun Evidence of Inefficiency

The high degree of fragmentation points to a market that operates inefficiently because the returns to scale in matching patients and donors are likely to be high. However, it is possible, at least in principle, that the market is highly fragmented but close to efficient. For example, it may be that hospitals operate at an approximately efficient scale, so that the loss due to market fragmentation is small. We now present smoking-gun evidence that this is not the case by showing that hospitals often conduct exchanges that are clearly inefficient from a social perspective.

One easily detectable inefficiency is a transplant between an O blood type donor and a non-O blood type patient. As explained in Roth et al. (2007) and in Section 2, O donors are scarce and O patients are in abundance. In a large enough market such as the U.S., optimal matches should only transplant O donors to O patients because O patients cannot accept other blood

[29] Even an extremely conservative lower bound allows us to conclude that at least half the kidney exchanges were facilitated outside the national platforms, and instead by individual hospitals. This lower bound is based on the fact that the other two national platforms (UNOS KPD and the APD) together accounted for about 10% of the market in 2014. Agarwal et al. (2018) report that these platforms performed a little over than 80 transplants in 2014 (out of 800 transplants nationwide).

[30] Our data for the NKR extends until December 1, 2014. This censoring may account for the slight drop in transplants in the last year of this figure.

[31] Source: https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/ (accessed December 21, 2017).

types.[32] The exception to this rule is for a highly sensitized patient, i.e. one with a very high PRA. The platform may want to use an O donor to transplant this patient if it is the only way to get this patient transplanted.

Figure 2 displays the fraction of O donors that are used to transplant non-O patients, categorized into NKR transplants, non-NKR across hospital transplants, and non-NKR within hospital transplants. Among NKR transplants, only 6.6% of O donors are used for non-O patients. In contrast, among non-NKR within hospital transplants, this percentage is 22.4%. This difference is statistically significant at the 0.1% level, and constitutes strong evidence that hospitals often perform inefficient matches outside the platform. The non-NKR across hospital transplants are in between the other two categories, but much closer to the NKR. This fact is consistent with most of these transplants being performed by other national platforms.

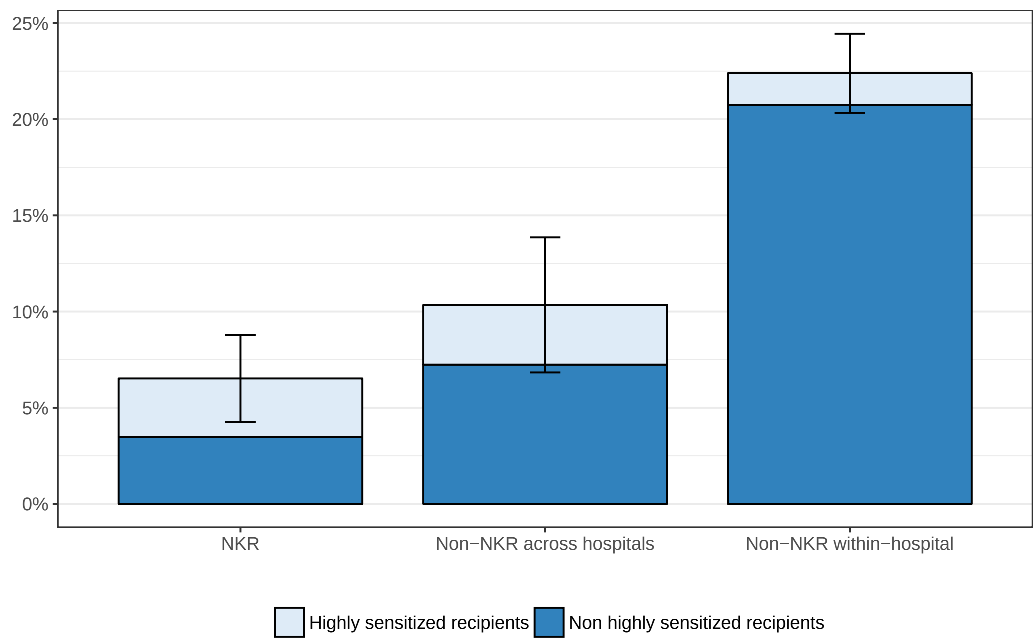


Figure 2: Smoking-gun evidence of hospitals performing inefficient matches

Notes: The bars display the percentage of transplanted O donors that were transplanted into non-O patients for different categories of transplants. NKR represents transplants facilitated by the NKR. Non-NKR across-hospital represents transplants that were not facilitated by the NKR, and where the donor hospital is different than the patient hospital. Non-NKR within hospital represents transplants that were not facilitated by the NKR, and where donor and patient hospitals are the same. The colors decompose this total into highly sensitized patients (PRA >90), and non-highly sensitized patients. The error bars depict 95% confidence intervals for the totals.

An alternative explanation to inefficient matching is that within hospital transplants are using

[32]Strictly speaking, the efficiency discussed here is in the sense of maximizing the total number of transplants. However, transplanting an O donor to a non-O patient is also likely to be Pareto inefficient. To see this, consider a pairwise exchange between two overdemanded A-O pairs. This exchange results in two transplants. It would be more efficient to transplant each of the A-O pairs to an underdemanded O-A pair, which otherwise would be left unmatched.

O donors to help highly sensitized patients that would otherwise remain untransplanted. However, Figure 2 shows that almost none of the potentially inefficient transplants in the non-NKR within hospital category involve highly sensitized patients. In contrast, about half of the potentially inefficient NKR transplants involve highly sensitized patients.

This exercise treats all transplants, whether conducted through the NKR or elsewhere, equally. It is possible that there are dimensions on which within hospital transplants are superior to national platforms. For example, a transplant through the NKR could involve a longer wait on dialysis or a lower quality donor. However, table 1 shows that patients who receive a transplant from the NKR typically wait for only 2 more months on dialysis than patients who receive a within hospital transplant. Given that the average patient waits is about 32 months, this difference represents an 8% longer waiting-time. This longer waiting time at the NKR should be expected because, as we discuss below, patients transplanted through the platform are harder to match on average. Further, it does not appear that there are differences in how desirable the donors might be to patients. The indicators of donor quality such as age, weight, height, BMI are extremely similar across platforms. One reason why patients considering the NKR need not worry about donor quality is that the NKR allows patients and doctors to specify acceptability criteria for the donor. It also allows patients to refuse transplants even after they have been proposed if the donor is considered unsuitable.

If each of these inefficient transplants comes at the cost of one other transplant, as in the Roth et al. (2007) model, then achieving the level of efficiency obtained by the NKR would have resulted in about 250 additional transplants between 2008 and 2014. The advantage of considering only the clearly inefficient transplants is that the results provide transparent, smoking-gun evidence of inefficiency. The total inefficiency, of course, can be much larger.

3.3 Why don't Hospitals Participate? Hospital Behavior and Evidence of Agency Problems

We have shown that the kidney exchange market is highly fragmented, with many hospitals performing matches outside the national platforms. Moreover, this fragmentation leads to real efficiency losses because internal matches performed by hospitals are often socially inefficient. These results lead to the question of why hospitals do not participate in national platforms to a greater extent. We address this issue by first documenting key facts about hospital behavior. Then, we discuss different hypotheses that can explain this behavior. Our results indicate that hospitals do not purely maximize the number of transplanted patients. Instead, hospitals seem to maximize complex and heterogeneous objectives, including but not limited to profits and patient welfare.

3.3.1 Descriptive Evidence

We start by describing the relationship of hospital size with the extensive and intensive margins of participation in the NKR. We measure hospital size as the number of transplants

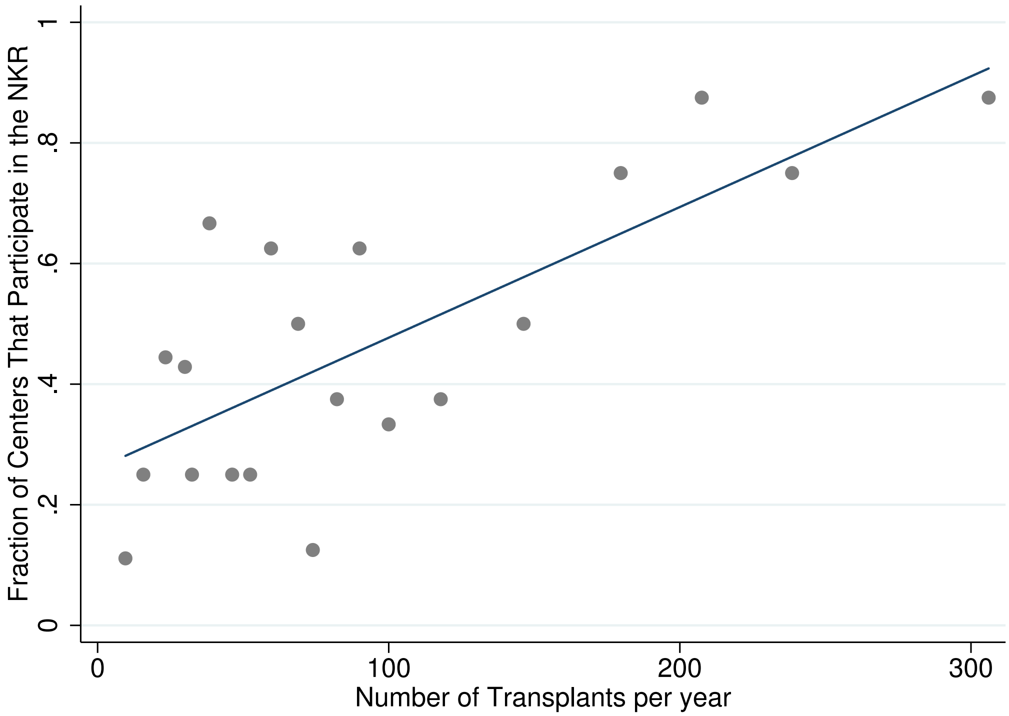


Figure 3: Heterogeneity in participation in the NKR

conducted per year, including deceased donor and direct living donor kidney transplants.[33] Figure 3 depicts the extensive margin of participation amongst hospitals conducting kidney exchange transplants. A hospital is considered an NKR participant if it has ever submitted a patient or a donor to the NKR. The figure is a binned scatterplot of the fraction of hospitals that participate in the NKR versus hospital size. Figure 4 depicts the intensive margin of participation. The vertical axis in this scatterplot is the fraction of kidney exchange transplants that a hospital performs through the NKR.

The figures reveal four key facts about participation. First, both the extensive and intensive margins are important drivers of market fragmentation. Only 46.6% of hospitals participate in the NKR. And, out of those that participate, only 53.1% of transplants are conducted through the NKR. Second, larger hospitals are considerably more likely to participate in the NKR. The probability of participating at all is about 80% for a hospital that performs approximately 250 transplants per year, but only about 35% for a hospital that performs about 50 transplants per year (figure 3). Third, conditional on participating, large hospitals conduct more of their matches outside the platform (figure 4). Although size is positively correlated with the fraction of kidney exchange transplants performed in the NKR, the relationship is negative if we focus exclusively on hospitals that participate at all (figure 4). Fourth, there is a high

[33] This broad measure of size limits the endogenous effect of participation in NKR on hospital size because deceased donor and direct living donor transplants form the bulk of kidney transplants conducted by a hospital. Moreover, during our sample period, the total number of kidney transplants has remained stable relative to the growth in kidney exchange.

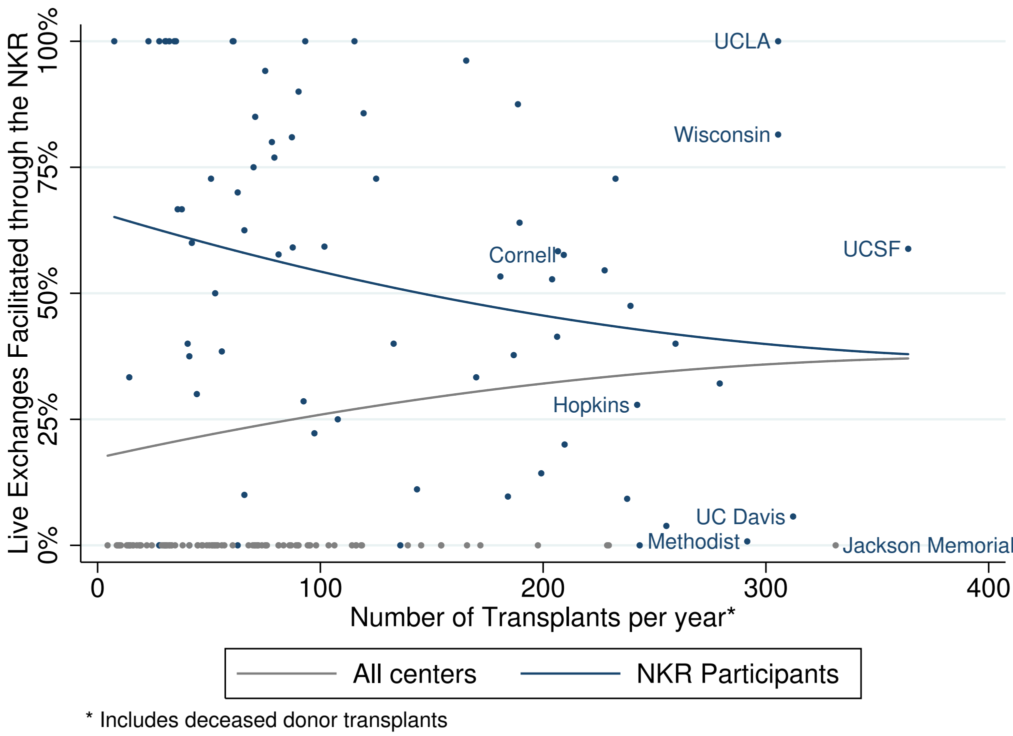


Figure 4: Reliance on the NKR for live-donor exchanges

degree of heterogeneity in intensive margin participation. Even among hospitals with similar size, participation varies considerably (figure 4). For example, amongst the five transplant hospitals that perform more than 300 transplants per year, one does not participate at all (Jackson Memorial), one participates close to zero percent (UC Davis Medical Center), two participate in the 50-60% range (UCSF Medical Center and the University of Wisconsin Hospital), and one has participation of more than 80% (UCLA Medical Center).

In addition to the decision to participate, the data provides information on the characteristics of patients that are submitted to the NKR, and the characteristics of patients transplanted by each hospital categorized by how the transplant was facilitated. Tables 1 and 2 reveal three main facts. First, the NKR receives submissions that are very hard to match compared to the general population (table 2). The blood type of both altruistic and paired donors is skewed away from O donors and in favor of A donors relative to the US population. The deceased donor population has about 45% of O donors and 40% A donors. In contrast, patients in pairs are disproportionately likely to have blood-type O (58.6%) and their related donors are unlikely to have blood-type O (31.5%). Only a small fraction of pairs (13.8%), are overdemanded. Interestingly, unpaired patients are much more likely to have an easy-to-match blood-type with the majority having an A blood-type. The average PRA for patients registered with the NKR is high. At a mean PRA of 48.8%, the average patient in the NKR is tissue-type incompatible with approximately half of the reference donor population. Second, the NKR transplants patients that are considerably harder to match than patients transplanted by single hospitals (table 1). Approximately 40% of the patients transplanted through the NKR were blood-type O and 41% of donors were blood-type O. The PRA of the patients transplanted through the NKR is approximately 35% and about one in six patients have a PRA above 90%. These statistics are similar for cross-hospital kidney exchanges not facilitated by the NKR. In contrast, among within hospital kidney exchanges outside the NKR, half of the donors are blood-type O, but only 40% of the patients are blood-type O. The average PRA of patients transplanted through within hospital exchanges is only 18%. This is almost half the mean PRA for patients transplanted through the NKR. Third, NKR and internal transplants look similar in measures of donor quality that do not affect compatibility, such as weight, body mass index, and age (table 1). This fact supports our equal treatment of all transplants, whether they are facilitated though the NKR or not.

