AEA Papers and Proceedings 2018, 108: 334–340
https://doi.org/10.1257/pandp.20181077

INNER WORKINGS OF ORGAN MARKETS AND ORGAN ALLOCATION[‡]

What Matters for the Productivity of Kidney Exchange?[†]

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

Almost 100,000 patients are currently waiting for a lifesaving kidney transplant. Buying organs is illegal in the United States. Kidney Exchange (KE), therefore, presents a unique opportunity for patients with a living but incompatible donor (Roth, Sönmez, and Ünver 2004). In 2017, KE facilitated approximately 15 percent of all living donor transplants in the United States.[1] In addition to increasing the quality and length of life, each transplant saves several hundreds of thousands of dollars on health-care expenditure over remaining on dialysis.

The goal of this paper is to describe the technology with which patients and donors are matched in KE, and to understand what drives the productivity of these platforms. Mechanisms used by various KE platforms have been based on insights from matching theory, but their implementations incorporate unmodeled real-world considerations in varied ways. As we discuss below, the various KE platforms in the United States take different approaches to resolving these logistical issues. Next, we use administrative data from the three largest national KE platforms to quantify how these features affect the fraction of patients and donors that are matched.

I. The Production Function Approach

We study this market using the empirical production function approach developed in Agarwal et al. (2017). This framework views a KE platform as a neoclassical firm—it procures inputs (patients and donors) from hospitals and uses them to produce an output (transplants). Formally, the platform uses the vector of inputs $\mathbf{q} = (q_i)_{i=1,\ldots,I}$, where q_i is the quantity of type i submissions, and produces $f(\mathbf{q})$ transplants. The rewards for hospital submissions are the transplants that it is allocated. This reformulation is based on institutional features of the market, specifically the reasons why transplants are a good numeraire in this market. Agarwal et al. (2017) estimate this production function using data from the National Kidney Register (NKR).

This view has been implicit in the early literature on KE. Roth, Sönmez, and Ünver (2007) calculate a production function in a sufficiently large market without frictions and only patient-donor pairs. The relevant types in the limiting economy depend only on the blood types of the patient and donor, and therefore $I = 16$. They derive a linear limit production function—over-demanded pairs (a pair in which the patient is blood-type compatible with the related donor

[‡]*Discussants:* Utku Ünver, Boston College; Glen Weyl, Microsoft Research; Benjamin R. Handel, University of California-Berkeley.

*Agarwal: MIT and NBER (email: agarwaln@mit.edu); Ashlagi: Stanford University (email: iashlagi@stanford.edu); Azevedo: Wharton, University of Pennsylvania (email: eazevedo@wharton.upenn.edu); Featherstone: Wharton, University of Pennsylvania (email: claytonf@wharton.upenn.edu); Karaduman: MIT (email: omerk@mit.edu). We thank Utku Ünver for a helpful discussion of this paper. Agarwal and Ashlagi acknowledge support from the National Science Foundation grants SES-1729090 and SES-1254768, respectively. The data reported here have been supplied by the United Network for Organ Sharing as the contractor for the Organ Procurement and Transplantation Network. 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 US Government. Results on the UNOS KPD program is based on OPTN data as of November 10, 2017.

[†]Go to https://doi.org/10.1257/pandp.20181077 to visit the article page for additional materials and author disclosure statement(s).

[1] We count all transplants categorized as paired and anonymous donations as kidney exchanges. Source: https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/.

but they have a different blood type) generate two additional transplants when they join the platform, while under-demanded pairs (pairs in which the patients and donors are not blood-type compatible) generate 0 transplants.

Indeed, many papers in this literature are motivated by the goal of making KE more productive. In our framework, one can view many results in this literature as improving a design feature A that affects the production function, $f(\mathbf{q};A)$. For example, many papers consider how the size of cycles and chains impact productivity. Others directly study how matching algorithms impact productivity. Other research has been devoted to finding ways to improve the composition of types, that is, change $\mathbf{q}$, to make KE more productive. Sönmez and Ünver (2013) survey these results.

The improvements identified by these studies usually require deep institutional insights or theoretical work, with the production function implicitly governing the resulting benefits. Indeed, the marginal products for (immunologically easy to match) over-demanded and under-demanded pairs derived by Roth, Sönmez, and Ünver (2007) are qualitatively similar to the ones estimated by Agarwal et al. (2017). These estimates are based on data from the NKR, and detailed knowledge of the logistics and algorithms used by the platform. This alignment of answers is reassuring for both theoretical and empirical analyses of this market.

