LEMMA Working Paper

n° 2023-06

Unpaired Kidney Exchange : Overcoming Double Coincidence of Wants without Money

Mohammad Akbarpour

Stanford University

Julien Combe

Ecole Polytechnique, CREST

Yinghua He

Rice University

Victor Hiller

Université Paris-Panthéon-Assas, LEMMA

Robert Shimer

University of Chicago

Olivier Tercieux

Paris School of Economics, CNRS

LEMMA EA 4442 - 4 rue Blaise Desgoffe 75006 Paris

Unpaired Kidney Exchange: Overcoming Double Coincidence of Wants without Money*

Mohammad Akbarpour† Julien Combe‡ YingHua He§ Victor Hiller¶
Robert Shimer‖ Olivier Tercieux**

February 4, 2023

Abstract

For an incompatible patient-donor pair, kidney exchanges often forbid receipt-before-donation (the patient receives a kidney before the donor donates) and donation-before-receipt, causing a double-coincidence-of-wants problem. We study an algorithm, the *Unpaired* kidney exchange algorithm, which eliminates this problem. In a dynamic matching model, we show that waiting time of patients under the Unpaired is close to optimal and substantially shorter than widely used algorithms. Using a rich administrative dataset from France, we show that Unpaired achieves a match rate of 63 percent and an average waiting time of 176 days for transplanted patients. The (infeasible) optimal algorithm is only slightly better (64 percent and 144 days); widely used algorithms deliver less than 40 percent and at least 232 days. We discuss a range of solutions that can address the potential practical incentive challenges of the Unpaired. In particular, we extend our analysis to an environment where a deceased donor waitlist can be integrated to improve the performance of algorithms. We show that our theoretical and empirical comparisons continue to hold. Finally, based on these analyses, we propose a practical version of the Unpaired algorithm.

Keywords: Kidney exchange, medium of exchange, dynamic matching.

JEL classification codes: D47, C78, E00

*We thank Nikhil Agarwal, Nick Arnosti, Itai Ashlagi, Eric Budish, Paul Milgrom, Michael Ostrovsky, Parag Pathak, Alvin Roth, Paulo Somaini, Tayfun Sonmez, Utku Unver and several seminar participants for valuable comments. Combe is grateful for the support from the ERC Grant No. 682417, Frontiers in Design - Frontiers in Mechanism Design: Methodology and Applications. Hiller is grateful for the support from ANR grant PAIRED_KIDNEY_DONATION (ANR-17-CE36-0004-01) and Labex MME-DII (ANR11-LBX-0023-01). We are indebted to the French biomedicine agency for providing us the access to the data, in particular Christian Jacquelinet, Benoit Audry and Marie-Alice Macher. All errors remain our own.

†Graduate School of Business, Stanford University. Email: mohamwad@stanford.edu
‡CREST, Ecole Polytechnique. Email: julien.combe@polytechnique.edu
§Rice University. Email: yinghua.he@rice.edu
¶Université Paris-Panthéon-Assas, LEMMA. Email: victor.hiller@u-paris2.fr
‖Kenneth C. Griffin Department of Economics, University of Chicago. Email: shimer@uchicago.edu
**CNRS and Paris School of Economics. Email: tercieux@pse.ens.fr

1 Introduction

Transplantation is the treatment of choice for kidney failure. Yet, all around the world, many people struggle with dialysis while enduring long waits for transplantation and imposing substantial healthcare costs on society,[1] with many thousands of patients dying each year, all due to the shortage of compatible organs. *Kidney exchange* is a recent innovation addressing this issue, especially when there are incompatible patient-donor pairs. An incompatible pair is formed when a donor is willing to donate a kidney to a patient but is unable to do so because she is biologically incompatible with the patient. Between two such pairs, if the donor of each pair is compatible with the patient in the other pair, the two pairs can exchange donor kidneys. While ingenious, this leads to the well-known "double-coincidence-of-wants" problem—you not only have to have the kidney that I want, but also have to want the kidney that I have (Jevons, 1885).

To overcome the double-coincidence of wants problem, we propose a new matching algorithm—*Unpaired* kidney exchange. In essence, we create a marketplace where patient i can receive the (compatible) kidney of donor j, even if donor i's kidney is not compatible with patient j. When such a trade happens, patient j will be categorized as an unpaired patient, meaning that she has the right to receive a kidney in the future. At the same time, donor i be categorized as an unpaired donor, meaning that her kidney can be given to some other patient in future.

To convince policymakers to adopt this algorithm, we shall provide an answer for at least the following three questions: First, is this algorithm going to meaningfully outperform the currently used algorithms? By answering this question, our analysis will also quantify the benefits of relaxing the simultaneity constraints embedded in most kidney exchange algorithms. Second, given that it allows for receiving a kidney before donating one and vice versa, is it an incentive-compatible proposal? And third, is it morally acceptable or does it involve repugnance considerations?

This paper provides a comprehensive theoretical and empirical investigation of the Unpaired algorithm to answer the first and second questions, and briefly discusses the third one. Section 2 develops a dynamic kidney exchange model with two types of patients. Patient-donor pairs arrive at some rate n. A fraction λ of patients are *hard-to-match* and the rest, $1-\lambda$, are *easy-to-match*. Hard-to-match and easy-to-match patients are compatible with a random donor with probabilities p_H and p_E, respectively, where $p_H < p_E$. This two-type assumption is a reasonable approximation to the continuous but bimodal distribution of match probability among the patients in kidney exchange in the U.S. and in France.[2] Patients and donors stay until they are matched. The planner, everything else equal, wishes to match patients with donors as quickly as possible. Hence, the main objective of interest is minimizing the average waiting time of patients.

[1] The U.S. Medicare's dialysis cost in 2020 was nearly 1 percent of the entire federal budget.Dialysis cost is obtained by summing spending on dialysis (https://usrds-adr.niddk.nih.gov/2022/end-stage-renal-disease/9-healthcare-expenditures-for-persons-with-esrd) and spending on drugs for ESRD patients (https://usrds-adr.niddk.nih.gov/2022/end-stage-renal-disease/10-prescription-drug-coverage-in-patients-with-esrd). The U.S. federal budget for 2020 is available here: https://www.cbo.gov/publication/57170.

[2] This is shown at least graphically in Ashlagi et al. (2019) for the case of the U.S. For France, we run the DIP test (Hartigan and Hartigan, 1985) and fail to reject the hypothesis of bimodality.

The Unpaired algorithm works as follows: whenever a new patient-donor pair arrives, match the patient to a compatible donor (if any), and match the donor to a compatible patient (if any), breaking ties in favor of hard-to-match patients.

To evaluate the (relative) performance of the Unpaired algorithm, we study three alternative matching algorithms. The first—the *Pairwise* algorithm—matches two patient-donor pairs whenever they are pairwise compatible. The second—the *Chain* algorithm—starts with a finite number of altruistic donors and matches patients with donors whenever there exists a chain of donations starting with an altruistic donor. These two algorithms and some combination of the two correspond to the state-of-the-art algorithms used in most countries. The third—the *Optimal* algorithm—minimizes patients' average waiting time in the class of all matching algorithms.

We prove the following results, all in the regime that p_H is small: First, the Unpaired algorithm substantially outperforms the Pairwise algorithm. In particular, if the majority of patients are hard-to-match (i.e., $\lambda > 1/2$), the ratio of the waiting times under the two algorithms is $O(1/p_H)$. Even if there are more easy-to-match pairs ($\lambda < 1/2$), Unpaired still outperforms Pairwise. For instance, when only 30 percent of patients are hard-to-match, the waiting time of hard-to-match patients under Pairwise is at least twice as long as under Unpaired.

Second, the Unpaired algorithm outperforms the Chain algorithm; in particular, if the fraction of hard-to-match patients (λ) is large, the Chain algorithm's performance becomes substantially worse than Unpaired. For instance, if 60 percent of patients are hard-to-match, the Unpaired algorithm matches hard-to-match patients nearly twice as fast as Chain.[3]

Finally, we compare the Unpaired algorithm with the Optimal algorithm. We prove that the Optimal algorithm's waiting time is at least 50 percent of that under the Unpaired algorithm. Note that the Unpaired algorithm matches patients and donors greedily, while the Optimal algorithm is forward-looking and can in principle wait to thicken the market. This result, nevertheless, shows that the additional gains from thickening the market are relatively small.

After presenting our theoretical results, we empirically investigate the performance and challenges of the Unpaired algorithm in Section 3. This is necessary because our theoretical model ignores many real-world details of the kidney exchange problem. For instance, a patient's biological compatibility with a donor depends on blood type and tissue type compatibilities; the compatibility realizations are unlikely to be i.i.d. across patient-donor pairs in practice. The empirical analysis imposes no such assumptions on compatibilities. Instead, they directly come from data.

We rely on a dataset provided by the Agency of Biomedicine (*Agence de la Biomédecine*), a government agency that oversees all organ transplants in France. It covers the period of December 2013 to February 2018, including all transplants with deceased and living donors kidneys. We identify a pool of *incompatible* pairs who are most likely to participate in a kidney exchange program (KEP), i.e., the 78 pairs who participated in France's KEP and another 508 pairs who went through

[3] We emphasize that this result (and only this result) is shown under the assumption that an easy-to-match patient is compatible with all donors ($p_E = 1$). As noted in Ashlagi et al. (2019), the Markov chain induced by the Chain algorithm is hard to analyze. However, our simulations and those in Ashlagi et al. (2019) show that the average waiting time under the Chain algorithm when $p_E = 1$ is a good approximation to the average waiting time for $p_E < 1$; moreover, we theoretically show that the waiting time under Unpaired does not depend on p_E.

incompatible transplantation facilitated by desensitization. We then sample with replacement from the pool to generate dynamic markets of different sizes (i.e., different arrival rates of patient-donor pairs).[4]

We run simulations to compare four algorithms: Pairwise, Chain, Unpaired, and Optimal. We evaluate an algorithm's performance by the transplant rate (the fraction of the patients in the simulation sample receiving a transplant in the simulation period), as well as the average waiting time of transplanted patients. Because computing the Optimal algorithm requires additional assumptions on the data generating process, we simulate an even better-than-optimal alternative, the *Omniscient* algorithm, which assumes that the planner has perfect foresight about all arrivals in our sample period. Given this perfect information, the planner simply minimizes the average waiting time of patients over our sample period. The waiting time of the Omniscient is a lower-bound for the waiting time of *any* algorithm, including the Optimal algorithm.

Consistent with our theoretical results, the simulations show that Unpaired performs much better than the Pairwise and Chain algorithms. In particular, both Pairwise and Chain have a transplant rate below 40 percent, while Unpaired obtains 63 percent. Perhaps more surprisingly, Unpaired's transplant rate is almost equal to the Omniscient's rate, which is 64 percent. The same pattern holds for waiting times of transplanted patients—248 days for Pairwise, 232 days for Chain, 176 days for Unpaired, and 144 days for Omniscient. We show that these findings are not driven by the small size of the French KEP, as we find similar results for a wide range of market sizes.

After showing that the Unpaired algorithm can perform substantially better than the typically used algorithms, we turn into our second main question: Is it incentive compatible? The Unpaired algorithm comes with two practically relevant incentive concerns. First, because of donation-before-receipt, a patient whose intended donor already gave his kidney may wait for a long time *after* her paired donor's donation. In our main simulation of a market the size of the French KEP (on average 83 pairs), the median waiting time among the 29 unpaired patients after their donors' donation is 245 days. A pair may find it unacceptable to donate a kidney in exchange for a kidney that may arrive so late. Second, due to receipt-before-donation, a donor may wait for a long time after her paired patient's transplantation, increasing the chance that she reneges or becomes unfit to donate. In our main simulation, the median waiting time among the 26 unpaired donors after their patients' transplant is 339 days.

Similar challenges also exist in some current practices that are related to the Unpaired algorithm.[5] The first concern exists in the "voucher" programs in the U.S. (Veale et al., 2017). Donors in these programs donate their kidneys in exchange for a future kidney promise. Notably, some of their paired patients are not in an urgent need for a kidney, while some may not ever need one—their voucher will be used only if a member of their family needs a kidney. Yet, many donors are willing to participate. In addition, our simulations show that the long waits of unpaired patients are largely due to the small market size, as they decrease significantly with size in our simulations. This makes us cautiously optimistic that the first concern may not be binding in practice.

[4] We discuss robustness checks with respect to the pool from which we sample our pairs in footnote 29.

[5] See Section 5 for a detailed discussion of the connections.

The second concern—that unpaired donors may renege—is a challenge in the current practice of the Chain algorithm, where a donor donates only if his paired patient has already received a kidney. Data shows that such donors rarely renege. Through simulations, we show that even renege rates as high as 10 times of that assumed in the medical literature (Gentry et al., 2009), or 30 times of the rate documented among bridge donors in a chain (Cowan et al., 2017), do not significantly affect the performance of Unpaired. This also makes us optimistic that the Unpaired algorithm may not encounter serious practical risk.

Having said that, the incentive challenges discussed may still be seen as obstacles, especially for small markets in which they are more prominent. Hence, we next propose a practical solution by taking advantage of the flexibility of the Unpaired algorithm to introduce modifications that can (almost) fully address these concerns. The key idea is to use the kidneys supplied to the *deceased donor list* (DDL). We propose a modified Unpaired algorithm—Unpaired with DDL—under which patients who do not get matched with a compatible living donor upon joining the KEP will be offered both arriving kidneys from living donors as well as arriving DDL kidneys. To ensure this proposal does not hurt patients who are waiting in the DDL, where a KEP patient is matched to a DDL kidney, a living donor waiting in the KEP will have to donate his kidney back to a patient waiting on the DDL. The last part of the paper analyzes this algorithm, theoretically and empirically.[6]

First, we extend our dynamic matching model to allow for arrivals of deceased donors. We show that all the theoretical results of Section 2 comparing Unpaired, Pairwise, and Optimum extend to this new environment where all algorithms have access to DDL kidneys. An interesting theoretical property of the Unpaired with DDL is related to the comparative static of increasing the arrival rate of pairs, keeping the arrival of DDL kidney fixed. In this case, there can be more competition for DDL kidneys from paired patients, and thus the waiting times of patients may become longer. We theoretically prove that the waiting time of patients under Unpaired with DDL *decreases* when the arrival rate of pairs increases. Importantly, this is *not* necessarily the case for Pairwise with DDL: we show that the waiting time of patients under Pairwise with DDL may increase when the arrival rate of pairs increases. Hence, if the good performance of Unpaired successively attracts more pairs into the system, the performance will not be jeopardized.

Next, we empirically investigate the Unpaired with DDL. We take the arrival of DDL kidneys in our data as given and only offer high-quality DDL kidneys (based on the commonly used *kidney donor profile index*) to patients. Both Pairwise and Unpaired algorithms have a significantly improved performance in their versions with DDL: in a market similar to the French KEP, the mean waiting time of patients is reduced by about 88 to 91 percent. Consistent with our theoretical results, the Unpaired still performs better than the Pairwise, and is very close to the Omniscient. In addition, Unpaired with DDL is more favorable than Pairwise with DDL to O patients who typically have the longest waiting times (Glander et al., 2010).

Our empirical counterfactuals also show that the two aforementioned practical challenges are

[6] As we will see the use of DDL kidneys will help alleviating the two aforementioned incentive issues. While we do not formally investigate this, it also serves another purpose which is to incentivize patient-donor pairs to join the KEP.

successfully addressed: the median waiting time for unpaired patients and donors is 50 days and 65 days, respectively. Again, these waiting times decrease sharply as market size grows.

There is, however, a negative unintended consequence for the Unpaired with DDL algorithm. Precisely because this algorithm is fairly successful in matching patients quickly to a living or deceased donor, it does not provide enough incentives for patients to find easy-to-match donors. A patient who has two potential donors in the family with blood types O and AB gets no reward for bringing the O donor, since in any case she is likely to be matched to a deceased donor quickly. And, in fact, she may prefer to bring the AB donor to reduce the probability of her donor ending up donating a kidney.

This concern motivates us to consider a version of Unpaired with DDL that we propose as the final and most practically plausible solution, the *Unpaired with DDL with delay* δ : In the Unpaired with DDL algorithm, each patient is required to wait for δ months before receiving any DDL kidney offers, but a patient whose donor has already donated can receive DDL kidney offers immediately. This modified version provides incentives for patients to find a donor who is likely to donate soon to a patient in the KEP (e.g., an O donor who is likely to be compatible with many patients), so that they can receive high-quality DDL kidneys earlier.

In our simulations, we find that the algorithm with $\delta = 6$ months can address all the practical concerns that are discussed above. In a market similar to the French KEP, (i) it matches 55 percent of patients with a living donor, hence increasing the incentive to bring a good donor to get matched earlier, (ii) a median unpaired patient only waits for 6 days before receiving a kidney, and (iii) a median unpaired donor waits 39 days in donating a kidney.

We close this section by discussing the third main question: Is our proposal morally acceptable? To answer that, let us first remind that economists have previously proposed a solution to the kidney shortage: legalizing the exchange of money for kidneys (Becker and Elias, 2007). Such a market would resolve the double coincidence problem by allowing a donor to sell her kidney to any patient and permitting a patient to buy a kidney from any donor. Second, a kidney market will likely increase the supply of donors.[7] While a kidney market is appealing to some economists, many people find it repugnant. Roth (2007) lists three reasons why a kidney marketplace might be repugnant. First, giving a kidney to a loved one is intrinsically good, while giving one for money may be morally wrong because it objectifies the human body. Second, it is likely that disproportionately many poor people would sell a kidney, and this may be viewed as coercive. Third, a marketplace for kidneys can be a slippery slope into more ethically dubious arrangements, for example, those in which debtors could be forced to give a kidney in bankruptcy proceedings. We do not take a stance on whether a marketplace for kidneys is repugnant, but simply note that it is illegal everywhere in the world except Iran today, and that changing the law is probably not politically feasible in most countries.

[7]Indeed, in the only legal kidney market in the world in Iran, there is a large supply (nearly 45% of kidney donors are paid living donors), which in turn has led to low transplant waiting times for kidney patients (less than one year). A drawback of the market is that it has almost completely crowded out family donations. See Akbarpour et al. (2019) for a detailed analysis of the Iranian market for kidneys.

Recall that the Unpaired algorithm creates a marketplace where patient i can receive the kidney of donor j, even if donor i's kidney is not compatible with patient j. When such a trade happens, one interpretation is that pair j receives a $+1$ token, meaning that the patient j has the right to receive a kidney in the future. At the same time, pair i receives a -1 token, meaning that donor i's kidney can be given to some other patient in future.[8] However, we believe the Unpaired algorithm, by design, avoids encouraging kidney donation for pecuniary benefits, thereby avoiding the repugnance concerns associated with a kidney market. Donors give kidneys because they love someone who needs one, not because of money. Poor people cannot sell a kidney. Creditors cannot demand a kidney in return for discharging a debt. Technically, the key difference between money and tokens is fungibility. The fungibility of money allows for potentially repugnant uses. The fact that tokens are non-fungible and attached to a specific patient or donor means that society can control how it is used to avoid concerns about repugnance.

1.1 Related work

The economics literature on kidney exchange starts with Roth et al. (2004). In a subsequent paper, Roth et al. (2007) demonstrate the efficiency gains of creating a large kidney exchange, as well as those from allowing 3-way or larger cycles.

The double-coincidence-of-wants problem has been a known challenge since the beginning of kidney exchange. We now review the two approaches that have been used to tackle it in practice.

The first approach is to create a sufficiently thick market:

> *And we will show that, even without a medium of exchange, if the market is thick enough, the problem of the coincidence of wants can be substantially ameliorated by the organization of an appropriate clearinghouse. (Roth et al., 2007)*

A KEP grows when more incompatible pairs join the market. More recently, Sönmez et al. (2020) propose an incentivized system for *compatible* pairs to participate in an exchange. More specifically, the system "rewards" compatible pairs participating in the KEP with a high priority on the DDL once they need a repeat transplant in case of a kidney failure. This will not only increase the market thickness, but also change the composition of patients and donors in the KEP. In turn, this may help blood-type O patients who have a hard time finding a match.

Similar to the literature, our simulation confirms these results. Every algorithm's performance improves when the market is thicker. Yet, except for Unpaired, none of the state-of-the-art algorithms are close to the Optimal algorithm, while the advantages of Unpaired remain even when the KEP is three times larger than the largest one in the world (i.e., the NKR in the U.S.) and when some compatible pairs participate in the KEP. In this sense, the Unpaired algorithm is a way to

[8] A donor giving his kidney today provides a favor in exchange of which the associated patient obtains a $+1$ token. This token entitles the patient to receiving a kidney/favor at some future date. This defines an exchange rate between current donation and future donation. This is reminiscent of "chip strategies" introduced by Möbius (2001) in the context of a favor-exchange model.

reach the full potential of a KEP of any given size with any other incentive schemes (such as the incentivized-exchange policy of Sönmez et al. (2020)).

The second approach is to authorize non-simultaneous exchanges by allowing receipt-before-donation or donation-before-receipt. Non-simultaneous altruistic donor chains (Roth et al., 2006) allow for receipt-before-donation. A chain is initiated by an altruistic donor who donates a kidney to a patient whose paired donor then donates to another patient, and so on. Transplants may happen simultaneously or sequentially. Nowadays, such chains account for a large fraction of kidney exchange transplants. By allowing receipt-before-donation, a chain need not form a closed loop, and thus, can alleviate the problem of double-coincidence-of-wants:

> *Developing the capability to arrange trades in longer cycles and chains helps overcome this [double coincidences of wants] barrier... In the case of kidney exchange, long non-simultaneous chains of the sort proposed in Roth et al. (2006) are proving increasingly important. (Ashlagi et al., 2012)*

While clever, the Chain algorithm confronts three practical challenges. First, its efficiency is limited by the number of available altruistic donors. Second, in places where altruistic donation is illegal (e.g., France and Germany), this algorithm is infeasible. Last but not the least, even with a reasonable number of altruistic donors, it goes only half-way in solving the double coincidence problem, because donation-before-receipt is not allowed.