3.3.2 Implications for Hospital Behavior

These facts have important implications for different hypotheses about hospital behavior. In the discussion that follows, we approximate total patient welfare with the total number of transplants. As we argued in Section 2, costs of kidney exchange are small relative to the benefits of transplantation.

The first hypothesis is that that hospitals maximize the total welfare of all patients in the system, regardless of which hospital a patient belongs to. This hypothesis is strongly rejected by several features of the data. Most clearly, this hypothesis is inconsistent with the smoking-gun evidence of socially inefficient matches (figure 2).

A second hypothesis is that hospitals maximize the welfare of only their own patients. This hypothesis was investigated theoretically by Ashlagi and Roth (2014). Their main prediction is that hospitals will try to match as many of their patients internally as possible, and register remaining patients in kidney exchange platforms. This hypothesis fits some qualitative patterns in the data. For example, conditional on participating, larger hospitals perform fewer transplants through the NKR because these hospitals have more opportunities to match patients outside the platform (figure 4). However, the hypothesis does not explain many other important features of the data. For example, many hospitals do not participate in the NKR at all (figure 3), even though all hospitals are likely to have patients who cannot be matched.[34] Moreover, many small hospitals do not participate in the NKR, even though these hospitals are precisely the ones most unlikely to find matches outside the platform due to their size. These patterns are consistent with a model where hospitals respond to fixed costs of participating in kidney exchange platforms, although these costs are small from a social perspective.

A third hypothesis is that hospitals are profit maximizers. This hypothesis is consistent, for example, with the fact that small hospitals are less likely to participate in the NKR (figure 3) because the fixed logistical costs of participation may not compensate for the gains in profits from additional transplants. However, this theory alone is stretched to explain

[34] Recall that overdemanded pairs are typically scarce. We will see in Section 5 that even the NKR is able to match only 50% of its donors.

the large variation in the degree of participation, especially among large hospitals. For example, Cornell Medical Center is a large hospital that has a high rate of participation in NKR. Its behavior fits neither the prediction of a pure profit maximization motive, nor of the motive to purely maximize the welfare of a hospital's own patients. Interviews with transplant coordinators at Cornell reported in Ellison (2014) suggest that a primary reason for participating is the view that contributing to a national kidney exchange platform is important.

Taken together, the evidence on hospital participation alone suggests that hospitals maximize complex and heterogeneous objectives. This finding is consistent both with the anecdotal evidence on kidney exchange reviewed in Section 2, and with typical findings about the behavior of healthcare providers. For example Clemens and Gottlieb (2014) investigate how providers respond to changes in Medicare reimbursement rates, and Kolstad (2013) investigates how surgeons respond to the introduction of publicly observable report cards. These studies find that providers respond to incentives, but are not purely motivated by profits, and also take patient welfare into account.

The facts about selection also indicate that these two theories, profit maximization and maximizing the welfare of their own patients, can explain the behavior of many hospitals. A common implication of these theories is that the pairs submitted to national platforms are negatively selected, in the sense of being hard to match. In both cases, a hospital only submits a pair to a platform if an internal match is not possible. Under most plausible assumptions about the matching technology, the submitted patients will be harder to match than those not submitted. Unfortunately, we cannot directly test this prediction because we do not have data on the entire pool of patients available to individual hospitals. Nonetheless, as we discussed above, the data shows that submitted pairs are much harder to match than the general population. Moreover, the NKR transplants patients and donors that are harder to match. If the NKR and the hospitals are equally capable of matching hard-to-match patients, then these two facts together suggest that hospitals are selecting hard-to-match types when submitting patients to the NKR. However, it is likely that some of the selection is driven by NKR being significantly better at finding kidney exchanges for hard to match patients with a given pool. Nonetheless, it is reassuring that the results on selection do not falsify the two theories that best fit the participation behavior.

Taken together, these findings have two important implications. First, there is clear evidence of agency problems, as we defined broadly in Section 2. The data rejects the hypothesis that hospitals purely and rationally maximize the welfare of their own patients. Second, there is no simple model that describes the behavior of all hospitals. These two findings represent a theoretical and empirical challenge because we cannot rely on a simple stylized model of hospital behavior to propose practical market design solutions, or to perform welfare analyses. Moreover, they point to agency problems as being a potentially important driver of inefficiency, that cannot just be assumed away. Both implications are important elements in our analysis, which we now develop.

4 Theory

The evidence above shows that kidney exchange markets are fragmented, and that this fragmentation leads to real efficiency losses. We now develop a model to explain how fragmentation and inefficiency arise, and to pin down the sources of market failure. The model also shows what sufficient statistics have to be estimated to empirically measure the inefficiency, and to design better mechanisms.

The basic idea is that the kidney exchange market is fundamentally similar to a traditional market. Namely, a kidney exchange clearinghouse is a platform that procures submissions (donors and patients) from hospitals, and rewards these hospitals with transplants. The platform's ability to produce transplants from submissions is given by a production function, which describes how many transplants can be performed with a given pool of submissions.

4.1 Model

A monopolistic kidney exchange platform procures submissions from hospitals, and rewards hospitals with transplants. The platform's ability to produce transplants is described by a **production function** f. We consider **types of submissions** indexed by $i = 1, \ldots, I$. Given a vector of **quantities** $\boldsymbol{q} = (q_i)_{i=1,\ldots,I}$, with q_i being the quantity of type i submissions, the platform can produce $f(\boldsymbol{q})$ transplants. The model can be interpreted either as static or as a steady-state from a dynamic model. We will use the steady-state interpretation in the empirical analysis (section 5). All variables are measured in flows, such as transplants per year, and are real numbers.

The production function summarizes what matches are possible. It is easy to understand it with a concrete example. Roth et al. (2007) calculated the production function in a simple model, that we described in Section 2. Roth et al. (2007) assumed that only blood type compatibility matters, so their model has $I = 16$ types. The production function is linear, and given by equation (1). The marginal product equals 2 for over-demanded pairs, 0 for under-demanded pairs and 1 for self-demanded pairs.

Much of the literature on kidney exchange is focused on the production function, even though we are the first to use this term. Previous work has studied how the production function depends, for example, on the maximum length of swaps (Roth et al., 2007), on the participation of compatible pairs (Sönmez and Ünver, 2014), and on the use of different algorithms (Akbarpour et al., 2016). In Agarwal et al. (2018), We take a complementary approach. Instead of deriving results about the production function, we take the production function as given, and use it to derive results in an economic model. For example, Theorem 1 shows how optimal mechanisms depend on key aspects of the production function, such as returns to scale.

The theoretical analysis applies both to simple production functions, such as those in the theoretical literature, and to more realistic production functions that incorporate institutional details. An advantage of this approach is that we can use an empirical model for the production function that is estimated using the data from a platform. Section 5 engages in this exercises. Amongst other details, our empirical production function allows submissions to differ by whether they are patient-donor pairs, altruistic donors or unpaired patients, and by a host of variables including blood types, antigens, and antibodies. Thus, the number of types I is potentially large.

The platform produces transplants using submissions provided by **hospitals** indexed by $h = 1, \ldots, H$. Hospitals are rewarded for these submissions with transplants. We assume that these rewards are linear and anonymous in submissions. That is, there exists a **vector of rewards** $\boldsymbol{p} = (p_i)_{i=1,\ldots,I}$ such that a hospital that submits a flow $\boldsymbol{q}^h$ receives a flow $\boldsymbol{p} \cdot \boldsymbol{q}^h$ of transplants. This linear reward schedule is a good approximation to current platforms' rules because current matching algorithms maximize a weighted sum of the number of matches without considering the entire pool of patients and donors submitted by the hospital. When a hospital submits an additional pair, the probability that the platform matches a different pair from the same hospital does not significantly change. Therefore, the current reward for submitting a type i pair is equal to the probability, p_i, that the pair is matched.

We assume that hospital preferences are quasilinear in the number of transplants they receive from the platform and the private cost of their submissions. Formally, hospital h maximizes

$$\boldsymbol{p} \cdot \boldsymbol{q}^h - C^h(\boldsymbol{q}^h),$$

where $C^h(\boldsymbol{q}^h)$ is the **private cost** of submissions in transplant units. One particular case is when hospitals maximize the number of their own patients that are transplanted, in which case $C^h(\boldsymbol{q}^h)$ is the number of within hospital transplants that the hospital would have to forfeit in order to submit $\boldsymbol{q^h}$.

We use transplants as the numeraire good because the platform can transfer transplants between hospitals at almost no efficiency cost. These transfers can take place by picking which hospitals' underdemanded pairs and unpaired patients the platform chooses to transplant. Efficiency losses are low because of the abundance of these types of patients and pairs in the platform.

The final piece of the model is a utilitarian welfare notion. Welfare is defined over an **allocation** $(\boldsymbol{q}^h)_{h=1,\ldots,H}$ that specifies the quantity of pairs supplied by each hospital. We will use two such notions. The first notion is **hospital welfare** $W^H(\boldsymbol{q}^1, \ldots, \boldsymbol{q}^H)$, which is the total welfare measured from the point of view of hospitals. Formally, hospital welfare is defined as

$$\begin{aligned} W^H(\boldsymbol{q}^1, \ldots, \boldsymbol{q}^H) &= \sum_{h=1}^{H} \boldsymbol{p} \cdot \boldsymbol{q}^h - C^h(\boldsymbol{q}^h) \\ &= f(\boldsymbol{q}) - \sum_{h=1}^{H} C^h(\boldsymbol{q}^h), \end{aligned} \tag{2}$$

where $\boldsymbol{q}$ is the aggregate quantity. Hospital welfare measures efficiency according to hospitals' objectives. This is a compelling notion of wefare if the goal is to help key market participants (hospitals in this case) achieve their objectives.

Hospital welfare is not compelling if there are agency problems. That is, if hospitals do not purely maximize a reasonable measure of the welfare of the patients and insurers they represent. As discussed in Sections 2.1 and 3, there is anecdotal and empirical evidence of agency problems. For this reason, we also consider a utilitarian welfare measure, that we term total welfare, and measure in units of transplants.

To formalize this idea, define $SC^h(\boldsymbol{q}^h)$ as the **social cost** for hospital h to supply a vector $\boldsymbol{q}^h$ submissions. If there are agency problems, then social and private costs are different, and there is an **externality** from hospital h's submissions given by

$$E^h(\boldsymbol{q}^h) = C^h(\boldsymbol{q}^h) - SC^h(\boldsymbol{q}^h).$$

For example, E^h is positive if hospital h acts as though the financial and logistical costs of participating in kidney exchange platforms are significant relative to their private value of a transplant. As discussed in Section 2.1, these costs are negligible relative to the social value of a transplant.

Given an allocation, let $E(\boldsymbol{q}^1, \ldots, \boldsymbol{q}^H)$ be the **aggregate externality**, which equals the sum of externalities from all hospitals. The externality represents the benefits to other stakeholders that are not internalized by hospitals. While we will use the terminology of agency problems, because the key stakeholders are patients and insurers that the hospitals represent, this wedge includes any possible deviation between hospital welfare and social welfare. These deviations may also be due to hospitals not behaving optimally for behavioral reasons. In the particular case where there are no agency problems, we have $\boldsymbol{\partial E} = \boldsymbol{0}$.

Define **total welfare** to be

$$SW(\boldsymbol{q}^1, \ldots, \boldsymbol{q}^H) = f(\boldsymbol{q}) - \sum_{h=1}^{H} SC^h(\boldsymbol{q}^h).$$

Thus, the difference between social and hospital welfare is equal to the aggregate externality E. Finally, define the **aggregate cost function** $C(\boldsymbol{q})$ as the minimum of the sum of costs of all hospitals subject to the total quantity submitted being equal to $\boldsymbol{q}$. For simplicity, we assume that the production, cost, social cost, and aggregate cost functions are defined over all non-negative real vectors and are smooth; aggregate costs are strictly convex; quantities are column vectors; and vectors of rewards and gradients are row vectors.

This model finesses a number of important issues, such as efficiency costs of transferring transplants, the choice of a particular welfare function, and the case of multiple competing platforms. We will return to these issues after deriving the main conclusions from the model.

4.2 Illustrative example

This model clarifies that kidney exchange is, in many ways, similar to a traditional market. It shows that the problem of regulating a kidney exchange platform is similar to the Ramsey (1927)-Boiteux (1956) problem of regulating a multi-product firm. The key difference is that, instead of a firm producing many products, a kidney exchange platform procures multiple types of submissions from hospitals. Our setting is also closely related to the platforms literature, where platforms maximize a private or social goal by setting incentives for participants (Rochet and Tirole, 2003; Weyl, 2010). This connection requires some abstraction because kidney exchange markets do not involve monetary transfers, and because the relevant numeraire are transplants. In this section we consider a concrete example to clarify this connection and the key assumptions. The example is only for illustration, and is not directly used in the subsequent analysis.

Let $K^h(\boldsymbol{q}^h)$ be the monetary costs borne by hospital h of sending $\boldsymbol{q}^h$ submissions to a kidney exchange platform. This cost can include fees from the platform, costs of rearranging the hospital schedule around the platform, and the cost of hiring additional transplant coordinators (see Section 2.1). Let $T^h(\boldsymbol{q}^h)$ be the flow of internal kidney exchange transplants that hospital h foregoes when supplying $\boldsymbol{q}^h$ to the platform.

Hospitals value each transplant at v dollars. This value includes profits, and the monetary value that hospitals place on transplanting their patients. Gross revenues from a transplant are of the order of \$150,000 (Held et al., 2016). For illustrative purposes, take v to be \$50,000, which represents a generous 50% mark-up on costs. In transplant units, hospital utility equals the number of transplants minus the monetary costs divided by the value per transplant, i.e. hospital h's welfare is

$$W^h = \boldsymbol{p} \cdot \boldsymbol{q}^h - T^h(\boldsymbol{q}^h) - \frac{K^h(\boldsymbol{q}^h)}{v}.$$

This hospital welfare function fits our model by specifying

$$C^h(\boldsymbol{q}^h) = T^h(\boldsymbol{q}^h) + \frac{K^h(\boldsymbol{q}^h)}{v}.$$

Society values each transplant at V dollars. This social value accounts the value of transplants to insurers, to taxpayers, to patients, and to other patients that benefit from reductions in the deceased donor waitlist. (Held et al., 2016) estimate that health-care cost savings alone from the average transplant is around \$300,000, and that the gains in the quality and length of life is valued at over \$1,000,000.[35] Thus, it is reasonable to estimate that V is in the ballpark of \$1,300,000. Taken together, there is a substantial wedge between the social and private value of a transplant because V is more than an order of magnitude larger than v. Moreover, the additional social costs of participating in a platform $K^h(\boldsymbol{q}^h)$ are negligible compared to the benefits of performing additional transplants. In this notation, social welfare in transplant units equals

$$SW^h = \boldsymbol{p} \cdot \boldsymbol{q}^h - T^h(\boldsymbol{q}^h) - \frac{K^h(\boldsymbol{q}^h)}{V}.$$

[35] Some patients who receive a kidney exchange transplant would otherwise receive a kidney from a deceased donor. But, in each of those cases, another patient in the waitlist receives this kidney. Therefore, the social benefit of each kidney exchange transplant should still be the same as the gain from a single transplant.

Because $f(\boldsymbol{q}) = \sum_h \boldsymbol{p} \cdot \boldsymbol{q}^h$, this social welfare function fits our model if we specify

$$SC^h(\boldsymbol{q}^h) = T^h(\boldsymbol{q}^h) + \frac{K^h(\boldsymbol{q}^h)}{V} \approx T^h(\boldsymbol{q}).$$

The externality term equals

$$E^h(\boldsymbol{q}^h) = C^h(\boldsymbol{q}^h) - SC^h(\boldsymbol{q}^h) = \left(\frac{1}{v} - \frac{1}{V}\right) \cdot K(\boldsymbol{q}^h) \approx \frac{K(\boldsymbol{q}^h)}{v}.$$

That is, the externality is approximately the difference, measured in transplant units, of how much more hospitals care about the costs of participating in a kidney exchange platform compared to society.

