Global Kidney Chains*

Afshin Nikzad,† Mohammad Akbarpour,‡
Michael A. Rees,§ Alvin E. Roth¶

Abstract

Kidney failure is a worldwide scourge, made more lethal by the shortage of transplants. We propose a new way to organize kidney exchange chains internationally, between middle-income countries with financial barriers to transplantation and high-income countries with many hard-to-match patients and patient-donor pairs facing lengthy dialysis. The proposal involves chains of exchange that begin in the middle-income country and end in the high-income country. We also propose a new way of financing such chains, using savings to U.S. healthcare payers.

JEL Code: D47.

*This work was partially supported by NSF grant 1061932 to the NBER. We thank Frank McCormick and Philip Held for helpful comments.

†Department of Economics, University of Southern California. afshin.nikzad@usc.edu

‡Graduate School of Business, Stanford University. mohamwad@stanford.edu

§Alliance for Paired Kidney Donation and University of Toledo Medical Center, Toledo, OH. Michael.Rees2@utoledo.edu

¶Department of Economics, Stanford University. alroth@stanford.edu

1 Introduction

Transplantation is the treatment of choice for kidney failure, but many barriers prevent transplantation. The barriers in economically developed countries differ from those in developing countries, but transplantation is woefully undersupplied relative to the prevalence of kidney failure in both richer and poorer nations.

In much of the developing world, transplantation or other treatment such as dialysis is largely unavailable due to financial constraints, and kidney failure is a death sentence [Coresh and Jafar 2015, Liyanage et al. 2015, Harris et al. 2020, International Society of Nephrology 2019]. In the developed world many people struggle with dialysis while enduring long waits for transplantation, with many thousands dying each year while waiting, due to the shortage of transplantable organs.[1]

This paper considers how these two problems—lack of access to transplantation or dialysis, and the shortage of transplantable organs—each can help to alleviate the other.

Kidney exchange, which has become a standard form of transplantation in the United States and is growing elsewhere, can increase the number of transplants available from willing living donors who are otherwise unable to achieve their wish to donate a kidney, by arranging exchanges in which the patient in each patient-donor pair receives a compatible kidney from another patient's donor (see, e.g., [Roth et al. 2004, Wallis et al. 2011, Biró et al. 2017]). The capacity of American kidney exchanges has steadily increased as exchange design has grown from simple exchanges between pairs to include larger cycles of exchange [Saidman et al. 2006] and transplant chains begun by a nondirected donor (see, e.g., [Roth et al. 2006, Rees et al. 2009, Ashlagi et al. 2011b;a, Anderson et al. 2015]. Chains can include many patient-donor pairs. One factor limiting the growth of kidney exchange in the United States, Europe, and elsewhere is that many patients are *highly sensitized*, i.e., they have developed antibodies to many human proteins—so they need a very large kidney exchange pool to have a good chance of finding a pair with whom they can exchange a compatible kidney.

A recent proposal, called Global Kidney Exchange (GKE), has U.S. health organizations inviting foreign patients with financial restrictions who have willing donors to participate in American kidney exchange, just as American patients do, for free [Rees et al. 2017a, Bozek et al. 2018]. The medical treatment (including all pre- and post-surgical care of the foreign

[1] In the United States there are over 90,000 people officially registered on the waiting list for a deceased-donor kidney transplant. In 2018 there were only 14,725 deceased-donor transplants, and 4,049 patients on the waiting list died while waiting, while another 4,189 were removed from the list when they became too sick to transplant. https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/

patient and donor) can be financed by the savings to the American healthcare system when a patient is transplanted and (thus) taken off dialysis, a much more expensive treatment.

Although GKE has generated considerable support [Minerva et al. 2019, Singer et al. 2020, Ambagtsheer et al. 2020] it has also generated critics, opposition and rebuttals (see, e.g., [Delmonico and Ascher 2017, Rees et al. 2017b, Marino et al. 2017, Roth et al. 2017; 2020, Ashlagi et al. 2013, López-Fraga and Domínguez-Gil 2020, Valera and Carrasco 2020, Danovitch 2020]).

Two criticisms have not so far been directly addressed through the *design* of global exchanges (cf. [Ro 2019]). The first is that transplanting foreign patients in the United States does not increase the number of transplants carried out by hospitals in their home countries. The second, related criticism is that when foreign patient-donor pairs participate in American kidney exchange chains, more American than foreign patients are transplanted, and this can appear inequitable.[2]

This paper proposes a new design, Global Kidney Chains (GKC), which would address these criticisms by building kidney exchange chains that originate in a country with modern transplant capabilities, but with financial barriers to transplantation. A foreign non-directed donor could start a kidney exchange chain in the foreign country (that could not otherwise be financed), with the donor of the last pair in the foreign chain (the bridge donor) donating to an American. The costs needed to cover the foreign patients and donors (at home and in the United States) would be paid by the savings to the American healthcare system from transplanting American patients who would otherwise have remained on dialysis.

In practice several foreign transplants could be funded by starting a long chain of transplants with a non-directed donor in the foreign location, with a bridge donor traveling to the United States to continue a chain that would include several hard-to-match American patients and conclude with donation to a patient on the American deceased-donor waiting list.

Some financial engineering is needed to make this program self-financing. The money saved by transplanting an American (or citizen of another high-income country) is saved on dialysis, while the additional costs will be incurred at transplant centers, in the foreign country and (for the donor who travels) in the United States[3] Since such a program will decrease time spent waiting for a transplant, it will also decrease dialysis costs, and we need

[2]And perceived inequity can contribute to the perceived repugnance of a transaction, cf. [Roth 2007].

[3]In countries with single-payer healthcare, this may be feasible more simply than in the United States as long as the savings exceed the costs that would otherwise be incurred, since savings and costs come from the same budget.

to show that the U.S. savings on dialysis remain sufficient to finance the additional foreign transplants even when the program operates on a very large scale, as many patients and patient-donor pairs in high- and middle-income countries gain access.

A concern for the American healthcare system is that it would require legislative changes to enable Medicare (which pays a very large share of dialysis costs) to finance transplantation of foreign patients, despite the savings that would accrue to Medicare through earlier transplantation of Americans. Fundamental legislative changes are relatively easy to imagine and advocate for, but difficult to implement.[4]

Instead, we propose here a financial design that could be implemented in the United States without further legislation. Private insurers in the United States are responsible for paying for the first 33 months of their patients' dialysis, with Medicare activated only after that. Since transplantation is considerably cheaper than 33 months of dialysis (and average dialysis times are considerably longer),[5] insurance companies and self-insuring American companies also experience savings from prompt transplantation, sufficient to pay for additional transplants. We propose the costs for foreign patients be paid from this pool of savings.

These savings are greatest for patients who have been on dialysis for the shortest time, and so a final novelty in the design we propose is that the queue for these American patients to receive a living donor kidney transplant from the bridge donor will have an approximate last-in-first-out queue discipline. This will be important in allowing the savings to remain large even when GKC operates on a scale that substantially reduces average time on dialysis.[6]

Global kidney chains thus involve three novel design features: kidney exchange chains that cross borders, financing by private payers (such as consortia of self-insured companies), and an approximately last-in-first-out queue discipline.

Note that it has been the policy of organizations like the World Health Organization to recommend that countries build self-sufficiency in transplantation. However, no country has

[4]Even simple legislative anomalies have remained despite attempts over many years to fix them. For example, Medicare pays for both dialysis and transplantation for end-stage renal disease (ESRD) patients, but for some patients it pays only for three years of immunosuppresive drugs following transplantation, despite the cost savings that accrue from helping patients avoid rejection and having to resume dialysis (which Medicare then has to pay for).

[5]In 2018, more than 60 percent of deceased-donor kidney recipients had waited on dialysis for more than three years (and more than 40 percent had waited more than five years) https://srtr.transplant.hrsa.gov/annual_reports/2018/Kidney.aspx#KI_9_char_adult_tx_clin, Table KI 9. Clinical characteristics of adult kidney transplant recipients, 2018

[6]Note the contrast with how *deceased* donor organs are traditionally allocated, in which at least tie-breaking priority is given to those who have been waiting the longest.

yet done so: even the wealthiest countries have more patients in need of transplants than they have transplantable organs. And for middle-income countries that cannot finance transplantation for all their citizens who need it, this recommendation of self-sufficiency is simply the advice to wait until the country becomes wealthy, which is an effective death sentence for their contemporary patients who cannot be treated. This paper, instead, considers how we might seek to ameliorate this global health problem with a global solution.

2 The model with short chains

The model that we will present is intended to represent a simple minimal realization of GKC, with the shortest possible U.S.-side chain, involving a single U.S. transplant—and hence the smallest U.S. savings. Specifically, a kidney exchange chain begins in the foreign location. The donor from the last patient-donor pair, i.e., the bridge donor, travels to the United States, and immediately donates to an American in a pool of patients expecting long dialysis, ending the chain. This allows our estimated cost savings to be conservative, and avoids the need to model explicitly the uncertainties associated with assembling a longer American chain initiated by the foreign bridge donor.[7]

We consider a population of domestic (American) patients with a long expected duration of dialysis who are covered by private insurance for the first 33 months of dialysis, and who might receive a kidney from a foreign bridge donor who is part of the last patient-donor pair in a chain of transplants conducted in the foreign country.

To build intuition, we start by considering as an example a (too) simple deterministic model.

Example 2.1. *Suppose that one domestic patient arrives to the pool every day. A patient departs the pool if she is not matched (i.e., has not received a transplant) after* 33 *months. Patients undergo dialysis while waiting in the pool, which costs* D *per patient per day, incurred by the private payer. A foreign bridge donor arrives to the pool every* n *days, starting at time* 0*, where* $n > 1$ *is an integer. Any donor is compatible with every patient, and hence can be matched to any patient. The total cost of every match is* S*, incurred by the private payer.*

According to the Last-In, First-Out (LIFO) allocation policy, an arriving foreign bridge

[7] Another simplification we make is to ignore the possibility that one of the patients expecting long dialysis might unexpectedly receive a deceased-donor kidney. It will become clear later why this shouldn't materially change the results.

donor is matched to the compatible patient with the most recent arrival time present in the pool. Thus under the LIFO policy, every donor is matched upon arrival to a patient who arrived at the same time, and thus has waiting time zero. Therefore, decreasing n (i.e., increasing the arrival rate of foreign bridge donors) reduces the average waiting time for domestic patients under the LIFO policy, by increasing the proportion of patients who have zero waiting time. Hence, when S is sufficiently small relative to D, reducing n decreases the total cost incurred by the private payer per unit of time under the LIFO policy.