Ausubel and Morrill (2014) introduce the idea of "sequential kidney exchange" that allows donation-before-receipt but not receipt-before-donation, opposite to the Chain algorithm. They study this in an overlapping generations model. In this sense, our Unpaired algorithm combines Chain and sequential kidney exchange by allowing both donation-before-receipt and receipt-before-donation.

Similar to the Unpaired algorithm is the voucher program that has been adopted by multiple hospitals in the U.S. (Veale et al., 2017). This program allows donation-before-receipt; in particular, a donor can donate and receive a voucher that her paired recipient can use to receive a kidney in future. We compare this program with the Unpaired algorithm in Section 5.

Our theoretical model is related to those of dynamic kidney exchange. Ünver (2010) studies a model of dynamic exchange with blood-type considerations. Akbarpour et al. (2020) consider a dynamic kidney exchange model with stochastic departures and show that optimal timing can be highly valuable; their focus, however, is only on pairwise exchanges. The two-type model studied here builds on the model of Ashlagi et al. (2019), where they compare Chain and Pairwise. To the best of our knowledge, our paper is one of the first papers offering a dynamic setting including DDL kidneys into KEP.[9] This setting may be useful for further research as well.

2 Theoretical Analysis of the Unpaired Algorithm

[9] A noticeable exception is Sönmez et al. (2018) who characterize—in a continuum model—match rates with deceased/living donors as a function of some policies implemented in the KEP.

2.1 Model

We now introduce a continuous-time, infinite-horizon model of a dynamic kidney exchange market.

Arrivals and types of patients. Incompatible patient-donor pairs arrive at the market according to a Poisson process with rate n. There are two types of patients: *hard-to-match* and *easy-to-match*. We refer to these types as H and E, respectively. A fraction $\lambda > 0$ of patients are hard-to-match and a fraction $(1-\lambda)$ are easy-to-match. An H patient is compatible with any donor with probability p_H, and an E patient is compatible with any donor with probability p_E. We discuss the plausibility of this assumption in Section 2.3.[10]

For any $t \geq 0$, let V_t^p and V_t^d be the set of patients and donors in the market at time t, respectively, and $S_t = |V_t^p|$ and $Z_t = |V_t^d|$. Define $\mathcal{E}_t \subseteq V_t^p \times V_t^d$ as the set of compatible patient-donor pairs and $\mathcal{G}_t = (V_t^p, V_t^d, \mathcal{E}_t)$ as the (bipartite) *compatibility graph* at time t. We refer to $\mathcal{E}_t$ as the set of *edges*. When a new incompatible patient-donor pair $v_i = (p_i, d_i)$ arrives at time t, edges are formed between p_i and all compatible donors in V_t^d, as well as between d_i and all compatible patients in V_t^p.

Matching algorithms. A set of edges (possibly empty) is a *matching* if no two edges share the same endpoints. A *matching algorithm*, at any time t, selects a matching M_t in the graph $\mathcal{G}_t$. The endpoints of the edges in M_t leave the market immediately. This definition of a matching algorithm does not require a donor d_i and her paired patient p_i to be in the same matching. Thus, there are algorithms that are illegal in some countries. For instance, the usual pairwise kidney exchange—which is the only legal form of exchange in France—substantially limits the set of possible matchings: A *pairwise compatibility* happens when two incompatible patient-donor pairs v_i and v_j are cross-compatible; that is, there are an edge between p_i and d_j and another edge between p_j and d_i. In pairwise kidney exchange, only pairwise compatible pairs can be matched.

A matching algorithm induces a stochastic process over the number of patients of each type remaining in the system. In this study, we restrict our attention to matching algorithms inducing a stochastic process with a unique invariant distribution. We consider three myopic matching algorithms, with the first two, the Pairwise and the Chain algorithms, from the literature (Akbarpour et al., 2020; Ashlagi et al., 2019) and the third, the *Unpaired* algorithm, being the core contribution of this study. We formally prove in Appendix C.5 that the Unpaired algorithm has a unique invariant distribution (the same argument can be used for Chains and Pairwise, see also Ashlagi et al. (2019)).[11]

[10] In our model, agents do not take any decision on whether to join the pool of pairs. In practice, when joining the pool agents tradeoff their waiting time to get a transplant with their outside options (e.g. desensitization). While the analysis of participation decisions is highly non-trivial, we are optimistic that the superiority of the Unpaired algorithm that we introduce in this paper over Pairwise and Chains will remain valid. See our discussion in the conclusion section.

[11] As is well-known, at an informal level, a Markov chain over a finite state space has a unique invariant distribution if it is possible to eventually get from every state to every other state with positive probability. In particular, starting from any state, with positive probability, the process will be back to the very same state with positive probability. As will become clear, this must be the case for the Markov chains induced by the algorithms we study. One issue though is that the state space of our Markov chains are countably infinite. So to ensure existence of a unique invariant

Definition 2.1 (Pairwise)**.** *If any new patient-donor pair v_i enters the market at time t, then match them with any cross-compatible patient-donor pair (if any), breaking ties in favor of hard-to-match patients.*[12]

For a pair, the Pairwise algorithm forbids *donation-before-receipt* (a donor donates before the paired patient receives a kidney) and *receipt-before-donation* (a patient receives a kidney before the paired donor donates). Therefore, it creates the problem of double-coincidence-of-wants.

We now introduce the *Chain* algorithm. Note first that it is feasible only in settings where altruistic donors exist. The definition is taken from Ashlagi et al. (2019) and implicitly extends the setting to allow for the existence of an altruistic donor at the beginning of time.

Definition 2.2 (Chain)**.** *There is a bridge or altruistic donor in the market at any given time.*[13] *Consider a newly arriving pair $v_1 = (p_1, d_1)$. If p_1 does not have an edge to the bridge (or altruistic) donor, then no match is formed. Otherwise, a chain-segment begins with matching the bridge (or altruistic) donor with p_1 and advances as follows. First, we search for an H patient that has an edge to d_1; if there are multiple such H patients, we select one uniformly at random; otherwise, we search for an E patient that has an edge to d_1 (again breaking ties uniformly at random). With the paired donor of the selected patient available to be matched, the process repeats among the pairs without a selected patient, until we select a patient whose paired donor is incompatible with all never-selected patients, forming a disjoint path. All patients and donors in the disjoint path leave the market, and the paired donor of the last selected patient becomes a bridge donor.*

Essentially, upon the arrival of a new patient-donor pair, the Chain algorithm identifies a chain in a greedy fashion. This policy does not necessarily pick the longest chain (since it is searching in a greedy fashion), but it does find a maximal size chain, i.e., a chain that is not properly contained in a longer chain, and at the same time gives priority to H patients. In that respect, we follow Ashlagi et al. (2019); we further discuss this in Remark 2.11.

For a pair, the Chain algorithm still forbids donation-before-receipt but allows receipt-before-donation if the patient is compatible with some donor in the chain. We are now ready to introduce the *Unpaired* algorithm that allows donation-before-receipt and receipt-before-donation.

Definition 2.3 (Unpaired)**.** *If any new patient-donor pair $v_i = (p_i, d_i)$ enters the market at time t, match p_i to a compatible donor (if any), breaking ties arbitrarily, and match d_i to a compatible patient (if any), breaking ties in favor of hard-to-match patients.*

The Unpaired algorithm allows both donation-before-receipt and receipt-before-donation. Whenever a pair engages in donation-before-receipt, the algorithm will match the pair's patient, defined

distribution, one has to further guarantee that the expected amount of time to return to a state given that the chain started in that state has finite first moment. Appendix C.5 proves this appealing to a sufficient condition provided in Meyn and Tweedie (1993).

[12] Our theoretical results do not depend on the way we break ties within types. In particular, this implies that the choice of the queueing discipline (first-come-first-served or others) within type is inconsequential for our theoretical results.

[13] As in Ashlagi et al. (2019), we may assume that there are finitely many $d \geq 1$ bridge or altruistic donors at any given time. The results will remain essentially the same. See footnote 22 for details.

as an *unpaired patient*, with a compatible kidney in the future. Similarly, if a pair does receipt-before-donation, the algorithm will ask the pair's donor, defined as an *unpaired donor*, to donate a kidney in the future. In other words, for a pair, the algorithm searches for a match for the patient while independently finding a match for the donor, as if they were unpaired. Hence, we call it the Unpaired algorithm.

By relaxing the timing constraints on kidney donation and receipt for a pair, the Unpaired algorithm may create two incentive issues. First, donation-before-receipt creates unpaired patients who need to wait to receive a compatible kidney after their donors' donation; if the wait is expected to be long, they may opt out of donation-before-receipt or quit the KEP all together. Second, receipt-before-donation results in unpaired donors who will wait to donate after their paired patients' transplantation; they may renege or become unfit to donate if the wait is long. In Section 3.4, we present an extensive empirical investigation of the consequences of these issues as well as several solutions addressing them. In short, as shown in the empirical analysis, these issues diminish with market size; even when some pairs opt out of donation-before-receipt or when some unpaired donors renege, the performance advantage of the Unpaired algorithm remains; still we acknowledge that these incentive issues may still be a concern for a policymaker, in particular, in a small KEP. In Section 4, we offer practical solutions which can effectively address these issues.

Objective. Patients and donors stay in the market until they are matched. For a patient p_i who enters the market at time t_0 and gets matched at time t_1, let $w(p_i) = t_1 - t_0$ be p_i's waiting time. Our objective is to minimize the average waiting time at the invariant distribution (recall that we focus on matching algorithms inducing stochastic processes that have a unique invariant distribution). By Little's law, this is equivalent to minimizing the average number of patients in the system.[14,15] Let $\mathbf{W}(ALG)$ denote the average waiting time for a given matching algorithm ALG in steady state. This is equal to $\lambda \mathbf{W}_H(ALG) + (1-\lambda)\mathbf{W}_E(ALG)$, where $\mathbf{W}_H(ALG)$ and $\mathbf{W}_E(ALG)$ denote the expected waiting time of hard-to-match and easy-to-match patients, respectively.

Optimal solution. In the following, we sometimes compare the performance of a matching algorithm to an optimal algorithm that is a theoretical benchmark but practically infeasible. We define the Optimal algorithm as the one achieving the smallest average waiting time that can be achieved by a matching algorithm. Unlike the other algorithms we study, the optimal algorithm need not be greedy, i.e., it may delay matching a patient/donor if they may help pairs arriving in the future. Formally, we define the average waiting time achieved by the Optimal algorithm, $\mathbf{W}(Optimal)$, as

$$\inf \mathbf{W}(ALG)$$

[14]Little's law states that the long-term average number of agents in a stationary system is equal to the long-term average effective arrival rate multiplied by the average time that an agent spends in the system.

[15]Note that, unlike in Akbarpour et al. (2020), our patients and donors do not depart. This makes our analysis less tedious, without creating any difference in the objective function. Our goal here is to minimize the total waiting time, whereas in a model with departures the goal is a mix of waiting time and deaths. With linear waiting cost and Poisson departures, Little's law implies that both of these objectives are minimized by minimizing the pool size.

where the infimum is taken over all matching algorithms (inducing a stochastic process with a unique invariant distribution). Therefore, the optimal algorithm also induces an invariant distribution.

The optimal algorithm depends on the whole "network structure" of patients and donors; that is, which patient is compatible to which donor at each point in time. Since the space of such networks grows exponentially, the optimal algorithm is generally intractable. However, it serves as a valuable benchmark.

It is intuitive that the Unpaired algorithm performs better than the Pairwise and even Chains—after all, it allows both receipt-before-donation and donation-before-receipt, imposing fewer constraints on the algorithms. While this intuition turns out to be theoretically valid, it is not as obvious as it may seem. In fact, in some realizations of the system, the Pairwise and Chain algorithms outperform the Unpaired. We illustrate this with two examples: Example 2.4 is one such scenario under which Pairwise can lead to significantly lower waiting time than Unpaired. Similarly, Example 2.5 shows that the Chain algorithm can lead to significantly lower waiting times than Unpaired. These examples illustrate that Unpaired may match agents "too quickly" while more transplants could be obtained by waiting for further arrivals. However, our main results will show that taking the average waiting times over all realizations of the stochastic process, the average waiting time of Unpaired is significantly lower than those of the widely used practical algorithms.

Example 2.4 (Pairwise *vs* Unpaired)**.** *Let $v_1, \ldots, v_4$, be the sequence of pairs who arrive until time T under one draw of the Poisson process (v_1 arrives first, v_2 second and so on). The compatibility graph is given in Figure 1 (panel A). Under this realization, Unpaired matches patients p_1, p_2 and p_4. To see why, note that when v_2 arrives, p_2 gets matched to d_1, and when v_3 arrives, d_3 donates to p_1, and when v_4 arrives, d_2 donates to p_4. On the contrary, under Pairwise v_1 and v_3 will perform a kidney exchange when v_3 arrives, so will v_2 and v_4 when v_4 arrives. Hence, Pairwise matches all the patients, performing better than Unpaired. By choosing appropriately the arrival times of future arriving pairs, one can easily show that the average waiting time can be lower under Pairwise than under Unpaired.*

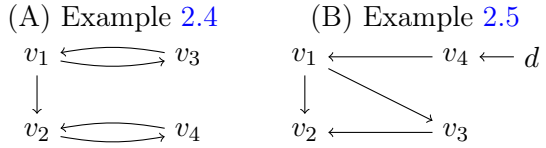

Figure 1: Compatibility graphs in Examples 2.4 and 2.5

Example 2.5 (Chain *vs* Unpaired)**.** *Let $v_1, \ldots, v_4$, be the sequence of pairs who arrive until time T under one draw of the Poisson process and let d be the altruistic donor who is in the market from the beginning. The compatibility graph is given in Figure 1 (panel B). Under this realization, for Unpaired, when v_2 arrives, d_1 donates to p_2, when v_3 arrives, nothing happens since p_2 was already matched earlier, and when v_4 arrives, d_4 donates to p_1. Hence, patients 1 and 2 get matched. Under*

the Chain algorithm, no matching will happen until v_4 arrives. At that point, a chain is activated. Recall that the Chain algorithm selects a maximal chain in a greedy fashion so that several chains can be implemented with positive probability. One such chain is the longest chain where d donates to p_4, d_4 donates to p_1, d_1 donates to p_3 and d_3 donates to p_2. So all patients get matched in this longest chain. The other maximal chain skips v_3 and so involves 3 patients being grafted. Thus, the expected number of grafts under Chain is strictly higher than under Unpaired. By choosing appropriately the arrival times of future arriving pairs, one can easily show that the average waiting time can be lower under Chain than under Unpaired.

2.2 Theoretical Results

This section compares the average waiting time of patients at steady state under the Unpaired algorithm with those under the Pairwise, Chain and optimal algorithms. We state our results for the regime where p_E is held constant and $p_H \to 0$. This regime should not be interpreted literally; as discussed in Section 2.3, the type distribution has a binomial shape, and $p_H \to 0$ captures the idea that a substantial fraction of patients have exceedingly low compatibility probabilities. In addition, this assumption makes the analysis theoretically tractable by removing nuisance terms. This is a standard method in random graph theory, starting from Erdős and Rényi (1960), and has been applied in the analysis of kidney exchange graphs recently (e.g., Akbarpour et al. (2020) and Ashlagi et al. (2019); for a more comprehensive discussion, see Ashlagi and Roth (2020)). Of course, as $p_H \to 0$, the average waiting time of hard-to-match patients explodes. Hence, we focus on their "normalized" waiting times, i.e., $p_H \mathbf{W}_H(Unpaired)$ (and perform a similar exercise for easy-to-match patients). Working with normalized waiting times will prove useful to compare waiting times across algorithms when p_H is small. The following proposition characterizes normalized waiting times for easy and hard-to-match patients under the Unpaired algorithm.

Proposition 2.6. *Under the Unpaired algorithm, the average waiting time of hard-to-match patients, $\mathbf{W}_H(Unpaired)$, and that of easy-to-match patients, $\mathbf{W}_E(Unpaired)$, satisfy*

$$\lim_{p_H \to 0} p_H \mathbf{W}_H(Unpaired) = \frac{\ln(1+\lambda)}{\lambda \cdot n} \text{ and } \lim_{p_H \to 0} p_H \mathbf{W}_E(Unpaired) = 0.$$

Hence,

$$\lim_{p_H \to 0} p_H \mathbf{W}(Unpaired) = \frac{\ln(1+\lambda)}{n}.$$

We prove this proposition in Appendix A. Here, we provide a sketch of the main idea behind the proof.

Proof overview. The key step behind the proof of this theorem is to carefully study the structure of the stochastic system induced by the Unpaired algorithm. First, note that the number of donors in the system is always equal to the number of patients. Thus, the state of the system can be tracked by the number of easy-to-match and hard-to-match patients that are currently in the system. For

simplicity, let us first sketch the proof of the proposition for the case that $p_E = 1$, which means that whenever an easy-to-match patient arrives, she will get matched immediately.[16] In this case, the number of hard-to-match patients is a sufficient statistic for the state of the system. For sake of simplicity, we also normalize the arrival rate of pairs to $n = 1$.

Suppose the total number of hard-to-match patients in the system is k. When a patient-donor pair arrives, three events can change the state of the system[17]:

1. The patient is hard-to-match, and neither the patient nor the donor are compatible to anyone in the pool. In this case, the system moves from state k to state $k+1$. This happens at rate $\lambda((1-p_H)^k)^2$.

2. The patient is hard-to-match, and both the donor and the patient are compatible to someone in the pool, in which case the system moves to state $k-1$. This happens at rate $\lambda(1-(1-p_H)^k)^2$.

3. The patient is easy-to-match, and both the donor and the patient are compatible to someone in the pool. Again, the system moves to state $k-1$. Since $p_E = 1$, this happens at rate $(1-\lambda)(1-(1-p_H)^k)$.

Thus, we are dealing with a standard birth-death Markov chain, where the birth event has rate $\lambda((1-p_H)^k)^2$ and the death event has rate $\lambda(1-(1-p_H)^k)^2 + (1-\lambda)(1-(1-p_H)^k)$. In the full proof, we show that $\mathbb{E}(k)$ is highly concentrated around the state where the birth and death forces balance[18]; *i.e.,* $\mathbb{E}(k) \simeq k^*$, where k^* is the solution to:

$$\lambda((1-p_H)^k)^2 = \lambda(1-(1-p_H)^k)^2 + (1-\lambda)(1-(1-p_H)^k).$$

Algebraic manipulations for the case where $p_H \to 0$ lead to $k^* \simeq \frac{\ln(1+\lambda)}{p_H}$. By Little's law, the expected waiting time in a queue is equal to the pool size divided by the arrival rate:

$$\mathbb{E}(\mathbf{W}_H) = \mathbb{E}(k)/\lambda \simeq \frac{k^*}{\lambda} \simeq \frac{\ln(1+\lambda)}{\lambda p_H}$$

When $p_E < 1$, we must keep track of both the number of hard-to-match and easy-to-match patients, which makes the analysis more tedious. Since easy-to-match agents only exert a negative externality on hard-to-match patients under Unpaired (by taking kidneys that would otherwise be assigned to hard-to-match patients), when p_E gets smaller, this can only help hard-to-match patients. Hence, intuitively, when $p_E < 1$, the expected number of hard-to-match patients remaining in the

[16] The only exception is when there are no other donors in the system. Thus, it is possible to be in the state that there is only one easy-to-match patient and one donor in the system, but as our full analysis in Appendix A shows, this is exponentially unlikely.

[17] Note that another kind of event can happen, the patient can be compatible to someone and the donor not compatible to anyone (or the donor can be compatible to someone and the donor not to anyone). However, those events do not modify k.

[18] Our proof shows this using the standard global balance condition characterizing invariant distributions of Markov chains. While intuitive, the feature that the process concentrates on a state equalizing birth and death rate is non-trivial. This is not a general feature and one can build simple birth-death processes under which this property fails.

system should be lower-bounded by what we obtain when $p_E = 1$. We show that this bound is actually tight. There is a simple intuition for why the behavior of the system is similar when $p_E < 1$ as when $p_E = 1$: as $p_H \to 0$, the number of patients (and thus donors) in the system explodes. If there are k donors in the system, a new arriving easy-to-match patient is compatible to some donor in the system with probability at least $1 - (1 - p_E)^k$, which goes to 1 for sufficiently large k. Thus, easy-to-match patients will get matched quickly as long as p_E is a constant.[19]

Proposition 2.6 reveals several interesting comparative statics. First, the average waiting time of hard-to-match patients is decreasing in λ. A larger λ means a higher arrival rate for hard-to-match patients and a lower one for easy-to-match patients. In other words, easy-to-match patients exert a negative externality on hard-to-match ones. Indeed, upon arrival, an easy-to-match patient is almost sure to be compatible with an existing donor. The departure of a donor makes the market smaller, reducing the opportunities for future hard-to-match patients and thus increasing their waiting time. As we shall see below, this is in contrast to Pairwise and Chain under which easy-to-match patients can potentially help hard-to-match patients by increasing their likelihood of being cross-compatible or increase the likelihood of initiating a chain-segment (see Figure 2).

As we already mentioned in our Examples 2.4 and 2.5, matching patients with donors quickly, as does the Unpaired algorithm, may not always be beneficial. Indeed, our examples showed that the average waiting time of patients under Unpaired can be higher than under Pairwise or Chains for some realizations of our Poisson process. Theorem 2.7 and Theorem 2.10 state that, when taking expectations with respect to possible realizations of our process, this is not the case anymore.

Theorem 2.7 (Unpaired vs. Pairwise)**.** *If $\lambda > \frac{1}{2}$, then:*

$$\lim_{p_H \to 0} p_H \frac{\mathbf{W}(Pairwise)}{\mathbf{W}(Unpaired)} = \frac{\ln(2\lambda)}{\ln(1+\lambda)},$$

and if $\lambda < \frac{1}{2}$, then:

$$\lim_{p_H \to 0} \frac{\mathbf{W}(Pairwise)}{\mathbf{W}(Unpaired)} = \frac{1}{p_E} \frac{\ln\left(\frac{1-\lambda}{1-2\lambda}\right)}{\ln(1+\lambda)} > 1.$$

Proof. See Appendix B.1. □

This theorem states that, when p_H gets small, the average waiting time under the Unpaired algorithm is less than that under the Pairwise algorithm, irrespective of the value of λ. Moreover, when there are more hard-to-match patients ($\lambda > \frac{1}{2}$), the gap between the two algorithms goes to infinity at rate $\frac{1}{p_H}$. When there are strictly more easy-to-match patients ($\lambda < \frac{1}{2}$), the gap is not always as large, but still Unpaired achieves a shorter average waiting time than Pairwise.