We stress that the term agency problems that we use for the wedge E^h should be understood in a broad sense. We use this term because a large share of the benefits likely accrue to patients and insurers who contract with the hospital, and are not internalized by hospitals. But, even in our example, E^h includes benefits to other third parties, so that one may wish to use term it externalities instead of agency problems. In fact, the general model allows for many other wedges, such as hospitals being unaware of kidney exchange or misperceiving its benefits and costs.

To develop intuition for the magnitude of these wedges, assume that the monetary cost is linear in the number of submissions, i.e. $K^h(\boldsymbol{q}^h) = k \cdot \|\boldsymbol{q}^h\|_1$. The agency wedge is

$$E^h(\boldsymbol{q}^h) \approx \frac{k}{v} \cdot \|\boldsymbol{q}^h\|_1.$$

It depends on the cost of participating in a platform that are borne by the hospitals as a percentage of the private value of a transplant, and not the social value. Therefore, the agency wedge is substantial, even though the costs of participating in a platform are small relative to V. For example, if k is \$10,000 and v is \$40,000, then the wedge per submission is $k/v =$ 0.25 transplants per submission. This wedge is substantial because hospitals compare it to probabilities of matching in the rewards vector $\boldsymbol{p}$. It is large in this example precisely because hospitals receive a very small share of the social benefits of transplants, but bear many of the costs. Therefore, costs that are negligible from a societal perspective are magnified when divided by v. Moreover, this wedge would not exist if hospitals were reimbursed for the costs of participating in kidney exchange platforms, $K^h\left(q^h\right)$.

Figure 5 presents a graphical illustration of this example to clarify the two sources of inefficiency in our model: agency problems, and inefficient platform incentives. The horizontal axis plots aggregate supply $\boldsymbol{q}$. The vertical axis plots marginal products, social costs, and social benefits, assuming that hospitals choose individual supply optimally given a rewards vector. The current vector of rewards, which is equal to the probability of matching each pair, is denoted by $\boldsymbol{p}_0$. The current quantity supplied given these rewards is $\boldsymbol{q}_0$.

The figure shows that the current supply is inefficient from both the hospital and the social perspective. The hospital-optimal quantity is $\boldsymbol{q}^*$ because it equates $\boldsymbol{\nabla f}$ with marginal private costs. Thus, the first inefficiency is that the platform gives inefficient incentives, $\boldsymbol{p} \neq \boldsymbol{\nabla f}$. The second inefficiency is that there are agency problems because private cost and social cost differ, $C^h \neq SC^h$, or alternatively, $\boldsymbol{\partial E} \neq 0$. The socially efficient quantity, $\boldsymbol{q}^{**}$, can be achieved by using efficient incentives, $\boldsymbol{p} = \boldsymbol{\nabla f}$, and solving agency problems so that $\boldsymbol{\partial E} = 0$. In this example, agency problems can be solved by reimbursing hospitals for the costs of kidney exchange through the platform, i.e. $K^h\left(q^h\right)$.

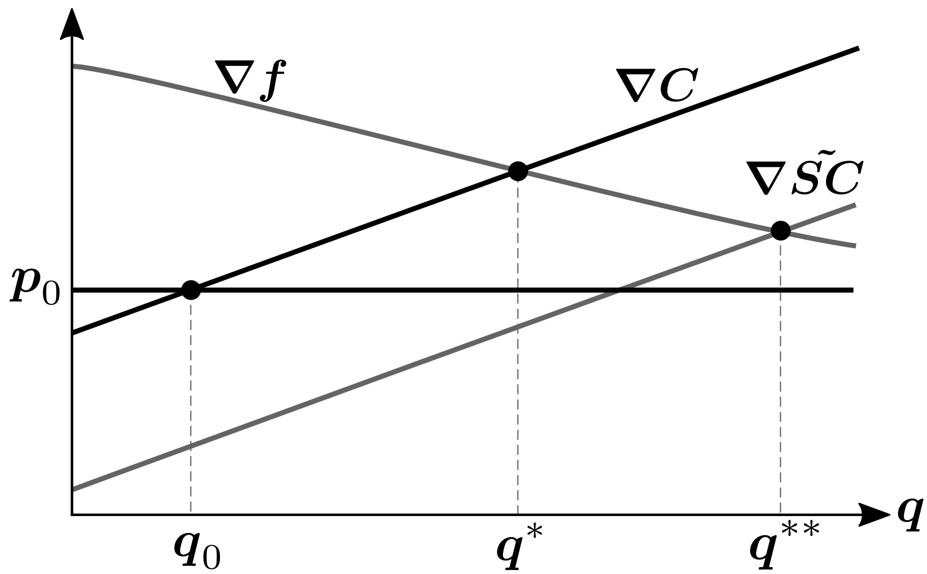


Figure 5: The two Sources of Market Failure

Notes: The horizontal axis represents aggregate quantity of submissions into the kidney exchange platform. The curves represent the marginal product of submissions $\boldsymbol{\nabla f}$, the marginal private cost of submissions $\boldsymbol{\nabla C}$ from hospital's perspective, and the marginal social cost of submissions $\boldsymbol{\nabla \tilde{SC}}$. The figure depicts the current quantity $\boldsymbol{q}_0$, with agency problems and a suboptimal mechanism, the quantity $\boldsymbol{q}^*$ from a hospital-optimal mechanism but with agency problems, and the first-bets quantity $\boldsymbol{q}^{**}$ with an efficient mechanism, and no agency problems, so that hospital and social incentives are aligned, $C^h = SC^h$. The social cost as a function of aggregate supply is defined assuming that the relevant $\boldsymbol{q}$ vectors are strictly positive, and that hospitals optimize given linear rewards, so that $\tilde{SC}(\boldsymbol{q}) = \sum_{h \in H} SC^h(S^1(\boldsymbol{\nabla C}(\boldsymbol{q})), \ldots, S^H(\boldsymbol{\nabla C}(\boldsymbol{q})))$. For simplicity, the figure assumes that individual supply is uniquely defined function S^h of rewards and that the platform has approxmately constant returns to scale.

This intuitive explanation finessed two subtleties, as will be made clear by the formal results. First, efficient platform incentives are only approximately equal to marginal products. A platform cannot set incentives equal to the marginal product because there are increasing returns to scale in kidney exchange, and therefore, marginal products exceed average products (Theorem 1). However, estimates in Section 5 will show that this adjustment term is negligible for the National Kidney Registry. Second, it is not possible to reach the first-best by only improving the mechanism if there are agency problems. Acheiving the first best requires setting rewards equal to the marginal products plus the externalities. But, even with constant returns to scale, there are only enough transplants to implement reward hospitals that equal the marginal product, making it impossible to reward hospitals to correct for externalities. These arguments suggest a two-pronged approach: design optimal mechanisms with the market design perspective of helping hospitals achieve their collective goals, and simultaneously implement policies to solve the agency problems.

4.3 Optimal mechanisms

We can now describe optimal mechanisms. The following theorem collects the main insights.

Theorem 1 (Optimal Mechanisms)**.** *Consider a vector of rewards $\boldsymbol{p}$ and allocation $(\boldsymbol{q}^h)_{h=1,\ldots,H}$ with strictly positive aggregate quantity $\boldsymbol{q}$ that maximizes hospital welfare subject to all hospitals choosing supply optimally given $\boldsymbol{p}$, and subject to not promising more transplants than produced. Then:*

1. *The platform rewards each type of submission with its marginal product minus an adjustment term,*
$$\boldsymbol{p} = \boldsymbol{\nabla f}(\boldsymbol{q}) - \boldsymbol{A},$$
where
$$\boldsymbol{A} = \frac{\boldsymbol{\nabla f} \cdot \boldsymbol{q} - f}{\boldsymbol{q}' \cdot \boldsymbol{D^2 C} \cdot \boldsymbol{q}} \cdot \boldsymbol{q}'\boldsymbol{D^2 C}.$$

2. *If the production function has constant returns to scale, then this allocation attains the first-best hospital welfare, and the reward for each type of submission is exactly equal to the marginal product.*

3. *If, in addition, there are no agency problems, in the sense that $\boldsymbol{\partial E} = \boldsymbol{0}$, then this allocation also maximizes total welfare.*

The theorem focuses on mechanisms that maximize hospital welfare. The first part shows that the reward for each submission in an optimal mechanism is approximately equal to its marginal product. The intuition for this result is simple if we ignore the constraint that the platform cannot promise more transplants than it produces. The platform is similar to a firm that produces a consumption good (transplants), using intermediate goods (submissions). The supply of intermediate goods is efficient when prices $\boldsymbol{p}$ are equal to marginal products $\boldsymbol{\nabla f}$. The proof is identical for kidney exchange platforms even though there are no prices paid to acquire submissions. The first order condition for the first-best aggregate supply is $\boldsymbol{\nabla C} = \boldsymbol{\nabla f}$. The marginal cost curve, which governs hospital incentives, equals the supply curve and therefore, optimal rewards are $\boldsymbol{p} = \boldsymbol{\nabla f}$.

The only complication is the constraint that a platform cannot promise more transplants than it produces. This constraint is binding if f exhibits increasing returns to scale because the number of transplants produced, $f(\boldsymbol{q})$, is less than the number of transplants that would have to be paid if rewards were equal to the marginal product, $\boldsymbol{\nabla f}(\boldsymbol{q}) \cdot \boldsymbol{q}$. It is easy to visualize this case if $\boldsymbol{q}$ is one-dimensional because the tangents to the production function cross the vertical axis below zero. Therefor, a platform has to shade its rewards to hospitals relative to the marginal products. The optimal level of shading for each type of submission is given by the adjustment term $\boldsymbol{A}$. The platform should shade more aggressively on submissions with less elastic supply. In fact, our formula is identical to the Ramsey (1927)-Boiteux (1956) formula for an optimal linear commodity tax. For example, if the cross-elasticities of supply are zero, we obtain an inverse-elasticity rule for the optimal shading, as in Ramsey's work and in the Lerner index from optimal monopoly pricing.

The theorem suggests that current platform rules are very unlikely to be efficient. Instead of choosing an optimal reward, current platform rules implicitly set a reward for a submission that is equal to the probability that the submission is transplanted. Therefore, there is most likely a wedge between the social and private benefits of submissions. Under this reward scheme, a hospital faces the dilemma of choosing between serving their own patients or providing a service to the system as a whole. A clear example of this dilemma, described in Section 2.2, is that of a hospital with two overdemanded pairs. This hospital could match them internally instead of submitting them to a platform, but doing so would cause the type of inefficiency documented in Section 3.

The second part of the theorem shows that, when returns to scale are constant, the optimal mechanism rewards submissions exactly according to marginal products. The adjustment term in this case equals zero, and optimal rewards achieve first-best hospital welfare. As we will show in Section 5, this case is empirically relevant because the NKR is well within the region of approximately constant returns to scale. Therefore, optimal mechanisms can be calculated in practice by estimating marginal products because they not depend on supply elasticities.

Moreover, there is no need to consider nonlinear rewards because we can achieve first-best welfare by rewarding hospitals linearly with their submissions. One approach for using these results in practice is to introduce a simple dynamic points mechanism. For example, for each submission, a platform can credit a hospital points equal to the marginal product. Then, a point can be subtracted whenever a hospital transplants a patient. The platform performs optimal matches with a constraint that no balance falls below a certain level. Naturally, there are important theoretical issues related to implementing incentives in this kind of mechanism without compromising efficiency (Hauser and Hopenhayn, 2008). We will return to these issues in Section 6.

The third part of the theorem states that if $\boldsymbol{f}$ exhibits constant returns to scale and there are no agency problems, then the optimal mechanism achieves first-best social welfare. This result clarifies that there are two possible sources of inefficiency in our model: inefficient platform incentives, and agency problems. Platform incentives are inefficient if rewards deviate from marginal products, $\boldsymbol{p} \neq \boldsymbol{\nabla f}$. In the platforms literature, this problem is usually attributed to wedges between the goals of the platform and of society (Rochet and Tirole, 2003; Armstrong, 2006; Weyl, 2010). Agency problems exist if hospitals do not fully internalize the welfare of the parties that they represent, i.e. $\boldsymbol{\partial E} \neq \boldsymbol{0}$ (Jensen and Meckling, 1976). The market functions efficiently if platform incentives are optimal ($\boldsymbol{p} = \boldsymbol{\nabla f}$) and there are no agency problems ($\boldsymbol{\partial E} = \mathbf{0}$). This suggests that the inefficiency documented in Section 3 may be due to a combination of inefficient platform incentives and agency problems.

Figure 5 depicted these two types of market failure under some regularity conditions. The current aggregate supply is $\boldsymbol{q}_0$, which is determined by rewards that equal matching probabilities. If a platform switches to using an efficient mechanism, aggregate supply would move to $\boldsymbol{q}^*$. If agency problems are also solved, the market would move to the first-best aggregate supply $\boldsymbol{q}^{**}$. The deadweight loss at any of these points is given by a (multi-dimensional) Harberger triangle, between the marginal product and the marginal social cost curves.

The upshot of this analysis is that, much like in more traditional markets, many key questions about kidney exchange depend on the production function, which we turn to next.

5 Production Function Estimates and Results

We now estimate the production function using data from the largest kidney exchnage platform, the NKR. We use these estimates it to measure the magnitude of the inefficiencies, and to develop simple policy responses. The key results come from three analyses. First, we can measure total inefficiency by estimating the returns to scale, and estimating how many more transplants would be performed if production was moved to the efficient scale. This is similar to the misallocation literature in macroeconomics (Hsieh and Klenow (2009)). Second, we can design optimal mechanisms using marginal products. This is similar to the derivation of optimal policy formulas based on a small set of statistics commonly used in public finance (Dixit and Sandmo (1977); Saez (2001); Chetty (2009)). Finally, we can use our price-theoretic framework to measure the gains from optimal mechanisms. This is similar to the standard Harberger triangle analysis. Using these results, we can gauge the importance of policies that address agency problems versus platform incentives.

5.1 Estimation

There is a long tradition in economics of estimating production functions using data on inputs and outputs from several firms. The key econometric challenges in this literature are endogeneity in the chosen inputs, and selection in the set of operating firms (see Marschak and Andrews, 1944; Olley and Pakes, 1996). Unfortunately, this approach is not feasible in our setting for three reasons. First, these methods are best suited for low-dimensional production functions that only depend on a few inputs such as capital and labor, but suffer from a curse of dimensionality if there are many types of inputs. In our case, the vector of inputs is high-dimensional because submissions can vary in many ways. Second, commonly used functional forms such as Cobb-Douglas are unlikely to fit our data well. Third, the standard methods depend on a panel dataset with inputs and outputs of multiple firms, and exogenous variation of inputs. However, we only have data from a single large platform and with no clear exogenous variation in inputs.

Fortunately, our setting is well suited to an alternative approach to estimating the production function. We have detailed institutional knowledge of the operational procedures and algorithms used by kidney exchange platforms. One of us (Ashlagi) has developed the matching software for several platforms, and worked with the NKR. Moreover, we have detailed data on the operations of the NKR, and on the composition and biological compatibility of its patient pool. Therefore, we develop a detailed simulation model of a kidney exchange platform.

We simulate the various steps involved in organizing kidney exchange to evaluate the number of transplants, $f(\boldsymbol{q};\theta)$, that can be produced with any set of inputs, $\boldsymbol{q}$, and parameters θ. We use our knowledge of this production process and technology to specify this model. A subset of parameters can be estimated directly using administrative data while others are calibrated to fit key outcomes. We start by describing the operations of a kidney exchange platform such as the NKR, the set of unknown parameters, and how these parameters are set.[36] There are four key events that take place, each governed by a set of parameters:

1. **Submissions, $\boldsymbol{q}$:** Hospitals submit patients and donors, either individually or in pairs, to the platform. These submissions are added to the current pool of patients and donors already registered with the exchange. Patients and doctors, at this time, can submit minimal acceptance criteria for a donor.