We therefore view this empirical approach as complementary to the theory on KE design by providing a quantitative counterpart. It allows us to investigate the magnitude of trade-offs identified in the theory using estimates that are finely tuned to the institutional environment and the engineering details of KE markets. In addition, the approach may help us to identify and develop solutions to the most important hurdles currently facing KE.

II. Kidney Exchange Platforms

A. *Logistics and Frictions in Kidney Exchange*

The three largest multi-hospital platforms in the United States are the Alliance for Paired Donation (APD), the United Network for Organ Sharing kidney (UNOS), and the National Kidney Registry (NKR). In addition, there are some regional and many single center platforms, with Methodist Hospital in San Antonio being the largest.

These KE platforms have a pool of registered patients and donors, most of them paired. These patients and donors are submitted to the platforms by various member hospitals. The platforms periodically run algorithms to match patients and donors for KE. These exchanges take the form of either cycles, involving only pairs of biologically incompatible patients and donors, or chains that are initiated by an altruistic donor with no related patient. Cycles are typically limited to two or three pairs due to logistical constraints, while chains can be longer.

We now discuss key logistical details that can influence the fraction of its patients that a platform is successful at transplanting.

Submissions.—Participation in a KE platform is not mandatory. Hospitals are the key decision-makers that select which pairs to submit to the platform. They may participate in multiple platforms. The types of patients and donors submitted to a platform can determine the total fraction transplanted. For example, a platform that has many altruistic donors can use chains, and therefore will likely be able to match more patients than a platform with fewer altruistic donors. Even within pairs, the blood-types and immune sensitivity is likely to be important. Moreover, patients and donors often leave platforms before they match, either because they receive a transplant elsewhere, or because the patient passes away or becomes untransplantable.

Matching Procedures.—Most national platforms use optimization algorithms to propose exchanges in an existing pool. Some platforms place priorities to various transplants. These optimization algorithms are usually myopic and triggered *periodically*, for example daily, weekly, bi-weekly or longer. Platforms also need to decide whether and when to end the chain by using a donor from a pair to transplant a patient on the deceased donor waiting list. In principle, the chain can be continued by using this last donor as a *bridge donor* to initiate a new chain.

Consumating Matches.—Patients registered at the exchange specify minimum acceptance criteria for donors (e.g., age, BMI) and are required to exhaustively list antibodies. Nonetheless, before proposals from the algorithms proceed to transplantation, (i) patients (and their doctors) must agree to the transplant, and (ii) a final tissue-type (crossmatch) test must be conducted to limit the chances of organ rejection.

These failures result in frictions that effectively make the market thin. First, processes such as medical tests require time and cause delays of days to a couple weeks. In the interim, the patients and donors in the proposed match cannot be matched with others in the pool, effectively making the pool smaller. Second, even though a patient-donor pair may be biologically compatible, they may not be *transplantable*, making the effective compatibility graph thinner. Therefore, the platform operates under incomplete information about the transplants that can be carried out and only learns over time.

B. *Differences across Platforms*

Table 1 describes two important ways in which the implementation in the three largest national platforms differ: the constraint on the chain segment length and the frequency of matching. These rules have evolved over time. Platforms in the United States have moved toward matching more frequently and have experimented with different priorities that are assigned to various types of patients.

The platforms also differ in the logistics of consumating proposed matches. For example, the APD maintains a laboratory with blood samples so that it can conduct final compatibility tests, called a crossmatch, in-house and on demand. In contrast, hospitals participating in the NKR and UNOS have to ship blood samples, and obtaining the results from medical tests can take several days to a couple weeks. Patients and donors are expected to decide upon a proposed match within one, two, and four days at the NKR, APD, and UNOS, respectively. These periods were longer in the past.

Refusal rates for proposed matches also vary across platforms. The chance that a proposed match is declined can be as high as 30 percent at some platforms, but is closer to 20 percent for the NKR. The accuracy of crossmatch (tissue-type) tests also varies because proposed matches are only based on a "virtual crossmatch" that uses the reported antibodies of the patient (Ashlagi et al. 2017; Dickerson, Procaccia, and Sandholm 2013). Some platforms require more information than others at registration. The APD, which has a blood lab, and single center exchange programs can circumvent some of these issues by performing medical tests in-house.