[19] For a new arriving easy-to-match patient, the likelihood of being compatible to some donor in the system goes to 1 as p_H vanishes. However, conditional on the small probability event that this patient does not get matched upon arriving, given the priority rule under Unpaired, he will have to wait for an arriving donor to be incompatible with all hard-to-match patients remaining in the system. Given that the number of hard-to-match patients in the system explodes, one may expect the conditional waiting time to be very long. However, at an intuitive level, the rate of arrival of such donors should be bounded above by $n(1-p_H)^{k^*}$ which converges to a constant as p_H vanishes. This suggests that the conditional waiting time does not explode. We make this precise in our argument in Appendix C.

To gain some intuition for the first part of the result, let us focus on the special case that there are only hard-to-match patients ($\lambda = 1$). Consider a given patient who is waiting in a pool of size k. New donors arrive with rate n and they are compatible to our patient with probability p_H, which means that the arrival rate of a compatible donor for this patient is np_H. Since all agents are *ex ante* symmetric, our patient, under the Unpaired algorithm, will get matched at rate proportional to np_H/k. Now consider the Pairwise algorithm. A patient will get matched if a cross-compatibility happens, so our patients match rate is proportional to np_H^2/k. Thus, for any given pool size, Unpaired matches agents $1/p_H$ times faster than Pairwise, which in turn means that the waiting time under Unpaired is $1/p_H$ times smaller.[20]

Furthermore, the next theorem shows that the Unpaired algorithm's performance is not too far from the Optimal algorithm or, at least, much closer to it than Chain or Pairwise.

Theorem 2.8 (Unpaired vs. Optimal)**.** *For any λ, we have:*

$$\lim_{p_H \to 0} \frac{\mathbf{W}(Unpaired)}{\mathbf{W}(Optimal)} \leq 2\frac{\ln(1+\lambda)}{\lambda} \leq 2.$$

We prove this theorem in Appendix B.2. The main idea behind the proof is to provide a lower bound on the expected waiting time that applies to any possible matching algorithm, and hence, to Optimal. The essence of the argument can be given in the one-type model. So let us set $\lambda = 1$, normalize the arrival rate to $n = 1$ and fix an arbitrary matching algorithm ALG. Given a fixed pool size k, an arriving patient can be matched right way (in which case, his waiting time is 0) or he can join the pool. The latter event occurs with probability $(1-p_H)^k$. In this event, in order to get matched, the patient will have to wait for an arriving compatible donor (this is necessary but, of course, not sufficient) which occurs at rate p_H. Thus, the *conditional* expected waiting time of patients is bounded below by[21]

$$(1-p_H)^k \frac{1}{p_H} \geq (1-kp_H)\frac{1}{p_H}.$$

Taking expectations over possible values of k, we get as a lower bound on the *unconditional* expected waiting time of patients

$$\mathbf{W}(ALG) \geq \frac{1}{p_H} - \mathbb{E}[k]$$

where $\mathbb{E}[k]$ stands for the expected pool size. Now, the final step of the proof simply consists in noting that, by Little's law, the expected pool size divided by the arrival rate of patients equals the expected waiting time, i.e., $\mathbf{W}(ALG) = \mathbb{E}[k]$. Thus, we obtain the following lower bound on the

[20] Ashlagi et al. (2019)—from whom we borrow the analytical expression of the waiting time of patients under the Pairwise algorithm—were not able to derive a closed form solution when the arrival rate of H patients and E patients are the same. This is why we do not report any result for $\lambda = 1/2$. However, they provide simulations suggesting that, in that case, the expected waiting time of H patients scales with $1/p_H$.

[21] In the inequality, we use the Bernoulli inequality, which states that for any $x \leq 1$ and any $n \geq 1$, $(1-x)^n \geq 1-xn$.

expected waiting time of the algorithm

$$\mathbf{W}(ALG) \geq \frac{1}{2p_H}.$$

This lower-bound together with Proposition 2.6 gives us Theorem 2.8 when $\lambda = 1$. One can easily extend this argument to the two-type model to obtain $\frac{\lambda}{2np_H}$ as a lower bound and derive Theorem 2.8 in this general case.

This theorem shows that the waiting time of patients under the Unpaired algorithm is not too far from that under the Optimal algorithm. This is not true for Pairwise as implied by Theorem 2.7, especially when $\lambda > \frac{1}{2}$. Note that Unpaired is purely greedy and computationally simple, while the Optimal algorithm is forward looking and potentially computationally complex. Yet, the Optimal algorithm does not substantially improve upon Unpaired. This result can be further strengthened: As made precise in the next remark, if the Optimal algorithm is restricted to provide easy-to-match patients with an average waiting time as short as their waiting time under the Unpaired algorithm, the waiting time under Unpaired is at most 38 percent more than under this restricted Optimal algorithm. This is a natural restriction if the objective is to improve, at least weakly, the situation of all patients (i.e., decrease the waiting time of hard-to-match patients without increasing the waiting time of easy-to-match patients).

Remark 2.9. *The bound in Theorem 2.8 can be improved under this additional constraint on the waiting time of easy-to-match patients. Appendix B.2 shows that for any matching algorithm ALG satisfying* $\mathbf{W}_E(ALG) \leq \mathbf{W}_E(Unpaired)$,

$$\lim_{p_H \to 0} \frac{\mathbf{W}(Unpaired)}{\mathbf{W}(Optimal)} \leq (1+\lambda)\frac{\ln(1+\lambda)}{\lambda} \leq 2\ln(2) \simeq 1.38.$$

So far, we have shown that the Unpaired algorithm performs substantially better than Pairwise, while being much closer to the Optimal algorithm. In practice, the Chain algorithm also plays an important role in matching patient-donor pairs. Our next result compares Unpaired with Chain.

Theorem 2.10 (Unpaired vs. Chain)**.** *For any* λ*, if* $p_E = 1$*, we have:*

$$\lim_{p_H \to 0} \frac{\mathbf{W}(Chain)}{\mathbf{W}(Unpaired)} = -\frac{\ln(1-\lambda)}{\ln(1+\lambda)} \geq 1.$$

Proof. See Appendix B.3. □

Since $-\frac{\ln(1-\lambda)}{\ln(1+\lambda)}$ is greater than one, when p_H gets small, the average waiting time under Unpaired is smaller than under the Chain algorithm irrespective of the value of λ. Note also that $-\frac{\ln(1-\lambda)}{\ln(1+\lambda)}$ is an increasing function of λ and goes to infinity as $\lambda \to 1$. In other words, the Unpaired algorithm performs increasingly better than the Chain algorithm as the fraction of hard-to-match patients increases; as this fraction approaches one, Chain delivers an average waiting time that is infinitely longer than what Unpaired achieves.

Below is an intuition of this theorem. The Chain algorithm performs better when the probability of starting a new chain-segment is higher, which makes easy-to-match patients critical. When a pair with an easy-to-match patient arrives, the patient will be matched with the bridge donor and advance the chain-segment with probability p_E. This probability reduces to p_H, which is vanishing, if the arriving patient is hard-to-match. Note that the donor in the newly arriving pair cannot be considered for matching unless the paired patient finds a match. Therefore, when only a small minority of arrivals have an easy-to-match patient, the probability of starting a new chain-segment is small and Chain performs poorly. In contrast, the Unpaired algorithm does not crucially depend on the type of the arriving patient, because it allows donation-before-receipt.[22]

Theorem 2.10 comes with an additional technical assumption: to provide a closed-form solution for the Chain algorithm, we assume $p_E = 1$. Without this assumption, the Markov chain induced by the Chain algorithm seems too complicated to be analyzed. For the same reason, Ashlagi et al. (2019) do not provide a formal proof for the performance of Chain when $p_E < 1$. Our numerical simulations, as well as those in Ashlagi et al. (2019), indicate that the closed-form solution for $p_E = 1$ approximates well the simulated average waiting time for $p_E < 1$. Based on these results, we conjecture that Theorem 2.10 holds even when $p_E < 1$.

Remark 2.11. *The Chain algorithm identifies a chain in a greedy fashion. It does not necessarily pick the longest chain. The Optimal Chain algorithm (selecting the longest chain) has been studied in a one-type model by Anderson et al. (2017). In that model, it is easy to show that the waiting time of patients under Optimal Chain is always greater than that of Unpaired. Indeed, under Optimal Chain, an arriving patient has probability p_H to be matched with the bridge donor right away. With the complement probability, this patient will be unmatched and enter the pool. In that event, she will have to wait for an arriving patient to be compatible with the bridge donor (which is necessary to initiate a chain-segment), which occurs with rate p_H. Thus, in expectation, for small values of p_H, this patient will have a waiting time bounded below by $1/p_H$. This is larger than $\ln(2)/p_H$, the waiting time of patients under the Unpaired algorithm (see Proposition 2.6).*

We summarize our theoretical results in Figure 2 with $p_E = 1$ and $n = 1$. For a given algorithm, taking p_H to zero, a line in the figure depicts the limiting average (normalized) waiting time for hard-to-match patients (panel A) and that for all patients (panel B), as a function of the fraction of hard-to-match patients ($\lambda \in [0, 1]$). In terms of these waiting times, the difference between Unpaired and Optimal is bounded and relatively small for all λ (Theorem 2.8); the difference between Unpaired

[22] If there are finitely many bridge donors at each point in time, the same argument applies. Indeed, as p_H vanishes, the likelihood that an arriving hard-to-match patient is compatible with a bridge donor vanishes. Therefore, when only a small minority of arrivals are easy-to-match patient, Chain still performs poorly and Theorem 2.10 still holds, as mentioned in footnote 13. In addition, note that nothing changes for an arriving easy-to-match patient since, by assumption $p_E = 1$, and so more bridge donors is not helpful to start a chain segment in case of such an arrival. Thus, more generally, the comparison between Unpaired and Chains remains the same when the number of bridge donors is larger than 1. Of course, with $p_E < 1$, more bridge donors may be useful to allow easy-to-match patients to start new chain segments. However, at an intuitive level, we are considering the most favorable case for Chains when assuming that $p_E = 1$ where each arrival of an easy-to-match patient starts a chain segment. Further, simulations in Ashlagi et al. (2019) as well as ours show that the average waiting time under the Chain algorithm when $p_E = 1$ is similar to the average waiting time for $p_E < 1$.

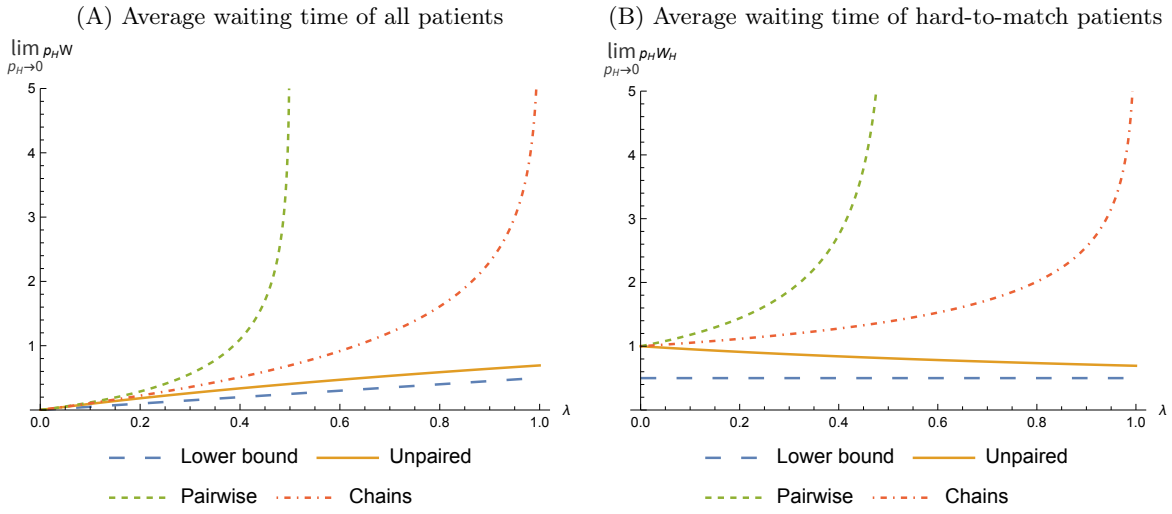

Figure 2: Waiting Time under Each Algorithm and the Fraction of Hard-to-match Patients (λ)

Notes: Given $p_E = 1$ and $n = 1$, the lines show that, as a function of the fraction of hard-to-match patients (λ), the waiting times under the Pairwise, Chain, and Unpaired algorithms, as well as a lower bound for the Optimal algorithm.

and Pairwise increases with λ and explodes when $\lambda > \frac{1}{2}$ (Theorem 2.7); last, the difference between Unpaired and Chain increases with λ and goes to infinite when $\lambda = 1$ (Theorem 2.10). The results reported in Figure 2 allows us to quantify the benefits of relaxing the simultaneity constraints embedded in the Pairwise algorithm. They show that those benefits may be substantial, especially when the share of hard-to-match patients is large. We will evaluate the empirical relevance of these results in Section 3.

2.3 Discussion of the Assumptions

We now discuss some of the assumptions and modeling choices that we have adopted.

First, our model assumes two patient types, each of which has a different probability of being compatible with a random donor. Below, we show some evidence that this can be a reasonable approximation in some settings. A real-life kidney exchange problem certainly has more than two types of patients, because a patient's biological compatibility with a donor depends on blood type and tissue type compatibilities. There are finitely many blood types with a known compatibility relation, but tissue compatibilities involve a more subtle comparison of the antibodies of a patient with the antigens in a donor's tissue. This information is usually summarized by a continuous measure, Panel Reactive Antibodies (PRA), quantifying the probability that a patient is tissue type *incompatible* with a random blood type compatible donor. The higher a patient's PRA, the more difficult it is to find a compatible donor for the patient. The distribution of PRA among patients is bimodal with high concentrations of patients at very low and very high PRA values in the U.S. (Anderson et al., 2017). Panel A of Figure 3 shows a similar bimodal pattern for the patients in the French kidney exchange program (KEP) that are used in our simulation analyses in Section 3.

Adding blood type compatibility still induces a bimodal distribution for the probability of overall biological compatibility in our data (panel B).

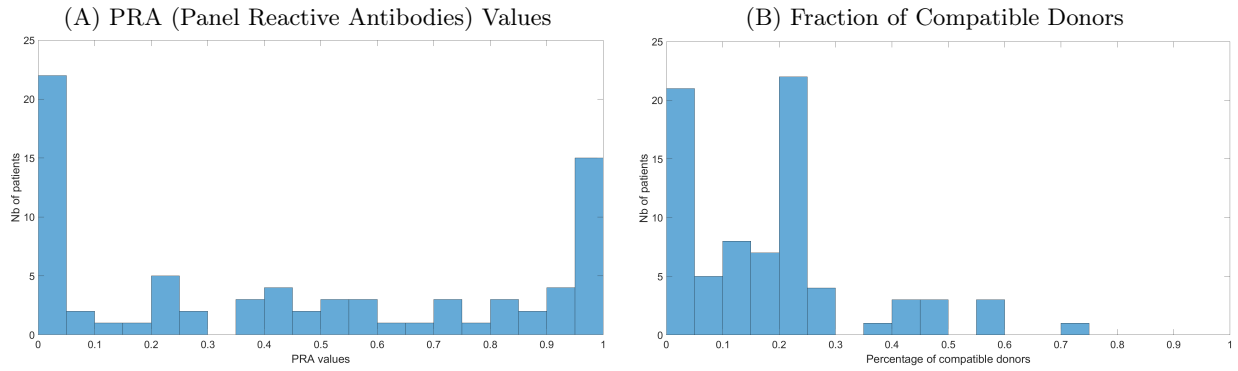

Figure 3: Distribution of Patient Types in the French KEP

Notes: Calculated by the authors from the 78 patient-donor pairs who ever participated in the French kidney exchange program (KEP) during December 2013 to February 2018. In panel A, the higher a PRA, the more difficult it is for a patient to find a compatible donor. In panel B, a patient is compatible with a donor if they are both blood type and tissue type compatible. Using the DIP test in Hartigan and Hartigan (1985), for both distributions, we fail to reject the hypothesis of bimodality. See Table 1 for more summary statistics on these patients.

Second, there is no death in our model, which can be overly simplifying for patients waiting for kidney transplant.[23] However, patients participating in a KEP are usually in better health conditions. In the sample period of more than four years, none of the patients who participated in the French KEP died while waiting for a kidney. Some of them did leave the KEP for a deceased donor kidney or an incompatible living donor kidney after going through desensitization (see Section 3.1). Introducing a pair exit decision would add a layer of complexity in our model, but we conjecture that our main results would hold. Intuitively, the longer the waiting time under an algorithm, the higher the probability that a patient exits. Compared with Pairwise and Chain, the Unpaired algorithm has a shorter average waiting time, and thus a model with exit may reinforce the advantage of the Unpaired algorithm. Moreover, our simulations in Section J show that our results are robust even when we allow pairs to exit.

Last, our main theoretical results are limiting results for $p_H \to 0$, with p_H being the probability of a hard-to-match patient being compatible with a random donor. Working with limit results gives us more analytical tractability, but it also implies that our results provide a good approximation only when p_H is sufficiently low. In our data, a patient with a PRA above 85 percent, who is considered hard-to-match by convention, is compatible with only 1.5 percent of the living donors. In contrast, a patient with a PRA below 85 percent is compatible with 24.5 percent of the living donors. Admittedly, our theoretical results do not give any insight about how small p_H should be for the results to be a good approximation.Ultimately, this is an empirical question and our empirical analysis in Section 3 confirms the relevance of our results in real life.

[23] In the U.S., in 2014, 4,761 patients died while waiting for a kidney transplant; another 3,668 patients became too sick to receive a kidney transplant. Source: National Kidney Donation (www.kidney.org).

3 Simulation Analyses Using French Data

Our theoretical model investigates the steady-state performance of the algorithms in a stylized model. There are two main reasons for why an empirical investigation is crucial. First, even if all features of the environment (e.g., arrival rate) remain steady over time, it will take some time for a given market to reach the steady state. Policymakers are certainly interested in an algorithm's short-run performance as well, and our theoretical results are silent with respect to that. Second, while the two-type assumption of our model is not unreasonable as a first-order approximation, the real world is indeed substantially more complex. For instance, a patient's biological compatibility with a donor depends on both blood type and tissue type compatibilities and thus, for a given patient, the compatibility realizations are not *i.i.d.* across donors.

With these considerations, we assess the algorithms with an administrative dataset on kidney transplants in France for a period of 1644 days. We observe all 586 incompatible pairs that either have completed or are still waiting to complete the transplantation in France in that period. These pairs then serve as the "population pool" from which we randomly draw pairs to participate in the algorithm in consideration.

We proceed as follows. Section 3.1 describes the institutional background and our data. In Section 3.2, we detail some definitions and assumptions that are necessary in our simulations. Section 3.3 presents the performance of the four algorithms (Pairwise, Chain, Optimal, and Unpaired) in our baseline simulation. We highlight that the empirical results are consistent with the theoretical predictions. Moreover, focusing on the waiting times of patients/donors after being unpaired, we acknowledge that the two potential incentive issues of Unpaired appear to be a concern in practice. However, in Section 3.4, we show that one of the issues does not affect the performance of Unpaired, while the other one becomes negligible as the market size grows. Moreover, we shall show in Section 4 that this latter issue can also be made negligible in small markets such as the French KEP when we integrate the possibility of getting deceased donor kidneys under Unpaired.

3.1 Institutional Background and Data

Our analysis relies on administrative data from France provided by the Agency of Biomedicine (*Agence de la Biomédecine*, ABM), a government agency that oversees all organ transplants in France. Our data covers the period of December 13, 2013 to February 23, 2018, or 1644 days in total, including all transplants with deceased or living donor kidneys, as well as discarded kidneys from deceased donors.

In France, when a patient is diagnosed as requiring a kidney transplant, her doctor must register her at the national deceased donor list (DDL) to join the waitlist. If a patient finds an incompatible living donor, she may either join the kidney exchange or go through a desensitization procedure whereby she can receive an incompatible kidney from her donor. Together, they become the pool of incompatible pairs from which we will randomly draw pairs to create different markets. Below, we provide institutional details.

Kidney Exchange Program. France's kidney exchange program (KEP) started in 2013, following the revision of the bioethics law (*loi de bioéthique*) that regulates the medical practices in France. Unlike the U.S., France has a single KEP at the national level, administered by the ABM. By law, any exchange of living donors must be done through the KEP.

The KEP executes a match run every three months. Over the time period we study, only 2-way pairwise exchanges were allowed, while non-directed kidney donations and chains were prohibited.[24] The KEP's objective is to maximize the total number of transplants in each match run. In total, there are 78 pairs participating in the 15 match runs in our sample period.[25] A detailed summary can be found in Combe et al. (2019).