In contrast, under the First-In, First-Out (FIFO) allocation policy, a foreign bridge donor is matched to the patient with the earliest time of arrival present in the pool. So under the FIFO policy, every donor is matched upon arrival to a patient who arrived 33 *months earlier. Therefore, decreasing n (i.e., increasing the arrival rate of foreign bridge donors) does* not *change the average waiting time under the FIFO policy, as long as $n > 1$. Hence reducing n* increases *the cost under the FIFO policy, as it adds surgery costs but does not subtract any dialysis costs.*[8]

This simple example shows, on the one hand, that GKC has the capacity to be self financing on a scale approaching that of the arrival rate of domestic patients who can expect dialysis throughout the time they are covered by private insurance. It also shows the importance of the LIFO policy versus the FIFO policy: the savings on dialysis are realized under the LIFO policy but not under the FIFO policy. This is because, as long as the arrival rate of donors is smaller than that of patients, there will always be patients who are supported by the private payer for the full first 33 months of dialysis, so that the FIFO policy, while transplanting the same number of patients as the LIFO policy (and incurring the same additional costs for surgeries), does not reduce the payer's dialysis costs.

This simple example doesn't consider the essential stochastic nature of arrivals, departures, lengths of foreign chains, etc. We next consider a formal model with which we can show that the intuition obtained from the example is robust to these features.

2.1 A formal model

Domestic patients arrive according to a Poisson process with rate m. Each patient stays in the pool for $\zeta > 0$ units of time, and then departs. The foreign bridge donors arrive according to a Poisson process with rate λm, where λ is the policy parameter to be set. We

[8]Note that if we had added deceased donation to the model, some domestic patients would have received a deceased-donor kidney under the national FIFO policy. These would also be patients who had waited the longest, and for whom there would be little or no savings on dialysis.

suppose that a bridge donor b is the last donor in a chain involving l_b foreign patients. The random variable l_b is drawn independently for every bridge donor b from a distribution F with mean μ.

We suppose that each bridge donor is *compatible* to each patient independently with probability $r > 0$.

The planner adopts a LIFO allocation policy. When a foreign bridge donor arrives, he or she is matched to the compatible patient with the latest time of arrival. In case no patient is in the pool, the donor departs immediately.

Let D denote the domestic cost of dialysis per patient per unit of time. Also, let S_d and S_f denote the costs per domestic and foreign kidney transplant, respectively. Define $\mathbf{C}(m, \lambda)$ to be the average, per domestic patient per unit of time, of total healthcare costs when the policy parameter is λ. (Hence, $\mathbf{C}(m, \lambda)$ accounts for all the costs of foreign patients as well.) The next result shows that, for every finite m, increasing λ by any positive $\epsilon < 1 - \lambda$ that is not "too small" reduces $\mathbf{C}(\lambda)$ if $\zeta D > S_d + \mu S_f$. (Note that ζD is the dialysis cost for a patient who does not receive a transplant.)

Theorem 2.2. *Suppose that $\zeta D > S_d + \mu S_f$. Then, for any fixed m, increasing λ by any positive $\epsilon < 1 - \lambda$ decreases $\mathbf{C}(m, \lambda)$ if $\epsilon > \gamma \cdot \frac{\zeta D}{\zeta D - S_d - \mu S_f}$, where $\gamma \leq \frac{(1+r)\log m}{rm} + \frac{2}{m}$.*

This theorem requires ϵ to be not "too small": The right-hand side of the constraint that ensures ϵ is not too small is the parameter γ multiplied by a constant independent of m.[9] Observe that γ approaches zero with rate $\frac{\log m}{m}$ as m grows large. Hence, as m grows large, the lower bound on ϵ becomes essentially nonbinding. This observation is formalized in the next theorem, which also computes the derivative of the average healthcare costs with respect to λ, asymptotically. Define $\mathbf{C}(\lambda) = \lim_{m\to\infty} \mathbf{C}(m, \lambda)$.

Theorem 2.3. *Under the LIFO policy, for any $\lambda \in (0, 1)$, $\mathbf{C}'(\lambda) = S_d + \mu S_f - \zeta D$.*

A back-of-the-envelope account of these theorems goes as follows. Whenever a patient is about to enter the system under the LIFO policy, she is matched with probability close to λ. Thus, an increase in λ reduces her expected dialysis cost approximately by $\lambda\zeta D$, and increases the expected transplant costs by $\lambda(S_d + \mu S_f)$, incurred by her own transplant and the preceding chain. Thus, $\mathbf{C}'(\lambda) = S_d + \mu S_f - \zeta D$.

For the average healthcare costs to decrease with the arrival rate of bridge donors, both of the theorems above require $\zeta D > S_d + \mu S_f$. We next evaluate this condition using the

[9] The lower bound required on ϵ is due to our proof approach, which finds upper and lower bounds on the average queue length, rather than finding the exact value. We conjecture that this lower bound is dismissible.

estimated values for its parameters provided by [Held et al. 2016] and [Paloyo 2019]. In the United States, dialysis costs about $250,000 per patient year for a commercial payer, and a transplant costs about $100,000, followed by the cost of immunosuppressive medications and follow-up care, which is about $30,000 per year. We account for 10 years of immunosuppressives and follow-up care, and hence set $S_d = \$400,000$. We account for the transplant cost of a foreign patient, using the Philippines for our example, by a $12,000 surgery cost plus 10 years of immunosuppressives medications and follow-up care costing about $6,000 per year. Hence, we set $S_f = \$72,000$. For these parameter values, the condition $\zeta D > S_d + \mu S_f$ is satisfied if $\mu \leq 4.86$.

Hence, according to the above estimates, the LIFO policy reduces the total healthcare costs when the average length of foreign chains is not larger than 4.86.

We note that a simple static back-of-the-envelope argument based on current dialysis and surgery costs is insufficient for establishing these results: such an argument ignores the counterfactual costs, which depend on the dynamics. For instance, suppose the planner switches to the FIFO policy. In that case, increasing λ by any $\epsilon < 1 - \lambda$ will *increase* total costs! That is, as long as the number of GKC bridge donors is below what would be needed to secure transplants for *all* domestic patients, under the FIFO policy all domestic patients will have a waiting time of ζ, whether they get transplanted or not. So increasing λ would not affect the average dialysis costs paid by the private insurer, although it would add the transplant costs of the foreign patients. The next theorem captures this effect.

Theorem 2.4. *Under the FIFO policy, for any* $\lambda \in (0,1)$, $\mathbf{C}'(\lambda) > 0$.

Hence, the cost-benefit analysis crucially depends on market dynamics.

3 Concluding remarks

Since the beginning of the twenty-first century, kidney exchange at scale has developed from a largely academic idea intially implemented at a small scale [Roth et al. 2004; 2005] to a standard mode of transplantation in the United States (with well over 1,000 exchange transplants in 2019) and in several other countries. This has been an important development, with many milestones along the way, including, crucially, developments in the design and implementation of kidney exchange chains. But these accomplishments have been victories in a war that we are losing. At the turn of the century there were in the neighborhood of 40,000

patients on the U.S. waitlist for deceased-donor organs, and today there are close to 100,000.[10] The situation is similar elsewhere in the wealthy world. Over the same period, there has been a growth of kidney disease as a cause of death around the world (as developing countries have made progress in combating infectious disease) and there have begun to be high-quality transplant centers in middle-income as well as in rich countries, which nevertheless face obstacles—including important financial obstacles—to increasing the number of transplants they are able to deliver.[11]

Before the development of kidney exchange, the organization of transplantation developed largely within the national boundaries of wealthy countries. It was primarily focused on deceased-donor transplants, and the scarcity of organs meant that the concentration of effort within single countries did not have a large impact on the total number of transplants achieved.[12] With the growth of kidney exchange, there are now some preliminary explorations of coordinating across borders between countries with existing kidney exchange programs, primarily concentrating on looking for exchanges between hard-to-match pairs who have been left unmatched in the within-country kidney exchange. Global kidney exchange opens up this possibility to a much larger part of the world, including countries in which unmatched patient-donor pairs may have had financial rather than immunological barriers, and so may be easier to match with hard-to-match pairs. And because kidney exchange chains have amplified kidney exchange wherever they have been implemented, Global Exchange Chains offers a way to bring these advantages to a much larger group of patients and donors.[13]

[10]The increase in the number of patients on the waiting list is not entirely bad news related to the growing incidence of kidney disease; it also represents progress in keeping kidney patients alive longer before they receive a transplant. And it reflects increases in traffic safety, which reduces the number of deceased-donor transplants from victims of automobile accidents.

[11]Harris et al. write: "It is estimated that the number of people dying globally with ESRD for want of kidney replacement therapy is up to 3 times the number who receive it. Kidney transplantation meets only a small fraction of the therapeutic need. Finally, about 188 million people experience catastrophic health expenditure annually as a result of kidney diseases across low- and middle-income countries, the greatest of any disease group."

[12]There are well-established efforts to share deceased-donor kidneys across national borders in limited circumstances.

[13]If our concern in this paper were only with American patients, global kidney chains, with their costs of care for international patients, would likely be more expensive per patient, initially, than other ways of increasing, on the margin, the number of donor kidneys available to Americans. (These avenues should also be, and are being pursued, of course). Some of these—like increasing the number of deceased-donor kidneys—don't have the potential to cover the full need for organs (because only a tiny fraction of deaths occur in a way that makes organs potentially recoverable for transplant), but each life saved is precious, and each viable organ is very valuable. Other avenues, like increasing the number of living donors by providing greater incentives to donate, may be repugnant and illegal under current law in the United States and elsewhere. It seems likely that financial *dis*incentives to donation could be reduced under current law, however, and

Notice that if an international exchange works perfectly—i.e., when all of the patients and donors involved have successful surgeries, excellent follow-up care, and are all restored to active, long-lasting good health—then it will be easy to see the exchange as just another example of the success of standard kidney exchange in which all patients are from the same country. But if the pair from the developing country were to return home and have bad health outcomes, it would look a lot like badly arranged black market transactions, which are justly condemned. So to make kidney exchange work between developed and developing countries, exceptional care will have to be delivered to the developing-country donors and patients, particularly since patients in poor countries—like their compatriots who have never suffered from kidney disease—can be expected to have somewhat worse health outcomes than otherwise comparable people in rich countries, no matter what efforts are made to give them the best possible post-operative care. Consequently a first element of a successful design for GKC is the choice of reliable international partners, able to provide excellent care for patients and donors, both prospectively and post-operatively.