   We estimate the vector of current submission rates, $\boldsymbol{q}_0$, by resampling from the administrative NKR data. Specifically, we estimate the total submissions per day as an empirical average, and then draw specific patient/donor/pair characteristics and minimum acceptance criteria from the dataset with replacement. The time-average of these simulation draws yields $\hat{q}_{0,i}$ for any finite partition of individual patients, donors or pairs into types, $i = 1, \ldots, I$. An identical resampling process allows us to simulate arrivals for any alternative rate of submissions $\boldsymbol{q} \neq \boldsymbol{q}_0$.

2. **Transplant Proposal:** Each day, the NKR identifies an optimal weighted set of potential exchanges within the stock of patients and donors registered with the platform. This algorithm incorporates four constraints. First, none of the proposed transplants should be (known to be) biologically incompatible or ruled out by pre-set acceptance criteria. These constraints are directly observed in the data. Second, no donor or recipient can be involved in more than one transplant. Third, a donor that is part of a pair is only asked to donate an organ if her intended recipient has been proposed a transplant. Finally, kidney exchange platforms limit the cycle size because of logistical difficulties in organizing many simultaneous surgeries.[37]

[36] See Anderson et al. (2015) for a description of the operation of the APD.

[37] Formally, the NKR maximizes $\sum_{jk} c_{jk} w_{jk} x_{jk}$ by picking $x_{jk} \in \{0,1\}$, where $x_{jk} = 1$ denotes a proposed transplant from donor k to patient j; w_{jk} is the weight accorded to each such transplant by the NKR; and $c_{jk} = 1$ if a transplant from k to j is feasible (biologically compatible and acceptable) and 0 otherwise. This problem is subject to three additional constraints. First, no donor or patient is involved in more than one transplant, i.e. $\sum_j x_{jk} \leq 1$ and $\sum_k x_{jk} \leq 1$. Second, if donor k and patient j belong to a pair, then $x_{j'k} = 1$ for some j' only if $x_{jk'} = 1$ for some donor k'. To write the third constraint, note that a cycle of length n is

The parameters of this algorithm are the weights w_{jk} used by the NKR for a transplant involving donor k and patient j, and the maximum cycle size. Consistent with NKR policy and observed data, we prohibit all cycles of length four or greater. The weights are known to one of the authors (Ashlagi). These weights prioritize transplants that are unlikely in an attempt to utilize hard-to-match donors and transplant hard-to-match patients whenever possible. They typically only break ties between two matches with the same number of transplants in favor of retaining patients and donors that are likely to match in the future.[38]

3. **Final Review and Transplantation:** Proposed transplants are reviewed by doctors, patients and donors and approved before a transplant is performed. Both approval and biological testing can take several days. Moreover, patients and donors in proposed transplants that are under review on a given day are excluded from the maximal matching algorithm on that day. This step also involves a final set of blood-tests to ensure biological compatibility.[39] Cycles in which any patient refuses or is found to be incompatible with her proposed donor are abandoned. NKR usually abandons chains in which either the second patient cannot be transplanted. For other chains, all proposals until the first failure are consummated. The donor belonging to the final patient-donor pair in such a chain may initiate new chains in the future much like an altruistic donor. This donor is often referred to as the "bridge" donor. Consistent with NKR policy, unpaired patients are prioritized according to the net difference between altruistic donors and unpaired patients previously transplanted by the patient's center.

   This step results in frictions within the system. The parameters that govern these frictions are the time required for each of the two approval steps, the probability that a proposed transplant is abandoned in each step, and the duration for which a bridge donor is retained in the pool before donating her kidney to a patient on the deceased donor list.

   Unfortunately, we do not have detailed data on which transplants were refused, how often transplants were aborted due to biological testing and how long each review phase takes.[40] Additionally, the NKR does not seem to have clear-cut algorithmic policies on how to use bridge donors. Chains would be indefinitely long if bridge donors are allowed to initiate new chains forever, but too short if bridge donors were not used.

an ordered tuple, $(j_1, j_2, \ldots, j_n)$ where $x_{j_k j_{k+1}} = 1$ for $k < n$ and $x_{j_n j_1} = 1$. We impose the constraint $n \leq 3$. Because there are a very large number of cycle length constraints, we first solve a relaxed problem without this last constraint and iteratively add the constraints to prohibit large cycles. Appendix C provides further details on the algorithm.

[38] Appendix C contains further details on the specific weights that are used.

[39] These failures are recorded by setting $c_{jk} = 0$ for future iterations if the donor k was refused by patient j.

[40] Some smaller platforms have reduced these problems. Methodist Hospital in San Antonio, the largest single-hospital platform, has centralized decisions and biological testing. Similarly, the Alliance for Paired Donation maintains a blood lab to save time in shipping blood samples between transplant centers. Agarwal et al. (2018) uses our production function approach to evaluate how these practices affect the productivity of a kidney exchange platform.

Although cases of donors reneging are rare (Cowan et al., 2017), platforms try to transplant bridge-donors quickly, to an unpaired patient if necessary, to avoid these cases.

We calibrate these parameters by simulating our model finding values that most closely replicate the match probabilities, durations, chain lengths and pool size observed in our data. We match these statistics by the following submission types: altruistic donor, patient-donor pairs and unpaired patients.[41] These quantities are precisely the rewards that the NKR implicitly sets for hospitals, $\boldsymbol{p}_0$, and therefore a key component of our economic model.

Our simulations suggest that a two-week period for both the acceptance and the biological testing phase and a failure rate of one-fifth for each phase best fit the observed transplant rates, and chain lengths observed in the dataset. Reducing the failure rates in simulations primarily increases chain length and transplantation rates, while reducing the duration of either phase increases the transplantation rates without having a large effect on chain length. For the bridge donor policy, we find that a hold-period of 30 days best fits the data.

Details on the fit of our calibrated parameters is provided in Appendix C.5.1. Further, Appendix D repeats all of our analyses under alternative parameters to ensure robustness of our results.

4. **Departure:** Patients and donors often depart the NKR without a transplant. A patient may leave the platform because she patient dies, becomes too sick to transplant or receives a kidney transplant elsewhere. Therefore, we need to estimate the probability that a patient or a donor leaves the NKR without a transplant.

   We estimate a model of departures using the registration and transplantation dates (if transplanted) for each patient and donor. Additionally, we use regular snapshots of the set of patients and donors registered at the NKR to determine how long the patient or donor was registered in the NKR without a transplant. We estimate an exponential hazards model for the departure process using maximum likelihood.[42] The departure rates in the model depend on the fraction of donors (patients) ever registered with the NKR that are compatible with a patient (donor), blood-type dummies for the donor and the patient, and the patient and donor's ages at registration. Appendix C.2.2 presents the estimates for the model.

[41] In principle, we could have estimated these parameters using simulated minimum distance. However, a simulation for each parameter value can take weeks making optimization over the parameter set infeasible.

[42] Specifically, the departure rate for registration j is given by $\lambda_{g_j} \exp(z_j \beta)$, where g_j denotes whether j is an altruistic donor, a patient-donor pair or an unpaired patient; λ_{g_j} is a group-specific constant departure risk; z_j denotes a vector of characteristics for j; and β is a conformable vector of coefficients. We use maximum likelihood using the (censored) observations of departure times for each registration in the NKR. Censoring in our dataset can occur because we only observe a lower bound for the departure time if j was transplanted or remained in the NKR pool at the end of our sample period.

To summarize, each of the four steps has a set of parameters associated with it. The parameters governing steps 1 and 4 are directly estimated from the NKR data; the parameters involved in step 2 are known; and the parameters from step 3 are calibrated to fit observed chain lengths, and transplantation probabilities for various patient and donors. We are able circumvent the econometric issues that are the focus of the literature on estimating production functions by using an engineering approach based on detailed institutional knowledge and administrative data on the processes involved in organizing kidney exchange.

This procedure allows us to evaluate NKR's production function for any vector of inputs $\boldsymbol{q}$ by simulating each of these events for each calendar day. Given any initial pool of patients and donors in the NKR, these simulations generate a Markov chain with a sequence of registrations, transplants, and departures. We initialize the NKR pool with the set of patients and donors registered on April 1, 2012, and burn-in 2,000 simulation days in each run. The dependence on the initial pool eventually fades away. Let $y_{jk}^{*,t} = 1$ if a kidney from donor k is transplanted to patient j. We compute the time average of the total number of transplants to estimate f for each simulation s:

$$\hat{f}_s\left(\hat{q}\right) = \frac{1}{T}\sum_{t=1}^{T} \left|y_s^{*,t}\right|,$$

where T is the total number of days simulated and $|y_s^{*,t}|$ is the total number of transplants in period t in simulation s. In what follows, we report estimates based on an average of 100 simulations.

5.2 Scale, Inefficiency and Optimal Rewards

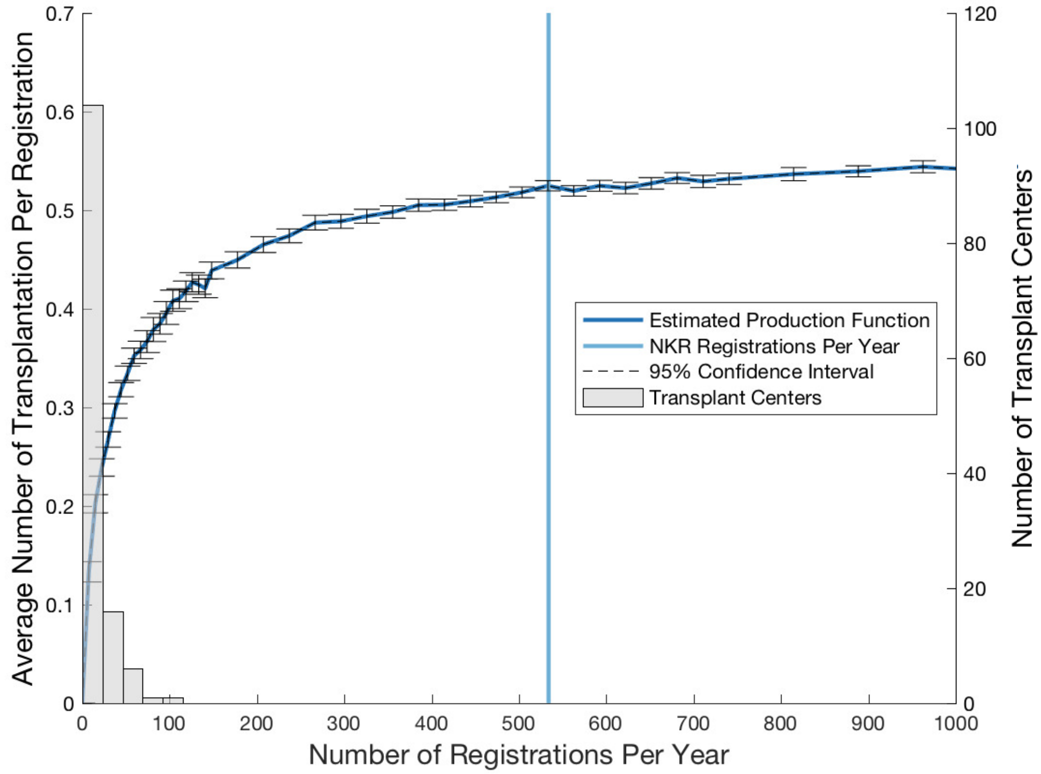
This section uses the estimated production function to measure the amount of inefficiency, the importance of the two sources of market failure, and to design better mechanisms. We use the total number of kidney exchange transplants (including both in platforms and in hospitals) as our measure of social welfare, SW. For simplicity, we begin by reporting the results under our main specifications, and then discuss detailed robustness analyses.

5.2.1 Returns to scale

We first document the returns to scale of the production function. That is, how the average product changes with the size of a platform. We evaluated the production function for pools of submissions $\boldsymbol{q}$ with the same composition as NKR, but with different scales as measured by the total flow of altruists and pairs per year. Figure 6 depicts average products, equal to $f(\boldsymbol{q})/\|\boldsymbol{q}\|_1$, as a function of the total flow, $\|\boldsymbol{q}\|_1$.

The estimates show a remarkable pattern. Although the returns to scale always increase, they reach a plateau fairly quickly. With a scale of 534 donor arrivals per year, the NKR is well within the region of approximately constant returns to scale. It has an average product of 0.52, which varies only marginally once the scale is sufficiently large. A platform that

Figure 6: Production Efficiency versus Scale



Notes: The line-plot represents the average product of a kidney exchange platform versus its scale. The histogram is based on the estimated scale of various hospitals. The left-vertical axis represents average products, defined as the share of pairs and altruists who are transplanted. The right-vertical axis is the scale for the histogram. The horizontal axis represents scale, measured as the yearly arrival rate of pairs and altruists. The plot uses the baseline parameters and the pool composition from NKR.

is half the size of NKR has an average product of 0.49, while a platform with double the size has an average product of 0.54. Therefore, the market would operate at a high level of efficiency even if there exist a handful of competing platforms. These estimates suggest that mergers of sufficiently large platforms would have small effects on efficiency.

We can use these estimates to calculate whether individual hospital platforms, which collectively account for a majority of kidney exchange transplants, operate at an efficient scale. A challenge with this exercise is that we observe neither the number nor the composition of patients and donors available to a hospital. To make progress, assume, for the moment, that hospitals have the same production technology and composition as the NKR. Further, assume that hospitals conducting within-hospital transplants do not participate in the NKR. Under these assumptions, one can use the observed rate of kidney exchange transplants at individual hospitals to infer the scale for each hospital. Specifically, let y^h be the flow of within hospital kidney exchange transplants conducted at hospital h. The scale of hospital

h is then $q^h = \hat{f}^{-1}\left(y^h\right)$ where $\hat{f}$ is the relationship between NKR's scale and total output, based on the estimated production function.

This exercise suggests that individual hospitals operate far below the efficient scale. The histogram in Figure 6 shows the estimated distribution of hospital scale. The median hospital has a scale of 10. The 90th percentile is 30. The largest, Methodist Hospital in San Antonio, has a scale of 114. The average product at these efficient scales is, respectively, 0.16, 0.27 and 0.41. Thus, even the largest single-hospital platform is not large enough to operate at an efficient scale. The implied efficiency losses are considerable even for the largest single-hospital program. These results are consistent with the evidence presented in Section 3.2 that hospitals often perform matches that are inefficient from a social perspective.

5.2.2 Misallocation: inefficiency due to small production scale

We start by using the baseline approach in the previous section to estimate of the inefficiency due to market fragmentation. That is, we estimate how many additional transplants would be performed if the entire kidney exchange market at the efficiency of NKR's scale. We use a hospital's estimated scale, $q^h = \hat{f}^{-1}\left(y^h\right)$, to calculate the difference in the average product between the hospital and NKR. Because NKR operates at constant returns to scale, this difference multiplied by the hospital scale is the total number of transplants that are lost due to the hospital conducting kidney exchange at an inefficiently small scale. The aggregate transplants lost equal the total deadweight loss because our social welfare function is the total number of transplants nationwide. The estimated deadweight loss presented in Table 3 shows that 550.0 transplants are lost per year due to market fragmentation (panel A, column (1)). This number is large relative to the 800 transplants that are conducted through kidney exchange each year.

This baseline approach is simple, but suffers from four potential biases. First, the composition of submissions in hospitals may differ from that in NKR. We assess robustness to this assumption by estimating the inefficiency using patient and donor compositions based on submissions from three different groups of hospitals: all hospitals, hospitals in the top quartile of participation rate, and hospitals in the bottom quartile.[43] Because hospitals in the top quartile of participation rate should be submitting a less selected pool of patients and donors, robustness of the estimates should provide a sense of whether compositional differences are introducing large biases in our approach. Columns (2) and (3) present estimates under alternative assumptions on the composition of patients and donors that are available to the hospital. A comparison of the estimates suggest that the overall inefficiency is not particularly sensitive to these compositional differences. One reason for this robustness is that tissue-type compatibility is more likely to be the main barrier to kidney exchange the scale of a hospital a compared to the scale at which a large platform such as the NKR operates (Roth et al., 2007).

[43] We measure participation rate as the number of donors submitted to the NKR as a fraction of donors submitted to the NKR or transplanted in a within hospital kidney exchange.