Table 1: Kidney Patients and Donors: Summary Statistics

	Incompatible Pairs				
	KEP Pairs		Desensitization Pairs		DDL Kidneys[e]
	Patient	Donor	Patient	Donor	
	(1)	(2)	(3)	(4)	(5)
# of observations	78	78	508	508	13,036
Patient grafted/Donor donated	69%[b]	35%	100%	100%	97%
Age[a]	46.1	48.1	42.7	45.8	54.9
	(12.9)	(10.5)	(14.1)	(11.8)	(18.5)
Female	47%[c]	49%[d]	37%	62%	43%
Blood type					
A	31%	51%	23%	49%	44%
B	10%	18%	16%	16%	10%
O	56%	23%	59%	29%	43%
AB	3%	8%	2%	6%	4%
Sensitization Status[f]					
Hypersensitized	27%		24%		
Sensitized	50%		47%		
Non-sensitized	23%		29%		
ABO Compatible within the pair	42%		44%		
HLA Compatible within the pair	32%		48%		

Notes: This table presents characteristics of kidney patients and donors in France from December 2013 to February 2018. Columns (1) and (2) are on the 78 patient-donor pairs who ever participated in the KEP. Columns (3) and (4) are on pairs who did desensitization. Column (5) describes all the DDL kidneys in the sample period, column (6) focuses on those qualified for at least one KEP patient, and column (7) includes those qualified for at least one one patient in KEP or desensitization pairs. The definition of being a "qualified" DDL kidney is in footnote 32.

[a] An age in the table is calculated on January 1, 2012, except for a DDL kidney (which is calculated on its retrieval day).

[b] Patients in the KEP can receive a transplant outside the KEP (e.g., from DDL or desensitization), and a donor can donate outside the KEP (i.e., by desensitization). There are 12 pairs (15.4%) engaged in an exchange in the KEP in the sample.

[c] This percentage is calculated among the 70 patients with non-missing gender information.

[d] This percentage is calculated among the 68 donors in the KEP with non-missing gender information.

[e] A DDL donor may provide two kidneys, and the statistics in column (5) are calculated at the individual kidney level.

[f] A patient's sensitization status measures how likely it is for her to find a deceased or living donor kidney that is compatible. The exact definition is provided in Appendix H.1.

Columns (1) and (2) of Table 1 present more statistics on the 78 KEP pairs. Some patients receive a kidney from the DDL or a living donor kidney outside the KEP, leading to 69 percent of them receiving a transplant. 35 percent of the donors donate, with some of them donating outside

[24] See Section 5 for a discussion of the recent reform of the rules governing the french KEP.

[25] In the sample period, December 2013 to February 2018, the first match run happened in December 2013. There were only three match runs in each of the years 2014, 2015 and 2017 and four match-runs in 2016. Additionally, our data covers one match run in February 2018. On average, a pair stays for 3.4 match runs, and a match run has 17.5 participating pairs.

the KEP. Many of the patients have blood type O (56 percent) or are hypersensitized (27 percent), indicating that a large fraction of them are hard to match. The most common blood type among the donors is A (51 percent), while only 23 percent of them have type O blood. Among all the pairs, 42 percent are blood type compatible, and 32 percent are human leukocyte antigen (HLA) compatible.

Desensitization Pairs. Desensitization is an immunosuppressive treatment that can eliminate immunological compatibility constraints. Once treated, a patient is able to receive a transplant from an incompatible donor. For a brief review of desensitization, please see Andersson and Kratz (2020) and Heo et al. (2018) as well as the references therein.

In general, incompatible transplants facilitated by desensitization are more expensive than compatible ones,[26] while leading to poorer patient outcomes.[27] Nonetheless, desensitization is a popular choice for incompatible pairs in France. During our sample period, there are 508 incompatible pairs that take this option without trying the KEP. These transplants, as well as the associated patients and donors, are recorded by the ABM.

Columns (3) and (4) of Table 1 present summary statistics on desensitization pairs. Compared with the donors, the patients are less likely to be a female (37 percent vs. 62 percent), are of similar ages (45.9 vs. 45.8 years old), and are more likely to have type O blood (59 percent vs. 29 percent). 24 percent of these patients are hypersensitized. Relative to the KEP pairs, the desensitization pairs are more likely to be blood type compatible (44 percent vs. 42 percent) or HLA compatible (48 percent vs. 32 percent).

Deceased Donor Kidneys. Some of later evaluations will involve deceased donor kidneys list, which we refer to as DDL kidneys. Our sample has $13,036$ such kidneys offered to patients on the DDL.[28] Column (5) of Table 1 describes all the DDL kidneys in our sample period. On average, they are 54.9 years old, and 43 percent are from a female donor. The top two blood types are A (44 percent) and O (43 percent).

3.2 Definitions and Assumptions in the Simulations

The following definitions, assumptions, and data preparations are needed in our simulations.

[26] Compared with a compatible living donor transplant, on average, a blood-type incompatible transplant is $100,000 more expensive, and an HLA incompatible transplant is $180,000 more expensive (Axelrod et al., 2018).

[27] In a recent survey, Scurt et al. (2019) conclude that blood-type incompatible transplants result in an excess of mortality and loss of kidney grafts compared to compatible transplants. The outcomes of HLA incompatible transplants are even worse (Marfo et al., 2011).

[28] On any given day in our sample period, there are about $7,000$ patients waiting on the (active) DDL. In our sample period, there are 389 deceased donors whose kidneys are offered to the DDL patients but discarded in the end due to either refusals or last-minute cancellations. For some of the deceased donors, we do not know how many kidneys are available for the DDL; in that case, we assume such a donor has only one kidney available.

Simulating markets. There are 586 KEP and desensitization pairs form our pool of pairs.[29] We measure time period by days vary the size of the market via daily Poisson arrival rate of pairs n. For a given size n, we first draw from the Poisson distribution the number of arriving pairs on each day and then randomly draw that number of pairs from the pool with replacement.[30] We run each simulated market under a given algorithm for 1644 days, corresponding to the number of days in our sample period. To evaluate an algorithm in a market of size n, we run 1000 sets of independent simulations and report the average across them.

We consider eight different sizes, $n \in \{0.05, 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2\}$.[31] In the following, we focus on $n = 0.05$ which corresponds to the French KEP. Many larger real-life KEPs are included in our simulations. For example, the Spanish KEP is roughly $n = 0.2$; the UK KEP is $n = 0.4$; and the NKR, the largest in the world, is roughly $n = 1$ (Agarwal et al., 2018; Biró et al., 2019).

Compatibility between a patient and a donor. Patient p_i and donor d_j are compatible unless they are either blood type incompatible or HLA incompatible. We have sufficient information to determine the compatibility between any patient and any donor in the data. Specifically, we compare p_i and d_j's blood types to determine their blood type compatibility; if d_j has at least one antigen that is unacceptable to p_i, p_i is HLA incompatible with d_j.

Hard-to-match. A patient is defined to be hard-to-match if she is hypersensitized, i.e., tissue-type incompatible with more than 85 percent of the 13,622 donors, living or deceased, in our data. There are 21 such patients from the KEP pairs and another 120 from the desensitization pairs (Table 1). As mentioned in Section 2.3, even when we abstract from blood type incompatibility, the probability of those patients finding a compatible kidney is low. Additionally, we separately report statistics on O patients, who are also likely to be hard to match.

"Waiting rooms" for unpaired patients (P) and donors (D). The Unpaired algorithm may have two incentive issues involving unpaired patients and donors due to donation-before-receipt and receipt-before-donation. Namely, a patient may not be willing to let her donor donate before she receives a kidney and a donor may renege if her patient has already received a transplant from

[29] Hence, our simulation only includes incompatible pairs. As shown in Table 1, the distribution of characteristics of the KEP pairs are different from those of the desensitization pairs, due to self-selection of the pairs into the KEP or desensitization. As a robustness check, we run several additional simulations: resampling only from the KEP pairs and resampling from a simulated U.S. pool (NKR and APKD) based on marginal distributions reported in Ashlagi and Roth (2020). In all these simulations, we find similar patterns in the performance of the algorithms. Relatedly, some countries allow compatible pairs to participate in a KEP, and policymakers can even incentivize compatible pairs to participate and improve a KEP's performance (Sönmez et al., 2018). The participation of compatible pairs will not only increase the market size but also change its composition because compatible pairs are, on average, easier to match than incompatible ones. As another robustness check, we add compatible pairs into the simulation. Again, we find similar patterns in the algorithms' performance.

[30] As in the literature, we assume that the participation decision of a pair is exogenous and not affected by the performance of the algorithm in use. An efficient algorithm tends to encourage more pairs to participate, while every algorithm performs better in a larger market (as we shall see later). Hence, our assumption may underestimate the actual performance of a more efficient algorithm, in our case, Unpaired.

[31] We also run simulations for $n \in \{2.5, 3, 3.5, 4\}$, whose results, although not reported, also confirm our findings.

someone else. Some definitions will be useful in our analysis: we let P be the "waiting room" for unpaired patients who are waiting for a kidney after their paired donors have donated; D is the "waiting room" for unpaired donors who wait to donate after their paired patients have received a transplant.

Tie-breaking in the algorithms. When selecting among multiple compatible donors/patients, an algorithm needs a tie-breaking rule. In our definitions (Definitions 2.1, 2.2, and 2.3), we allow for arbitrary tie-breaking rules within types (i.e., we favor hard-to-match patients but allow for arbitrary tie-breaking rules within patients of the same type; similarly, we allow for arbitrary tie-breaking rules among donors), as the theoretical results do not depend on that. Our simulation, however, complements the definitions with the following rule: when selecting among multiple donors, Unpaired favors those in D and breaks any remaining ties by their waiting time; when it chooses among multiple patients, hard-to-match patients enjoy the highest priority, and any remaining ties are broken first by whether a patient is in P and then by their waiting time. Pairwise and Chain also use this tie-breaking rule, although Pairwise involves neither P nor D and Chain has no P.

3.3 Evaluating the Algorithms: Baseline Simulations

We focus on a market with a daily arrival rate of $n = 0.05$, similar to the French KEP. We consider a baseline in which there is neither pair exiting nor donor reneging. Presumably, the exit rate of pairs is endogenous and would be rarer in the high-performing algorithm such as Unpaired. We start with no exit and allow an exogenous exit rate in a robustness check in Section J. Donor reneging in practice is shown to be low (Cowan et al., 2017), although we shall relax this no reneging assumption shortly.

We apply an algorithm to each of the 1000 simulated markets. For a given simulation, pairs are ordered by arrival date: $i = 1, \ldots, n_\ell$. Let $a(i)$ and $e(i)$ be the dates of arrival and exit of pair i, respectively. Let T be the end of our simulation time horizon; obviously, $T \geq a(n_\ell)$. There is no exit, or equivalently $e(i) = T$ for all i. That is, before the last day (T), once a pair arrives, the patient leaves only if she receives a kidney, and the donor leaves only if she donates a kidney.

We simulate Pairwise and Unpaired by following their definitions and applying the aforementioned tie-breaking rule (Section 3.2). To initiate Chain in the simulation, we select a DDL kidney as an altruistic donor.[32] We allow for multiple bridge donors in Section J. Note that, following common practices, pairwise exchanges are still allowed in Chain in the simulation, contrary to our theoretical analysis.[33]

[32] To ensure that a DDL kidney is "high quality" for a given patient, we require that the DDL kidney be compatible with the patient and have a Kidney Donor Profile Index (KDPI, lower is better), a risk index of post-transplant graft failure, below the Living Kidney Donor Profile Index (LKDPI) of the patient's paired incompatible donor. The LKDPI is an index for living donor kidneys corresponding to the KDPI, and they are of the same scale. See Appendix H for more details on KDPI and LKDPI. Among the DDL kidneys meeting this selection criterion for at least one KEP patient, we randomly pick one, regardless of its arrival date. We assume that this DDL kidney arrives on the first simulated arrival day, $a(1)$, and remains until being transplanted or until the end of the simulation. We redraw a new DDL kidney for each set of simulated dates.

[33] Based on simulations of the theoretical model, the steady-state average waiting time of this algorithm (allowing

A best-case, infeasible algorithm: Omniscient. In Section 2, we used the optimal algorithm to provide the best-case scenario for average waiting time. However, this algorithm is not computationally feasible due to, among other reasons, the exponential size of the state space. As such, instead of simulating the optimal algorithm, we consider a better-than-optimal algorithm, the *Omniscient* algorithm. This algorithm assumes that the designer has full information about all the patients and donors arriving in our simulation, including their arrival/exit dates and their characteristics. The designer then minimizes the average waiting time of all patients up to date T (i.e., the end of our simulation). [34]Therefore, in terms of this average waiting time, it dominates all algorithms, including Unpaired. While Omniscient is practically infeasible, it helps us evaluate the potential of Unpaired.

Simulating Omniscient is feasible because it does not need assumptions on pair arrivals beyond our simulation time horizon. We detail the definition and implementation of the Omniscient algorithm in Appendix G, which for the baseline simulations corresponds to a standard polynomial algorithm to compute a minimum weight bipartite matching.

Results. Our simulation results in columns (1)-(5) of Table 2 are in line with our theoretical findings in Section 2.2. The transplant rate under Unpaired, very similar to what Omniscient achieves, is significantly higher than those under Pairwise and Chain; the waiting time under Unpaired is substantially lower than Pairwise and Chain (cf. Theorem 2.7 and Theorem 2.10). Although Omniscient achieves the lowest waiting time, it does not significantly outperform Unpaired (cf. Theorem 2.8). We also run the same simulations for various market sizes and find the similar patterns; see Appendix Figure I.1.

We first report transplant rates: Unpaired is close to Omniscient (63 vs. 64 percent) and far above the other two algorithms (33 percent under Pairwise and 36 percent under Chain). The same conclusion is true among hard-to-match patients.[35]

Conditional on transplantation, the average waiting time is 176 days under Unpaired, which is close to the 144 days under Omniscient and far lower than the 248 days under Pairwise and the 232 days under Chain. However, among hypersensitized patients, this conditional waiting time is lower under either Pairwise or Chain than Unpaired. The main reason is that Pairwise and Chain each only achieves less than a half of the transplants under Unpaired.

This observation motivates us to consider an alternative waiting time measure: the average censored waiting time for all patients, which also includes the waiting time of those who are not transplanted at the end of the simulation period. As pairs arriving throughout the simulation period are randomly drawn from the same pool, this new measure creates some balance between transplant

for both chains and pairwise cycles) is almost identical to that of Chain. The simulations are available upon request.

[34]The optimal algorithm used in the theory section simply minimizes waiting time over all possible algorithms (with a unique invariant distribution). Importantly, an algorithm is a mapping from current compatibility graph into matchings. Hence, it does not condition current decisions to future compatibility graphs. Put differently, under this optimal algorithm, the designer does not know for sure the future. The omniscient assumes that all the future is known to the designer.

[35]The transplant rate of either hypersensitized or O patients under any algorithm is significantly lower than that of other patients. This is consistent with our assumption that they are harder to match than others.

Table 2: Performance of Different Algorithms

	Pairwise (2-way) Exchange (1)	2-way & 3-way Exchanges (2)	Chain & Pairwise Exchange (3)	Unpaired Exchange (4)	Omniscient (5)	With DDL & $\delta = 6$ months	
						Pairwise (6)	Unpaired (7)
Transplants							
% patients receiving transplant	33%	40%	36%	63%	64%	85%	93%
hypersensitized patients	*18%*	*23%*	*18%*	*40%*	*41%*	*74%*	*82%*
O patients	*24%*	*30%*	*26%*	*46%*	*49%*	*84%*	*91%*
% transplants from living donors	100%	100%	100%	100%	100%	24%	55%
Average waiting time (days)							
Patients receiving transplant	248	240	232	176	144	169	72
hypersensitized patients	*234*	*292*	*232*	*281*	*231*	*239*	*145*
O patients	*362*	*307*	*357*	*311*	*228*	*182*	*83*
All patients (censored)	617	565	579	350	325	193	91
hypersensitized patients	*737*	*704*	*736*	*574*	*559*	*297*	*203*
O patients	*688*	*643*	*671*	*486*	*445*	*208*	*106*
Patients going through P							
Total number	-	-	-	29	32	-	29
hypersensitized patients	-	-	-	*11*	*12*	-	*13*
O patients	-	-	-	*24*	*25*	-	*21*
Waiting time of patients in P							
Median	-	-	-	245	384	-	6
hypersensitized patients	-	-	-	*517*	*584*	-	*54*
O patients	-	-	-	*237*	*396*	-	*6*
Donors going through D							
Total number	-	-	4	26	31	-	36
AB donors	-	-	*1*	*4*	*5*	-	*4*
Waiting time of donors in D							
Median	-	-	207	339	417	-	39
AB donors	-	-	*368*	*618*	*614*	-	*48*

Notes: The statistics are from the 1000 sets of simulations, each of which contains independent draws of pairs with a daily arrival rate of 0.05 (roughly the size of France's KEP). There are on average 83 incompatible pairs, among which 20 pairs have a hypersensitized patient and 48 have an O patient. The waiting time for a patient or a donor may be censored from above if she has not received or donated a kidney by the last date of the simulation. The same censoring applies to the number of days in P or D. P and D are waiting rooms for unpaired patients and donors, respectively. Pairwise (2-way exchange) (column 1) is defined in Definition 2.1, column (2) allows 3-way exchanges in addition to 2-way, Chain is defined in Definition 2.2 and is combined with Pairwise (column 3). In column (4), Unpaired is defined in Definition 2.3. Omniscient (column 5) uses full information on all pairs in the sample period to match patients and donors to minimize the total censored waiting time. Patients in columns (6) and (7) have to wait for six months before receiving a DDL kidney offer (unless their donor has already donated under Unpaired).

rate and waiting time conditional on transplantation. Omniscient (column 5) minimizes average censored waiting time and reaches 325 days, while Unpaired (column 4) achieves a similar level, 350 days. Pairwise and Chain (columns 1 and 3) perform significantly worse, delivering a waiting time at least 65 percent longer than Unpaired. This pattern also holds among hard-to-match patients (defined as either hypersensitized or O patients).

In practice, a KEP may allow both 2-way and 3-way pairwise exchanges. Including this possibility, column (2) shows that this more flexible algorithm performs better than Pairwise but remains

significantly worse than Unpaired.[36]

In sum, our theoretical analysis shows that imposing simultaneity constraints in a kidney exchange algorithm can be costly, while our simulation reveals that such costs are indeed substantial in realistic settings.

3.4 Two Potential Practical Challenges

Table 2 confirms that Unpaired has two potential practical challenges. First, that receipt-before-donation makes some donors enter D and possibly wait for a long time. The median waiting time of such donors is 339 days. Such a long wait can bring a non-negligible chance of losing an unpaired donor, because she may refuse or become unfit to donate. The second challenge arises because we allow donation-before-receipt, which means that some patients enter P after their donors have donated; the median waiting time of the patients who go through P is 245 days. Hypersensitized patients wait even longer: they have a median wait time of 517 days. Pairs may refuse to donate before receiving a kidney if they expect such long wait-times.

To analyze the first challenge, we simulate Unpaired, but allow for each donor who enters D to renege at some rate. Gentry et al. (2009) use a monthly renege rate of 5%, which is substantially higher than what is documented by Cowan et al. (2017) in a dataset from NKR in the US.[37]

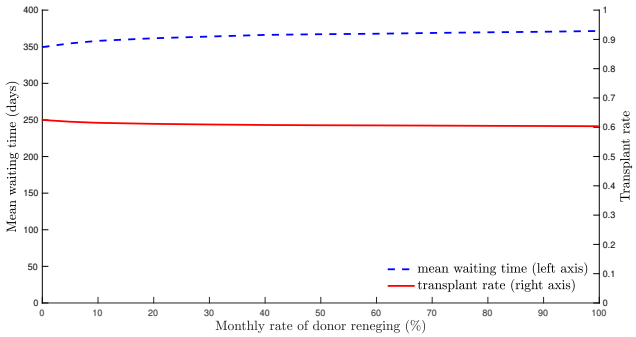

Figure 4: Waiting Time and Transplant Rate under Unpaired When Donors May Renege

Notes: The size of the market in terms of daily arrival rate is $n = 0.05$, corresponding to the French KEP. For each donor who enters D, there is an independent daily reneging probability, from which we can calculate a monthly reneging probability. We consider a wide range of monthly rates, 5 percent and another 10 values from 10 to 99.90 percent. When the monthly reneging rate is zero, the results are the same as column (4) of Table 2.

Figure 4 shows that the possibility of reneging barely changes the performance of the Unpaired: If *no* donor reneges, then patients' mean waiting time and transplant rate are 349 days and 62.5

[36]The Unpaired algorithm eliminates the timing constraints on the donation and receipt but, compared to the 2-way and 3-way pairwise exchanges, it also allows exchange cycles of arbitrary sizes. One may naturally wonder whether the performances of Unpaired relies on the relaxation of the constraints on the cycle sizes. To answer this question, we simulated pairwise exchanges allowing cycles of arbitrary size. Its performance is still significantly worse than Unpaired: the transplant rate is 44 percent (vs. 63 percent under Unpaired) and the (censored) average waiting time is 539 days (vs. 350 days under Unpaired).

[37]In their dataset, among the 1244 bridge donors in a chain over about 7.5 years, 1.6 percent of the donors did not donate for some reasons, e.g., donor health problems; only 0.5 percent of them elected not to proceed with donation.

percent, respectively. In the worst case scenario, if (almost) *all* donors who enter D renege in a month, then these numbers change to 371.5 days and 60.4 percent.[38]

The effect of reneging is negligible for a variety of reasons. First and foremost, most donors do not enter D queue to begin with. Table 2 shows that, when there is no reneging, only 31 percent of donors ever enter D. Second, when a donor in D reneges, it reduces the chance that another donor enters D, thus reducing the risk of additional reneges in the future. This endogenous effect mitigates the risk of reneging for Unpaired. Indeed, the results show that while the number of donations from donors in D decreases, the one from donors who are still paired with their associated patient increases. Finally, donors who eventually renege tend to be less valuable than the average donor, as they are precisely the donors who tend to wait longer in the D queue.

We now discuss the second practical challenge: How does the possibility of patients waiting for a long time in the P queue affects the performance of Unpaired? Would they still opt to donate a kidney before receiving one?[39] We first note that donating before receiving still increases the chance of receiving a kidney relative to waiting for a cross-compatible match. Thus, the question is whether the increased chance is enough to compensate for the risk of donating a kidney and never receiving a kidney back. Data from the Advanced Donation Program implemented by NKR (see Flechner et al. (2015) and Section 5) suggest that many individuals are willing to donate a kidney with a promise of receiving a kidney years after.