The other three design elements, proposed and explored in this paper, involve: starting a chain in a foreign country and having a bridge donor continue it in the United States; using a LIFO queue policy on the pool of patients assembled by, e.g., a coalition of self-insured companies responsible for paying for their care; and having those savings finance the additional costs (compared to an entirely domestic chain), in both countries. As we have shown, such a program could operate at a significant scale, comparable to the number of domestic patients presently beginning lengthy dialysis annually. Global Kidney Chains thus appear to present a scalable approach to cross-border kidney exchange, and to increasing the availability of transplantation globally.

To summarize, the story of kidney transplantation as a market design problem has been one of increasing the scope of kidney transplantation, step by step. Before kidney exchange, many potential donors were unable to donate, because they were incompatible with their intended recipients.

Kidney exchange makes it possible for such donors to achieve transplants for their intended recipients despite their incompatibility, and so increases the supply of transplantable kidneys. But many hard-to-match pairs are still unmatched by kidney exchange, and, in the

limited steps in this direction are included under the recent Executive Order (https://www.whitehouse.gov/presidential-actions/executive-order-advancing-american-kidney-health/). Each of these avenues is well worth exploring, and each has the prospect of saving lives and medical costs, but none of them seems to offer a scale that could end the growth in the deceased-donor waiting list, and even reduce it and none would offer the prospect of extending the benefits of transplantation to international patients while also furthering domestic American goals.

developing world, kidney failure remains a death sentence for all but the wealthy.

Global kidney exchange, still in its infancy, seeks to address these two problems by finding ways to invite patient-donor pairs to participate in American kidney exchange, from middle-income countries with medical systems that are adequate for transplant patients except for adequate finances. The additional expenses can be paid for by savings from more expensive dialysis that would otherwise be the fate of hard-to-match American patients. One of the advantages of GKE is that patient-donor pairs whose barrier to transplantation is financial may not themselves be immunologically difficult to match, which means that they can more easily match hard-to-match pairs who otherwise have difficulty finding a match. But GKE typically matches one foreign pair at a time, and the need for transplantation in the developing world is vast.

Global kidney chains, proposed here, extend the benefits of GKE to greater numbers of patient-donor pairs from middle-income developing countries, by starting exchange chains where they live, and continuing them in the U.S or other wealthy countries with well-established kidney exchanges. This can be financially self-supporting, because transplantation is cheaper in middle-income countries with adequate hospitals, so more of their patients can be transplanted with the savings that come to the American healthcare system by transplanting a patient who would otherwise face long dialysis. We argue here that the scale of such transplants could be very large, so that GKC can become at least a first step toward providing a global solution to the global problem of kidney failure.

References

Frederike Ambagtsheer, Bernadette Haase-Kromwijk, Frank JMF Dor, Greg Moorlock, Franco Citterio, Thierry Berney, and Emma K Massey. Global kidney exchange: Opportunity or exploitation? an ELPAT/ESOT appraisal. *Transplant International*, 2020.

Ross Anderson, Itai Ashlagi, David Gamarnik, and Alvin E Roth. Finding long chains in kidney exchange using the traveling salesman problem. *Proceedings of the National Academy of Sciences*, 112(3):663–668, 2015.

Itai Ashlagi, Duncan S Gilchrist, Alvin E Roth, and Michael A Rees. Nead chains in transplantation. *American Journal of transplantation*, 11(12):2780–2781, 2011a.

Itai Ashlagi, Duncan S Gilchrist, Alvin E Roth, and Michael A Rees. Nonsimultaneous chains

and dominos in kidney-paired donation—revisited. *American Journal of transplantation*, 11(5):984–994, 2011b.

Itai Ashlagi, Patrick Jaillet, and Vahideh H. Manshadi. Kidney exchange in dynamic sparse heterogenous pools. In *EC*, pages 25–26, 2013.

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

Danielle N Bozek, Ty B Dunn, Christian S Kuhr, Christopher L Marsh, Jeffrey Rogers, Susan E Rees, Laura Basagoitia, Robert J Brunner, Alvin E Roth, Obi Ekwenna, et al. Complete chain of the first global kidney exchange transplant and 3-yr follow-up. *European Urology Focus*, 4(2):190–197, 2018.

Clement Canonne. A short note on Poisson tail bounds. 2019. http://www.cs.columbia.edu/~ccanonne/files/misc/2017-poissonconcentration.pdf.

Josef Coresh and Tazeen H Jafar. Disparities in worldwide treatment of kidney failure. *The Lancet*, 385(9981):1926–1928, 2015. doi: 10.1016/S0140-6736(14)61890-0. URL http://dx.doi.org/10.1016/S0140-6736(14)61890-0.

Gabriel M Danovitch. The global kidney exchange programme. *The Lancet*, 395(10235): 1485, 2020.

Francis L. Delmonico and Nancy L. Ascher. Opposition to irresponsible global kidney exchange. *American Journal of Transplantation*, 17(10):2745–2746, 2017. doi: 10.1111/ajt.14473. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ajt.14473.

David CH Harris, Simon J Davies, Fredric O Finkelstein, and Vivekanand Jha. The second global kidney health summit outputs: developing a strategic plan to increase access to integrated end-stage kidney disease care worldwide. *Kidney International Supplements*, 10(1):e1–e2, 2020.

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

International Society of Nephrology. https://www.theisn.org/focus/ckd#health-atlas, 2019.

T Liyanage, T Ninomiya, V Jha, B Neal, HM Patrice, I Okpechi, MH Zhao, J Lv, AX Garg, J Knight, A Rodgers, M Gallagher, S Kotwal, A Cass, and V Perkovic. Worldwide access to treatment for end-stage kidney disease: a systematic review. *The Lancet*, 385:1975–1982, 16–22 May 2015.

Marta López-Fraga and Beatriz Domínguez-Gil. The global kidney exchange programme. *The Lancet*, 395(10235):1484, 2020.

Ignazio R Marino, Alvin E Roth, Michael A Rees, and Cataldo Doria. Open dialogue between professionals with different opinions builds the best policy. *American Journal of Transplantation*, 17(10):2749–2749, 2017.

Francesca Minerva, Julian Savulescu, and Peter Singer. The ethics of the global kidney exchange programme. *The Lancet*, 394(10210):1775–1778, 2019.

Siegfredo Ricerra Paloyo. Private communication, 2019.

M. A. Rees, T. B. Dunn, C. S. Kuhr, C. L. Marsh, J. Rogers, S. E. Rees, A. Cicero, L. J. Reece, A. E. Roth, O. Ekwenna, D. E. Fumo, K. D. Krawiec, J. E. Kopke, S. Jain, M. Tan, and S. R. Paloyo. Kidney exchange to overcome financial barriers to kidney transplantation. *American Journal of Transplantation*, 17(3):782–790, 2017a. ISSN 1600-6143. doi: 10.1111/ajt.14106. URL http://dx.doi.org/10.1111/ajt.14106.

M. A. Rees, S. R. Paloyo, A. E. Roth, K. D. Krawiec, O. Ekwenna, C. L. Marsh, A. J. Wenig, and T. B. Dunn. Global kidney exchange: Financially incompatible pairs are not transplantable compatible pairs. *American Journal of Transplantation*, 17(10):2743–2744, 2017b. doi: 10.1111/ajt.14451. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ajt.14451.

Michael A. Rees, 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, Tuomas Sandholm, M. Utku Ünver, and Robert A. Montgomery. A nonsimultaneous, extended, altruistic-donor chain. *New England Journal of Medicine*, 360(11):1096–1101, 2009. doi: 10.1056/NEJMoa0803645. URL http://dx.doi.org/10.1056/NEJMoa0803645. PMID: 19279341.

Christine Ro. Why the global kidney exchange remains controversial. *Forbes*, 17, 2019. URL https://www.forbes.com/sites/christinero/2019/12/15/why-the-global-kidney-exchange-remains-controversial/#6d0531845bcc.

Sheldon M Ross. *Introduction to probability models*. Academic press, 2014.

Alvin E Roth. Repugnance as a constraint on markets. *Journal of Economic perspectives*, 21(3):37–58, 2007.

Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. Kidney exchange. *The Quarterly Journal of Economics*, 119(2):457–488, 2004.

Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. Pairwise kidney exchange. *Journal of Economic theory*, 125(2):151–188, 2005.

Alvin E Roth, Tayfun Sönmez, M Utku Ünver, Francis L Delmonico, and Susan L Saidman. Utilizing list exchange and nondirected donation through 'chain'paired kidney donations. *American Journal of transplantation*, 6(11):2694–2705, 2006.

Alvin E. Roth, Kimberly D. Krawiec, Siegfredo Paloyo, Ekwenna Obi, Marsh Christopher L., Alexandra J. Wenig, Ty B. Dunn, and Michael A. Rees. People should not be banned from transplantation only because of their country of origin. *American Journal of Transplantation*, 17(10):2747–2748, 2017. doi: 10.1111/ajt.14485. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ajt.14485.

Alvin E. Roth, Ignazio R. Marino, Obi Ekwenna, Ty B. Dunn, Siegfredo R. Paloyo, Miguel Tan, Ricardo Correa-Rotter, Christian S. Kuhr, Christopher L. Marsh, Jorge Ortiz, Giuliano Testa, Puneet Sindhwani, Dorry L. Segev, Jeffrey Rogers, Jeffrey D. Punch, Rachel C. Forbes, Michael A. Zimmerman, Matthew J. Ellis, Aparna Rege, Laura Basagoitia, Kimberly D. Krawiec, and Michael A. Rees. Global kidney exchange should expand wisely. *Transplant International*, n/a(n/a), 2020. doi: 10.1111/tri.13656. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/tri.13656.