Table 1—Operational Differences across Platforms

	Max chain (segment) length	Match frequency
APD	4	Daily
UNOS	4	Bi-weekly
NKR	∞	Daily

Most platforms now request hospitals to prespecify unacceptable donor characteristics to limit refusals after an offer has been made and sometimes impose penalties for non-compliance. This is done to reduce the failure rate, which can significantly affect productivity. Platforms take different approaches to resolving these issues. For example, a high matching frequency allows the platform to learn the acceptable transplants more quickly.

Platforms in other countries also differ along these dimensions, but single national platforms are more common than in the United States. Canada, United Kingdom, Netherlands, and Australia have a national platform with mandatory participation. These national platforms identify exchanges only every three or four months in contrast to the very frequent matching in the United States (Ferrari et al. 2014). This long interval allows these national platforms re-optimize after proposals have failed. France, Poland, and Portugal do not organize chains because altruistic donation is illegal. We refer the reader to Biró et al. (2017) for a more comprehensive survey of KE practices in Europe.

Table 2 summarizes the number of patient-donor pairs and altruistic donors that register in each of the three national exchanges and the number of transplants in these exchanges. The NKR is the largest in terms of the number of pairs, altruistic donors, and number of transplants. The table points to its significant advantage, particularly over UNOS, in terms of the number of altruistic donors as important in its ability to facilitate a large number of transplants. These donors allow a platform to initiate chains that are very useful when organizing KE.

Table 3 summarizes the types of pairs and donors submitted to each of the three platforms,

TABLE 2—REGISTRATIONS AND TRANSPLANTS IN APD, UNOS, AND NKR

	2012	2013	2014
APD			
Pairs	140	171	155
Altruistic donors	2	5	5
Transplants	31	39	37
UNOS			
Pairs	266	297	344
Altruistic donors	16	10	4
Transplants	10	52	46
NKR			
Pairs	354	468	443
Altruistic donors	43	70	51
Transplants	181	307	276

Notes: Submissions and transplants conducted between 2012 and 2014. For the NKR, the statistics do not include submissions and transplants prior to April 2, 2012 and after December 4, 2014.

and the number of participating hospitals between 2012 and 2014. There are fewer O donors than O patients since many O donors are compatible with their intended recipients and are not interested in KE. This makes O donors particularly scarce and valuable, particularly when the patient is not blood type O. This combination of blood types makes the pair overdemanded. Conversely, O patients are in abundance and there are many under-demanded pairs in KE. Moreover, all platforms have a very high fraction of patients that have very sensitive immune systems.

III. Measuring the Drivers of Productivity

This section presents simulations to assess the importance of the composition of the pool and the implementation decisions discussed above for platform productivity. We use simulations that vary these dimensions from the baseline empirical production function developed in Agarwal et al. (2017), which is based on the data from and practices of the NKR.

The figures below plot the average product, $f(\mathbf{q}, A)/|\mathbf{q}|$, because the production function described in Section I is high dimensional. We only count the total number of donors registered in the platform when we calculate $|\mathbf{q}|$ because hospitals have a very large number of patients without a related donor waiting on the deceased donor list. Therefore, the average product is identical to the fraction of donors transplanted. The exercises below show how the average product varies with platform size and various features of a KE platform.

We refer the reader to Agarwal et al. (2017) for details on the simulation. Briefly, the production function is based on the practices of the NKR; the rate of submissions of various types, $\mathbf{q}$, and the departure rates are estimated using the data from the NKR; and frictions in consummating matches are calibrated to match the transplant rate. In the base case, each of the two phases of post proposal acceptances incurs a delay of 14 days, each proposed transplant has a failure rate of 20 percent in each phase, and the algorithm is run daily.[2]

A. *Frequency of Matching*

As mentioned above, platforms differ on how frequently they run their matching algorithms. Figure 1 presents the baseline estimates from the NKR in which matches are run daily, and then moves to lower frequencies of every three days, weekly, and bi-weekly. The baseline estimates based on daily matches indicate that platform scale matters. Agarwal et al. (2017) show that these returns to scale are an important driver of overall efficiency in the US market for KE.

Less frequent matching can result in more possible matches by creating a thicker pool, but can also result in patients and donors departing unmatched in the interim. Remarkably, the figure shows that matching daily performs the best. Intervals of up to a week yield similar results, and the differences widen in large pools. In fact, for very large pools, bi-weekly matches result in about 35 percent of donors being matched while daily matches result in over 50 percent of donors being matched.