In addition, the long wait-time of patients in P is largely due to the small size of the French KEP. Once we do simulations on larger market sizes, we observe that the waiting time in P shrinks substantially, as depicted in Figure 5. It shows that the median waiting time of patients in the P queue substantially decreases as the market size gets bigger. When $n = 0.2$ (size of the Spanish KEP), the median waiting time of unpaired patients drops to 115 days. When the market is $n = 0.4$ (the UK KEP size), it is as low as 76 days. If we have a market of the size of the NKR, the median waiting of unpaired patients is 39 days.

4 Integrating Deceased Donors into the KEP

While the results in the previous section make us cautiously optimistic that patient-donor pairs will opt into donate-before-receipt in a relatively large KEP, policymakers in countries with a small KEP, such as France, may still have legitimate concerns with the two incentive issues. We now propose a new algorithm that, by integrating deceased donors into the unpaired exchange program, substantially reduces the waiting time of the unpaired patients (as well as the unpaired donors). In Section 4.1, we introduce and extend the theoretical results of Section 2.2 to this new setting. Then, in Section 4.2, we empirically analyze this idea and we make a final proposal for a practical

[38] These results are for $n = 0.05$ arrival rate, a size similar to the French KEP, and we obtain the same negligible effect of donor reneging for a larger market with $n = 0.2$, corresponding to the size of the Spanish KEP.

[39] In Table J.3 in Appendix J, we report the results of simulations assuming that some types of patients opt-out from donation-before-receipt. We simulate Unpaired and our practical proposal defined in Section 4.2 under two scenarios: (i) all hypersensitized patients opt-out and (ii) only patients with a PRA above 0.98 opt-out. In these simulations, we still confirm that Unpaired systematically outperforms the other algorithms and is close to Omniscient.

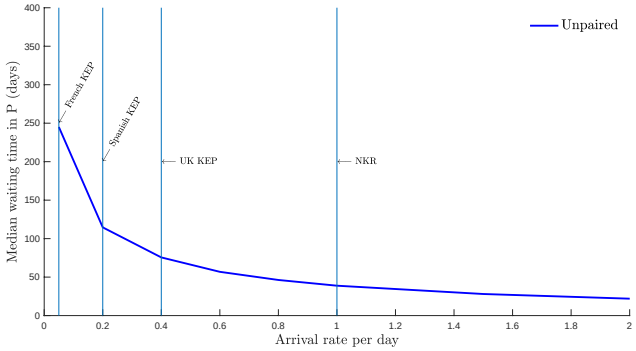

Figure 5: Median Waiting Time of Unpaired Patients in P for Various Market Sizes

Notes: This figure shows the performance of Unpaired in markets of eight different sizes, $n \in \{0.05, 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2\}$. The vertical lines indicate the size of some real-life KEPs.

implementation of the unpaired algorithm.

4.1 Integrating Deceased Donors: Theory

We now introduce the extended model. Deceased donors arrive in the market at rate μ. A patient is compatible with a deceased or a living donor with the same independent probability (p_H for an H patient and p_E for an E patient). Moreover, we assume that patients are indifferent between receiving a graft from a compatible living or a compatible deceased donor.[40]

To integrate Unpaired and Pairwise with DDL, we will allow DDL kidneys to be offered to patients in the KEP, while living donations still follow the same rules as the standard Unpaired/Pairwise algorithm. We assume that when a DDL kidney is matched with a patient in the KEP, one living donor waiting in the KEP donates her kidney to a patient on the DDL By doing this, we ensure that our proposed algorithm will not hurt patients waiting on the DDL. In such a case, we simply say that the donor is *removed* from the KEP.[41]

Definition 4.1 (Unpaired with DDL)**.** *If any new patient-donor pair $v_i = (p_i, d_i)$ enters the market at time t, match p_i to a compatible donor (if any), breaking ties arbitrarily, and match d_i to a compatible patient (if any), breaking ties in favor of hard-to-match patients. If any deceased donor dd_i arrives at t, match dd_i to a compatible patient (if any), breaking ties in favor of hard-to-match patients. When dd_i is matched with a patient, one living donor is selected at random and removed from the market. If dd_i is incompatible with all the patients in the market, remove dd_i.*

In the above definition, we assume that an incoming DDL kidney that is incompatible with all current patients in the KEP will not be offered to the KEP. This is reasonable since each DDL

[40] In our empirical analysis, we will consider only kidneys from deceased donors that are comparable to living donors in terms of quality. Hence, in the theoretical analysis, we implicitly assume that only kidneys from deceased donors of "high quality" will be offered to patients in the KEP. Thus, we should think of μ as the rate of arrival of "high quality" DDL kidneys.

[41] For the theoretical analysis, the identity of the donor removed from the market does not matter.

kidney in reality is only available for a very short period of time.[42] Note that we are implicitly assuming that we can immediately find a patient on the DDL compatible with our living donor. This is a weak assumption since in virtually all countries, the size of the KEP is marginal relative to the number of patients waiting for a deceased donors.[43]

The following result generalizes Proposition 2.6 by characterizing average waiting times of patients under Unpaired with DDL.

Proposition 4.2. *Under the Unpaired with DDL algorithm, for the average waiting time of hard-to-match patients and that of easy-to-match patients, we have:*

$$\lim_{p_H \to 0} p_H \mathbf{W}_H(Unpaired\ DDL) = \frac{\ln\left(n(1+\lambda)+\mu\right) - \ln\left(n+\mu\right)}{\lambda \cdot n}$$

and $\lim_{p_H \to 0} p_H \mathbf{W}_E(Unpaired\ DDL) = 0$. *Hence,*

$$\lim_{p_H \to 0} p_H \mathbf{W}(Unpaired\ DDL) = \frac{\ln\left(n(1+\lambda)+\mu\right) - \ln\left(n+\mu\right)}{n}.$$

We prove this result in the Appendix C.

Differentiating with respect to μ clarifies that the waiting time of (hard-to-match patients) is decreasing in the arrival rate of deceased donors. The new inflow from deceased donors gives more opportunities to match patients quickly.[44]

Integrating deceased donors into Pairwise and Optimum. Of course, the inflow of deceased donors can help *any* algorithm. It is, therefore, unclear whether our previous results comparing unpaired, pairwise, and optimum will continue to hold: Is unpaired with DDL still substantially better than pairwise with DDL and close to the optimal algorithm with DDL? To answer this question, we first define a natural version of pairwise with DDL algorithm.

As under unpaired (not to exert any negative externalities on patients waiting for DDL kidneys), we assume that whenever a DDL kidney is assigned to a patient in the KEP, one living donor is

[42] In practice, deceased donor kidneys are proposed, by order of priority, to compatible patients who are waiting on the DDL. Deceased-donor kidneys are exposed to the so-called cold ischemia time (from flush to out-of-ice) and it is well-documented that each additional hour of cold ischemia time significantly increases the risk of graft failure and mortality following renal transplantation. Hence, in practice, kidneys from deceased donors are being offered to patients for a short period of time.

[43] In January 2013, there were 4,500 patients waiting for a deceased donor in the New York Organ Donor Network (NYRT) (Agarwal et al., 2021). Over the period 2012-2020, the average length of the (active) DDL was around 5,500 patients in the UK and 8,150 patients in France. Since a compatible living donor is, on average, of better quality than many compatible deceased donor (Massie et al., 2016) it is reasonable to think that a living donor proposed to the DDL will find a compatible patient who is willing to accept her kidney.

[44] In our theoretical investigations, we do not distinguish between paired and unpaired patients. Nevertheless, when the algorithm treats paired and unpaired patients in a symmetric way, both types of patients would benefit from the inflow of deceased donors. Indeed, at an intuitive level, when the algorithm treats paired and unpaired patients symmetrically, the waiting time of unpaired patients corresponds to the waiting time of an arriving patient conditional on the event that he is not matched upon arriving (recall that, by definition, unpaired patients are not matched upon arriving). Using our main result that the unconditional waiting time decreases in μ, one can show that the same holds true for this conditional waiting time.

removed from the market. As before, the interpretation is that this living donor gives his kidney to a patient waiting on the DDL. Under Pairwise, however, the selected living donor naturally corresponds to the intended living donor of the patient getting matched to the deceased donor (in order not to create any unpaired patient/donor).

Definition 4.3 (Pairwise with DDL)**.** *If any new patient-donor pair v_i enters the market at time t, then match them with any cross-compatible patient-donor pair (if any), breaking ties in favor of hard-to-match patients. If any deceased donor dd_i arrives at t, match dd_i to a compatible patient (if any), breaking ties in favor of hard-to-match patients. When dd_i is matched with a patient, the living donor associated with the matched patient is removed from the market. If dd_i is incompatible with all the patients in the market, remove dd_i.*

The following result, proved in the Appendix E, characterizes the waiting time of the Pairwise with DDL algorithm.[45]

Proposition 4.4. *Under the Pairwise with DDL algorithm, for the waiting time of easy-to-match patients, we have:*

$$\lim_{p_H \to 0} p_H \mathbf{W}_E(Pairwise) = 0.$$

For the waiting time of hard-to-match patient,

1. *If $\mu > n(2\lambda - 1)$, we have:*

$$\lim_{p_H \to 0} p_H \mathbf{W}_H(Pairwise) = \frac{c}{\lambda \cdot n}.$$

where c solves

$$n(1-\lambda)e^{-cp_E} + \mu e^{-c} = n(1-2\lambda) + \mu.$$

2. *If $\mu < n(2\lambda - 1)$, we have:*

$$\lim_{p_H \to 0} p_H^2 \mathbf{W}_H(Pairwise) = \frac{\ln(2\lambda n) - \ln(n+\mu)}{\lambda \cdot n}.$$

When the rate of arrival of deceased donor is large, the performance is mainly determined by deceased donors, and thus the exchange algorithm is less important. Proposition 4.4 formalizes this intuition: When μ is high, the average waiting time of Pairwise with DDL is close to Unpaired with DDL, in the sense that they are both proportional to $1/p_H$. Note, however, that even in this case, the waiting time of Pairwise is always larger than Unpaired (see Figure 6 panel A for an illustration).

On the other hand, when μ is not too large ($\mu < n(2\lambda - 1)$), then the results of Theorem 2.7 extend: the scaling of the average waiting time under Pairwise is $1/p_H^2$, an order of magnitude higher than Unpaired. Panel B of Figure 6 illustrates this result.

[45] We did not find a closed form expression for the waiting time of the Pairwise with DDL algorithm in the knife-edge case where $\mu = n(2\lambda - 1)$.

Finally, in the Appendix D, we define and analyze a modified version of the Optimal algorithm that may use deceased donors.[46] We prove that the Unpaired algorithm's performance is close to the Optimal algorithm—even closer than what we obtained under $\mu = 0$ in Theorem 2.8.

Taken together, these findings show that even though the availability of deceased donors improves the performance of all algorithms, our previous theoretical results remain qualitatively similar. This is clear from comparing Figure 2 and Figure 6.

Remark 4.5. *In our context, one can also naturally define the Chain with DDL algorithm.*[47] *We do not formally characterize the expected waiting times of Chains with DDL. Chain algorithm is challenging to study (even with no DDL, Ashlagi et al. (2019) had to assume that $p_E = 1$). However, in the one-type model (i.e., $\lambda = 1$), one can easily show that Unpaired with DDL outperforms Chain with DDL. Indeed, under Chain with DDL, an arriving patient has probability p_H to be matched with the altruistic/bridge donor right away. With the complement probability (which tends to 1 as p_H vanishes), this patient will be unmatched and enter the pool. In that event, she will have to wait (1) either for an arriving patient to be compatible with the altruistic/bridge donor (which is necessary to initiate a chain-segment), which occurs with rate np_H; (2) or for a compatible deceased donor to arrive which occurs with rate μp_H. Thus, in expectation, this patient will have a waiting time bounded from below by $\frac{1}{(n+\mu)p_H}$. This is larger than $\ln\left(\frac{2n+\mu}{n+\mu}\right)/(np_H)$, the waiting time of patients under the Unpaired with DDL algorithm (see Proposition 4.2).*[48,49]

Impact of the market size. What happens to these performances of the algorithms that employ DDLs when the arrival rates of pairs increases? On one hand, an increase in the arrival rate of patient-donor pairs n (while fixing μ, the arrival rate of deceased donors) can increase the demand for deceased donors, which in turn can increase the waiting time of patients. On the other hand, an increase in the arrival rate of pairs makes the market thicker, which in turn can decrease the waiting time. Which force is more powerful is *ex ante* non-obvious.

A simple inspection of Proposition 4.2 shows that the second effect always dominates the first effect for the Unpaired with DDL—an increase in the arrival rate of pairs always deceases the waiting

[46] Essentially, the Optimal algorithm with DDL is the one that—among all matching algorithms—achieves the minimal average waiting time when using both kidneys from living and deceased donors *under the constraint that*, each time a deceased donor is matched to a patient, a living donor is removed from the system. The motivation for imposing this constraint is the same as under Unpaired/Pairwise: we do not want our algorithms to impose any negative externalities on patients waiting for a deceased donor. The interpretation is thus that the living donor removed gives his kidney to one of the (numerous) patients waiting for a deceased donor. A formal definition of Optimal with DDL algorithm is given in Section D of the Appendix where, generalizing Theorem 2.8, we provide a lower bound on the average waiting time achieved under this version of the Optimal algorithm.

[47] Informally, under Chain with DDL, patients in the pool will be offered kidneys from chain-segment (sparked by arriving patients compatible with the altruistic/bridge donor) as well as arriving DDL kidneys. In the latter event, the intended living donor will have to donate his kidney to a patient waiting on the DDL.

[48] That $\ln\left(\frac{2n+\mu}{n+\mu}\right)/n \leq 1/(n+\mu)$ holds since this can be rewritten as $\ln(n/(n+\mu)+1) \leq n/(n+\mu)$ which holds true since $n/(n+\mu)$ is a positive number smaller than 1.

[49] Further note that the argument applies both to the case where the chain algorithm identifies a chain-segment in a greedy fashion or optimally.

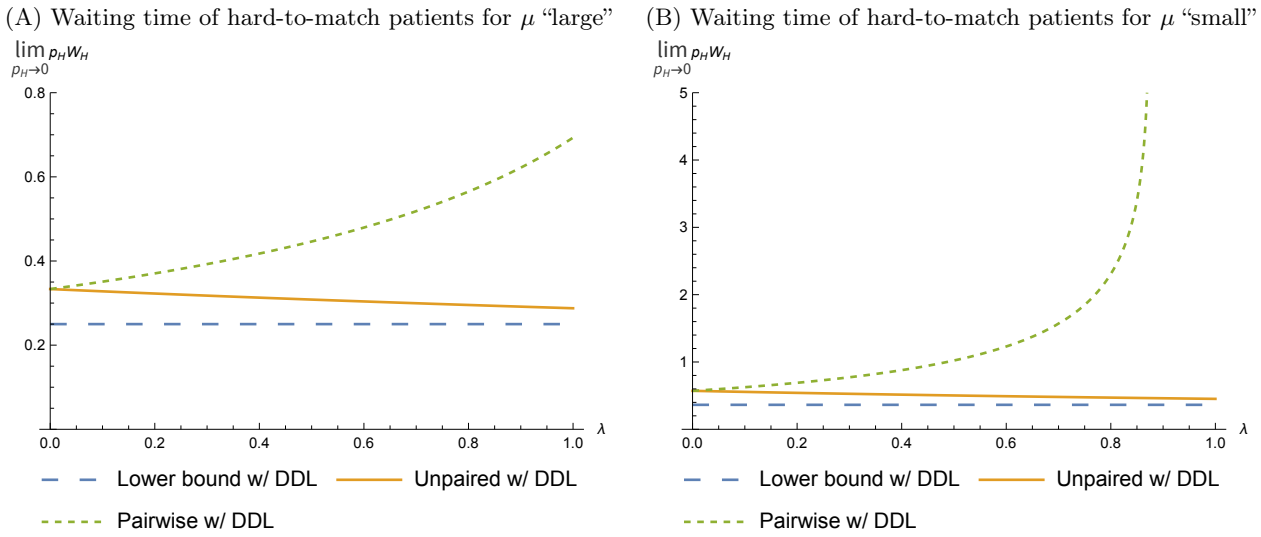

Figure 6: Waiting Time under Each Algorithm and the Arrival Rate of Hard-to-match Patients (λ)

Notes: Given $p_E = 1$, $n = 1$, the above graph shows for $\mu = 2$ for panel (A) and $\mu = 3/4$ for panel (B), as a function of the arrival rate of hard-to-match patients (λ), the average waiting times of hard-to-match patients under Pairwise with DDL, Unpaired with DDL algorithms, as well as a lower bound of the average waiting time achieved under the Optimal algorithm with DDL.

time.[50] This seemingly obvious comparative static becomes less obvious when one notes that this is *not* the case for Pairwise with DDL; that is, an increase in n can either increase or decrease the waiting time, depending on the parameters.

To get a simple intuition for this observation, consider a patient p currently waiting in the system. A new arriving patient will 'compete' with patient p for deceased donors. However, the new patient comes into the system with an incompatible donor, who is as likely to be compatible to p as any deceased donor. Thus, the new donor more than compensates for the reduction in the probability of getting matched to a deceased donor. This simple intuition though does not apply to the Pairwise: the likelihood of being compatible with an arriving deceased donor is larger than the likelihood of being *cross-compatible* with an arriving pair. This is particularly problematic when many arriving patients are hard-to-match. As illustrated in Figure 7, the negative force can dominate the positive force when λ is large. One can show that for low values of λ, the positive force can dominate, whereas for intermediate values of λ, the comparison is more tricky and the waiting time under Pairwise can be a U-shaped function of n.[51]

[50] As discussed in footnote 44, if paired and unpaired patients are treated in a symmetric way by our algorithm, the decreasing relationship between waiting time and n should hold true for both paired and unpaired patients. We indeed observe this pattern for unpaired patients in our simulations (see Section 4.2).

[51] We formally show in Section F of the appendix that for large values of λ the waiting time under Pairwise with DDL is increasing in n in the most favorable regime for Pairwise, i.e., when $\mu > n(2\lambda - 1)$.

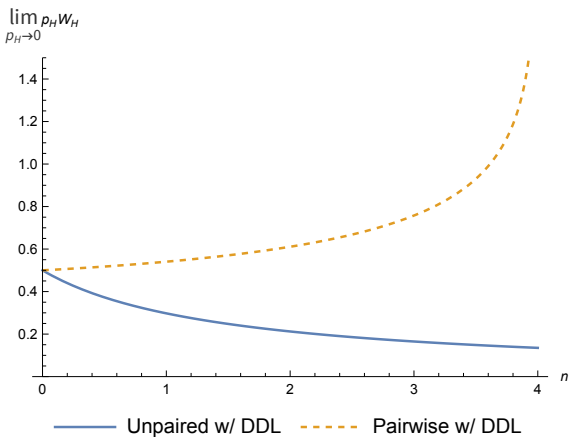

Figure 7: Waiting time of hard-to-match patients as a function of the arrival rate of pairs (n).

Notes: Given $p_E = 1$, $\lambda = 3/4$ and $\mu = 2$, the above graph shows, the average waiting times of hard-to-match patients, as a function of the arrival rate of pairs (n).

4.2 Integrating Deceased Donors: Empirics

Using the same data from France as in Section 3, we now evaluate the performance of Unpaired with DDL and compare it with Pairwise with DDL.[52]

A practical challenge of this algorithm is that DDL kidneys typically have lower quality than living kidneys. If we were to offer any type of DDL kidneys, some pairs would receive a kidney that is of lower quality than the kidney they provide (in expectation), which then may discourage them joining the system to begin with. We address this issue by selecting, for each patient, a DDL kidney that has a KDPI below the LKDPI of the patient's paired incompatible donor.[53] As we explained in footnote 32, KDPI and LKDPI are two comparable quality measures for DDL and living donor kidneys, respectively.

While we control and vary the market size by the daily Poisson arrival rate of pairs n, we keep the arrival of DDL kidneys fixed. This captures the fact that, while the arrival rate of patient-donor pairs changes as a KEP is more or less successful, the arrival of DDL kidneys is typically exogenous. We also assume that each DDL kidney arrives on its actual arrival date in the data and is available for transplant on that date only.

Assessment of algorithms with DDL. We first evaluate Unpaired and Pairwise with DDL algorithms. Figure 8 (panel A) shows the results for various market sizes. Relative to Pairwise and Unpaired without DDL (Appendix Figure I.1), both algorithms have a significantly improved performance: the mean waiting time of patients is reduced by about 88 to 91 percent. Yet, Unpaired

[52] In our simulations, under Pairwise with DDL we remove a paired patient's associated living donor from the market when that patient receives a DDL kidney. This is feasible since all patients are paired. Under Unpaired with DDL, we do this as long as this is feasible. For unpaired patients (whose donor has already donated), we remove one unpaired donor from the market. This is feasible because there always exist an equal number of unpaired patients and donors.