Susan L Saidman, 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*, 81(5):773–782, 2006.

Peter Singer, Francesca Minerva, and Julian Savulescu. The global kidney exchange programme–authors' reply. *The Lancet*, 395(10235):1485–1486, 2020.

Luca Valera and María Alejandra Carrasco. The global kidney exchange programme. *The Lancet*, 395(10235):1484–1485, 2020.

C. Bradley Wallis, Kannan P. Samy, Alvin E. Roth, and Michael A. Rees. Kidney paired donation. *Nephrology Dialysis Transplantation*, 26(7):2091–2099, 2011. doi: 10.1093/ndt/gfr155. URL http://ndt.oxfordjournals.org/content/26/7/2091.abstract.

Wikipedia. Catalan number. https://en.wikipedia.org/wiki/Catalan_number, June 2020.

Appendix

A Preliminary definitions

For brevity, we refer to domestic patients as *patients* and foreign bridge donors as *donors* throughout the appendices. We use the terms *transplanted* and *matched* interchangeably; e.g., by saying that *a donor is matched to a patient* we mean the patient has received a transplant from that donor.

The *departure time* of a patient is ζ units of time after her arrival time.

A.1 Graph theory

For every graph G, we let $V(G)$ denote the set of its nodes and $E(G)$ denote the set of its edges. An *independent* set in a graph G is a subset $S \subseteq V(G)$ such that no two nodes in S are adjacent in G. A *matching* in G is a subset of $E(G)$ such that no two edges in the subset have a common node. The *size* of the matching is the number of its edges.

We typically denote a bipartite graph by $G(X, Y)$ where X, Y denote the set of nodes on each side of G. (That is, $V(G) = X \cup Y$, and both X, Y are independents sets in G.)

A.2 Asymptotic notions

We say a statement $\mathcal{S}(i)$ holds for *sufficiently large* i if there exists i_0 such that $\mathcal{S}(i)$ holds for all $i > i_0$.

Let $E(i)$ be an event parameterized by a positive integer i. We say that $E(i)$ occurs with *high probability as i grows large* if $\lim_{i\to\infty} \mathbb{P}[E(i)] = 1$. We often let the parameter i be m, the arrival rate of patients. When this is clearly known from the context, we simply say that $E(m)$ occurs *with high probability* or, briefly, $E(m)$ occurs *whp*.

Furthermore, we say that $E(i)$ occurs *with very high probability as i grows large* if there exists $\alpha > 1$ such that $\lim_{i\to\infty} \frac{1-\mathbb{P}[E(i)]}{e^{-(\ln i)^{\alpha}}} = 0$. We often let the parameter i be m, the arrival rate of patients. When this is clearly known from the context, we simply say that $E(m)$ occurs *with very high probability* or, briefly, $E(m)$ occurs *wvhp*. We say an event $E(m)$ holds *with very low probability* (*wvlp*) when $\overline{E(m)}$ holds wvhp.

For any two functions $f, g : \mathbb{R}_+ \to \mathbb{R}_+$ we adopt the notation $f = o(g)$ when for every positive constant ϵ there exists a constant i_ϵ such that $f(i) \leq \epsilon g(i)$ holds for all $i > i_\epsilon$.

We define $f = O(g)$ if there exist positive constants i_0, Δ such that $f(i) \leq g(i)\Delta$ holds

for all $i > i_0$. We write $g = \Theta(f)$ when $f = O(g)$ and $g = O(f)$, and $g = \Omega(f)$ when $f = O(g)$.

B Proof of Theorem 2.3

First, we prove a theorem for a simpler setting in Section B.1, which we will then use to prove Theorem 2.3 in Section B.2.

B.1 A simple setting

We first analyze a simplified version of the model in Section 2.1, where every donor is compatible with every patient.

We consider a LIFO queue with a Poisson arrival rate of C for *patients* (who represent domestic patients). Each patient waits in the queue for one unit of time, and abandons the queue if not matched by then. Donors arrive according to a Poisson process with rate S. Upon the arrival of a donor, if the queue is non-empty, the patient who arrived last is matched to the donor and removed from the queue.

Theorem B.1. *The time-average length of the queue is at least $C - S$. Furthermore, the average length is at most $C - S + \log(\frac{C}{C-S})$ if $S \leq C - 1$, and at most $1 + \log C$ otherwise.*

The proof of Theorem B.1 is in two steps. Step 1 proves the lower bound and Step 2 the upper bound on the average length of the queue.

B.1.1 Preliminaries

Consider a stochastic process similar to the simple setting of Section B.1 (described above), but without abandonments. In this alternative process, the patients never leave the queue unless matched (and everything else remains identical to the process with abandonments). Let $\mathcal{P}, \mathcal{P}'$ respectively denote the stochastic processes with and without abandonment. We suppose that both $\mathcal{P}, \mathcal{P}'$ are started with an empty queue at time 0. (All results hold with any other initial finite queue.)

$\mathcal{P}'$ is a continuous-time random walk on nonnegative integers: at each arrival event, the length of the queue is increased by 1 with probability $\frac{S}{C+S}$, and is decreased by 1 with probability $\frac{C}{C+S}$ (if the queue is non-empty). Define $\delta = 1 - S/C$, i.e., $S = C(1 - \delta)$.

Denote the length of the queue at time t in $\mathcal{P}, \mathcal{P}'$ respectively by $l(t), l'(t)$. Define

$$w = \lim_{T\to\infty} \int_0^T l(t)/T \, \mathrm{d}t, \tag{B.1}$$

i.e., w is the average queue length (over time) in $\mathcal{P}$.

Claim B.2. *The limit defined by* (B.1) *exists.*

Proof. The limit exists because $\mathcal{P}$ returns to the state where the queue is empty in finite time with probability 1, and the average return time is finite.

To see why, observe that the return time of $\mathcal{P}$ to the empty state is stochastically dominated by the return time of the process that has the same arrivals and departures but makes no matches at all. Then, see that the process that makes no matches at all visits the empty state with probability at least e^{-m} in the time interval $(t, t+1)$, for any arbitrary $t > 0$. The reason is that the number of births b in a Poisson process with rate m has distribution $P(b) = e^{-m}\frac{m^b}{b!}$. This means that in $\mathcal{P}$ the probability that the empty state is visited by time $t > 0$ for an integer t is at least $(1 - e^{-m})^t$. Therefore, $\mathcal{P}$ returns to the empty state in finite time with probability 1, and the average return time is finite.

We use this to complete the proof. Let X_i denote the joint distribution of the random variables t_i, l_i where t_i denotes the time it takes for $\mathcal{P}$ to return to the empty state again after its i-th visit to that state, and $l_i = \int_{t_i}^{t_{i+1}} l(t)$. Both t_i, l_i have finite means (unconditionally), and $X_i = X_j$ for any i, j. Hence, the renewal reward theorem [Ross 2014] applies, and implies that

$$\lim_{T\to\infty} \int_0^T l(t)/T = \frac{\mathbb{E}_{X_1}[l_1]}{\mathbb{E}_{X_1}[t_1]}.$$

□

B.1.2 Step 1: The lower bound

To establish the lower bound on w, we first analyze the stochastic process $\mathcal{P}'$. Let p denote the probability of a just-arrived patient being matched in $\mathcal{P}'$ in finite time. Observe that the length of the queue in $\mathcal{P}'$ is a continuous-time random walk on nonnegative integers (where at each arrival event, the length of the queue increases by 1 with probability $\frac{C}{C+S}$, and otherwise decreases by 1 when the queue is nonempty.) Hence, p equals the probability that the random walk returns to the origin in finite time. In Lemma B.3 we will show that $p = S/C$. We then use this lemma to prove in Lemma B.4 that any patient is matched

with probability at most p in $\mathcal{P}$. The latter fact, together with Little's law, implies that $w \geq C(1-p) = C - S$, which would conclude Step 1. Thus, to complete this step, it remains to prove Lemmas B.3 and B.4.

Lemma B.3. $p = S/C$.

Proof. There are two possible scenarios in which a patient c who has just arrived could be matched: (i) the arrival of a donor right after the arrival of c, and (ii) the arrival of a patient c' right after the arrival of c. The probability of scenario (i) is $\frac{C}{C+S}$. The probability that c is matched in scenario (ii) is $\frac{S}{C+S} \cdot p^2$; it is the probability of arrival of a new patient c' times the probability that c' is matched, times the probability that c herself is matched after that (which is assumed to be p). Thus, we have the following equation:

$$p = \frac{S}{C+S} + \frac{C}{C+S} \cdot p^2,$$

the solution to which is $p = S/C$. □

Lemma B.4. *In the process $\mathcal{P}$ (i.e., the process with abandonments), any patient is matched with probability at most p.*

Proof. Fix an arbitrary patient c who has just arrived. At that point, define a coupling of the process with abandonments ($\mathcal{P}$) and the process without abandonments ($\mathcal{P}'$) as follows. $\mathcal{P}'$ has the same initial state as $\mathcal{P}$ at the time of the arrival of c. From then on, for each stochastic event that happens in $\mathcal{P}$ (the arrival of the patient or donor), let the same event happen in $\mathcal{P}'$. We claim that in any sample path of the coupling, the following is true: c is matched in $\mathcal{P}$ only if she is matched in $\mathcal{P}'$. Note that to complete the proof of the lemma, we just need to prove this claim.