Second, while some hospitals are "islands" that perform virtually all kidney exchange transplants within their hospital, other hospitals perform both internal and external matches. Our baseline approach assumes that all hospitals are islands. The bias from violations of this assumption does not have a clear direction. We address this issue by disaggregating the efficiency losses by whether a hospital participates in the NKR, and by the fraction of the hospital's paired kidney exchanges that are conducted through the NKR. If we restrict attention only to the 177 hospitals that do not participate in NKR, the efficiency loss in Column (1) is 244.5 transplants per year. Within the set of NKR participants, the 17 hospitals that participate in NKR and are in the lowest quartile of fraction of transplants performed in NKR alone contribute to an efficiency loss of 112.2 transplants per year.

Table 3: Total Efficiency Loss

	Number of Centers	Total Number of Kidney Exchange Transplants per year	Total Within Hospital Kidney Exchange Transplants	Efficiency Loss: Additional Kidney Exchange Transplants		
				(1)	(2)	(3)
Panel A: All Centers						
All Centers	256	791.0	462.0	550.0	406.9	561.5
Panel B: By center size (number of PKEs per year)						
Top Quartile	42	596.4	360.5	286.2	201.8	313.2
2nd Quartile	47	136.3	65.9	135.8	108.3	140.1
3rd Quartile	39	45.3	29.2	90.7	71.9	87.8
Bottom Quartile	35	13.1	6.4	37.4	24.9	20.4
Panel C: By NKR Participation						
Yes	78	575.8	294.9	305.5	219.9	322.8
No	177	213.4	162.4	244.5	187.0	238.7
Panel D: By NKR Participation Rate (Fraction of PKEs facilitated through the NKR)						
Top Quartile	17	66.26	8.6	19.7	15.1	19.7
2nd Quartile	17	100.3	27.7	53.8	41.1	51.8
3rd Quartile	17	196.5	97.7	97.1	69.8	105.1
Bottom Quartile	17	217.5	165.5	112.2	78.0	123.1

Notes: Column (1) assumes that the typical transplant center has a composition of patient-donor pairs and altruistic donors given by the average registration in the NKR. Column (2) assumes the composition in transplant centers using only the centers with the top quartile of participation rates in the NKR. Column (3) assumes a composition based on centers with the lowest quartile of participation rates.

Third, hospitals may use a different matching technology than NKR. For example, Bingaman et al. (2012) report that Methodist Hospital in San Antonio (perhaps the most sophisticated single-hospital program) initially used a Microsoft Access Database, and that their algorithm was "stratified by ABO compatibility and then by HLA compatibility." Such algorithms are less efficient than the linear-programming algorithms used by the NKR.[44] On the other hand, single-hospital programs face simpler logistical constraints, which may increase their

[44]In 2013, Methodist Hospital in San Antonio adopted software written by one of us (Ashlagi).

productivity vis-à-vis our estimates. The direction of this bias is not signed in general, but it is more likely that single-hospital platforms are less efficient than our estimated production function.

Fourth, our baseline approach is likely conservative for the overall inefficiency of the market because it keeps the patients and donors that are interested in kidney exchange fixed. However, this flow is endogenous, and affects the magnitude of the deadweight loss. In general, the direction of this bias is ambiguous. The chief concern is that hospitals value transplants at less than the social value, and due to administrative costs, expend inefficiently low effort in recruiting patients and donors. If incentives were optimal, hospitals may try to recruit a greater number of and more valuable donors into kidney exchange. Our approach does not account for this margin because we do not observe recruitment effort and is therefore likely to underestimate the overall inefficiency of the market.

Table 3 also provides tentative insights into which types of hospitals concentrate most of the inefficiency. Consider column (1) and, for the purposes of this decomposition, ignore the biases discussed above. Even though large hospitals perform internal exchanges more efficiently, they account for most of the inefficiency because their market share is higher. Indeed, 52.0% of the losses come from hospitals in the top quartile of number of kidney exchange transplants. Moreover, both the intensive and extensive margins of participation are important. About a little more than half of the efficiency losses are due to hospitals that do not participate in the NKR at all. Among hospitals that do participate in the NKR, a large share of the efficiency loss is due to the hospitals with low participation.

To summarize, although the baseline estimate of 550.0 lost transplants is potentially biased, a battery of robustness exercises suggest that the deadweight loss from market fragmentation is large. These losses arise from all types of hospitals. The most conservative estimates based only on hospitals that do not participate in the NKR place this loss at 244.5. Moreover, these estimates do not appear to be sensitive to potential compositional differences in the kidney exchange pool. Table D3 in Appendix D further evaluates the robustness of these results to alternative choices for the parameters of the production function that were calibrated. Across various specifications, we continue to find that an estimate of 180 lost transplants based on non-participants is conservative. These results are consistent with our descriptive finding that hospitals often perform inefficient matches.

5.2.3 Marginal products and inefficiency of current mechanisms

We now test whether current platform rules are efficient. Theorem 1 shows that, if the current rules are optimal, then current rewards for different submission types, $\boldsymbol{p}_0$, should equal optimal rewards, $\boldsymbol{p}^* = \boldsymbol{\nabla f} - \boldsymbol{A}$. We will test this equality at the composition and rate of submissions, $\boldsymbol{q_0}$, for the NKR during our sample period.

Current rewards, $\boldsymbol{p}_0$, equal the probabilities of matching each kind of submission. These probabilities can be easily estimated from our simulations, and the estimated probabilities closely match the probabilities in the data (see Appendix C.5.1). Marginal products $\boldsymbol{\nabla f}$ can be estimated by numerically differentiating the production function with respect to its 1429 dimensions.

In principle, calculating the adjustment term requires estimates of the elasticity matrix of the supply of submissions. The supply elasticity matrix is high-dimensional, making it difficult to estimate using observed submissions to the NKR. Fortunately, because returns to scale are approximately constant for NKR's size, this adjustment term is small. Therefore, optimal rewards are approximately equal to marginal products. Formally, Theorem 1 implies that the quantity-weighted average of the adjustment term is

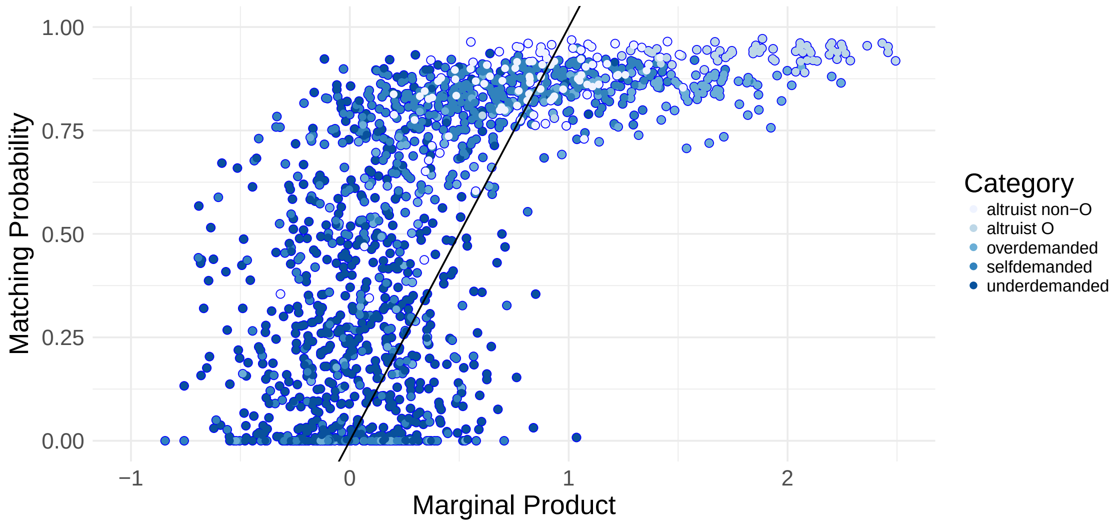
$$\boldsymbol{A} \cdot \frac{\boldsymbol{q}}{\|\boldsymbol{q}\|_1} = \boldsymbol{\nabla f} \cdot \frac{\boldsymbol{q}}{\|\boldsymbol{q}\|_1} - \frac{f}{\|\boldsymbol{q}\|_1}.$$

That is, the average level of shading is the difference between the average of marginal product and the average product. We estimate that this level of shading has an absolute value of 2.16×10^{-4}. In what follows, we simply approximate optimal rewards with marginal products because shading is not a major concern.

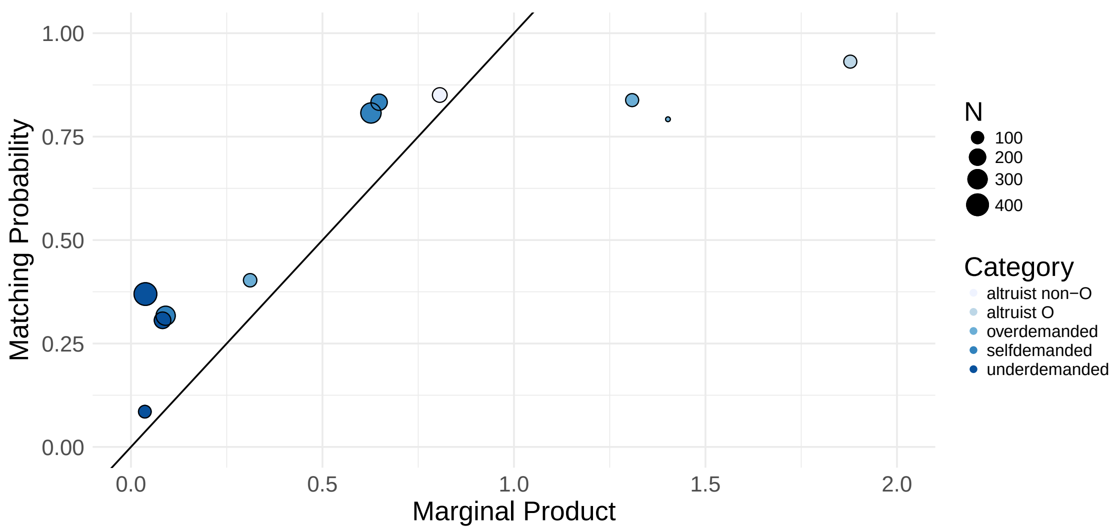
Figure 7a plots current rewards (the probabilities of matching $\boldsymbol{p}_0$) versus optimal rewards (marginal products $\boldsymbol{\nabla f}$). The marginal product of each of the 1429 submissions is measured with noise because of simulation error. Figure 7b aggregates these estimates by categories that constructed based on existing results on the productivity of various types. These aggregated marginal products and match probabilites are estimated more precisely.

The marginal products are qualitatively similar to the Roth et al. (2007) theoretical predictions discussed in Section 2. Consistent with our estimates, Roth et al. (2007) delivers a marginal product of 0 for underdemanded pairs. But, our estimates differ for some other types. For example, the model in Roth et al. (2007) delivers a marginal product of 2 for overdemanded pairs. We estimate that an overdemanded pair with low sensitization has a marginal product of 1.40. A reason for this difference is that these pairs only get matched with probability 0.79 according to our data. Our empirical model also refines the predictions from the theoretical models by showing how marginal products vary with sensitization. For example, the marginal product of overdemanded and self-demanded pairs is considerably lower if these pairs are sensitized. These finer results are important if the goal is to design practical mechanisms.

Both figures show that there is a large wedge between the current rewards and optimal rewards. If current rewards were optimal, all points on these two figures would be on the 45-degree line. Altruists and overdemanded pairs with low PRA are far below this line. Overdemanded pairs with low sensitization have marginal products of 1.40, but the probability of matching them is only 0.79. Even more extreme, altruistic O donors have a marginal product of 1.88, but a probability of matching of only 0.93. Therefore, hospitals are not rewarded enough for submitting these types, and it may not be surprising to see relatively few submissions of these types in the NKR. Other types of submissions are drastically over-priced. Underdemanded pairs with low sensitization have marginal products of approximately 0.04, but have a substantial probability of being matched of around 0.37. These differences suggest that the platform can do considerably better, by increasing rewards to the productive and undervalued submissions while reducing rewards to the unproductive submissions.



(a)



(b)

Figure 7: Private versus Socially Optimal Rewards for Submission Types

Notes: The vertical axis is the probability of a submission being matched, which are the private rewards that hospitals receive according to current exchange rules. The horizontal axis plots the marginal product of a submission, which equals the social contribution of the submission in terms of transplants. Each point correspond to a submission in the data. Probabilities of matching and marginal products are calculated in the baseline simulation. Marginal products are measured with substantial noise at the individual level because, due to computational reasons, each individual derivative uses a small number of simulation days. In aggregated version different dots of the same color correspond to the different PRA levels. Figure 7a shrinks the estimated marginal products and match probabilities towards to the group means following the procedure recommended by Morris (1983).

5.2.4 Optimality of simple mechanisms

The marginal products suggest that current platform rules are far from optimal, and that point mechanisms like the ones described in Sections 4 are likely to do better. We now show that marginal products are highly predictable using a small number of patient and donor categories. This suggests that very simple mechanisms, that can be plausibly used in practice, are close to optimal. For example, one could use the categories described in Figures 7a and 7b. A mechanism that assigns points based on these categories can be explained to participants with a simple table.

We take a more systematic approach by measuring how well individual characteristics predict marginal products using a regression tree. This approach is well suited to systematically utilize the numerical and categorical variables in a data-driven manner, and yields output that is easy to explain to market participants. Table 4 (depicted also in Figure C11) shows the results from the best cross-validated predictor for the marginal products using a standard algorithm. We allowed the tree to depend on the patient's PRA, submission type (altruistic, patient-donor pair, unpaired patient), and ABO blood type. The cross-validation procedure chose a simple tree, which suggests that marginal products are well approximated with a simple function of characteristics. Indeed, the (appropriately shrunk) within-category standard deviation in match probabilities and marginal products is relatively small (Table 4), suggesting that this simple tree captures most of the variation in marginal products.

Another implication of this result is that there are simple mechanisms that can be explained to market participants, that also give optimal rewards. For example, the last column in Table 4 suggests the number of points that a hospital would be awarded when one of its submissions is transplantated based on our approach. The platform could impose a lower bound on the balance a hospital must have before one of its submissions is transplanted. A hospital that has hit this bount would only be able to transplant either an altruistic donor or a overdemanded pair that does not have a highly sensitized patient (Non-O Patient, O Donor, PRA$<$94%). In psychology, there is evidence that professionals in several areas are likely to use data-driven advice when it is presented with simple decision trees (Gigerenzer and Goldstein, 1996; Gigerenzer and Kurzenhaeuser, 2005). We postpone a more detailed discussion of other implementation details of point systems to Section 6.

5.2.5 Welfare gains from simple optimal mechanisms

We now estimate the gain in welfare from moving to the simple optimal mechanism described above. This gain is equal to the deadweight loss that can be avoided by rewarding hospitals optimally as in Theorem 1. We begin by considering the gain in hospital welfare, and later consider the gain in total welfare.