[53] In Appendix H.4, we present the selection process of Deceased-Donor Kidneys in details. Moreover, we show in Appendix J that our results are robust to the use of a more demanding screening criterion.

still performs better than Pairwise, and is fairly close to the Omniscient. For the market size of the French KEP, for instance, the mean waiting time of Unpaired with DDL is around 43 days, the Omniscient is 38 days, and the Pairwise with DDL is 54 days. Unpaired with DDL also diminishes the waiting time of unpaired patients. For the market size of the French KEP, the median waiting time in P falls to 50 days (vs. 245 days for Unpaired without DDL). Hence, even in small markets, the main practical challenge associated with the Unpaired algorithm is drastically reduced thanks to the integration of the DDL. This is even more the case in larger markets. As shown in Figure I.2 (Panel A), time in P decreases sharply as the arrival rate of pairs increases.[54]

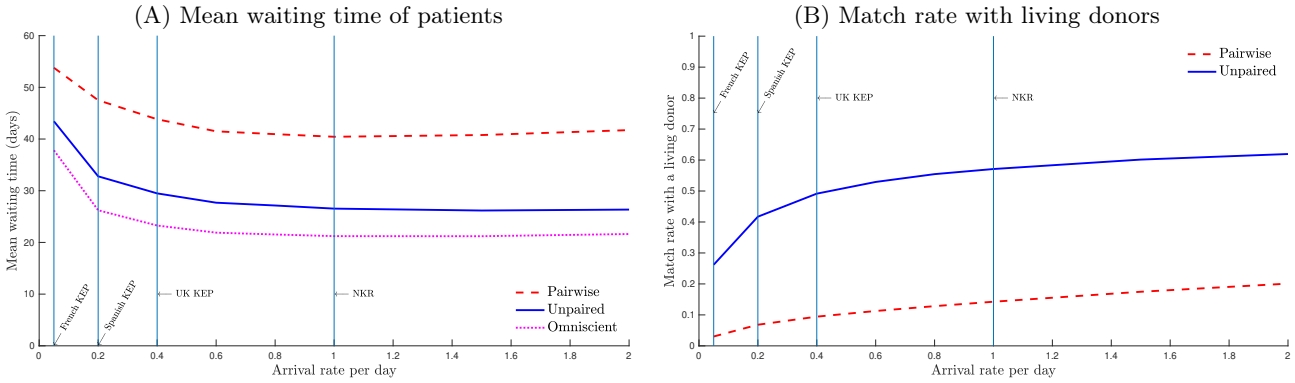

Figure 8: Performance of Pairwise/Unpaired with DDL: Various Market Sizes

Notes: This figure shows the performance of the three algorithms with DDL in markets of eight different sizes, $n \in \{0.05, 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2\}$. The vertical lines indicate the size of some real-life KEPs. In each simulated market, Omniscient is allowed to use the DDL kidneys that are actually taken by Unpaired with DDL.

The improved performance of Unpaired and Pairwise with DDL comes, at least partially, from the patients' access to DDL kidneys. As depicted in Figure 8 (panel B), for both algorithms, the share of patients receiving a kidney from a living donor increases with market size and is significantly higher under Unpaired with DDL compared to Pairwise with DDL for all market sizes. Moreover, this match rate with living donors is close to zero under Pairwise with DDL for a market size corresponding to the French KEP ($n = 0.05$). Hence, in that configuration, the system will de facto reduce to what is known as *list exchange*.[55]

A low match rate with living donors can have important fairness consequences such as hurting blood type O patients who already have the longest waiting time.[56] Indeed, many incompatible pairs have an O patient and a non-O donor.[57] Then, with a low match rate with living donors, many pairs will donate a non-O kidney to the DDL while they will eventually obtain an O-donor.

[54] We find the same pattern for the median waiting time of Unpaired donors. It falls to 65 days (vs. 339 days for Unpaired without DDL) for $n = 0.05$ and it is decreasing in n, reaching 38 days for $n = 2$.

[55] List exchange allows a living donor to donate to a patient on the DDL and, in return, her paired patient obtains a high priority on the DDL (Delmonico et al., 2004). See also Section 5 for further details on List exchange.

[56] Glander et al. (2010) report the waiting time of patients in the US, showing that O patients wait the longest. The same pattern is observed in France as reported here (see Table R9): https://www.agence-biomedecine.fr/annexes/bilan2017/donnees/organes/06-rein/synthese.htm#tR9.

[57] Among the 586 incompatible pairs of our data set, almost 60% have an O patient while more than 60% have a non-O donor (see Table 1).

This problem constitutes a major objection against list exchange (Ross and Woodle, 2000). The impact for O patients of algorithms like Unpaired and Pairwise with DDL is twofold. On the one hand, O patients waiting on the DDL are likely to suffer a loss (measured by the difference between the number of O deceased donors offered to patients in the program and the number of O living donors given back to the DDL). On the other hand, many O patients would have access to a graft in the program thanks to those algorithms. In the sequel, we focus on the overall gains (or losses) for O patients (measured by the difference between the number of O patients grafted in the system and the loss for O patients waiting on the DDL). As reported on Table I.2 (in Appendix I), the loss for O patients waiting on the DDL is systematically higher under Pairwise with DDL than under Unpaired with DDL. Moreover, the overall gains generated by Unpaired with DDL are always positive and always higher than those generated by Pairwise with DDL. Hence, compared to Pairwise with DDL, the good performances of the Unpaired with DDL algorithm do not come at the expense of O patients.

As pointed out by Zenios et al. (2001), the fairness issue for O-patients might also be mitigated if patients were incentivized to bring an O rather than a non-O donor. However, in algorithms offering a high priority for DDL kidneys for patients participating to the KEP, these incentives are weak since patients expect to receive quickly a DDL offer regardless of the type of donor they bring to the system. This is confirmed in our simulations. For both Pairwise and Unpaired with DDL, we observe that patients bringing an O donor wait longer time than patients bringing a non-O donor (see Table I.2).

Final proposal. The above discussion motivates us to consider a version of the Unpaired with DDL, as defined below, as the most practically plausible solution (we also define further a related version of Pairwise with DDL):

Definition 4.6 (Unpaired with DDL with δ Delay)**.** *Each patient is required to wait for δ months before receiving any DDL kidney offers. Patients can always receive living donor offers based on the rules of the Unpaired algorithm. Patients whose donors have already donated can receive DDL kidney offers at any time.*

This modified algorithm works like the standard Unpaired with DDL algorithm, but requires patients to wait for at least δ months before receiving any DDL kidneys *if their donor has not already donated.* This will provide incentives for patients to find a donor who is likely to donate soon to a patient in the KEP (e.g., an O donor who is likely to be compatible with many patients), so that they can receive high quality DDL kidneys as soon as possible.[58]

One can also similarly define *Pairwise with DDL with δ Delay*, where a patient is also required to wait for δ months before receiving any DDL kidney offers but living donations are based on the

[58] Table I.2 reports the average waiting time of patients as a function of the blood type of the donor they bring to the system. As already mentioned, when $\delta = 0$, on average the patients with an O donor wait the longest. However, when $\delta = 6$ this result is reversed, patients having an O donor wait the shortest.

rules of the Pairwise algorithm.[59] This algorithm can be seen as a practical version of Pairwise with DDL which can be a good alternative if policymakers have strong preferences for avoiding unpairing patients and donors.

Notice that if $\delta = 0$, these algorithms turn into the standard Unpaired with DDL/Pairwise with DDL. Thus, the key question is, for practical purposes, what is the right value of δ? On one hand, a smaller delay parameter creates an incentive for patients to enter with low-value donors. On the other hand, a larger delay parameter disincentivizes some patients to join the system and, more importantly, imposes a waiting cost on patients who are under dialysis.

Finding the "optimal" value of δ is a context-dependent exercise that requires a comprehensive evaluation of various factors beyond the scope of this paper. Nevertheless, we simulate the system for a series of reasonable delay parameters. Here we only focus on $\delta = 6$ months, though results are qualitatively similar for $\delta = 3$ or 9 months as well as a wide range of δ values (we discuss robustness of results with respect to δ in Figure I.3 of the appendix).[60] In general, the match rate with living donors increases with δ, *ceteris paribus*. Hence, a KEP may also use historical data to select the lowest δ value to approximately achieve its desired match rate with living donors.

Figure 9 presents the results for various market sizes. Naturally, relative to $\delta = 0$ (Figure 8), the mean waiting time increases when $\delta = 6$ months (panel A). Despite the fact that we require patients to wait for 180 days before receiving a DDL kidney (unless their donors have already donated), the Unpaired algorithm manages to keep the waiting time under 90 days. In addition, panel B shows that the match rate with living donors is significantly increased. Unpaired with DDL matches 55 to 73 percent of patients with a living donor, a sizable increase relative to $\delta = 0$.

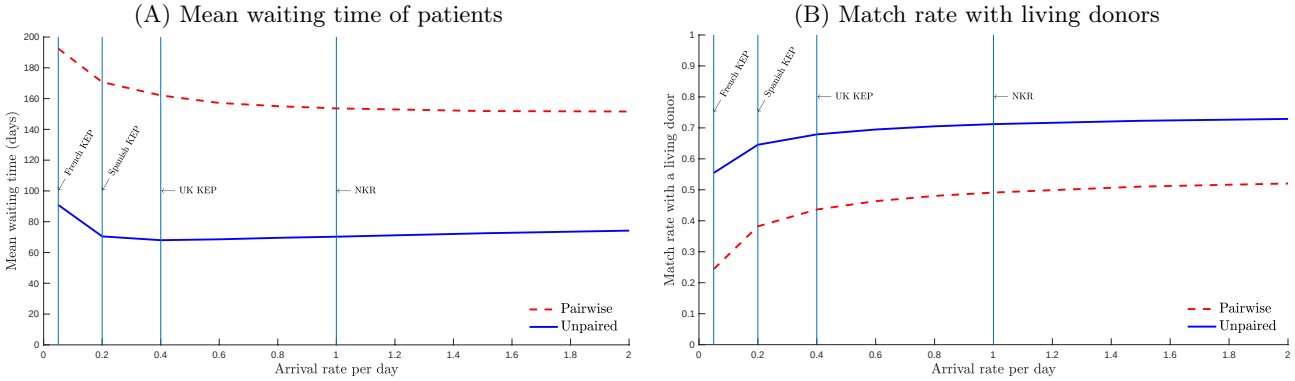

Figure 9: Performance of Pairwise/Unpaired with DDL & $\delta = 6$ Months: Various Market Sizes

Notes: This figure shows the performance of the three algorithms with DDL in markets of eight different sizes, $n \in \{0.05, 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2\}$. The vertical lines indicate the size of some real-life KEPs. Each patient to wait for $\delta = 6$ months before receiving any DDL kidney offers under each algorithm, while the patient can still receive living donor offers during that δ months.

[59] Pairwise with DDL (without delay) suffers from the same drawbacks as Unpaired with DDL in terms of fairness for the O patients waiting on the DDL and incentives to bring an O donor to join the KEP (see Section 4.2). Hence, Pairwise with DDL with δ Delay might be viewed as an implementable version of the Pairwise with DDL algorithm.

[60] We consider $\delta = 6$ months as a reasonable waiting time in the French context. As discussed in footnote 25, an incompatible pair usually participates in KEP for 3.4 match runs, amounting to waiting for 9-12 months.

For the market size $n = 0.05$, or a size similar to the French KEP, columns (6) and (7) of Table 2 show more statistics. In particular, for Unpaired with DLL and $\delta = 6$ months, the two potential issues discussed in Section 3.4 are almost negligible. A median unpaired patient only waits for 6 days before receiving a kidney, and a median unpaired donor waits 39 days in donating a kidney.[61] Even the hard-to-match patients (high PRA patients and O patients) do not need to wait long: median waiting time for such unpaired patients is below 55 days.

Figure 9 and Table 2 also allow us to compare the versions with delay of Unpaired with DDL and Pairwise with DDL. The performances of Pairwise, in terms of waiting time of patients, transplant rate, and match rate with living, are substantially worse than those of Unpaired. However, by construction, Pairwise entirely removes the two incentives issues because it does not create any unpaired patient/donor.

5 Practices Related to the Unpaired Algorithm

Our proposed Unpaired algorithm eliminates the timing constraints on the donation and receipt for any incompatible pair. Specifically, it allows *receipt-before-donation* (i.e., a patient can receive a kidney before her paired donor donates to some other patient) as well as *donation-before-receipt* (i.e., a donor can donate before her paired patient receives a transplant). Some recent practices also relax the timing constraints, and we discuss their connection to the Unpaired algorithm below.

Vouchers. A voucher program, introduced in Veale et al. (2017), allows donation-before-receipt. An example of such program is the Advanced Donation Program (ADP) implemented by NKR.[62] When an advance donation happens, the paired patient obtains a voucher that gives her a higher priority for a kidney in the future. In the ADP, a patient with a voucher will have a high priority for receiving a kidney from a donor who would otherwise end a chain.[63]

Combined with the Chain algorithm, the ADP allows both donation-before-receipt and receipt-before-donation, but with important restrictions. For example, a patient with a voucher is offered a donor kidney only when the donor's paired patient has already received a transplant in a chain and when the donor kidney is incompatible with all patients in the remaining pairs. Moreover, an advanced donor can only donate to patients whose donors have not donated. As a result, patients with a voucher in ADP tend to wait for a long time.[64]

[61] Interestingly, the introduction of a delay to the Unpaired with DDL algorithm reduces the waiting time of patients in P (see Figure I.2 for a comparison of waiting times in P for the algorithms without and with delay). Indeed, in the absence of any delay, more patients outside P receive a kidney from a deceased donor such that more living donors are given back to the DDL. This reduces the chance of getting a match for a patient in P.

[62] Flechner et al. (2015) report 10 advanced donations within NKR from August 2011 to August 2014. Since then, ADP has expanded. As of April 27, 2020, there have been about 500 advanced donations. Half of these donations are from donors whose paired incompatible patient is in urgent need of a kidney (see NKR's quarterly report on paired kidney exchange for Q1 2020; available at https://www.kidneyregistry.org/pages/c6/nkr_quarterly_reports).

[63] In the priority order, the patients with a voucher are right after former NKR donors in need of a kidney transplant and patients involved in real-time swap failures where the donor has donated but the patient did not receive a kidney; see Tenenbaum (2018).

[64] Among NKR's 10 advanced donations during August 2011 to August 2014, by the end of that period, 8 of the

In contrast, an unpaired patient in Unpaired can receive a kidney from any donor, paired or unpaired, while a donor can donate to any compatible patient, paired or unpaired. In this sense, the Unpaired algorithm generalizes advanced donation by relaxing the constraints on donation-before-receipt and integrating it with receipt-before-donation.

List Exchange. The Unpaired with DDL algorithm that we introduced combines the Unpaired algorithm with the DDL. Certain integrations of DDL and living donation are already observed in practice. Upon the approval of United Network for Organ Sharing, the New England region implemented a program called List Exchange. It allows a living donor to donate to a patient on the DDL and, in return, her paired patient obtains a high priority on the DDL (Delmonico et al., 2004). In other words, it allows donation-before-receipt for donations to the DDL.

As already discussed, one important objection against list-exchange lies in the fact that it can hurt blood-type O patients who already have the longest waiting time on the DDL (Ross and Woodle, 2000). As discussed in Section 4.2, Unpaired with DDL is less detrimental to O patients waiting on the DDL and more beneficial to O patients in general (including those participating to the KEP) compared to Pairwise with DDL (which essentially corresponds to List Exchange). Moreover, List Exchange does not incentivize patients to find a O donor since any healthy kidney will be quickly accepted by some patients on the DDL due to the huge excess demand. In contrast, our final proposal – Unpaired with DDL and delay – may increase the supply of highly sought-after kidneys, O kidneys in particular, because a patient with a O donor is likely to be matched earlier. Indeed, in this proposal, a paired patient can only receive from a deceased donor after few months while an unpaired patient is offered deceased donor kidneys immediatly. By definition, a patient becomes unpaired as soon as the paired donor has donated which is more likely to happen if this donor is of blood type O.

Deceased donor-initiated chain. Roth et al. (2004) contains a proposal on how to integrate List Exchange with kidney exchange programs: Instead of donating directly to the DDL, a donor could initiate a chain of transplant within the KEP, in exchange of a high priority on the DDL for her intended patient. This proposal is close to our Unpaired with DDL algorithm as it integrates the DDL and the KEP and it allows donation-before-receipt (the first donor of the chain donates before her patient receives from a deceased donor) and possibly receipt-before-donation (along a chain of transplant initiated by this first donor). The main difference between the two lies in the fact that, under Unpaired with DDL, the patient who was associated with the first donor of the chain not only benefit from a high priority on the DDL but can also receive a living donation. While the proposal made by Roth et al. (2004) has not been implemented so far, the latest revision of

10 patients had received a kidney 178 days on average after their donors had donated (Flechner et al., 2015). In our simulations of market sizes comparable to NKR, the Unpaired algorithm leads to an average waiting time in P, conditional on receiving a transplant, of 61 days. Moreover, Tenenbaum (2018) reports that ADP notifies hard-to-match patients (typically hypersensitized) that the waiting time after their donor has donated may often exceed 1–2 years. In our simulations the mean waiting time in P (conditionally on receiving a transplant) for hypersensitized patients is only 147 days.

the bioethics law in France allows a variant of it (Combe et al., 2022). Moreover, another kind of deceased donor-initiated chain has been at work in Italy since 2019 (Furian et al., 2019). Under this alternative design, patients from incompatible pairs benefit from a high priority on the DDL, and a donor of those pairs initiate a chain only after her intended patient has received a deceased donor kidney. As discussed in Section 4.2, all those mechanisms giving a high priority on the DDL for patients participating to the KEP mitigate the incentives that patients have to bring a O donor. One possibility to restore those incentives would be to impose a delay between the arrival date of an incompatible pair in the program and the date at which the patient of this pair benefits from a high priority on the DDL, as suggested by Wang et al. (2021).[65] This is in line with our final proposal *Unpaired with DDL and delay*.

In sum, the implementation of the voucher program and List Exchange in practice and the reform in France make us optimistic about the potential of Unpaired, as well as Unpaired with DDL, to promote kidney exchange. The innovations in our algorithm have been shown to be acceptable in practice, while our algorithms enjoy significant performance advantages relative to state-of-art algorithms.

6 Conclusion

We have proposed a new matching algorithm, Unpaired kidney exchange, and argued that it significantly improves upon the outcome of currently utilized state-of-the-art algorithms. The main reason is that Unpaired eliminates the common simultaneity constraints. In this sense, our results provide not only a new policy but also a tool to evaluate the costs of those constraints.

Eliminating those simultaneity constraints brings two practical challenges, the potentially long waiting times of a patient after her paired donor's donation and of a donor after her paired patient's transplantation. We have proposed several solutions and recommended a practical version, an integration of Unpaired with the deceased donor list, while a patient is eligible for a deceased-donor kidney after a pre-specified time or immediately after her paired donor's donation.

One thing we have not systematically examined is endogenous participation in the kidney exchange. In practice, seeing the algorithm in use, a patient-donor pair will decide if they want to participate in the exchange, a participating pair will decide if they want to quit and choose desensitization, and a patient may even be incentivized to find a different type of donor. Ignoring such endogenous responses, our results are likely to provide a lower bound of the advantages of Unpaired. When more pairs participate in a more efficient exchange, the performance of the exchange will be further enhanced because a larger market size in general improves the performance of every algorithm. We believe that a systematic analysis taking into account endogenous participation is a

[65] Wang et al. (2021) simulate the impact of authorizing deceased donor-initiated chain in the US. They write "*It may seem important to have a period of time before a new arrival to the kidney paired donation pool is eligible for a deceased donor chain-initiating kidneys transplant, so the delay time of 0, which is optimum in terms of the number of transplants achieved, may not be acceptable. A delay time of 3 months or possibly 6 months may be a reasonable compromise.*"

fruitful avenue for future research.[66]

When taking a new algorithm to practice, patients and donors may be skeptical. Here we are hopeful that the combination of theory and evidence in this paper will alleviate this skepticism. We have shown that if Unpaired or Unpaired with DDL had been employed in some existing markets, participants' outcomes would have been meaningfully improved. We have also explained theoretically why we should expect this to be the case. This combination of evidence and theory gives us confidence that any future application of the Unpaired algorithms will improve patient outcomes.

[66]The rational queueing literature (see Hassin (2016)) is dealing with decisions of participation in queueing systems where agents tradeoffs their waiting times to get served/matched with their outside option. There are only few attempts of adapting these frameworks to dynamic matching environments, see Baccara et al. (2020) and Che and Tercieux (2020).

References

Agarwal, Nikhil, Itai Ashlagi, Eduardo Azevedo, Clayton Featherstone, and Ömer Karaduman, "What Matters for the Productivity of Kidney Exchange?," in "AEA Papers and Proceedings," Vol. 108 2018, pp. 334–40.

_ , _ , Michael A Rees, Paulo Somaini, and Daniel Waldinger, "Equilibrium allocations under alternative waitlist designs: Evidence from deceased donor kidneys," *Econometrica*, 2021, *89* (1), 37–76.

Akbarpour, Mohammad, Farshad Fatemi, and Negar Matoorian, "The Iranian market for kidneys," *Working paper*, 2019.

_ , Shengwu Li, and Shayan Oveis Gharan, "Thickness and information in dynamic matching markets," *Journal of Political Economy*, 2020, *128* (3), 783–815.

Anderson, Ross, Itai Ashlagi, David Gamarnik, and Yash Kanoria, "Efficient dynamic barter exchange," *Operations Research*, 2017, *65* (6), 1446–1459.

Andersson, Tommy and Jörgen Kratz, "Pairwise kidney exchange over the blood group barrier," *The Review of Economic Studies*, 2020, *87* (3), 1091–1133.

Ashlagi, Itai and Alvin E Roth, "Kidney Exchange: an Operations Perspective," 2020.

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

_ , Maximilien Burq, Patrick Jaillet, and Vahideh Manshadi, "On matching and thickness in heterogeneous dynamic markets," *Operations Research*, 2019, *67* (4), 927–949.

Ausubel, Lawrence M and Thayer Morrill, "Sequential kidney exchange," *American Economic Journal: Microeconomics*, 2014, *6* (3), 265–85.

Axelrod, David A, Mark A Schnitzler, Huiling Xiao, William Irish, Elizabeth Tuttle-Newhall, Su-Hsin Chang, Bertram L Kasiske, Tarek Alhamad, and Krista L Lentine, "An economic assessment of contemporary kidney transplant practice," *American Journal of Transplantation*, 2018, *18* (5), 1168–1176.

Baccara, Mariagiovanna, SangMok Lee, and Leeat Yariv, "Optimal dynamic matching," *Theoretical Economics*, 2020, *15*, 1221–1278.

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

Bentall, A, LD Cornell, JM Gloor, WD Park, MJ Gandhi, JL Winters, MF Chedid, PG Dean, and Mark D Stegall, "Five-year outcomes in living donor kidney transplants with a positive crossmatch," *American Journal of Transplantation*, 2013, *13* (1), 76–85.