We prove a stronger claim: c is matched in $\mathcal{P}$ iff she is matched in $\mathcal{P}'$ in less than a unit of time. The key to proving this is that if, in $\mathcal{P}'$, c is matched after waiting one unit of time, then in $\mathcal{P}$, all unmatched patients who arrived after c and c herself have abandoned the queue (this is due to the LIFO policy and the fact that all patients wait for one unit of time). On the other hand, if, in $\mathcal{P}'$, a patient c is matched in less than a unit of time, then c is also matched in $\mathcal{P}$. □

B.1.3 Step 2: The upper bound

Consider the process without abandonments, $\mathcal{P}'$. The *matching time* of a patient in $\mathcal{P}'$ is the amount of time that the patient waits before being matched to a donor, and is ∞ if the

patient is never matched. Let $f : [0, \infty] \to \mathbb{R}_+$ denote the PDF of the matching time of a patient in $\mathcal{P}'$. (Note that since $p < 1$, we have $f(\infty) > 0$.) The following is an upper bound on w:

$$w/C \leq \int_0^1 t f(t)\,\mathrm{d}t + \int_{t\in[1,\infty)} f(t)\,\mathrm{d}t + (1-p)\cdot 1. \tag{B.2}$$

To see why the above inequality holds, we interpret its right-hand side as follows. For this interpretation, we need to consider the coupling of $\mathcal{P}, \mathcal{P}'$, as defined in the proof of Lemma B.4. The first summand in (B.2) accounts for the waiting time of patients who are matched in $\mathcal{P}$; we correspond this summand to the patients in $\mathcal{P}'$ who have waiting time less than 1. The second summand in (B.2) accounts for the waiting time of a subset of unmatched patients in $\mathcal{P}$, who correspond to the patients in $\mathcal{P}'$ with waiting times larger than 1. Note that such patients only wait for one unit of time in $\mathcal{P}$, and thus rather than multiplying the term inside the integral by t, we have multiplied it by 1. The third summand is an upper bound on the waiting time of the rest of the unmatched patients in $\mathcal{P}$, which correspond to the patients in $\mathcal{P}'$ who are not matched in finite time; these patients also wait only for one unit of time in $\mathcal{P}$.

Next, by providing a simple upper bound for the right-hand side of (B.2), we rewrite (B.2) as follows. Let $\mathsf{w}'(c)$ be a random variable denoting the waiting time of a patient c in $\mathcal{P}'$. Then,

$$\begin{aligned} w/C &\leq \int_{t\in[0,\infty)} t f(t)\,\mathrm{d}t + (1-p)\cdot 1 \\ &= p\cdot \mathbb{E}\left[\mathsf{w}'(c)\middle|\mathsf{w}'(c) < \infty\right] + (1-p)\cdot 1. \end{aligned} \tag{B.3}$$

In writing the above equality, we have used the fact that $\int_{t\in[0,\infty)} t f(t)\,\mathrm{d}t = p\,\mathbb{E}\left[\mathsf{w}'(c)\middle|\mathsf{w}'(c) < \infty\right]$, for any patient c.

Next, we show that $\mathbb{E}\left[\mathsf{w}'(c)\middle|\mathsf{w}'(c) < \infty\right] \leq \frac{\theta_\delta}{C+S}$, where $\theta_\delta = 2\log\left(\frac{1}{\delta}\right)$. If we prove this, then (B.3) implies that

$$w \leq \frac{p\theta_\delta \cdot C}{C+S} + C(1-p) = \frac{\theta_\delta S}{C+S} + C - S \leq \theta_\delta/2 + C - S, \tag{B.4}$$

which would prove the theorem. Thus, to complete the proof it remains to prove the following lemma.

Lemma B.5. $\mathbb{E}\left[\mathsf{w}'(c)\middle|\mathsf{w}'(c)<\infty\right] \leq \frac{\theta_\delta}{C+S}$.

Proof. Consider a random walk on integers that starts at the origin. At each step, the walk goes either to the left or the right, with probabilities $P_l = \frac{S}{C+S}$ and $P_r = \frac{C}{C+S}$, respectively. We use this random walk to prove the lemma. Let q be the probability that the random walk returns to the origin conditional on its first step being to the right. Then,

$$\mathbb{E}\left[\mathsf{w}'(c)\middle|\mathsf{w}'(c)<\infty\right] = P_r \cdot q \cdot \frac{1}{C+S}, \tag{B.5}$$

where in the right-hand side of (B.5), the first factor is the probability that the first step is to the right (which corresponds to the arrival of another patient c' right after the arrival of c), the second factor is the probability of returning to the origin (which corresponds to the probability of matching c conditioned on the arrival of c'), and the third factor accounts for the transition speed in the random walk (the rate of movement in the random walk is $C+S$). To complete the proof, we compute an upper bound on q, namely $\overline{q}$, and then set $\theta_\delta = \overline{q}P_r$, which will conclude the proof.

To compute q, we count the number of sample paths of length n which start with going to the right and return to the origin after n steps (n being even). Denote the number of such paths by C_n, and see that

$$q = \frac{1}{P_r} \cdot \sum_{n=2}^{\infty} (P_l P_r)^{n/2} C_n.$$

It is well-known that C_n equals the $n/2$-th Catalan number (e.g., see [Wikipedia 2020]):

$$C_n = \frac{1}{n/2} \cdot \binom{n-2}{n/2-1}.$$

Define $\beta_\delta = \frac{1-\delta}{(2-\delta)^2}$. Using the above two equalities and the fact that $S = C(1-\delta)$, we

can write:

$$
\begin{aligned}
q &= \frac{1}{P_r} \cdot \sum_{n\in\mathbb{N}, n \bmod 2=0}^{\infty} \left(\frac{CS}{(C+S)^2}\right)^{n/2} \cdot \frac{1}{n/2} \cdot \binom{n-2}{n/2-1} \\
&\leq \frac{1}{P_r} \cdot \sum_{n\in\mathbb{N}, n \bmod 2=0}^{\infty} \left(\frac{1-\delta}{(2-\delta)^2}\right)^{n/2} \cdot \frac{1}{n/2} \cdot \binom{n-2}{n/2-1} \\
&\leq \frac{1}{P_r} \cdot \sum_{n\in\mathbb{N}, n \bmod 2=0}^{\infty} \frac{1}{n/2} \cdot {\beta_\delta}^{n/2} \cdot 4^{n/2-1} \\
&= \frac{1}{2P_r} \cdot \sum_{n\in\mathbb{N}, n \bmod 2=0}^{\infty} \frac{1}{n} \cdot (4\beta_\delta)^{n/2} \\
&= \frac{1}{4P_r} \cdot \log\left(\frac{1}{1-4\beta_\delta}\right) \leq \frac{2}{P_r} \log\left(\frac{1}{\delta}\right), \qquad \text{(B.6)}
\end{aligned}
$$

where the last equality follows since $\beta_\delta < 1$ and the last inequality holds for all $\delta \in (0,1)$. Denote the right-hand side of (B.6) by $\overline{q}$.

To complete the proof for the case where $S \leq C-1$, we plug in the above bound on q in (B.5) and write

$$
\mathbb{E}\left[\mathsf{w}'(c) \middle| \mathsf{w}'(c) < \infty\right] \leq \frac{2\log\left(\frac{1}{\delta}\right)}{C+S}, \qquad \text{(B.7)}
$$

which is the promised claim.

To prove the claim for the case where $S \geq C-1$, we note that the average length of the queue when $S \geq C-1$ is bounded from above by the average length of the queue when $S = C-1$. (This holds by a straightforward coupling argument.) Hence, the upper bound provided in the first case for $S = C-1$ holds in the second case as well. This completes the proof. □

B.2 Using (B.9) to prove Theorem 2.3

In this proof we use the phrases *queue* and *pool*, interchangeably. Throughout the proof, by a donor we mean a foreign bridge donor. Let S denote the arrival rate of donors, i.e., $S = \lambda m$. Also, let $C = m$ denote the arrival rate of patients. This notation will allow us to apply the results from Section B.1 directly. For notational simplicity, throughout the proof we normalize $\zeta = 1$. This is just a rescaling of time and is without loss of generality.

Let $\mathcal{P}^*$ denote the stochastic process that governs the pool in Theorem 2.3. (That is, $\mathcal{P}^*$

governs the arrivals, departures, and the allocation policy.) We suppose that $\mathcal{P}^*$ starts with an empty pool at time 0. (All results hold with any other initial pool with a finite size.) Let $A^*(t)$ denote the set of patients present in the pool at time t in the process $\mathcal{P}^*$, and let $L^*(t) = |A^*(t)|$. Define

$$l^* = \lim_{T\to\infty} \int_0^T L^*(t)/T \, \mathrm{d}t.$$

An argument similar to the proof of Claim B.2 shows that the above limit exists. By Little's law, the *average waiting time per domestic patient* is l^*/m. Recall the variable w defined by (B.1), which denotes the average queue length (over time) in $\mathcal{P}$. Throughout this proof we denote w by l. Thus, the average waiting time per domestic patient is l/m in the process $\mathcal{P}$.

We recall the definition of process $\mathcal{P}$ from Section B.1.1: $\mathcal{P}$ is the same process as $\mathcal{P}^*$ with the difference that every donor is compatible with every patient in $\mathcal{P}$. A straightforward coupling of $\mathcal{P}, \mathcal{P}^*$ can be constructed such that, in every sample path of the coupling, $L(t) \leq L^*(t)$ holds for all t. This implies that $l \leq l^*$. Section B.1.2 showed that $l \geq C - S$. Hence,

$$l^* \geq C - S. \tag{B.8}$$

Next, we will show that

$$l^* \leq C - S + 2 + \frac{1+r}{r} \cdot \log(S+1). \tag{B.9}$$

We then use this bound to conclude the proof of the theorem.

B.2.1 The proof for (B.9)

We outline the proof before presenting it formally. The proof uses multiple coupling arguments. First, it couples the stochastic process $\mathcal{P}^*$ with a process $\mathcal{P}'$, then $\mathcal{P}'$ with a process $\mathcal{P}''$, and finally $\mathcal{P}''$ with a process $\mathcal{P}'''$. Each of these couplings simplifies $\mathcal{P}^*$ further. The process $\mathcal{P}'''$ will be very similar to the process $\mathcal{P}$, which we defined in Section B.1.1 to prove a simpler version of this theorem (Theorem B.1). We define the processes $\mathcal{P}', \mathcal{P}'', \mathcal{P}'''$ below.

In all of these processes there is a queue that, at any time t, contains all patients present at that time in decreasing order with respect to their arrival times. So, the patient at position 1 in the queue is the patient who arrived the latest among all patients present at time t, and so on.

The process $\mathcal{P}'$ is the same as the process $\mathcal{P}^*$, except that, in $\mathcal{P}'$, if the queue length is smaller than a predetermined parameter b, or if the arriving donor is not compatible with

the first b patients in the queue, then we *dismiss* that donor (i.e., we do not match that donor to any patient). The value of b will be specified later.