The calculation of deadweight loss is similar to a multi-dimensional version of the deadweight loss from linear commodity taxation. Figure 8 illustrates a two-dimensional projection. It depicts the current aggregate supply $\boldsymbol{q}_0$, the current rewards $\boldsymbol{p}_0$, the current marginal products $\boldsymbol{\nabla f}_0$, and the optimal aggregate supply $\boldsymbol{q}^*$. Intuitively, the hospital deadweight loss

Table 4: Regression Tree Summary Statistics

	N	Match Probability			Marginal Product			Points per Transplantation
		Mean	S.E.	Within Category Standard Deviation	Mean	S.E.	Within Category Standard Deviation	
Panel A: Altruistic Donors								
Non-O Donor	102	0.85	(0.01)	0.06	0.81	(0.03)	0.17	0.95
O Donor	62	0.93	(0.01)	0.00	1.88	(0.04)	0.23	2.02
Panel B: Patient-Donor Pairs								
O Patient, Non-O Donor	502	0.28	(0.01)	0.17	0.04	(0.02)	0.14	-0.87
O Patient, O Donor, PRA >= 82%	128	0.35	(0.01)	0.26	0.08	(0.03)	0.13	-0.77
O Patient, O Donor, PRA < 82%	111	0.79	(0.01)	0.03	0.64	(0.03)	0.19	-0.20
Non-O Patient, O Donor, PRA >= 94%	151	0.32	(0.01)	0.26	0.12	(0.03)	0.15	-0.63
Non-O Patient, non-O Donor, PRA < 94%	274	0.84	(0.01)	0.05	0.62	(0.02)	0.23	-0.25
Non-O Patient, O Donor, PRA < 94%	99	0.82	(0.01)	0.03	1.32	(0.03)	0.33	0.61
Panel C: Unpaired Patients								
Non-AB Patients	406	0.20	(0.01)	0.12	-0.01	(0.02)	0.12	-1.03
AB Patients	95	0.36	(0.02)	0.07	0.09	(0.04)	0.14	-0.74

Notes: Categories are determined by regression tree analysis to predict marginal products as a function of whether a submission is a pair or altruist, blood types, and the patient's PRA. Our procedure followed standard recommendations in Friedman et al. (2001). Specifically, we used 10-fold cross-validation to pick the penalty parameter on the number of nodes, required each leaf to have at least 20 observations and pruned a leaf if it does not increase the overall fit by at least 2%. The resulting tree is depicted in Figure C11. Standard errors for the simulations are calculated by following Chapter 12 of Robert and Casella (2004). The within category standard deviation is estimated using shrinkage methods recommended in Morris (1983).

$W^H(\boldsymbol{q}^*) - W^H(\boldsymbol{q}_0)$ equals the area between marginal product curve $\boldsymbol{\nabla f}$ and marginal cost curve $\boldsymbol{\nabla C}$ between $\boldsymbol{q}_0$ and $\boldsymbol{q}^*$. Formally, the deadweight loss is the integral of $\boldsymbol{\nabla f}(\boldsymbol{q}) - \boldsymbol{\nabla C}(\boldsymbol{q})$ as $\boldsymbol{q}$ goes from $\boldsymbol{q}_0$ to $\boldsymbol{q}^*$. This calculation is the multidimensional version of the Harberger triangle formula, that is, the area between the marginal benefit and marginal cost curves.

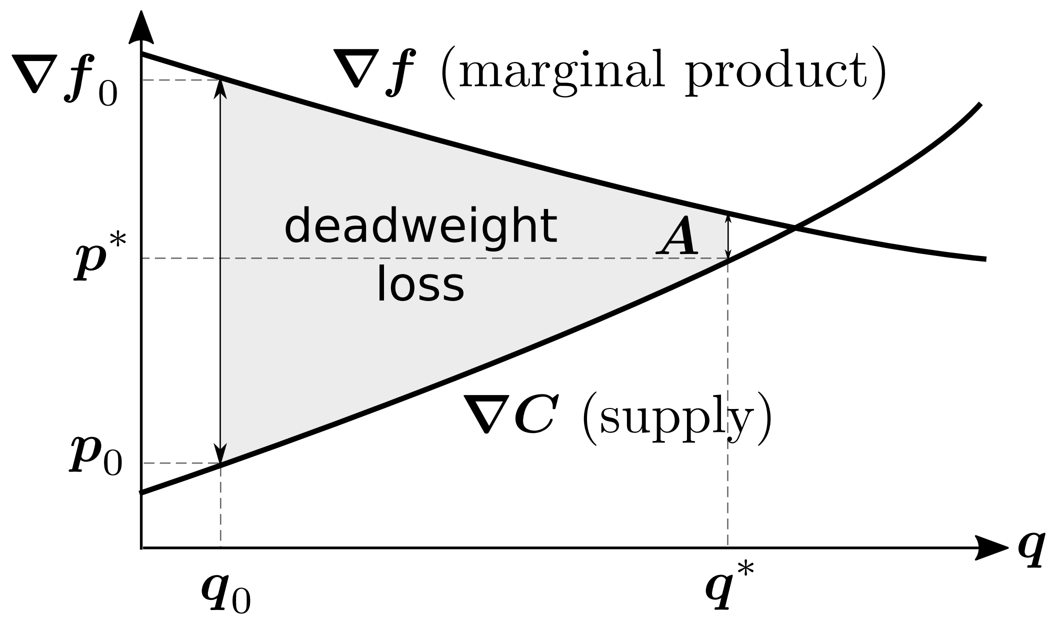


Figure 8: Hospital-Welfare Deadweight Loss from the Current Mechanism

Notes: The deadweight loss from the current mechanism is the shaded area between marginal products and the supply curve. Current rewards are $\boldsymbol{p}_0$, equal to the probability of matching each type of submission, while optimal rewards $\boldsymbol{p}^*$ equal marginal products. Current quantities $\boldsymbol{q}_0$ and rewards $\boldsymbol{p}_0$ are observed. Marginal products $\boldsymbol{\nabla f}$, including the current value $\boldsymbol{\nabla f}_0$ can be calculated from the production function. In contrast, the supply curve $\boldsymbol{\nabla C}$ and optimal rewards $\boldsymbol{p}^*$ and quantities $\boldsymbol{q}^*$ are not observed, and depend on the elasticity of supply. The figure represents multidimensional objects, so that this area is a path integral going from current rewards $\boldsymbol{p}_0$ to optimal rewards $\boldsymbol{p}^*$.

Again, the intuition is similar to that in linear commodity taxation. In linear commodity taxation, the deadweight loss is proportional to the square of the tax wedge. Hence, larger taxes lead to larger welfare losses. The following proposition shows that a similar approximation to the hospital deadweight loss holds in our setting.

Proposition 1. *Consider a strictly positive aggregate supply of pairs of* $\boldsymbol{q}_0$*, which is produced when hospitals choose supply optimally given rewards* $\boldsymbol{p}_0$*. Consider aggregate supply* $\boldsymbol{q}^*$ *and rewards* $\boldsymbol{p}^*$ *that maximize hospital welfare as in Theorem 1. Assume that the matrix* $\boldsymbol{D}^2\boldsymbol{C}(\boldsymbol{q}^*) - \boldsymbol{D}^2\boldsymbol{f}(\boldsymbol{q}^*)$ *is non-singular. Then the deadweight loss in hospital welfare at* $\boldsymbol{q}_0$ *is approximated by*

$$\frac{1}{2}(\boldsymbol{\nabla f}_0 - \boldsymbol{p}_0) \cdot (\boldsymbol{q}^* - \boldsymbol{q}_0)$$

Alternatively, the deadweight loss is approximated by

$$\frac{1}{2}(\boldsymbol{\nabla f}_0 - \boldsymbol{p}_0)[\boldsymbol{D}^2\boldsymbol{C}(\boldsymbol{q}_0) - \boldsymbol{D}^2\boldsymbol{f}(\boldsymbol{q}_0)]^{-1}(\boldsymbol{\nabla f}_0 - \boldsymbol{p}_0)', \tag{3}$$

with the errors bounded by expressions (A10) and (A11) in Appendix A.

These formulas are a multidimensional version of the standard approximation for the Harberger triangle in one-dimensional linear commodity taxation. The first formula is the multidimensional version of the one half base times height formula. The second formula is the equivalent of the one half of the tax wedge squared, times the inverse of the derivative of inverse supply minus the derivative of inverse demand. The second formula shows that the deadweight loss is one half of a quadratic expression in the wedge $\boldsymbol{\nabla f}_0 - \boldsymbol{p}_0$. The term $\boldsymbol{D}^2\boldsymbol{C}(\boldsymbol{q}_0)$ accounts for the fact that a more elastic supply leads to larger deadweight losses. The term $\boldsymbol{D}^2\boldsymbol{f}$ accounts for the change in marginal products in response to a change in $\boldsymbol{q}$. For example, the deadweight loss is lower if increasing the supply of overdemanded pairs results in these pairs becoming less useful.

The proposition shows that estimating the deadweight loss requires estimates of $\boldsymbol{\nabla f}_0 - \boldsymbol{p}_0$, and either $\boldsymbol{q}^* - \boldsymbol{q}_0$ or $\boldsymbol{D}^2\boldsymbol{C}(\boldsymbol{q}_0) - \boldsymbol{D}^2\boldsymbol{f}(\boldsymbol{q}_0)$. It is easy to estimate $\boldsymbol{\nabla f}_0$, $\boldsymbol{p}_0$ and $\boldsymbol{q}_0$ using the observed data and the estimated production function. Unfortunately, we do not have a good estimate of the hospital supply curve, and therefore we cannot directly estimate $\boldsymbol{q}^*$ or $\boldsymbol{D}^2\boldsymbol{C}(\boldsymbol{q}_0)$. Nevertheless, the large wedge between the current private and social incentives suggests that the deadweight loss is significant unless the elasticity of supply is extremely small.

We can use Proposition 1 to formalize this point. Equation (3) provides a path for estimating the deadweight loss under a broad range of assumptions about supply elasticities. We restrict attention to mechanisms that set vectors of rewards for the categories of submissions that are predictive of marginal products in the regression tree analysis above. We estimated the wedge $\boldsymbol{\nabla f}_0 - \boldsymbol{p}_0$ and the curvature matrix $\boldsymbol{D}^2\boldsymbol{f}$ for these categories using our production function. To use equation (3), we need to specify supply elasticities through the matrix $\boldsymbol{D}^2\boldsymbol{C}(\boldsymbol{q}_0)$. One challenge in directly specifying this quantity is that different types of submissions may respond differently to rewards. For example, the submission of hard-to-match types to the system may not substantially decrease when rewards are lowered because there are few other avenues for matching them. Our approach is to calculate the maximum deadweight loss under varying bounds on the maximum elasticity of any type of submission. It allows us to be agnostic about the elasticities of different types of submissions. The deadweight loss is zero when we assume that the maximum elasticity is zero because the submissions will not respond to the rewards system, and we will have that $(q^* = q_0)$. As we increase the bound on the elasticity, submissions respond and the maximum implied deadweight loss increases.

Further, we can repeat this exercise for varying assumptions on cross-elasticities.[45]

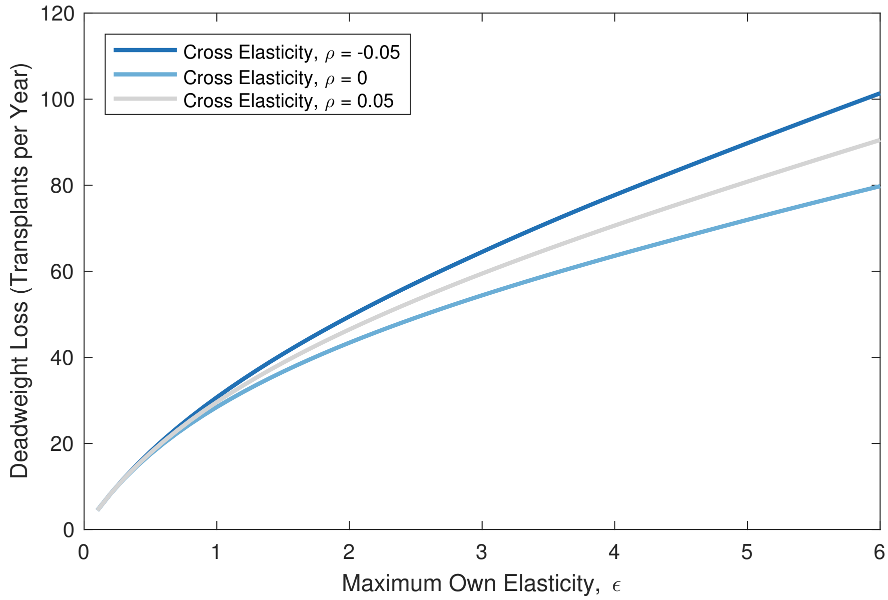


Figure 9: Hospital-Welfare Deadweight Loss from the Current Mechanism

Notes: The estimated hospital-welfare deadweight loss from the current mechanism, using the approximation from Proposition 1, as a function of the elasticity matrix of supply. Maximum own-elasticities are in the horizontal axis.

Figure 9 plots the maximum hospital deadweight loss for bounds on the own-price elasticities elasticities ranging from 0 to 6. The curve in the middle describes the results for zero cross-price elasticities and the other two curves present results for non-zero cross-price elasticities. The hospital deadweight loss is zero if supply is perfectly inelastic, and is increasing in the elasticity. For small elasticities, the hospital deadweight loss increases by about 30 transplants per year for a unit change in the maximum elasticity. The deadweight loss is significant for most of this range and above 40 transplants per year if the maximum elasticity is at least 2. For very high elasticities, the deadweight loss increases at a slower rate because of the curvature of the production function. The deadweight loss at an elasticity of 6 is only between 80 and 100 because the marginal products of the productive types that the optimal mechanism attracts decreases with supply. Although the results for large elasticities are

[45]Specifically, we solved the problem

$$\max_{D^2C(q_0)} \frac{1}{2}(\boldsymbol{\nabla f}_0 - \boldsymbol{p}_0)[\boldsymbol{D^2C}(\boldsymbol{q}_0) - \boldsymbol{D^2f}(\boldsymbol{q}_0)]^{-1}(\boldsymbol{\nabla f}_0 - \boldsymbol{p}_0)'$$

$$\text{s.t.} \left(\frac{\partial^2 C}{\partial^2 q_j}\right)^{-1} \frac{p_{0,j}}{q_{0,j}} \leq \varepsilon \text{ and } \left(\frac{\partial^2 C}{\partial^2 q_j}\right)^{-1} \frac{p_{0,j}}{q_{0,j}} \geq 0$$

$$\left(\frac{\partial^2 C}{\partial q_j \partial q_k}\right)^{-1} \frac{p_{0,k}}{q_{0,j}} = \rho \frac{\left(\frac{\partial^2 C}{\partial^2 q_j}\right)^{-1} \frac{p_{0,j}}{q_{0,j}} + \left(\frac{\partial^2 C}{\partial^2 q_k}\right)^{-1} \frac{p_{0,k}}{q_{0,k}}}{2}.$$

for each value of the bound on elasticities, ε.

subject to greater approximation error, it is unlikely that the deadweight losses come close to the efficiency loss relative to the first-best allocation, even for elasticities of about 6.

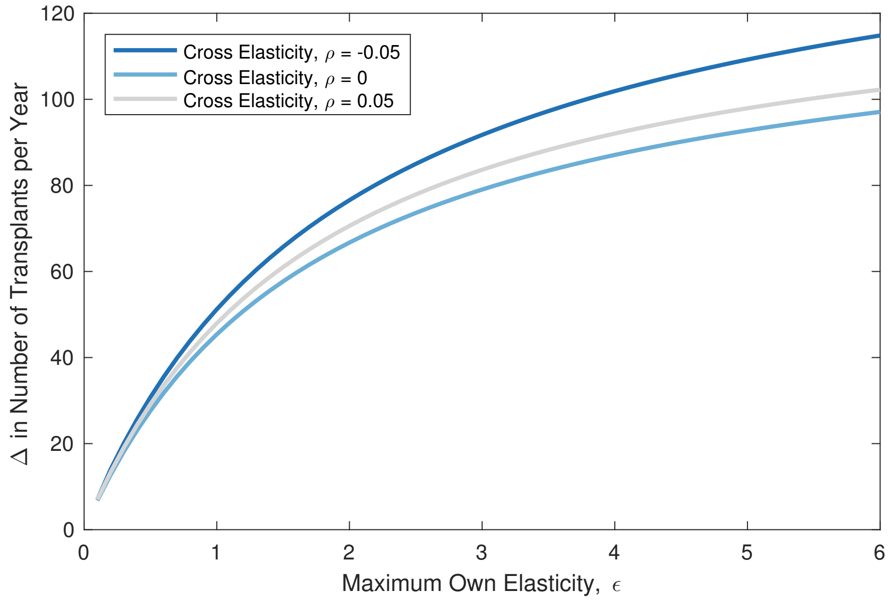


Figure 10: Transplants Lost in the Platform Due to the Current Mechanism

Notes: The estimated lost transplants from the current mechanism, using the approximation from Proposition 1, as a function of the elasticity matrix of supply. Maximum own-elasticities are in the horizontal axis.