Biró, Péter, Bernadette Haase-Kromwijk, Tommy Andersson, Eyjólfur Ingi Ásgeirsson, Tatiana Baltesová, Ioannis Boletis, Catarina Bolotinha, Gregor Bond, Georg Böhmig, Lisa Burnapp, Katarína Cechlárová, Paola Di Ciaccio, Jiri Fronek, Karine Hadaya, Aline Hemke, Christian Jacquelinet, Rachel Johnson, Rafal Kieszek, Dirk R. Kuypers, Ruthanne Leishman, Marie-Alice Macher, David Manlove, Georgia Menoudakou, Mikko Salonen, Bart Smeulders, Vito Sparacino, Frits C.R. Spieksma, María Oliva Valentín, Nic Wilson, and Joris van der Klundert, "Building Kidney Exchange Programmes in Europe—An Overview of Exchange Practice and Activities," *Transplantation*, 2019, *103* (7), 1514–1522.

Calvillo-Arbizu, Jorge, Miguel A. Pérez-Valdivia, Miguel A. Gentil-Govantes, Pablo Castro de-la Nuez, Auxiliadora Mazuecos-Blanca, Alberto Rodríguez-Benot, María C. Gracia-Guindo, Francisco Borrego-Utiel, Mercedes Cabello-Díaz, Rafael Bedoya-Pérez, Manuel Alonso-Gil, Mercedes Salgueira-Lazo, and Laura M. Roa-Romero, "Does the Kidney Donor Profile Index (KDPI) predict graft and patient survival in a Spanish population?," *Nefrología (English Edition)*, 2018, *38* (6), 587–595.

Che, Yeon-Koo and Olivier Tercieux, "Optimal Queue Design," *Working Paper*, 2020.

Combe, Julien, Victor Hiller, Olivier Tercieux, Benoît Audry, Jules Baudet, Géraldine Malaquin, François Kerbaul, Corinne Antoine, Marie-Alice Macher, Christian Jacquelinet et al., "Perspectives pour une évolution du programme de don croisé de reins en France," *Néphrologie & Thérapeutique*, 2022, *18* (4), 270–277.

_ , _ , _ , _ , Yinghua He, Christian Jacquelinet, and Marie-Alice Macher, "Outlook on the Kidney Paired Donation Program in France," *IPP Notes, Paris School of Economics*, 2019.

Cowan, N, HA Gritsch, Nima Nassiri, Joseph Sinacore, and J Veale, "Broken chains and reneging: A Review of 1748 kidney paired donation transplants," *American Journal of Transplantation*, 2017, *17* (9), 2451–2457.

Delmonico, Francis L, Paul E Morrissey, George S Lipkowitz, Jeffrey S Stoff, Jonathan Himmelfarb, William Harmon, Martha Pavlakis, Helen Mah, Jane Goguen, Richard Luskin et al., "Donor kidney exchanges," *American Journal of Transplantation*, 2004, *4* (10), 1628–1634.

Erdős, Paul and Alfréd Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, 1960, *5* (1), 17–60.

Flechner, SM, D Leeser, R Pelletier, M Morgievich, K Miller, L Thompson, S McGuire, J Sinacore, and G Hil, "The incorporation of an advanced donation program into kidney paired exchange: initial experience of the National Kidney Registry," *American Journal of Transplantation*, 2015, *15* (10), 2712–2717.

Furian, Lucrezia, Cristina Cornelio, Cristina Silvestre, Flavia Neri, Francesca Rossi, Paolo Rigotti, Emanuele Cozzi, and Antonio Nicolò, "Deceased donor–initiated chains: First report of a successful deliberate case and its ethical implications," *Transplantation*, 2019, *103* (10), 2196–2200.

Galichon, Alfred, *Optimal transport methods in economics*, Princeton University Press, 2018.

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

Glander, Petra, Klemens Budde, Danilo Schmidt, T Florian Fuller, Markus Giessing, Hans-Hellmut Neumayer, and Lutz Liefeldt, "The 'blood group O problem'in kidney transplantation—time to change?," *Nephrology Dialysis Transplantation*, 2010, *25* (6), 1998–2004.

Hartigan, John A and Pamela M Hartigan, "The dip test of unimodality," *The annals of Statistics*, 1985, pp. 70–84.

Hassin, Refael, *Rational Queueing*, CRC Press, 2016.

Heo, Eun Jeong, Sunghoon Hong, Youngsub Chun et al., "Kidney exchange with Immunosuppressants," *Economic Theory*, 2018, pp. 1–19.

Jevons, William Stanley, *Money and the Mechanism of Exchange*, Vol. 17, Kegan Paul, Trench, 1885.

Korte, Bernhard H, Jens Vygen, B Korte, and J Vygen, *Combinatorial optimization*, Vol. 1, Springer, 2011.

Lehner, Lukas Johannes, Anna Kleinsteuber, Fabian Halleck, Dmytro Khadzhynov, Eva Schrezenmeier, Michael Duerr, Kai-Uwe Eckardt, Klemens Budde, and Oliver Staeck, "Assessment of the kidney donor profile index in a European cohort," *Nephrology Dialysis Transplantation*, 2018, *33* (8), 1465–1472.

Marfo, Kwaku, Amy Lu, Min Ling, and Enver Akalin, "Desensitization protocols and their outcome," *Clinical Journal of the American Society of Nephrology*, 2011, *6* (4), 922–936.

Martello, Silvano and Paolo Toth, "Linear assignment problems," *Surveys in Combinatorial Optimization*, 1987, *132*, 259–282.

Massie, Allan B, Joseph Leanza, Lara M Fahmy, Eric KH Chow, Niraj M Desai, Xun Luo, Elizabeth A King, Mary G Bowring, and Dorry L Segev, "A risk index for living donor kidney transplantation," *American Journal of Transplantation*, 2016, *16* (7), 2077–2084.

Meyn, Sean P and Richard L Tweedie, "Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes," *Advances in Applied Probability*, 1993, *25* (3), 518–548.

Möbius, Markus, "Trading Favors," *Report*, 2001.

Organ Procurement and Transplant Network, "A Guide to Calculating and Interpreting the Kidney Donor Profile Index (KDPI)," May 2019. Accessed: July 11, 2019.

Rao, Panduranga S, Douglas E Schaubel, Mary K Guidinger, Kenneth A Andreoni, Robert A Wolfe, Robert M Merion, Friedrich K Port, and Randall S Sung, "A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index," *Transplantation*, 2009, *88* (2), 231–236.

Rehse, Gregor, Fabian Halleck, Dmytro Khadzhynov, Lukas J Lehner, Anna Kleinsteuber, Anja Staeck, Michael Duerr, Klemens Budde, and Oliver Staeck, "Validation of the Living Kidney Donor Profile Index in a European cohort and comparison of long-term outcomes with US results," *Nephrology Dialysis Transplantation*, 2018, *34* (6), 1063–1070.

Ross, Lainie Friedman and E Steve Woodle, "Ethical issues in increasing living kidney donations by expanding kidney paired exchange programs," *Transplantation*, 2000, *69* (8), 1539–1543.

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

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

_ , _ , and _ , "Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences," *American Economic Review*, 2007, *97* (3), 828–851.

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

Scurt, Florian G, Lara Ewert, Peter R Mertens, Hermann Haller, Bernhard MW Schmidt, and Christos Chatzikyrkou, "Clinical outcomes after ABO-incompatible renal transplantation: a systematic review and meta-analysis," *The Lancet*, 2019, *393* (10185), 2059–2072.

Sönmez, Tayfun, M Utku Ünver, and M Bumin Yenmez, "Incentivized kidney exchange," *American Economic Review*, 2020, *110* (7), 2198–2224.

_ , _ , _ et al., "Incentivized kidney exchange," *Boston College Department of Economics Boston College Working Papers in Economics*, 2018, *931*.

Tenenbaum, Evelyn M, "Swaps and Chains and Vouchers, Oh My!: Evaluating How Saving More Lives Impacts the Equitable Allocation of Live Donor Kidneys," *American journal of law & medicine*, 2018, *44* (1), 67–118.

Ünver, M Utku, "Dynamic kidney exchange," *The Review of Economic Studies*, 2010, *77* (1), 372–414.

Veale, Jeffrey L, Alexander M Capron, Nima Nassiri, Gabriel Danovitch, H Albin Gritsch, Amy Waterman, Joseph Del Pizzo, Jim C Hu, Marek Pycia, Suzanne McGuire, Marianand Charlton, and Sandip Kapur, "Vouchers for future kidney transplants to overcome chronological incompatibility between living donors and recipients," *Transplantation*, 2017, *101* (9), 2115–2119.

Wang, Wen, Michael A Rees, Alan B Leichtman, Peter X-K Song, Mathieu Bray, Valarie B Ashby, Tempie Shearon, Andrew Whiteman, and John D Kalbfleisch, "Deceased donors as nondirected donors in kidney paired donation," *American Journal of Transplantation*, 2021, *21* (1), 103–113.

Williams, David, *Probability with martingales*, Cambridge university press, 1991.

Zenios, Stefanos, E Steve Woodle, and Lainie Friedman Ross, "Primum non nocere: avoiding increased waiting times for individual racial and blood-type subsets of kidney wait list candidates in a living donor/cadaveric donor exchange program," *Transplantation*, 2001, *72* (4), 648–654.

A Proof of Proposition 2.6

In Proposition 4.2, we characterize the average waiting time of patients under the Unpaired with DDL algorithm defined in Section 4.1 (Definition 4.1). This algorithm corresponds to the Unpaired algorithm except from the fact that patients might receive a graft from deceased donors who arrive in the market at rate μ. It is clear from Definition 4.1 that the two algorithms are equivalent when there is no deceased donors (i.e. when $\mu = 0$). Hence the waiting times under the unpaired algorithm can be directly deduced from the waiting time under the unpaired with DDL algorithm – provided in Proposition 4.2 – fixing $\mu = 0$ (the proof of Proposition 4.2 is provided in Appendix C). The results of Proposition 2.6 directly follow.

B Other Proofs

B.1 Proof of Theorem 2.7

We deduce from the Theorem 1 and Lemma 4 in Ashlagi et al. (2019)[B.1] that

$$\lim_{p_H \to 0} p_H \mathbf{W}(Pairwise) = \begin{cases} \frac{\ln(2\lambda)}{np_H} & \text{if} \quad \lambda \geq \frac{1}{2} \\ \frac{\ln\left(\frac{1-\lambda}{2\lambda-1}\right)}{np_E} & \text{if} \quad \lambda < \frac{1}{2} \end{cases} \tag{B.1}$$

Using (B.1) and the average waiting time under unpaired derived in Proposition 2.6, we immediately get that

$$\lim_{p_H \to 0} \frac{\mathbf{W}(Pairwise)}{\mathbf{W}(Unpaired)} = \begin{cases} \frac{\ln(2\lambda)}{\ln(1+\lambda)} \cdot \frac{1}{p_H} & \text{if} \quad \lambda \geq \frac{1}{2} \\ \frac{\ln\left(\frac{1-\lambda}{1-2\lambda}\right)}{\ln(1+\lambda)} \cdot \frac{1}{p_E} & \text{if} \quad \lambda < \frac{1}{2} \end{cases}$$

B.2 Proof of Theorem 2.8 and Remark 2.9

The following result implies the statements in Theorem 2.8 as well as in Remark 2.9.

Proposition B.1. *Fix a matching algorithm ALG inducing a stochastic process with an invariant distribution. (1) We must have*

$$\lim_{p_H \to 0} \sup \frac{\mathbf{W}(Unpaired)}{\mathbf{W}(ALG)} \leq 2\frac{\ln(1+\lambda)}{\lambda}.$$

(2) Assume that $\lim_{p_H \to 0} p_H \mathbf{W}_E(ALG) = 0$. *Then,*

$$\lim_{p_H \to 0} \sup \frac{\mathbf{W}(Unpaired)}{\mathbf{W}(ALG)} \leq (1+\lambda)\frac{\ln(1+\lambda)}{\lambda}.$$

Let us denote the size of the pool by $\tilde{k}$. In the sequel, $\tilde{\mathbf{W}}(\text{ALG})$ is the random variable describing the average waiting time of an arriving patient. Note that a necessary condition for a patient to be

[B.1] Theorem 1 enunciates the average waiting time for hard-to-match patients while Lemma 4 proposes an upper bound for the waiting time of easy-to-match patients. From this Lemma we get that $\lim_{p_H \to 0} p_H \mathbf{W}_E(Pairwise) = 0$.

matched is that he is compatible with a donor in the pool upon arriving or, in case this does not occur, he is compatible with a donor in the future. In the former case, his waiting time is simply 0 while in the latter case, by the Poisson thinning property, the expected waiting time is $\frac{1}{np_T}$ if the patient is of type $T \in \{H, E\}$.[B.2] Hence, we obtain

$$\begin{aligned}
\mathbb{E}\left[\tilde{\mathbf{W}}(ALG)\,\middle|\,\tilde{k}=k\right] &\geq \lambda(1-p_H)^k\frac{1}{np_H} + (1-\lambda)\,(1-p_E)^k\frac{1}{np_E} \\
&\geq \lambda\left[(1-kp_H)\frac{1}{np_H}\right] + (1-\lambda)\left[(1-kp_E)\frac{1}{np_E}\right] \\
&= \frac{\lambda}{np_H} + \frac{1-\lambda}{np_E} - k\frac{1}{n}.
\end{aligned}$$

Thus, using the fact that, by Little's law, $\mathbf{W}(\mathrm{ALG}) = \frac{\mathbb{E}[\tilde{k}]}{n}$, we have

$$\begin{aligned}
\mathbf{W}(ALG) &= \mathbb{E}\left[\mathbb{E}\left[\tilde{\mathbf{W}}(ALG)\,\middle|\,\tilde{k}=k\right]\right] \\
&\geq \frac{\lambda}{np_H} + \frac{1-\lambda}{np_E} - \frac{1}{n}\mathbb{E}\left[\tilde{k}\right] \\
&= \frac{\lambda}{np_H} + \frac{1-\lambda}{np_E} - \mathbf{W}(ALG).
\end{aligned}$$

This gives us

$$\mathbf{W}(ALG) \geq \frac{\lambda}{2np_H} + \frac{1-\lambda}{2np_E}. \tag{B.2}$$

Now, we are in a position to prove the point (1) of the proposition. Indeed,

$$\lim_{p_H\to 0}\sup\frac{\mathbf{W}(Unpaired)}{\mathbf{W}(ALG)} = \frac{\lim_{p_H\to 0} p_H\mathbf{W}(Unpaired)}{\lim_{p_H\to 0}\inf p_H\mathbf{W}(ALG)} \leq \frac{\frac{\ln(1+\lambda)}{n}}{\frac{\lambda}{2n}} = 2\frac{\ln(1+\lambda)}{\lambda}.$$

where the inequality holds by Proposition 2.6 and together with Equation (B.2).

Now, to show point (2) of the proposition, further note that using a similar logic one can show that

$$\begin{aligned}
\mathbf{W}_H(ALG) &= \mathbb{E}\left[\mathbb{E}\left[\tilde{\mathbf{W}}_H(ALG)\,\middle|\,\tilde{k}=k\right]\right] \\
&\geq \frac{\lambda}{np_H} - \frac{1}{n}\mathbb{E}\left[\tilde{k}\right] \\
&= \frac{\lambda}{np_H} - \mathbf{W}(ALG) \\
&= \frac{\lambda}{np_H} - \lambda\mathbf{W}_H(ALG) - (1-\lambda)\mathbf{W}_E(ALG)
\end{aligned}$$

(where $\tilde{\mathbf{W}}_H(\mathrm{ALG})$ is the random variable describing the average waiting time of an arriving hard-

[B.2] The argument below works for any $p_E \leq 1$.

to-match patient) which yields

$$(1+\lambda)\mathbf{W}_H(ALG) \geq \frac{\lambda}{np_H} - (1-\lambda)\mathbf{W}_E(ALG)$$

Now, under the assumption that $\mathbf{W}_E(\text{ALG})p_H$ goes to 0 when p_H vanishes, obtain that

$$\lim_{p_H\to 0} p_H\mathbf{W}(ALG) = \lim_{p_H\to 0} p_H\mathbf{W}_H(ALG) \geq \frac{\lambda}{n(1+\lambda)}.$$

Hence,

$$\lim_{p_H\to 0} \sup \frac{\mathbf{W}(Unpaired)}{\mathbf{W}(ALG)} \leq \frac{\frac{\ln(1+\lambda)}{n}}{\frac{\lambda}{n(1+\lambda)}} = (1+\lambda)\frac{\ln(1+\lambda)}{\lambda}.$$

We now complete the proof of Theorem 2.8.

Completion of the proof of Theorem 2.8. Using the point (2) of Proposition B.1, we obtain that

$$\lim_{p_H\to 0} \sup \frac{\mathbf{W}(Unpaired)}{\mathbf{W}(Optimal)} \leq 2\frac{\ln(1+\lambda)}{\lambda}.$$

Indeed, assume that $\lim_{p_H\to 0} \sup \frac{\mathbf{W}(Unpaired)}{\mathbf{W}(Optimal)} > 2\frac{\ln(1+\lambda)}{\lambda}$. By definition, of OPT, there exists a sequence of matching algorithms $\{\text{ALG}_n\}_{n\geq 1}$ such that $\mathbf{W}(\text{ALG}_n) \to \mathbf{W}(\text{OPT})$. Then, this means that for n large enough, $\frac{\mathbf{W}(\text{Unpaired})}{\mathbf{W}(ALG_n)} > 2\frac{\ln(1+\lambda)}{\lambda}$, a contradiction with point (2) of the above proposition. □

B.3 Proof of Theorem 2.10

We know from the Proposition 1 in Ashlagi et al. (2019) that, when $p_E = 1$[B.3]

$$\lim_{p_H\to 0} p_H\mathbf{W}(Chain) = \frac{\ln\left(\frac{1}{1-\lambda}\right)}{n} \tag{B.3}$$

Using (B.3) and the waiting time under unpaired derived in Proposition 2.6, we immediately get that

$$\lim_{p_H\to 0} \frac{\mathbf{W}(Chain)}{\mathbf{W}(\text{Unpaired})} = \frac{\ln\left(\frac{1}{1-\lambda}\right)}{\ln(1+\lambda)} = -\frac{\ln(1-\lambda)}{\ln(1+\lambda)}$$

[B.3]Note that, when $p_E = 1$, an arriving easy-to-match patient is immediately matched by the bridge donor so that $\lim_{p_H\to 0} p_H\mathbf{W}(Chain) = \lim_{p_H\to 0} p_H\lambda\mathbf{W}_H(Chain)$.

C Proof of Proposition 4.2

This section is organized as follows. We first give the basic description of the Markov chain (over the number of patients of each type (k_H, k_E)) induced by the Unpaired algorithm as well as some basic definitions that will be used all along the proof (Section C.1). We define k_H^* as the number of hard-to-match patients which equalizes the transition rates from k_H^* to $k_H^* + 1$ and that from k_H^* to $k_H^* - 1$ assuming that no easy-to-match patients is in the pool. Then, the formal argument is presented and we prove that, at the invariant distribution, as p_H vanishes, the number of hard-to-match patients waiting in the system is highly concentrated around k_H^*. We split the proof into two blocks. In a first block, we show that the number of hard-to-match patients remaining in the system at the invariant distribution puts vanishing weight above k_H^* (Section C.2). In the second block (Section C.3), we prove the concentration result, i.e., show that this upper bound is actually tight. In order to prove the tightness result, we need to prove that with probability going to 1, the number of easy-to-match patients remaining in the pool is "small," i.e., we show that it is of order smaller than $1/p_H$.

Finally, we explain how we can use the bounds to obtain Proposition 2.6 (Section C.4).

C.1 Preliminaries

We make several preliminary remarks. First, under the unpaired exchange algorithm, one can easily check that the number of patients remaining in the system equals the number of donors remaining ($S_t = Z_t$ for all t). This is useful since we can simply focus on the evolution of the number of patients of each type remaining in the system.

We denote by Q the transition rate matrix over states $\mathbb{N} \times \mathbb{N}$. We will focus on the following transition rates:

$$
\begin{aligned}
Q([k_H, k_E], [k_H + 1, k_E]) &= n\lambda(1 - p_H)^{k_E + k_H}(1 - p_E)^{k_E}(1 - p_H)^{k_H} \\
Q([k_H, k_E], [k_H - 1, k_E]) &= n\Big\{\lambda[1 - (1 - p_H)^{k_E + k_H}][1 - (1 - p_H)^{k_H}] + \\
&\quad (1 - \lambda)[1 - (1 - p_E)^{k_E + k_H}][1 - (1 - p_H)^{k_H}]\Big\} \\
&\quad + \mu[1 - (1 - p_H)^{k_H}] \\
Q([k_H, k_E], [k_H, k_E + 1]) &= n(1 - \lambda)(1 - p_E)^{k_E + k_H}(1 - p_E)^{k_E}(1 - p_H)^{k_H} \\
Q([k_H, k_E], [k_H, k_E - 1]) &= n\Big\{\lambda[1 - (1 - p_H)^{k_E + k_H}](1 - p_H)^{k_H}[1 - (1 - p_E)^{k_E}] + \\
&\quad (1 - \lambda)[1 - (1 - p_E)^{k_E + k_H}](1 - p_H)^{k_H}[1 - (1 - p_E)^{k_E}]\Big\} \\
&\quad + \mu(1 - p_H)^{k_H}[1 - (1 - p_E)^{k_E}] \\
Q([k_H, k_E], [k_H + 1, k_E - 1]) &= n\lambda(1 - p_H)^{k_E + k_H}(1 - p_H)^{k_H}[1 - (1 - p_E)^{k_E}] \\
Q([k_H, k_E], [k_H - 1, k_E + 1]) &= n(1 - \lambda)(1 - p_E)^{k_E + k_H}[1 - (1 - p_H)^{k_H}]
\end{aligned}
$$

Let also first recall that the Global Balance Equations (GBE) are a set of equations that char-

acterize the invariant distribution of a Markov chain, when such a distribution exists. The above stochastic process is a Markov chain which has an invariant distribution as proved in Appendix C.5. In the sequel, we let π be this invariant distribution. The GBE can be stated as follows: for any subset $S \subset \mathbb{N} \times \mathbb{N}$, we must have:

$$\sum_{j \in S} \pi(j) \sum_{i \notin S} Q(j,i) = \sum_{i \notin S} \pi(i) \sum_{j \in S} Q(i,j) \tag{C.4}$$

Finally, let us define k_H^* as the real number ensuring

$$n\lambda(1-p_H)^{2k_H^*} = n\lambda\left[1-(1-p_H)^{k_H^*}\right]^2 + (n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H^*}\right]. \tag{C.5}$$

Simple algebra shows that

$$(n(1+\lambda)+\mu)(1-p_H)^{k_H^*} = n+\mu \tag{C.6}$$

and so

$$k_H^* = \frac{\ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right)}{-\ln(1-p_H)}.$$

In what follows, we sometimes use the notation π_H (resp. π_E) for the marginal of distribution π on the first (resp. second) dimension of the state space, i.e., $\pi_H(k_H) := \sum_{k_E=0}^{\infty} \pi(k_H, k_E)$.