In the process $\mathcal{P}''$, patients and donors arrive according to the same Poisson processes as in $\mathcal{P}^*$, i.e., Poisson processes with rates C, S, respectively. Similar to $\mathcal{P}^*$, patients depart after a unit of time. The matches, however, are made in a different way in $\mathcal{P}''$. In $\mathcal{P}''$, when a donor arrives, if the length of the queue is less than b, then the donor is dismissed. Otherwise, if the length of the queue is at least b, then the donor is matched to the patient at position b of the queue with probability $1-(1-r)^b$, and is dismissed with probability $(1-r)^b$. The value of b will be set later, so that $b \approx \frac{\log(S+1)}{r}$.

The process $\mathcal{P}'''$ is similar to $\mathcal{P}''$. In the process $\mathcal{P}'''$, the arrival rate of patients is C, and the arrival rate of donors is $\tilde{S} = S(1-(1-r)^b)$. Furthermore, whenever a donor arrives she is matched to the patient at position b of the queue if such a patient exists; otherwise, the donor is dismissed.

We suppose that all of the processes $\mathcal{P}^*, \mathcal{P}', \mathcal{P}'', \mathcal{P}'''$ start with an empty pool. (All results hold with any other initial composition with a finite length.) Let $A'(t)$ denote the set of patients present in the pool at time t in the process $\mathcal{P}'$, and let $L'(t) = |A'(t)|$. Define

$$l' = \lim_{T\to\infty} \int_0^T L'(t)/T \, \mathrm{d}t.$$

An argument similar to the proof of Claim B.2 shows that the above limit exists. Define $A''(t), L''(t), l''$ and $A'''(t), L'''(t), l'''$ similarly for processes $\mathcal{P}''$ and $\mathcal{P}'''$, respectively.

We complete the proof in four steps, by showing that $l \le l'$, $l' \le l''$, $l'' = l'''$, and that $l''' \le C - S + 1 + \frac{1+r}{r} \cdot \log(S+1)$. This will prove (B.9).

Step (i): $l \le l'$.

The proof uses the following coupling of $\mathcal{P}, \mathcal{P}'$. Suppose both processes have identical sequences of arrivals for donors and patients, and that the compatibilities of all patients with all donors are also identical. Then, in any sample path of the coupled process and at any time t we have $A_t \subseteq A'_t$. This implies that $l \le l'$.

Step (ii): $l' \le l''$.

First, we need a definition. The coupled process $\mathcal{Q} = (\mathcal{P}', \mathcal{P}'')$ will be defined in a way that the queue length in $\mathcal{P}''$ remains always at least as large as the queue length in $\mathcal{P}'$, i.e., $L'(t) \le L''(t)$ will be satisfied in all sample paths of the coupling. This would imply that

$l' \leq l''$. Thus, if we construct $\mathcal{Q}$ such that this property is satisfied, then this step would be complete.

The coupled process $\mathcal{Q}$ runs $\mathcal{P}', \mathcal{P}''$ using the same sequence of arrivals of donors and patients. The way the matches are made, however, differs in $\mathcal{P}'$ and $\mathcal{P}''$. Suppose a donor arrives at time t. Then, the donor is matched in the coupling according to one of the following cases:

Case 1. If $L''(t) < b$, then dismiss the donor in both $\mathcal{P}', \mathcal{P}''$.

Case 2. Otherwise, if $L'(t) < b$, then dismiss the donor in $\mathcal{P}'$, and attempt to match the donor as usual in $\mathcal{P}''$, i.e., match the donor to the b-th patient in the queue with probability $1 - (1 - r)^b$, and dismiss the donor with probability $(1 - r)^b$.

Case 3. Otherwise, attempt to match the donor as usual in $\mathcal{P}'$, i.e., attempt to match the donor to the first compatible patient with position no larger than b, and dismiss it otherwise. If the donor is matched in $\mathcal{P}'$, then match the donor to the b-th patient in the queue in $\mathcal{P}''$. Otherwise, dismiss the donor in both $\mathcal{P}', \mathcal{P}''$.

We will show that in any sample path of this coupling, and at any time $t > 0$ in that sample path, we have $L'(t) \leq L''(t)$. To see why, we need the following definitions. For two finite sets $X = \{x_1, \ldots, x_n\}$ and $X' = \{x'_1, \ldots, x'_m\}$, we say X *is weakly smaller than* X' *in the lexicographic order*, and denote it by $X \preceq X'$, if for any $s \geq 0$ we have

$$|\{x \in X : x \geq s\}| \leq |\{x \in X' : x \geq s\}|.$$

For the i-th patient waiting in the queue in the process $\mathcal{P}$, define the *remainder time* of a patient to be one minus the amount of time she has been waiting in the queue. Let $T(t), T'(t)$ respectively denote the sets of remainder times for the waiting patients at time t.

We will show that by the design of the coupling, for all $t > 0$ we have $T'(t) \preceq T''(t)$. The proof is by contradiction. Consider the first event, occurred at time $t > 0$, after which $T'(t) \not\preceq T''(t)$. This event cannot be the arrival or departure of a patient, given that t is the smallest number at which $T'(t) \not\preceq T''(t)$ holds. Therefore, it should be the arrival of a donor. But then, a verification of each of the three cases considered above in the definition of coupling again implies that there should be time $t^* < t$ at which $T'(t^*) \not\preceq T''(t^*)$ holds, which would be a contradiction.

Now, observe that $T'(t) \preceq T''(t)$ implies that $L'(t) \leq L''(t)$, and hence $l' \leq l''$. This completes Step (ii).

Step (iii): $l'' = l'''$.

To prove the claim in this step, we couple the stochastic process $\mathcal{P}''$ to $\mathcal{P}'''$. The coupling is defined as follows. In the coupled process, namely $(\mathcal{P}'', \mathcal{P}''')$, both $\mathcal{P}'', \mathcal{P}'''$ have the same sequence of arrivals of patients. The arrival of donors in $\mathcal{P}'', \mathcal{P}'''$ is determined by a Poisson process, namely $\mathcal{S}$, which has rate S. Any arrival of a donor in $\mathcal{S}$ corresponds to the arrival of a donor in $\mathcal{P}''$. However, arrivals of donors in $\mathcal{P}'''$ are defined differently: for any arrival of a donor in $\mathcal{S}$, a coin is flipped with success probability $1-(1-r)^b$. In case of success, the arrival in $\mathcal{S}$ would also correspond to the arrival of a donor in $\mathcal{P}'''$. In case of failure, no donor arrives in $\mathcal{P}'''$.

Upon the arrival of a donor in $\mathcal{S}$, the matching of donors to patients in $(\mathcal{P}'', \mathcal{P}''')$ occurs as follows:

Case 1. If $L''(t) < b$, then dismiss the donor in $\mathcal{P}''$. In $\mathcal{P}'''$, with probability $1-(1-r)^b$, a donor arrives, but is not matched to any patient.

Case 2. If $L''(t) \geq b$, then in $\mathcal{P}''$, with probability $1-(1-r)^b$, match the donor to the b-th patient in the queue. If the donor is matched in $\mathcal{P}''$, then in $\mathcal{P}'''$ match the donor to the b-th patient in the queue.

The above two cases define the matching of the donors to the patients in the coupling. Let $L''(t), L'''(t)$ respectively denote the length of the queue in processes $\mathcal{P}'', \mathcal{P}'''$ at time $t \geq 0$. Observe that, if $L''(0) = L'''(0) = 0$ holds, then by the definition of the coupling we have $L''(t) = L'''(t)$. This also implies that $l'' = l'''$.

Step (iv): $l''' \leq C - S + 2 + \frac{1+r}{r} \cdot \log(S+1)$.

Recall that $\tilde{S} = S(1-(1-r)^b)$. We consider two cases: $\tilde{S} \leq C-1$ and $\tilde{S} > C-1$.

The case of $\tilde{S} \leq C-1$. Theorem B.1 implies that

$$l''' \leq C - \tilde{S} + \log\left(\frac{C}{C-\tilde{S}}\right) + b.$$

Now, we set $b = \lceil \frac{\log(S+1)}{r} \rceil$. Therefore, the above inequality implies that if $S \leq C-1$, then

$$\begin{aligned} l''' &\leq C - S + 1 + \log\left(\frac{C}{C-S}\right) + \left\lceil \frac{\log(S+1)}{r} \right\rceil \\ &\leq C - S + 1 + \log(S+1) + \left\lceil \frac{\log(S+1)}{r} \right\rceil \\ &\leq C - S + 2 + \frac{1+r}{r} \cdot \log(S+1). \end{aligned} \tag{B.10}$$

The case of $\tilde{S} > C-1$. In this case, Theorem B.1 implies that

$$l''' \leq 1 + \log C \leq 1 + \log(S+1). \tag{B.11}$$

The bounds (B.10) and (B.11) together conclude this step.

B.2.2 Proof of Theorem 2.3

The bounds (B.8) and (B.9) combined imply that

$$C - S \leq l^* \leq C - S + 2 + \frac{1+r}{r} \cdot \log(S+1)$$

when $\zeta = 1$. We recall that we normalized ζ to 1; as discussed before, this is just a rescaling of time done for simplifying notation. Scaling the Poisson arrival rates C, S and the parameter ζ by the same factor does not change the steady state distribution of the number of agents in the queue. Thus, the latter bound on the average queue length l^* translates to the following bound for every $\zeta > 0$

$$(1-\lambda)m \leq l^*/\zeta \leq (1-\lambda)m + 2 + \frac{1+r}{r} \cdot \log(\lambda m + 1) \tag{B.12}$$

where we used $C = m$ and $S = \lambda m$. Multiplying both sides of the above bound by ζ/m implies that, as m approaches infinity, the average waiting time of domestic patients approaches $\zeta(1-\lambda)m$ as m approaches infinity, due to Little's law. Also, the above bound implies that the fraction of matched patients approaches 1 as m approaches infinity, since every unmatched patient remains in the queue ζ units of time. Thus, as m approaches infinity, the average healthcare cost per domestic patient per unit of time approaches

$$\zeta(1-\lambda)D + \lambda(S_d + \mu S_f),$$

where the first and the second summands respectively correspond to the average dialysis cost per domestic patient per unit of time and the average transplant cost per domestic patient per unit of time. Taking the derivative from the above expression with respective to λ proves the theorem.