The hospital deadweight losses calculated above is a compelling measure of the effects on social welfare if there are no agency problems. However, these calculations will understate the loss in total welfare if hospitals undervalue transplants. Figure 10 shows the total increase in transplants facilitated by the NKR if it adopts the optimal points system. To do this, we added the area under $\boldsymbol{\nabla C}$ to the hospital deadweight loss numbers calculated above (see Figure 8). Because an increase in transplants at the NKR will come at the cost of fewer transplants at hospitals, this calculation overstates the loss in total welfare from the current mechanism. Not surprisingly, the estimated losses are higher than the previous figure. A little over 40 transplants are lost if the maximum elasticity is 1. This number is between 95 and 115 for an elasticity of 6. Therefore, total deadweight loss is higher than hospital deadweight loss, but the two are qualitatively similar.

Taken together, these results imply that addressing the inefficient platform incentives has a large positive impact, unless the elasticity of supply is extremely low. While we do not have quasi-experimental evidence on the magnitude of elasticities, the evidence in Section 3 is typical of markets with elastic supply. Most hospitals only register a subset of their patients with the NKR, and many other hospitals do not participate. Observed behavior is consistent with hospitals responding to financial incentives. These observations are consistent with many hospitals being on the margin, suggesting that supply is at least moderately elastic. Therefore, optimal point mechanisms are not only simple, but are also likely to have a substantial effect on the total number of transplants.

5.2.6 Importance of agency problems and inefficient platform incentives

We now discuss the quantitative importance of each of the two market failures identified above in order to better guide policy. While we cannot decompose the effects of each market failure, the results give us useful information on whether these market failures are important.

First, the misallocation analysis yields a conservative lower bound for the deadweight loss of about 180 transplants per year. The actual deadweight loss is potentially much larger as most specifications yield numbers more than twice as large. Therefore, it must be the case that at least one market failure is quantitatively important.

Second, the Harberger triangle analysis shows that inefficient platform incentives are important as long as supply is not inelastic. The gain in hospital welfare from moving to an optimal mechanism is 35 transplants per year for an elasticity of 1, and 50 transplants per year for an elasticity of 2. Thus, unless supply is extremely inelastic, optimal mechanisms generate appreciable gains in hospital welfare. Moreover, if there are agency problems, the gains in social welfare are even higher because hospitals undervalue transplants. Specifically, hospital welfare deducts the transplant-denominated private cost of hospitals providing more submissions to the platform. When there are agency problems, these private costs are significantly inflated relative to social costs, resulting in a social welfare gain from optimal mechanisms that is higher than the gain in hospital welfare.

Taken together, these results imply that agency problems are important unless elasticities are extremely high. Under the hypothesis that there are no agency problems, hospital welfare equals total welfare, and the optimal mechanism reaches first-best welfare (Theorem 1). Thus, the total deadweight loss in the misallocation analysis must be completely accounted for by the deadweight loss in the Harberger triangle analysis. But, even for a high elasticity of 6, the Harberger triangle yields a deadweight loss of 120, still below our lower bound result of 180 from the misallocation analysis. The only way that these estimates can overlap is if we have both high elasticities, and the approximation in Proposition 1 is significantly downward biased. The bias in the approximation depends on how much the production function deviates from the quadratic Taylor series, so that the bias is high if $\boldsymbol{\nabla f}$ is extremely convex. Thus, the only way that we can attribute all of the deadweight loss to inefficient platform incentives is if we have a combination of high elasticities, $\boldsymbol{\nabla f}$ being sufficiently convex, and the downward biases in our lower bound calculations is small.

The upshot is that a policy that addresses either market failures is likely to be valuable, and generate gains in the order of hundreds of transplants per year. Except under extreme assumptions about the supply function, there are significant gains both from implementing more efficient mechanisms, and from solving agency problems.

6 Discussion and Robustness

6.1 Implementing a point mechanism

Our steady-state model shows that a mechanisms that rewards hospitals with marginal product is efficient. Moreover, our empirical results show that a simple, low-dimensional rewards system would likely achieve sizable efficiency gains. Unfortunately, our simplified steady-state model cannot compare different game forms, and evaluating a dynamic mechanism based on points requires a fully specified dynamic model. This raises the practical and theoretical question of how to design such a mechanism, a task that requires a detailed specification of rules. Incentives created by a detailed set of rules are often too complex to capture in a simple model. For this reason, the design of practical dynamic mechanisms is often based on a combination of theory and practical considerations. While resolving all of these details is beyond the scope of our paper, we now discuss the key theoretical and practical issues.

In both theory and practice, the most well-known mechanism for this kind of problem is what we refer to as a point system. One possible implementation is as follows. Each hospital has a balance of points, which starts at zero. When a hospital makes a submission, the hospital is credited an amount equal to the marginal product of the submission calculated in Table 4 above. When a hospital's patient receives a kidney, the hospital is debited one point. The platform finds optimal matches periodically, as done currently. But, to provide incentives, the platform takes points into account when selecting the match. For example, the platform may always select an optimal match, but choose among optimal matches to minimize the largest point balance or use a weighted lottery that favors hospitals with large balances.

One motivation for this kind of mechanism comes from the dynamic mechanism design literature. Möbius (2001), Hauser and Hopenhayn (2008), Friedman et al. (2006) and Guo and Hörner (2015) call this kind of mechanism a chips, scrips, or token mechanism. Möbius (2001), Hauser and Hopenhayn (2008), and Abdulkadiroğlu and Bagwell (2013) consider dynamic favor exchange, and Guo and Hörner (2015) considers provision of goods to a consumer with stochastic valuations. The general finding of this literature is that token mechanisms, as proposed by Mobius, do better than autarky, but not as well as an optimal dynamic mechanism. Crucially, token mechanisms are close to the first-best if players are patient and there are many time periods. Jackson and Sonnenschein's (2007) general results on linked decisions imply that the inefficiency of token mechanisms declines as square root of the number of periods (see Guo and Hörner, 2015). This is consistent with the literature on monetary economics, where money can achieve high levels of efficiency and trade even with simple institutions (Kiyotaki and Wright, 1989), even though optimal dynamic mechanisms can often improve on money (Kocherlakota, 1998). Thus, the theoretical literature suggests that, even though point systems are not exactly optimal, they are simple and achieve a high level of efficiency.

Another motivation for point mechanisms comes from succesful practical applications. The most obvious related example is fiat money. Money sustains high levels of trade, not only in theoretical models, but also in our everyday lives. Point systems have also been used in market design applications. For example, Prendergast (2017) describes how this kind of mechanism was successfully used to increase the efficiency of food distribution across food banks.

A key issue with applying point systems is that there are several "plumbing" decisions that must be made. Should the matching algorithm impose a strict bound on negative balances? For example the matching algorithm could have a constraint that no hospital's balance goes under, say, -5 points. If so, what is the optimal minimum balance constraint? A tight constraint provides stronger incentives to hospitals, but may reduce efficiency. Should points be credited when pairs are submitted, or should points be credited when pairs are transplanted (in which case the credit would equal marginal product divided by probability of matching)? Rewards at submission time are less noisy, but raise the risk that hospitals will make shill submissions. How often should marginal products be recalculated as the composition of patients and donors in the platform changes? Recalculating them often is complex and reduces transparency, but recalculating infrequently can reduce efficiency.

Understanding these design issues is important for future theoretical and applied research. One interesting direction is to develop theoretical models that inform the tradeoffs between these different rules. Another important direction is to implement these mechanisms in practical settings, to learn what works.

6.2 Maximizing social welfare

Theorem 1 describes mechanisms that maximize hospital welfare. A natural alternative would be to use mechanisms that maximize social welfare. These mechanisms are described in the following proposition.

Proposition 2 (Total Welfare Optimal Mechanisms)**.** *Consider a vector of rewards $\boldsymbol{p}$ and strictly positive aggregate quantity $\boldsymbol{q}$ that maximize social welfare subject to all hospitals choosing supply optimally given $\boldsymbol{p}$, and subject to not promising more transplants than produced. Assume that the production function has constant returns to scale, that private costs are strictly convex. Define the aggregate externality as a function of aggregate quantity $\tilde{\boldsymbol{E}}$ as in equation (A12), and assume that it is smooth at $\boldsymbol{q}$. Then:*

1. *The platform rewards each type of submission with its marginal product, plus an adjustment term,*
$$\boldsymbol{p} = \boldsymbol{\nabla f}(\boldsymbol{q}) + \boldsymbol{A^{SW}},$$
where
$$\boldsymbol{A^{SW}} = \frac{1}{1+\lambda^{SW}}\boldsymbol{\nabla}\tilde{\boldsymbol{E}}(\boldsymbol{q}) - \frac{\lambda^{SW}}{1+\lambda^{SW}}\boldsymbol{q'D^2C}(\boldsymbol{q}).$$
and
$$\lambda^{SW} = \frac{\boldsymbol{\nabla}\tilde{\boldsymbol{E}}(\boldsymbol{q})\cdot\boldsymbol{q}}{\boldsymbol{q'D^2C}(\boldsymbol{q})\boldsymbol{q}}.$$
2. *The adjustment term can be non-zero even with constant returns to scale.*

3. *The optimal rewards attain first-best social welfare if and only if the average externality at the optimum,* $\boldsymbol{\nabla}\tilde{\boldsymbol{E}}(\boldsymbol{q}) \cdot \boldsymbol{q}$*, is zero.*

Part 1 shows that the optimal mechanism rewards submissions by their marginal products plus an adjustment. The adjustment equals an externality term, which is greater for submissions that generate more externalities, minus a shading term, that depends on elasticities. In the first-best, hospitals are rewarded for their marginal contributions to the platform as well as any externalities. However, if there are not enough transplants to pay for the externalities, the planner has to shade rewards. As in optimal linear commodity taxation, it is better to shade rewards for submissions with more inelastic supply.

Part 2 shows that the key difference in this case, relative to Theorem 1, is that the adjustment term is not zero, even for constant returns to scale. Therefore, the optimal rewards depend on more information. To set optimal rewards, one has to know the externalities that are generated by each type of submission. It therefore requires knowledge of the types of submissions for which hospital objectives deviate most from social objectives. Moreover, it is necessary to know the elasticity matrix, to measure how much shading has to be done for each type of submission. Elasticities matter so long as the average externality is non-zero because it results in the multiplier λ^{SW} being non-zero, and an adjustment term that depends on elasticities. Finally, part 3 shows that the optimal reward vector and allocation does not attain first-best social welfare. Therefore, allocations that achieve first-best social welfare require non-linear and complex incentives for hospitals.

The upshot is that using only the kidney exchange mechanism to maximize social welfare, as opposed to hospital welfare, runs into important challenges. Optimal rewards are more complex, depend on more information, and are sensitive to changes in the incentives facing hospitals that can affect overall externalities. These results suggest that solving agency problems is an important complement to improving the design of the kidney exchange mechanism.

6.3 Competing platforms

One natural policy response to fragmentation and the increasing returns to scale estimated earlier is to mandate participation at a single platform, or to merge platforms. These recommendations raise questions about the optimal strategy for competing platforms, and the efficiency costs of imperfect competition. To address these issues, consider a platform that faces an inverse supply of submissions $\boldsymbol{P_S}(\boldsymbol{q})$. For simplicity, assume that the platform has an empire-building objective, where it maximizes the number of transplants $f(\boldsymbol{q})$ facilitated by the platform. The following proposition describes the optimal rewards.

Proposition 3 (Oligopolistic Platforms)**.** *Consider a platform facing a smooth inverse supply curve of submissions* $\boldsymbol{P_S}(\cdot)$*. Consider a vector of rewards* $\boldsymbol{p}$ *and strictly positive aggregate quantity* $\boldsymbol{q}$ *that maximize the number of transplants in the platform subject to not promising more transplants than produced. Assume that the production function has constant returns to scale. Then:*

1. *The platform rewards each type of submission with its marginal product, plus an adjustment term,*
$$\boldsymbol{p} = \boldsymbol{\nabla f}(\boldsymbol{q}) + \boldsymbol{A^C},$$
where
$$\boldsymbol{A^C} = \frac{\boldsymbol{q'DP_S}(\boldsymbol{q})\boldsymbol{q}}{f(\boldsymbol{q})}\boldsymbol{\nabla f}(\boldsymbol{q}) - \boldsymbol{q'DP_S}(\boldsymbol{q}).$$

2. *This adjustment term is non-zero even with constant returns to scale. In particular, rewards are different from the rewards in an optimal mechanism from Theorem 1.*

3. *If supply is perfectly elastic, so that the matrix $\boldsymbol{DP_S}$ is zero, then rewards equal marginal products, as in the optimal rewards in Theorem 1.*

The proposition shows that empire-building platforms deviate from setting rewards equal to marginal products. Instead, the platform subsidizes submissions that are very productive, and whose rewards have a larger effect on supply. To clarify this intuition, consider the case where supply has zero cross elasticities, and own-elasticities denoted ϵ_i. Then the optimal rewards formula simplifies to an analogue of the Lerner index formula:

$$\frac{\partial_i f - p_i}{p_i} = \frac{1}{\epsilon_i}\left(\frac{f - \partial_i f \cdot q_i}{f}\right).$$

This formula describes the optimal mark-down in rewards relative to marginal products. If there is only one type of submission, then the right hand side is trivially equal to zero because f exhibits constant returns to scale. In this case, the optimal rewards are equal the marginal product. When there are multiple types of submissions, then the quantity weighted average of $\partial_i f - p_i$ across submissions must be zero because the platform cannot promise rewards that exceed its product. The expression shows that the platform has incentives to skew the rewards: optimal markdowns are larger for submissions with low elasticities, and submission categories that are less productive on the margin.

The proposition implies that competing, empire-building platforms exploit their market power and set rewards inefficiently. The proposition also implies that platforms set efficient rewards if the market is very competitive. Optimal rewards are close to marginal products if supply is very elastic, i.e. if ϵ_i is close to infinity, or more generally, $\boldsymbol{DP_S}$ is close to zero.

7 Conclusion

This paper documents that the kidney exchange market is highly fragmented, and this fragmentation leads to a real efficiency loss of hundreds of transplants per year. The inefficiency arises due to two standard market failures. First, platforms use inefficient mechanisms that do not reward hospitals according to marginal products of their contributions. This problem induces hospitals to perform inefficient matches even if they seek to help their patients. Second, there are agency problems that make hospitals too sensitive to the costs of participating in kidney exchange platforms relative to the social value of transplants. Our analysis shows that both market failures are likely to be important, and that platforms could use simple alternative mechanisms to substantially increase efficiency.

These findings have both short-term policy implications, and broader implications for the design of kidney exchange markets. There are two clear short-term policy implications. First, existing platforms should experiment with point systems, which are likely to considerably increase efficiency. This recommendation is particularly actionable because it can be implemented by single platforms. Moreover, moving to a point system is likely to help individual platforms expand, and to be in their interest. We caution that deploying these point systems depends on several design details, which can be optimized by theoretical research and practical experiments. Second, third-party payers should subsidize kidney exchange, including going above the purely monetary costs such as exchange fees. The institutional environment and the data suggest that hospitals are likely responsive to monetary, administrative, and logistical costs of participating in kidney exchange platforms, and that this leads to significant welfare losses. Subsidies by Medicare and private payers are likely to alleviate this problem. Moreover, our analysis suggests that this two-pronged approach, that addresses the two market failures separately, is simpler, more robust, and has lower data requirements than mechanisms that address both market failures simultaneously.

Consistent with our results, market participants consider these issues important, and there are initiatives moving in the direction of these policy changes. Regarding point systems, the NKR has started experimenting with a "Center Liquidity Contribution Program" in 2017. Regarding the subsidies, some private insurers have started covering costs of participating in kidney exchange platforms. Our results indicate that there can be large gains from continuing to move in this direction. Other platforms can follow the NKR and try to provide incentives to hospitals. And all platforms can use optimal, data-driven mechanisms, along the lines of Theorem 1. Payers can also improve their subsidy policies by covering not only the explicit costs of kidney exchange, but also aligning the subsidies with the entire costs perceived by hospitals, including administrative and logistic. Future research can contribute to the design and evaluation of these policies.