This result is consistent with biological constraints involved in KE. There is always a large supply of under-demanded pairs in large exchanges. Therefore, an O donor that is submitted can immediately and efficiently be matched to one of the O patients in waiting. When the

[2] This delay is calibrated to fit the NKR outcomes, but is often shorter in practice. The results are similar for shorter delays but higher frictions. We estimate the average product using the time-average from a simulated Markov chain. We use 500,000 simulation days with a burn-in to ensure convergence.

Table 3—Composition of Various Kidney Exchange Platforms

	O Donors	O Patients	Over-demanded	Under-demanded	Highly sensitized	Hospitals
APD	34.8%	60.7%	15.2%	43.1%	33.0%	40
UNOS	33.8%	60.2%	15.1%	42.9%	33.6%	99
NKR	31.9%	58.6%	13.9%	41.9%	30.8%	75

Notes: A pair is over-demanded if the patient is blood-type compatible with the related donor but has a different blood type. Under-demanded pairs include O patients with non-O donors or AB donors with non-AB patients. A highly sensitized patient has at most a 10 percent chance to tissue-type match with a randomly chosen donor ($PRA \geq 90$). A hospital is counted if it submitted at least one pair during the sample period. Sample includes all patients and donors for UNOS and APD registered between January 1, 2012 and December 31, 2014. For the NKR, the statistics do not include submissions prior to April 2, 2012 and after December 4, 2014.

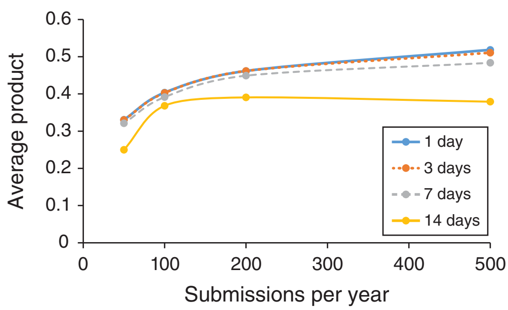

Figure 1. Effect of Match Frequency

Notes: Average fraction of donors transplanted. Each plot represents the number of days between two runs of the matching algorithm.

submission rate is low, the supply of under-demanded pairs is smaller and not every O donor can be matched upon submission to an O patient. But matching infrequently does not have large benefits because only a handful of pairs are submitted to a platform each weak. These conclusions are also consistent with results using data from the APD and the Methodist Hospital in San Antonio reported in Ashlagi et al. (2017).

Moreover, as discussed earlier, frequent proposals effectively allow the platform to test multiple possible transplants and resolve uncertainties about whether a patient is willing to accept a donor. This effect creates a particularly high cost of waiting in large platforms, and is reflected in the simulations with bi-weekly matching.

B. *Frictions*

We now assess the effects of reducing the frictions described in Section IIA. Figure 2 investigates the effects of shortening delays incurred by acceptance decisions and crossmatch tests. It also compares the baseline results with a two week wait for each of these phases with a world in which acceptances and crossmatch tests are all pre-resolved. These exercises are intended to understand the extent to which single-center platforms and those with in-house blood labs may be able to ease the logistics of coordinating acceptance decisions and medical tests.[3]

Figure 2 shows that these frictions are important, and can influence the productivity of a platform by close to 20 percent. Frequent matching and short delays are as good as pre-resolving potentially declined transplants. Indeed, platforms including the NKR and the APD are actively trying to reduce delays. These logistical differences have received little attention in the literature on the economics of KE, but translate to a substantial number of transplants.

C. *Pool Composition*

Agarwal et al. (2017) find that there is significant heterogeneity in the fraction of pairs that a hospital submits to the NKR. Indeed, hospitals that conduct most of their KEs through the NKR submit somewhat easier to match patients and donors.[4] Figure 3 shows that the composition of the patient-donor pool is an important driver of platform productivity. Indeed, patients and donors sampled only from hospitals with a high participation rate (top quartile) are easier

[3] The NKR recently instituted a policy requiring acceptance decisions within a day and crossmatch results within a week.

[4] Details available upon request.