C Proof of Theorem 2.2

Equation B.12 shows that the average length of the queue is bounded from below and above by $\zeta m(1-\lambda)$ and $\zeta m(1-\lambda+\overline{\gamma})$, respectively, where $\overline{\gamma} = \frac{(1+r)\log m}{rm} + \frac{2}{m}$. Hence, for an increase of ϵ in λ, the average (per domestic patient per unit of time) dialysis cost decreases by at least $\zeta(\epsilon - \overline{\gamma})D$, and the average surgery cost increases by at most $\epsilon(S_d + \mu S_f)$. Therefore,

$$\mathbf{C}(m, \lambda + \epsilon) - \mathbf{C}(m, \lambda) \leq -\zeta(\epsilon - \overline{\gamma})D + \epsilon(S_d + \mu S_f).$$

This proves the claim.

D Proof of Theorem 2.4

We first prove the following theorem.

Theorem D.1. *For every constant $\lambda < 1$, the steady state queue length under the FIFO policy is $\zeta m - o(m)$. Furthermore, the steady state number of donors that are matched per unit of time is $\lambda m - o(m)$.*

D.1 Proof of Theorem D.1

We note that scaling the Poisson arrival rates of patients and donors $(m, \lambda m)$ and the parameter ζ by the same factor does not change the steady state distribution of the number of agents in the queue. In particular, one can take this factor to be the constant $1/\zeta$. This scaling does not change the average queue length or the average number of agents matched per unit of time. Thus, to prove the theorem, without loss of generality we can assume that $\zeta = 1$.

Fix an arbitrary time t. Let $P = \{1, \ldots, m'\}$ denote the set of patients whose departure time belongs to the interval $(t, t+1)$. Also, let $D = \{d_1, \ldots, d_{m'}\}$ denote the set of departure times of these patients, with d_i denoting the departure time of patient i. Without loss of generality, suppose that $d_1 \leq \ldots \leq d_{m'}$.

Let $Q = \{1, \ldots, m''\}$ denote the set of donors who arrive between times t and $t+1$. Also, let $A = \{a_1, \ldots, a_{m''}\}$ denote the set of arrival times of these donors, with a_j denoting the arrival time of donor j. Without loss of generality, suppose that $a_1 \leq \ldots \leq a_{m''}$.

Let $\epsilon_m = \frac{1}{\log m}$. For every integer $i \geq 0$, let

$$D_i = \{d \in D : d \in (t + i\epsilon_m(1+\epsilon_m), t + (i+1)\epsilon_m(1+\epsilon_m))\}. \tag{D.1}$$

Similarly, for every integer $j \geq 0$, let

$$A_j = \{a \in A : a \in (t + j\frac{\epsilon_m(1-\epsilon_m)}{\lambda}, t + (j+1)\frac{\epsilon_m(1-\epsilon_m)}{\lambda})\}. \tag{D.2}$$

Define $\bar{i}$ to be the largest integer such that $(i+1)\epsilon_m(1+\epsilon_m) \leq 1$. Also, define $\bar{j}$ to be the largest integer such that $t + (j+1)\frac{\epsilon_m(1-\epsilon_m)}{\lambda} \leq 1$. Let I denote the set of nonnegative integers no larger than $\bar{i}$ and J denote the set of nonnegative integers no larger than $\bar{j}$.

Claim D.2. *For every nonnegative integer $i \leq \bar{i}$, wvhp it holds that*

$$m\epsilon_m(1 + \frac{\epsilon_m}{2}) \leq |D_i| \leq m\epsilon_m(1+\epsilon_m)^2.$$

Also, wvhp it holds that $|D_{\bar{i}+1}| \leq m\epsilon_m(1+\epsilon_m)^2$. For every nonnegative integer $j \leq \bar{j}$, wvhp it holds that

$$m\epsilon_m(1-\epsilon_m)^2 \leq |A_j| \leq m\epsilon_m(1 - \frac{\epsilon_m}{2}).$$

Also, wvhp it holds that $|A_{\bar{j}+1}| \leq m\epsilon_m(1 - \frac{\epsilon_m}{2})$.

Proof. We first recall the concentration bounds for Poisson distribution [Canonne 2019], which state that for a Poisson random variable X with mean μ,

$$\mathbb{P}\left[|X - \mu| > z\right] \leq 2e^{-\frac{z^2}{\mu+z}}. \tag{D.3}$$

We then note that for every $i \leq \bar{i}$, $|D_i|$ is distributed according to the Poisson distribution with mean $m\epsilon_m(1+\epsilon_m)$. This fact, together with the concentration bound (D.3), proves the claim for $i \leq \bar{i}$. To prove the claim for $i = \bar{i}+1$, it suffices to note that the distribution of $|D_{\bar{i}+1}|$ is stochastically dominated by the Poisson distribution with mean $m\epsilon_m(1+\epsilon_m)$. The concentration bound (D.3) then implies that $|D_{\bar{i}+1}| \leq m\epsilon_m(1+\epsilon_m)^2$ holds wvhp.

To prove the bound on $|A_j|$, we note that $|A_j|$ is distributed according to the Poisson distribution with mean $m\epsilon_m(1-\epsilon_m)$. This fact, together with the concentration bound (D.3),

implies that $m\epsilon_m(1-\epsilon_m)^2 \leq |A_j| \leq m\epsilon_m$ holds wvhp. To prove the claim for $j = \bar{j}+1$, it suffices to note that the distribution of $|A_{\bar{j}+1}|$ is stochastically dominated by the Poisson distribution with mean $m\epsilon_m(1-\epsilon_m)$. The concentration bound (D.3) then implies that $|A_{\bar{j}+1}| \leq m\epsilon_m$ holds wvhp. □

We state a few definitions before proceeding with the proof. For every $i \in I$, let P_i denote the set of patients whose departure times belong to the set D_i. Also, for every $j \in J$, let Q_j denote the set of donors whose departure times belong to the set A_j. Let R denote the set of patients who are present in the queue at time t. Also, let i_0 be the smallest index such that $P_{i_0} \subseteq R$. (Hence, for every $i \geq i_0$, every patient in P_i is present in the queue at time t.) Define $I^* = I \backslash \{0, \ldots, i_0 - 1\}$ and $P^* = \cup_{i \in I^*} P_i$. Also, let J^* denote J excluding its largest element; i.e., $J^* = J \backslash \{\bar{j}\}$. Define $Q^* = \cup_{j \in J^*} Q_j$.

Lemma D.3. *Under the FIFO policy, the following holds wvhp: at least* $\min\{|P^*|\left(\frac{1-\epsilon_m}{1+\epsilon_m}\right)^2, |Q^*|\}$ *matches are made between times* t *and* $t+1$.

Proof. Define $k = \min\{|I^*|, |J^*|\}$. We first observe that

$$k = \min\{|I^*|, |J^*|\} \leq \frac{1}{\epsilon_m(1+\epsilon_m)} \leq \log m. \tag{D.4}$$

Claim D.4. *Every donor in* $Q^0 \cup \ldots \cup Q^{k-1}$ *is matched to a patient in* P^**, wvhp.*

Proof. For every integer $j \geq 0$, let P'_j denote $P_{\bar{i}-k+j+1}$. Also, for every integer $j \geq 0$, let P''_j denote the set of all patients who arrive no later than the patient with the latest arrival time in P_j. By Claim D.2, for every nonnegative integer $i < k$,

$$|P'_i| - |Q_i| \geq m\epsilon_m^2 \tag{D.5}$$

holds wvhp.

For every i, j, we say that P'_i *occurs later than* Q_j if the arrival time of the donor with the latest arrival time in Q_j is sooner than the departure time of the patient with the earliest departure time in P'_i. We observe that, when m is sufficiently large, for every nonnegative integer $i < k$, P'_i occurs later than Q_i. This holds by (D.1) and (D.2).

Define $\mathcal{E}_i$ to be the event in which every donor in Q_i, upon her arrival, is compatible with a patient present in the queue belonging to P''_i. When a donor in Q_0 arrives, there are at least $|P'_0| - |Q_0|$ patients present in the queue belonging to P'_0. (This holds because P'_0

occurs later than Q_0.) Hence, the arriving donor will be matched to a patient in P_0'' with probability at least $1-(1-p)^{|P_0'|-|Q_0|}$. A union bound over all patients in P_0' then implies that

$$\mathbb{P}\left[\,\overline{\mathcal{E}_0}\,\right] \leq |Q_0|(1-p)^{|P_0'|-|Q_0|}. \tag{D.6}$$

On the other hand, $|P_0'|-|Q_0| \geq \epsilon_m^2 m$ holds wvhp, by (D.5). Also, $|Q_0| \leq \epsilon_m m$ holds wvhp, by Claim D.2. The two latter facts, together with (D.6), imply that $\mathcal{E}_0$ holds wvhp. That is, under the FIFO policy, wvhp every donor in Q_0 is matched to some patient in P_0''.

Conditional on event $\mathcal{E}_0$, event $\mathcal{E}_1$ holds wvhp. This is shown using the same argument as above, but for $\mathcal{E}_0, P_0, P_0', P_0'', Q_0$ replaced with $\mathcal{E}_1, P_1, P_1', P_1'', Q_1$, respectively. Similarly, for every $j < k-1$ conditional on event $\mathcal{E}_0 \wedge \ldots \wedge \mathcal{E}_j$, event $\mathcal{E}_{j+1}$ holds wvhp. Then, a union bound, together with (D.4), implies that event $\overline{\mathcal{E}_0} \vee \ldots \vee \overline{\mathcal{E}_{k-1}}$ holds wvlp. Therefore, $\mathcal{E}_0 \wedge \ldots \wedge \mathcal{E}_{k-1}$ holds wvhp. This means that, wvhp, every donor in $Q_0 \cup \ldots \cup Q_{k-1}$ is matched to a patient in P^*. □

Next, we note that

$$Q_0 \cup \ldots \cup Q_{|J^*|-1} = Q^*, \tag{D.7}$$

by definition. Also, by Claim D.2, for every $i \in I^*$ and $j \in J^*$, it holds wvhp that

$$\frac{|Q_j|}{|P_i|} \geq \left(\frac{1-\epsilon_m}{1+\epsilon_m}\right)^2,$$

which implies that

$$|Q_0 \cup \ldots \cup Q_{k-1}| \geq \left(\frac{1-\epsilon_m}{1+\epsilon_m}\right)^2 |P_{i_0} \cup \ldots \cup P_{i_0+k-1}| \tag{D.8}$$

holds wvhp.