More broadly, our results raise the question of whether or not to use heavy-handed regulation, such as mandating participation in a single platform. Indeed, the U.K., Netherlands, and Canada (De Klerk et al., 2005; Johnson et al., 2008; Malik and Cole, 2014) mandate participation at a single national program. At a first glance, this approach seems reasonable because of the increasing returns to scale in kidney exchange. However, this approach has the potential disadvantage of reducing competitive incentives for platforms that may otherwise encourage innovation. Indeed, many key innovations in kidney exchange happened through experimentation from the many different players in the market.[46] Moreover, our estimates of returns to scale suggest that it would be efficient to have a few large exchange in the US, and that most of the potential efficiency gain comes from moving the market from individual

[46]Examples include the introduction of non-simultaneous chains (Rees et al., 2009), the development Global Kidney Exchange which allows pairs from development countries to overcome financial barriers (Rees et al., 2017), voucher programs to increase donation for future priority (Veale et al., 2017; Wall et al., 2017), and other operational innovations that reduce frictions and improve matching algorithms.

hospitals to national exchanges. It is plausible that this can be achieved with the narrower policies that we discussed, and without mandating a single monopolistic exchange.

If a single exchange is not mandated, then an important future problem is whether to regulate a future, mature market. The results returns to scale suggest that kidney exchange platforms are likely to be a monopoly, or a concentrated oligopoly (Ellison and Fudenberg, 2003). It is possible that, as the market matures, all platforms adopt point systems. Theory then predicts that platforms have incentives to use inefficient rewards due to market power, to cream-skim the best submissions from competitors. If this becomes a problem, one possible solution is competition. But it may also be in the interest of platforms to coordinate on using efficient rewards.

Our approach in this paper is to integrate matching theory, through the production function, to standard economic models. This may be useful in other problems. In kidney exchange, an important issue is how to get additional patients and donors to participate. Existing proposals include recruiting compatible pairs and subsidizing transplants to pairs from developing countries (Sönmez and Ünver, 2014; Sönmez et al., 2017; Veale et al., 2017). The production function approach clarifies that these ideas are likely to be effective because they can attract submissions with high marginal value. Moreover, the production function can also inform these policies, by identifying which submissions have higher marginal value. More broadly, we use a data-driven approach to design and evaluate mechanism. This approach can be useful in the increasingly common settings where there is rich data, and where optimal market design depends on details that have to be measured empirically.

References

Abdulkadiroğlu, Atila and Kyle Bagwell, "Trust, reciprocity, and favors in cooperative relationships," *American economic journal: Microeconomics*, 2013, *5* (2), 213–259.

Abraham, David J, Avrim Blum, and Tuomas Sandholm, "Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges," in "Proceedings of the 8th ACM conference on Electronic commerce" ACM 2007, pp. 295–304.

Agarwal, Nikhil, Itai Ashlagi, Eduardo Azevedo, Clayton Featherstone, and Omer Karaduman, "What Matters for the Productivity of Kidney Exchange?," *Mimeo, MIT*, 2018.

Akbarpour, Mohammad, Shengwu Li, and Shayan Oveis Gharan, "Thickness and information in dynamic matching markets," 2016.

American Society of Transplant Surgeons, "Kidney Paired Donation: Community Perspectives and Best Practices," 2016.

Anderson, Ross, Itai Ashlagi, David Gamarnik, and Yash Kanoria, "A dynamic model of barter exchange," in "Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms" SIAM 2014, pp. 1925–1933.

_ , _ , _ , Michael Rees, Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver, "Kidney Exchange and the Alliance for Paired Donation: Operations Research Changes the Way Kidneys Are Transplanted," *Interfaces*, 2015, *45* (1), 26–42.

Armstrong, Mark, "Competition in two-sided markets," *The RAND Journal of Economics*, 2006, *37* (3), 668–691.

Ashlagi, Itai and Alvin E Roth, "Free riding and participation in large scale, multi-hospital kidney exchange," *Theoretical Economics*, 2014, *9* (3), 817–863.

_ , David Gamarnik, Michael A Rees, and Alvin E Roth, "The need for (long) chains in kidney exchange," Technical Report, National Bureau of Economic Research 2012.

_ , Maximillien Burq, Patrick Jaillet, and Vahideh Manshadi, "On matching and thickness in heterogeneous dynamic markets," *arXiv preprint arXiv:1606.03626*, 2016.

Becker, Gary S and Julio Jorge Elias, "Introducing incentives in the market for live and cadaveric organ donations," *The Journal of Economic Perspectives*, 2007, *21* (3), 3–24.

Bingaman, A W, F H Wright Jr, M Kapturczak, L Shen, S Vick, and C L Murphey, "Single-Center Kidney Paired Donation: The Methodist San Antonio Experience," *American Journal of Transplantation*, 2012, *12* (8), 2125–2132.

Boiteux, Marcel, "Sur la gestion des Monopoles Publics astreints a l'equilibre budgetaire," *Econometrica*, 1956, *24* (1), 22–40.

Chetty, Raj, "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods," *Annu. Rev. Econ.*, 2009, *1* (1), 451–488.

Clemens, Jeffrey and Joshua D Gottlieb, "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?," *American Economic Review*, 2014, *104* (4).

Cowan, Nigel C., H A Gritsch, Nima Nassiri, Joseph Sinacore, and Jeffrey L Veale, "Broken Chains and Reneging: A Review of 1748 Kidney Paired Donation Transplants.," *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, 2017, *17 9*, 2451–2457.

Danovitch, Gabriel M, *Handbook of kidney transplantation*, Lippincott Williams & Wilkins, 2009.

Dickerson, John P, Ariel D Procaccia, and Tuomas Sandholm, "Dynamic Matching via Weighted Myopia with Application to Kidney Exchange," in "AAAI" 2012.

Dixit, Avinash and Agnar Sandmo, "Some simplified formulae for optimal income taxation," *The Scandinavian Journal of Economics*, 1977, pp. 417–423.

Ellison, Blake, "A Systematic Review of Kidney Paired Donation: Applying Lessons from Historic and Contemporary Case Studies to Improve the US Model," *Mimeo, University of Pennsylvania*, 2014.

Ellison, Glenn and Drew Fudenberg, "Knife-Edge or Plateau: When Do Market Models Tip?," *The Quarterly Journal of Economics*, 2003, *118* (4), 1249–1278.

Farhi, Emmanuel and Xavier Gabaix, "Optimal Taxation with Behavioral Agents," *Mimeo, Harvard University*, 2017.

Friedman, Eric J, Joseph Y Halpern, and Ian Kash, "Efficiency and Nash equilibria in a scrip system for P2P networks," in "Proceedings of the 7th ACM conference on Electronic commerce" ACM 2006, pp. 140–149.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani, *The elements of statistical learning*, Vol. 1, Springer series in statistics New York, 2001.

Gentry, Sommer E, R A Montgomery, B J Swihart, and Segev D L, "The roles of dominos and nonsimultaneous chains in kidney paired donation," *American Journal of Transplantation*, 2009, *9* (6), 1330–1336.

Gigerenzer, Gerd and Daniel G Goldstein, "Reasoning the fast and frugal way: models of bounded rationality.," *Psychological review*, 1996, *103* (4), 650.

_ **and Stephanie Kurzenhaeuser**, "Fast and frugal heuristics in medical decision making," *Science and medicine in dialogue: Thinking through particulars and universals*, 2005, pp. 3–15.

Guo, Y and J Hörner, "Dynamic mechanisms with money," Technical Report, Mimeo, Yale University 2015.

Hauser, Christine and Hugo Hopenhayn, "Trading favors: Optimal exchange and forgiveness," *Collegio Carlo Alberto Carlo Alberto Notebooks*, 2008, *88.*

Held, Philip J, F McCormick, A Ojo, and John P Roberts, "A cost-benefit analysis of government compensation of kidney donors," *American Journal of Transplantation*, 2016, *16* (3), 877–885.

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 2009, *124* (4), 1403–1448.

Irwin, F D, A F Bonagura, S W Crawford, and M Foote, "Kidney paired donation: a payer perspective," *American Journal of Transplantation*, 2012, *12* (6), 1388–1391.

Jackson, Matthew O and Hugo F Sonnenschein, "Overcoming incentive constraints by linking decisions," *Econometrica*, 2007, *75* (1), 241–257.

Jensen, Michael C and William H Meckling, "Theory of the firm: Managerial behavior, agency costs and ownership structure," *Journal of financial economics*, 1976, *3* (4), 305–360.

Johnson, Rachel J, Joanne E Allen, Susan V Fuggle, J Andrew Bradley, Chris Rudge et al., "Early experience of paired living kidney donation in the United Kingdom," *Transplantation*, 2008, *86* (12), 1672–1677.

Kiyotaki, Nobuhiro and Randall Wright, "On money as a medium of exchange," *Journal of Political Economy*, 1989, *97* (4), 927–954.

Klerk, Marry De, Karin M Keizer, Frans HJ Claas, Marian Witvliet, Bernadette JJM Haase-Kromwijk, and Willem Weimar, "The Dutch national living donor kidney exchange program," *American Journal of Transplantation*, 2005, *5* (9), 2302–2305.

Kocherlakota, Narayana R, "Money is memory," *Journal of Economic Theory*, 1998, *81* (2), 232–251.

Kolstad, Jonathan T, "Information and quality when motivation is intrinsic: Evidence from surgeon report cards," *The American Economic Review*, 2013, *103* (7), 2875–2910.

Kwak, JY, OJ Kwon, Kwang Soo Lee, Chong Myung Kang, Hae Young Park, and JH Kim, "Exchange-donor program in renal transplantation: a single-center experience," *Transplantation Proceedings*, 1999, *31* (1), 344–345.

Malik, Shafi and Edward Cole, "Foundations and principles of the Canadian living donor paired exchange program," *Canadian journal of kidney health and disease*, 2014, *1* (1), 6.

Marschak, Jacob and William H Andrews, "Random simultaneous equations and the theory of production," *Econometrica*, 1944, pp. 143–205.

Massie, Allan B, Sommer E Gentry, Robert A Montgomery, Adam A Bingaman, and Dorry L Segev, "Center-Level Utilization of Kidney Paired Donation," *American Journal of Transplantation*, 2013, *13* (5), 1317–1322.

Möbius, Markus, "Trading favors," *Mimeo, Microsoft Research*, 2001.

Morris, Carl N, "Parametric empirical Bayes inference: theory and applications," *Journal of the American Statistical Association*, 1983, *78* (381), 47–55.

National Kidney Registry, "Member Center Terms& Conditions," http://www.kidneyregistry.org/docs/NKR_MC_Terms_Conditions.pdf 2016.

Olley, G Steven and Ariel Pakes, "The dynamics of productivity in the telecommunications equipment industry," *Econometrica*, 1996, *64* (6), 1263–1297.

Orandi, BJ, JM Garonzik-Wang, Allan B Massie, Andrea A Zachary, JR Montgomery, KJ Van Arendonk, Mark D Stegall, SC Jordan, J Oberholzer, TB Dunn et al., "Quantifying the risk of incompatible kidney transplantation: a multicenter study," *American Journal of Transplantation*, 2014, *14* (7), 1573–1580.

Pathak, Parag, "What Really Matters in Designing School Choice Mechanisms," *Advances in Economics and Econometrics, 11th World Congress of the Econometric Society*, 2017.

Prendergast, Canice, "The Allocation of Food to Food Banks," *University of Chicago, mimeo*, 2017.

Ramsey, Frank P, "A Contribution to the Theory of Taxation," *The Economic Journal*, 1927, *37* (145), 47–61.

Rees, Michael A, Jonathan E Kopke, Ronald P Pelletier, Dorry L Segev, Matthew E Rutter, Alfredo J Fabrega, Jeffrey Rogers, Oleh G Pankewycz, Janet Hiller, Alvin E Roth, and others, "A nonsimultaneous, extended, altruistic-donor chain," *New England Journal of Medicine*, 2009, *360* (11), 1096–1101.

_ , Mark A Schnitzler, EY Zavala, James A Cutler, Alvin E Roth, Frank D Irwin, Stephen W Crawford, and Alan B Leichtman, "Call to develop a standard acquisition charge model for kidney paired donation," *American Journal of transplantation*, 2012, *12* (6), 1392–1397.

_ , Ty B Dunn, Christian S Kuhr, Christopher L Marsh, Jeffrey Rogers, Susan E Rees, Alejandra Cicero, Laurie J Reece, Alvin E Roth, Obi Ekwenna et al., "Kidney exchange to overcome financial barriers to kidney transplantation," *American Journal of Transplantation*, 2017, *17* (3), 782–790.

Robert, Christian and George Casella, "Monte Carlo Statistical Methods Springer-Verlag," *New York*, 2004.

Rochet, Jean-Charles and Jean Tirole, "Platform competition in two-sided markets," *Journal of the european economic association*, 2003, *1* (4), 990–1029.

Roth, Alvin E, "Repugnance as a Constraint on Markets," *The Journal of Economic Perspectives*, 2007, *21* (3), 37–58.

_ and Elliott Peranson, "The redesign of the matching market for American physicians: Some engineering aspects of economic design," *The American Economic Review*, 1999, *89* (4), 748.

_ , Tayfun Sönmez, and M Utku Ünver, "Transplant Center Incentives in Kidney Exchange," *Mimeo, Boston College*, 2005.

_ , _ , and _ , "Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences," *The American economic review*, 2007, pp. 828–851.

_ , _ , _ , Francis L Delmonico, and Susan L Saidman, "Utilizing list exchange and nondirected donation through chain paired kidney donations," *American Journal of transplantation*, 2006, *6* (11), 2694–2705.

Russell, Stuart and Peter Norvig, *Artificial Intelligence: A Modern Approach*, second ed., Prentice Hall, 2003.

Saez, Emmanuel, "Using elasticities to derive optimal income tax rates," *The review of economic studies*, 2001, *68* (1), 205–229.

Saidman, Susan L, Alvin E Roth, Tayfun Sönmez, M Utku Ünver, and Francis L Delmonico, "Increasing the opportunity of live kidney donation by matching for two-and three-way exchanges," *Transplantation*, 2006, *81* (5), 773–782.

Sönmez, Tayfun and M Utku Ünver, "Altruistically unbalanced kidney exchange," *Journal of Economic Theory*, 2014, *152*, 105–129.

Sönmez, Tayfun, M Utku Unver, and M Bumin Yenmez, "Incentivized kidney exchange," Technical Report, Boston College Department of Economics 2017.

USRDS, United States Renal Data System, *2013 USRDS annual data report: Epidemiology of kidney disease in the United States*, Vol. 2, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2013.

_ , *2016 USRDS annual data report: Epidemiology of kidney disease in the United States*, Vol. 2, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2016.

Veale, Jeffrey L, Alexander M Capron, Nima Nassiri, Gabriel Danovitch, H Albin Gritsch, Amy Waterman, Joseph Del Pizzo, Jim C Hu, Marek Pycia, Suzanne McGuire et al., "Vouchers for Future Kidney Transplants to Overcome'Chronological Incompatibility'Between Living Donors and Recipients.," *Transplantation*, 2017.

Wall, Anji E, Jeffrey L Veale, and Marc L Melcher, "Advanced Donation Programs and Deceased Donor Initiated Chains-2 Innovations in Kidney Paired Donation.," *Transplantation*, 2017.

Weyl, E Glen, "A price theory of multi-sided platforms," *The American Economic Review*, 2010, *100* (4), 1642–1672.

Wolfe, Robert A, Valarie B Ashby, Edgar L Milford, Akinlolu O Ojo, Robert E Ettenger, Lawrence YC Agodoa, Philip J Held, and Friedrich K Port, "Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant," *New England Journal of Medicine*, 1999, *341* (23), 1725–1730.