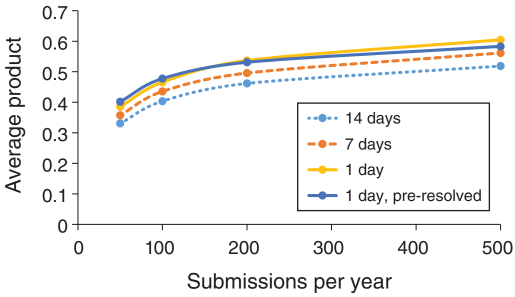

FIGURE 2. EFFECT OF FRICTIONS

Notes: Average fraction of donors transplanted. Each plot represents the number of days before a patient accepts an offer and a crossmatch is performed.

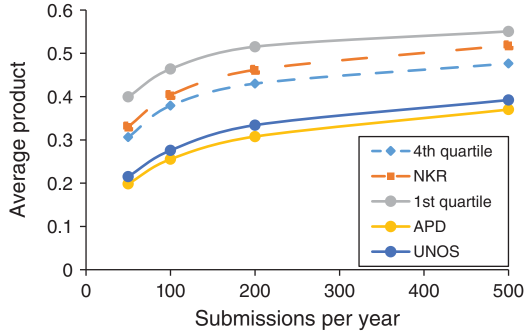

FIGURE 3. EFFECT OF POOL COMPOSITION

Notes: Average fraction of donors transplanted. Each plot represents a subset of hospitals, by their participation rate (including all hospitals), from which registrations are sampled.

to match than the general pool at the NKR. The APD and UNOS, in particular, have few altruistic donors (see Table 2), and can only transplant a much smaller share of all pairs.[5] These results are consistent with the hypotheses in Roth, Sönmez, and Ünver (2005) who suggest improving the pool composition by encouraging the participation of compatible pairs in KE.

IV. Conclusion

Kidney exchange is now responsible for a significant fraction of living donor transplants, but many challenges remain. Platforms implement different algorithms and many frictions reduce the total number of transplants. In addition, the total number of transplants may be impeded by a composition of patients and donors that is particularly hard to match.

This article illustrates that understanding the production function can help us identify the most important directions for improving the technology and logistics of KE. Engaging with these engineering and plumbing aspects are central to the endeavor implementing economic insights and theory into the real world (Roth 2002; Duflo 2017).

[5] These simulations sample donors and patients from the NKR using weights to mimic the distribution of types in the APD and UNOS. This simplifies our analysis because we can use the same set of medical characteristics for all simulations.

REFERENCES

Agarwal, Nikhil, Itai Ashlagi, Eduardo Azevedo, Clayton Featherstone, and Ömer Karaduman. 2017. "Market Failure in Kidney Exchange." Unpublished.

Ashlagi, Itai, Adam Bingaman, Maximilien Burq, Vahideh Manshadi, David Gamarnik, Cathi Murphey, Alvin E. Roth, Marc L. Melcher, and Michael A. Rees. 2017. "The Effect of Match-Run Frequencies on the Number of Transplants and Waiting Times in Kidney Exchange." *American Journal of Transplantation*. https://dx.doi.org/10.1111/ajt.14566.

Biró, Péter, Lisa Burnapp, Bernadette Haase, Aline Hemke, Rachel Johnson, Joris van de Klundert, and David Manlove. 2017. *First Handbook of the COST Action CA15210: European Network for Collaboration on Kidney Exchange Programmes (ENCKEP)*. Brussels: European Cooperation in Science and Technology.

Dickerson, John P., Ariel D. Procaccia, and Tuomas Sandholm. 2013. "Failure-Aware Kidney Exchange." *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*: 323–40.

Duflo, Esther. 2017. "The Economist as Plumber." National Bureau of Economic Research Working Paper 23213.

Ferrari, Paolo, Willem Weimar, Rachel J. Johnsn, Wai H. Lim, Kathryn J. Tinckam. 2014. "Kidney Paired Donation: Principles, Protocols and Programs." *Nephrology Dialysis Transplantation* 30 (8): 1276–85.

Roth, Alvin E. 2002. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica* 70 (4): 1341–78.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver. 2004. "Kidney Exchange." *Quarterly Journal of Economics* 119 (2): 457–88.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver. 2005. "A Kidney Exchange Clearinghouse in New England." *American Economic Review* 95 (2): 376–80.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver. 2007. "Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences." *American Economic Review* 97 (3): 828–51.

Sönmez, Tayfun, and M. Utku Ünver. 2013. "Market Design for Kidney Exchange." In *The Handbook of Market Design*, edited by Nir Vulkan, Alvin E. Roth, and Zvika Neeman, 93–137. New York: Oxford University Press.