Claim D.4, together with (D.7) and (D.8), implies that at least $\min\{|P^*|\left(\frac{1-\epsilon_m}{1+\epsilon_m}\right)^2, |Q^*|\}$ of the donors in Q^* are matched to a patient in P^*. This quantity is a lower bound on the number of matches made between times t and $t+1$. The claim is proved. □

Claim D.5. *It holds wvhp that* $|P^*| \geq X_t - m\epsilon_m(1+\epsilon_m)^2$ *and* $|Q^*| \geq m(\lambda - 3\epsilon_m)$.

Proof. We recall that R is the set of patients who are present in the queue at time t; i_0 is the smallest index such that $P_{i_0} \subseteq R$ (hence, for every $i \geq i_0$, every patient in P_i is present

in the queue at time t), $I^* = I \backslash \{0, \ldots, i_0 - 1\}$, and $P^* = \cup_{i \in I^*} P_i$. Therefore, by Claim D.2, wvhp it holds that

$$|P^*| \geq X_t - m\epsilon_m(1 + \epsilon_m)^2.$$

To prove the claim about Q^*, first recall that $Q = \{1, \ldots, m''\}$ is the set of donors whose arrival time belongs to the interval $(t, t+1)$. Since donors arrive according to a Poisson process with rate λm, then $|Q|$ is distributed according to the Poisson distribution with mean λm. Hence, by the concentration bound (D.3) for the Poisson distribution, wvhp it holds that

$$|Q| \geq \lambda m(1 - \epsilon_m). \tag{D.9}$$

Recall from (D.2) the definition of A_j, and that Q_j denotes the set of donors whose arrival times belong to the set A_j. Note that $Q = \cup_{j=0}^{\bar{j}-1} Q_j$ and $Q^* = \cup_{j=0}^{\bar{j}-1} Q_j$ hold by definition. Therefore,

$$|Q^*| = |Q| - |Q_{\bar{j}}| - |Q_{\bar{j}+1}|. \tag{D.10}$$

By Claim D.2, wvhp it holds that

$$|Q_{\bar{j}}|, |Q_{\bar{j}+1}| \leq \epsilon_m m. \tag{D.11}$$

Finally, we observe that (D.9), (D.10), and (D.11) together imply that

$$|Q^*| \geq \lambda m(1 - \epsilon_m) - 2\epsilon_m \geq m(\lambda - 3\epsilon_m).$$

This concludes the proof. □

Lemma D.6. *It holds wvhp that*

$$X_{t+1} \geq m(1 - \epsilon_m) - \max\{0, \lambda m(1 + \epsilon_m) - \min\{X_t(1 - 4\epsilon_m) - \epsilon_m m, m(\lambda - 3\epsilon_m)\}\}.$$

Proof. Let $\tilde{B}_t$ denote the set of donors with arrival time in $(t, t+1)$. Observe that the distribution of $|\tilde{B}_t|$ is the Poisson distribution with mean λm. Hence, by the concentration bound (D.3) for the Poisson distribution,

$$|\tilde{B}_t| \leq (1 + \epsilon)\lambda m \tag{D.12}$$

holds wvhp.

For every time s, let $\tilde{P}_s$ denote the set of patients with arrival time in $(s, s+1)$. Observe that the distribution of $|\tilde{P}_s|$ is the Poisson distribution with mean m. Hence, by the concentration bound (D.3) for the Poisson distribution,

$$|\tilde{P}_s| \geq (1-\epsilon)m \tag{D.13}$$

holds wvhp.

Lemma D.3 and Claim D.5 imply that at least

$$\min\{X_t(1-4\epsilon_m) - \epsilon_m m, m(\lambda - 3\epsilon_m)\}$$

of the donors in $\tilde{B}_t$ are matched to the patients in $\tilde{P}_t$. Hence, at most

$$\max\{0, |\tilde{B}_t| - \min\{X_t(1-4\epsilon_m) - \epsilon_m m, m(\lambda - 3\epsilon_m)\}\}$$

of the donors in $\tilde{B}_t$ are matched to the patients in $\tilde{P}_{t+1}$. Therefore, the length of the queue at time $t+1$ is at least

$$|\tilde{P}_{t+1}| - \max\{0, |\tilde{B}_t| - \min\{X_t(1-4\epsilon_m) - \epsilon_m m, m(\lambda - 3\epsilon_m)\}\}.$$

The above bound, together with (D.12) and (D.13), implies that

$$X_{t+1} \geq m(1-\epsilon_m) - \max\{0, \lambda m(1+\epsilon_m) - \min\{X_t(1-4\epsilon_m) - \epsilon_m m, m(\lambda - 3\epsilon_m)\}\}.$$

The claim is proved. □

Lemma D.7. *If $X_t \geq \frac{m(\lambda-2\epsilon_m)}{1-4\epsilon_m}$, then wvhp it holds that*

$$X_{t+1} \geq m(1-5\epsilon_m).$$

Otherwise, if $X_t < \frac{m(\lambda-2\epsilon_m)}{1-4\epsilon_m}$, then wvhp it holds that

$$X_{t+1} - X_t \geq m(1-\lambda-7\epsilon_m).$$

Proof. If $X_t \geq \frac{m(\lambda-2\epsilon_m)}{1-4\epsilon_m}$, then

$$\min\{X_t(1-4\epsilon_m) - \epsilon_m m, m(\lambda - 3\epsilon_m)\} = m(\lambda - 3\epsilon_m).$$

Lemma D.6 then implies that

$$\begin{aligned} X_{t+1} &\geq m(1-\epsilon_m) - \max\{0, \lambda m(1+\epsilon_m) - \min\{X_t(1-4\epsilon_m) - \epsilon_m m, m(\lambda - 3\epsilon_m)\}\} \\ &\geq m(1-\epsilon_m) - \max\{0, \lambda m(1+\epsilon_m) - m(\lambda - 3\epsilon_m)\} \\ &\geq m(1-\epsilon_m) - 4\epsilon_m = m(1-5\epsilon_m) \end{aligned}$$

holds wvhp. This proves the first claim.

On the other hand, if $X_t < \frac{m(\lambda-2\epsilon_m)}{1-4\epsilon_m}$, then

$$\min\{X_t(1-4\epsilon_m) - \epsilon_m m, m(\lambda - 3\epsilon_m)\} = X_t(1-4\epsilon_m) - \epsilon_m m.$$

Lemma D.6 then implies that

$$\begin{aligned} X_{t+1} &\geq m(1-\epsilon_m) - \max\{0, \lambda m(1+\epsilon_m) - \min\{X_t(1-4\epsilon_m) - \epsilon_m m, m(\lambda - 3\epsilon_m)\}\} \\ &\geq m(1-\epsilon_m) - \max\{0, \lambda m(1+\epsilon_m) - X_t(1-4\epsilon_m) + \epsilon_m m\} \\ &\geq m(1-\epsilon_m) - (\lambda m(1+\epsilon_m) - X_t(1-4\epsilon_m) + \epsilon_m m) \end{aligned}$$

holds wvhp, which implies that

$$\begin{aligned} X_{t+1} - X_t &\geq m(1-\epsilon_m) - \lambda m(1+\epsilon_m) - 5\epsilon_m m \\ &\geq m(1-\lambda-7\epsilon_m) \end{aligned}$$

holds wvhp. This proves the second claim.

□

Proof of Theorem D.1. Let $z = \lceil \frac{1}{1-\lambda-7\epsilon_m} \rceil$. Consider an arbitrary time t. We will show that, wvhp, $X_{t+z} \geq m(1-5\epsilon_m)$. Since t is chosen arbitrarily, this would imply that, at the steady state, the length of the queue at an arbitrary time is wvhp at least $m(1-5\epsilon_m)$. This would prove the claim of the theorem. Hence, to complete the proof, it suffices to show that $X_{t+z} \geq m(1-5\epsilon_m)$ holds wvhp.

For every nonnegative $i < z$, Lemma D.7 implies the following two facts: (i) If $X_{t+i} \geq \frac{m(\lambda-2\epsilon_m)}{1-4\epsilon_m}$, then $X_{t+i+1} \geq m(1-5\epsilon_m)$ holds wvhp, and (ii) if $X_{t+i} < \frac{m(\lambda-2\epsilon_m)}{1-4\epsilon_m}$, then $X_{t+i+1} - X_{t+i} \geq m(1-\lambda-7\epsilon_m)$ holds wvhp. The two latter facts, together with a union bound over all $i \in \{0, \ldots, z-1\}$, imply that $X_{t+z} \geq m(1-5\epsilon_m)$ holds wvhp. This shows that the steady state queue length is $m - o(m)$.

To complete the proof of the theorem, it remains to show that the steady state number of donors that are matched per unit of time is $\lambda m - o(m)$. To this end, we recall the above proof by which, for every time $t > z$, $X_t \geq m(1-5\epsilon_m)$ holds wvhp. Hence, a donor, upon arrival, is compatible with at least one patient in the queue with probability at least $1-(1-p)^{m(1-5\epsilon_m)}$. This means that the steady state number of donors that are matched per unit of time is $\lambda m - o(m)$. □

D.2 Proof of Theorem 2.4

Theorem D.1 implies that, for every positive constant $\lambda < 1$, the average waiting time of domestic patients approaches ζ as m approaches infinity, due to Little's law. In addition, the theorem implies that the probability that a patient is matched approaches $1/\lambda$ as m approaches infinity. Thus, the average healthcare cost per domestic patient per unit of time approaches $\zeta D + \lambda(S_d + \mu S_f)$, where the first summand and the second summand respectively correspond to the average dialysis cost per domestic patient per unit of time and the average transplant cost per domestic patient per unit of time. That this expression is increasing in λ proves the theorem.