C.2 Upper-bound result

In the sequel, we first prove the following result providing an upper-bound on the number of hard-to-match patients.

Proposition C.1. *Fix any $\delta > 0$,*

$$\pi_H\left\{k_H : \frac{k_H}{1/(-\ln(1-p_H))} \geq \ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right) + \delta\right\} \to 0$$

as p_H vanishes.

In order to show this, the following intermediary result is useful.

Lemma C.2. *The following must hold*

$$\frac{n\lambda\left[1-(1-p_H)^{k+1}\right]^2 + (n(1-\lambda)+\mu)\left[1-(1-p_H)^{k+1}\right]}{n\lambda(1-p_H)^{2k}} \geq \frac{n\lambda\left[1-(1-p_H)^{k+1}\right] + (n(1-\lambda)+\mu)\left[1-(1-p_H)^{k+1}\right]}{n\lambda(1-p_H)^{k}}$$

if $k \geq k_H^*$. The inequality holds in the other direction if $k \leq k_H^* - 1$.

Proof of Lemma C.2. Using simple algebra one can show that the inequality stated in Lemma C.2 is equivalent to

$$n+\mu \geq n\lambda(1-p_H)^{k+1} + (n+\mu)(1-p_H)^k.$$

If $k \geq k_H^*$, using (C.6), we have that $(n(1+\lambda)+\mu)(1-p_H)^k \leq n+\mu$. Since

$$(n(1+\lambda)+\mu)(1-p_H)^k \geq n\lambda(1-p_H)^{k+1} + (n+\mu)(1-p_H)^k,$$

we are getting the above inequality for $k \geq k_H^*$, as claimed. If $k \leq k_H^* - 1$, using (C.6) again, $(n(1+\lambda)+\mu)(1-p_H)^{k+1} \geq n+\mu$. Since

$$(n(1+\lambda)+\mu)(1-p_H)^{k+1} \leq n\lambda(1-p_H)^{k+1} + (n+\mu)(1-p_H)^k,$$

we are getting the reverse inequality for $k \leq k_H^* - 1$, as claimed. □

We can now show the following lemma.

Lemma C.3. *For any $\varepsilon > 0$, there exists a constant $\rho \in (0,1)$ such that, for any $p_H > 0$ and for any integer $k_H \geq k_H^*(1+\varepsilon)$*

$$\frac{\pi_H(k_H+1)}{\pi_H(k_H)} \leq \rho$$

Proof of Lemma C.3. Fix any $\varepsilon > 0$, and an arbitrary $k_H \geq k_H^*(1+\varepsilon)$ and let us consider the set $S = \{0,1,...,k_H\} \times \mathbb{N}$. The GBE (Equation (C.4)) gives us

$$\begin{aligned} & \sum_{k_E=0}^{\infty} \pi(k_H,k_E)\left[Q([k_H,k_E],[k_H+1,k_E]) + Q([k_H,k_E],[k_H+1,k_E-1])\right] \\ = & \sum_{k_E=0}^{\infty} \pi(k_H+1,k_E)\left[Q([k_H+1,k_E],[k_H,k_E]) + Q([k_H+1,k_E],[k_H,k_E+1])\right] \end{aligned}$$

Using the expressions of the transition rates, this can be rewritten as:

$$\begin{aligned} & \sum_{k_E=0}^{\infty} \pi(k_H,k_E)\left[n\lambda(1-p_H)^{k_E+k_H}(1-p_H)^{k_H}\right] \\ = & \sum_{k_E=0}^{\infty} \pi(k_H+1,k_E)\left[\begin{array}{c} n\lambda\left[1-(1-p_H)^{k_E+k_H+1}\right]\left[1-(1-p_H)^{k_H+1}\right] \\ +(n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H+1}\right] \end{array}\right]. \end{aligned}$$

Observing that the term in brackets in left-hand side is maximized at $k_E = 0$ while the term in brackets in the right-hand side is minimized at $k_E = 0$, we get

$$\begin{aligned} & \textstyle\sum_{k_E=0}^{\infty} \pi(k_H,k_E)\left[n\lambda(1-p_H)^{2k_H}\right] \\ \geq & \textstyle\sum_{k_E=0}^{\infty} \pi(k_H+1,k_E)\left[n\lambda\left[1-(1-p_H)^{k_H+1}\right]^2 + (n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H+1}\right]\right]. \end{aligned}$$

It implies that

$$\frac{\pi_H(k_H)}{\pi_H(k_H+1)} \geq \frac{n\lambda\left[1-(1-p_H)^{k_H+1}\right]^2 + (n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H+1}\right]}{n\lambda(1-p_H)^{2k_H}}. \tag{C.7}$$

where we recall that $\pi_H(k_H) = \sum_{k_E=0}^{\infty} \pi(k_H, k_E)$.

From the inequality above, we deduce:

$$\begin{aligned}
\frac{\pi_H(k_H+1)}{\pi_H(k_H)} &\leq \frac{n\lambda(1-p_H)^{2k_H}}{n\lambda\left[1-(1-p_H)^{k_H+1}\right]^2 + (n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H+1}\right]} \\
&\leq \frac{n\lambda(1-p_H)^{k_H}}{(n+\mu)\left[1-(1-p_H)^{k_H+1}\right]} \\
&\leq \frac{n\lambda(1-p_H)^{k_H^*(1+\varepsilon)}}{(n+\mu)\left[1-(1-p_H)^{k_H^*(1+\varepsilon)}\right]} \\
&= \frac{n\lambda\left(\frac{n+\mu}{(1+\lambda)n+\mu}\right)^{1+\varepsilon}}{(n+\mu)\left[1-\left(\frac{n+\mu}{(1+\lambda)n+\mu}\right)^{1+\varepsilon}\right]} := \rho \\
&< \frac{n\lambda\left(\frac{n+\mu}{(1+\lambda)n+\mu}\right)}{(n+\mu)\left[1-\left(\frac{n+\mu}{(1+\lambda)n+\mu}\right)\right]} = 1
\end{aligned}$$

where the first inequality comes from the Equation (C.7). The second inequality comes from Lemma C.2 and $k_H \geq (1+\epsilon)k_H^* \geq k_H^*$. The third inequality comes from the fact that $(1-p_H)^{k_H+1} \leq (1-p_H)^{k_H} \leq (1-p_H)^{(1+\epsilon)k_H^*}$. The first equality comes from equation (C.6).

Hence we obtain a positive constant, denoted by ρ, strictly smaller than one and independent of p_H, such that for all $k_H \geq k_H^*(1+\varepsilon)$: $\frac{\pi_H(k_H+1)}{\pi_H(k_H)} \leq \rho$. □

Using the result stated in Lemma C.3 we can show the following:

Lemma C.4. *For any $\varepsilon > 0$, there exists a constant $\rho \in (0,1)$ such that, for any $p_H > 0$ and for any integer $z > 0$:*

$$\pi_H\left\{k_H : k_H \geq k_H^*(1+\varepsilon) + z\right\} \leq \frac{\rho^z}{1-\rho}$$

Proof of Lemma C.4. We know from Lemma C.3 that for all $k_H \geq k_H^*(1+\varepsilon)$, $\pi_H(k_H+1) \leq \rho\pi_H(k_H)$ with $\rho \in (0,1)$.[C.4] Then by induction we get that:

$$\pi_H(k_H^*(1+\varepsilon)+i) \leq \rho^i \pi_H(k_H^*(1+\varepsilon)) \tag{C.8}$$

[C.4] In the sequel, for notational convenience, when we write $k_H^*(1+\varepsilon)$, we assume it is an integer. If this is not the case, the argument simply goes through replacing $k_H^*(1+\varepsilon)$ by $\lceil k_H^*(1+\varepsilon) \rceil$. Similar abuses of notations (where a real number has to be replaced by its floor or ceiling) will be used all along the proof.

It implies that

$$\begin{aligned}\pi_H\left\{k_H : k_H \geq k_H^*(1+\varepsilon)+z\right\} &= \sum_{i=z}^{+\infty} \pi_H(k_H^*(1+\varepsilon)+i) \\ &\leq \sum_{i=z}^{+\infty} \rho^i \pi_H(k_H^*(1+\varepsilon)) \\ &\leq \sum_{i=z}^{+\infty} \rho^i = \frac{\rho^z}{1-\rho}\end{aligned}$$

where the first inequality comes from the Equation (C.8), the second inequality immediately comes from the fact that $\pi_H(k_H^*(1+\varepsilon)) \leq 1$ and the last equality is obtained using the fact that $\rho < 1$. □

Completion of the proof of Proposition C.1. Fix any $\delta > 0$. We want to show that

$$\pi_H\left\{k_H : \frac{k_H}{1/(-\ln(1-p_H))} \geq \ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right) + \delta\right\} \to 0$$

as p_H vanishes. Let $z = 1/\sqrt{p_H}$ and fix $\varepsilon > 0$ and $p_H > 0$ small enough so that

$$\left[\ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right)\right](1+\varepsilon) + \sqrt{p_H} \leq \ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right) + \delta.$$

Hence, we obtain

$$\begin{aligned}&\pi_H\left\{k_H : \frac{k_H}{1/(-\ln(1-p_H))} \geq \ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right) + \delta\right\} \\ \leq\ &\pi_H\left\{k_H : \frac{k_H}{1/(-\ln(1-p_H))} \geq \left[\ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right)\right](1+\varepsilon) + \sqrt{p_H}\right\} \\ \leq\ &\pi_H\left\{k_H : k_H \geq k_H^*(1+\varepsilon) + \frac{\sqrt{p_H}}{-\ln(1-p_H)}\right\} \\ \leq\ &\frac{\rho^{\frac{\sqrt{p_H}}{-\ln(1-p_H)}}}{1-\rho} \to 0\end{aligned}$$

where the first inequality is ensured by our choice of ε and p_H while the last inequality is by Lemma C.4 and the convergence result holds since $\rho \in (0,1)$ (still by Lemma C.4) and since $-(1/p_H)\ln(1-p_H)$ goes to 1 as p_H vanishes and so $\frac{\sqrt{p_H}}{-\ln(1-p_H)}$ explodes as p_H vanishes. □

C.3 Lower-bound result

So far, we have provided an upper-bound on the number of hard-to-match patients in the pool. Recall that the upper-bound, i.e., k^*, is computed by equalizing the transition rates from k_H^* to k_H^*+1 and that from k_H^* to k_H^*-1 *assuming that no easy-to-match patients is in the pool.* Hence, mathematically, to show that this bound is tight it will be necessary to show that the number of

easy-to-match patients remaining in the pool is "small". At an intuitive level, if there was many easy-to-match patients remaining in the pool, then hard-to-match agents could be matched quickly and the upper-bound we obtained would be unlikely to be tight.

One issue to prove that the number of easy-to-match patients in the pool is small is the following. In the (small probability) event that an easy-to-match patient joins the pool, given the priority rule under Unpaired, he will have to wait for an arriving donor to be incompatible with all hard-to-match patients remaining in the system. Given that the number of hard-to-match patients in the system explodes, one may expect the conditional waiting time to be very long. However, we can use our upper-bound result (Proposition C.1) which bounds the rate at which the number of hard-to-match patients explodes to show that the number of easy-to-match patient remaining in the pool is small. This is what we show in Section C.3.1 below. Once this is proved we prove that our upper-bound is indeed tight in Section C.3.2.

C.3.1 An upper-bound on the number of easy-to-match patients

Lemma C.5. *There is an integer k_E^* such that, for any k and for any p_H small enough,*

$$\pi_E\{k_E : k_E \geq k_E^* + k\} \leq \frac{3+k}{(1-\hat{\rho})^2}\hat{\rho}^k$$

where $\hat{\rho} < 1$.

Proof of Lemma C.5. Fix an arbitrary $k_E \geq 0$ and let us consider the set $S = \mathbb{N}\times\{0,1,...,k_E\}$. Then, the GBE (Equation (C.4)) writes as:

$$\begin{aligned}
&\sum_{k_H=0}^{\infty} \pi(k_H,k_E)\left[Q([k_H,k_E],[k_H,k_E+1]) + Q([k_H,k_E],[k_H-1,k_E+1])\right] \\
= &\sum_{k_H=0}^{\infty} \pi(k_H,k_E+1)\left[Q([k_H,k_E+1],[k_H,k_E]) + Q([k_H,k_E+1],[k_H+1,k_E])\right]
\end{aligned}$$

Using the expressions of the transition rates, this can be rewritten as:

$$\begin{aligned}
&\sum_{k_H=0}^{\infty} \pi(k_H,k_E) n(1-\lambda) \left[\begin{array}{c} (1-p_E)^{k_E+k_H}(1-p_E)^{k_E}(1-p_H)^{k_H} \\ +(1-p_E)^{k_E+k_H}[1-(1-p_H)^{k_H}] \end{array}\right] \\
= &\sum_{k_H=0}^{\infty} \pi(k_H,k_E+1) \left[\begin{array}{c} n\lambda[1-(1-p_H)^{k_E+1+k_H}](1-p_H)^{k_H}[1-(1-p_E)^{k_E+1}] \\ +n(1-\lambda)[1-(1-p_E)^{k_E+1+k_H}](1-p_H)^{k_H}[1-(1-p_E)^{k_E+1}] \\ +\mu(1-p_H)^{k_H}[1-(1-p_E)^{k_E+1}] \\ +n\lambda(1-p_H)^{k_E+1+k_H}(1-p_H)^{k_H}[1-(1-p_E)^{k_E+1}] \end{array}\right]
\end{aligned}$$

This can be simplified to

$$\begin{aligned}
&\sum_{k_H=0}^{\infty} \pi(k_H, k_E) \left[n(1-\lambda)(1-p_E)^{k_E+k_H} \right] \\
\geq \quad &\sum_{k_H=0}^{\infty} \pi(k_H, k_E+1) \left[\begin{array}{c} n\lambda(1-p_H)^{k_H}[1-(1-p_E)^{k_E+1}]+ \\ n(1-\lambda)[1-(1-p_E)^{k_E+1+k_H}](1-p_H)^{k_H}[1-(1-p_E)^{k_E+1}] \\ +\mu(1-p_H)^{k_H}[1-(1-p_E)^{k_E+1}] \end{array} \right]
\end{aligned}$$

Observing that the expression in brackets in the left hand side is maximized at $k_H = 0$ and that the expression in brackets in right hand-side may be bounded below by disregarding the last two terms, we get that:

$$(1-\lambda)(1-p_E)^{k_E}\pi_E(k_E) \geq \sum_{k_H=0}^{\infty} \pi(k_H, k_E+1)\lambda(1-p_H)^{k_H}[1-(1-p_E)^{k_E+1}] \tag{C.9}$$

where π_E denotes the marginal of π on the number of easy-to-match patients in the pool, i.e., $\pi_E(k_E) = \sum_{k_H=0}^{\infty} \pi(k_H, k_E)$.

In the sequel, for each k_E and for any $\varepsilon > 0$, we define $S(k_E) := \{k_H : k_H \leq (1+\varepsilon)k_H^* + \ln(2)/p_H + k_E\}$ as well as $\pi_{S(k_E)}(k_E) := \sum_{k_H \in S(k_E)} \pi(k_H, k_E)$. We must have:

$$\begin{aligned}
\pi_E(k_E) - \pi_{S(k_E)}(k_E) &= \sum_{k_H \notin S(k_E)} \pi(k_H, k_E) \\
&\leq \sum_{k_E=0}^{\infty} \sum_{k_H \notin S(k_E)} \pi(k_H, k_E) = \sum_{k_H \notin S(k_E)} \pi_H(k_H).
\end{aligned}$$

Hence, by Lemma C.4, for any k_E,

$$\pi_E(k_E) - \pi_{S(k_E)}(k_E) \leq \frac{\rho^{\ln(2)/p_H + k_E}}{1-\rho} \tag{C.10}$$

where $\rho \in (0,1)$. Note that for each $k_H \in S(k_E)$, for p_H small enough, we have

$$(1-p_H)^{k_H} \geq (1-p_H)^{(1+\varepsilon)k_H^* + \ln(2)/p_H + k_E} \geq \frac{1}{3}\left(\frac{1}{1+\lambda}\right)^{1+\varepsilon}(1-p_H)^{k_E}$$

where the second inequality comes from the fact that as p_H vanishes, $(1-p_H)^{\ln(2)/p_H}$ converges to $\left(\frac{1}{e}\right)^{\ln(2)} = \frac{1}{2} > \frac{1}{3}$.

Given the above, let us rewrite Equation (C.9). The right-hand side can be lower-bounded by

$$\begin{aligned}
&\sum_{k_H \in S(k_E+1)}^{\infty} \pi(k_H, k_E+1)\lambda(1-p_H)^{k_H}[1-(1-p_E)^{k_E+1}] \\
\geq &\sum_{k_H \in S(k_E+1)}^{\infty} \pi(k_H, k_E+1)\lambda\frac{1}{3}\left(\frac{1}{1+\lambda}\right)^{1+\varepsilon}(1-p_H)^{k_E+1}[1-(1-p_E)^{k_E+1}]
\end{aligned}$$

Clearly, the left-hand side of Equation (C.9) can be rewritten as

$$(1-\lambda)(1-p_E)^{k_E}\pi_{S(k_E)}(k_E) + (1-\lambda)(1-p_E)^{k_E}\left(\pi_E(k_E) - \pi_{S(k_E)}(k_E)\right).$$

Hence, Equation (C.9) can be rewritten as

$$\begin{aligned}
&(1-\lambda)(1-p_E)^{k_E}\pi_{S(k_E)}(k_E) + (1-\lambda)(1-p_E)^{k_E}\left(\pi_E(k_E) - \pi_{S(k_E)}(k_E)\right) \\
\geq &\frac{1}{3}\left(\frac{1}{1+\lambda}\right)^{1+\varepsilon}\lambda(1-p_H)^{k_E+1}[1-(1-p_E)^{k_E+1}]\pi_{S(k_E+1)}(k_E+1)
\end{aligned}$$

Let us denote

$$\bar{\rho}(k_E) := \frac{(1-\lambda)(1-p_E)^{k_E}}{\frac{1}{3}\left(\frac{1}{1+\lambda}\right)^{1+\varepsilon}\lambda(1-p_H)^{k_E+1}[1-(1-p_E)^{k_E+1}]}.$$

The above inequality can be written as

$$\bar{\rho}(k_E)\pi_{S(k_E)}(k_E) + \bar{\rho}(k_E)\left(\pi_E(k_E) - \pi_{S(k_E)}(k_E)\right) \geq \pi_{S(k_E+1)}(k_E+1).$$

Since

$$\bar{\rho}(k_E) \leq \frac{1-\lambda}{\frac{\lambda}{3}\left(\frac{1}{1+\lambda}\right)^{1+\varepsilon}p_E(1-p_H)}\left(\frac{1-p_E}{1-p_H}\right)^{k_E},$$

it must be that for any p_H small enough, there is an integer k_E^* large enough (which does not depend on p_H) so that $\sup_{k_E \geq k_E^*}\bar{\rho}(k_E) =: \rho_* < 1$.[C.5] Hence, from the definition of ρ_* and from Equation (C.10) we obtain

$$\rho_*\pi_{S(k_E)}(k_E) + \rho_*\frac{\rho^{\ln(2)/p_H + k_E}}{1-\rho} \geq \pi_{S(k_E+1)}(k_E+1)$$

for any $k_E \geq k_E^*$. Clearly, for p_H small enough, $\frac{\rho^{\ln(2)/p_H}}{1-\rho} < 1$, and since $\rho_* < 1$, we have

$$\rho_*\pi_{S(k_E)}(k_E) + \rho^{k_E} \geq \pi_{S(k_E+1)}(k_E+1)$$

[C.5] k_E^* is simply defined as the smallest integer k_E ensuring $\frac{1-\lambda}{\frac{\lambda}{2}\left(\frac{1}{1+\lambda}\right)^{1+\varepsilon}p_E(1-p)}\left(\frac{1-p_E}{1-p}\right)^{k_E} < 1$ where p is an arbitrary number in $(0, p_E)$. We then require that p_H is smaller than p.

holds for any $k_E \geq k_E^*$. Now, setting $\hat{\rho} := \max(\rho_*, \rho) < 1$, we obtain

$$\hat{\rho}\pi_{S(k_E)}(k_E) + \hat{\rho}^{k_E} \geq \pi_{S(k_E+1)}(k_E+1).$$

Now, proceeding inductively, we can rewrite for any $k_E \geq k_E^*$

$$\hat{\rho}^i \pi_{S(k_E^*)}(k_E^*) + i\hat{\rho}^{k_E^*+i} \geq \pi_{S(k_E^*+i)}(k_E^*+i) \tag{C.11}$$

Now, we have

$$\begin{aligned}
\sum_{k_E \geq k_E^*+k}^{\infty} \pi_E(k_E) &= \sum_{i=k}^{\infty} \pi_E(k_E^*+i) \\
&= \sum_{i=k}^{\infty} \pi_{S(k_E^*+i)}(k_E^*+i) + \sum_{i=k}^{\infty} [\pi_E(k_E^*+i) - \pi_{S(k_E^*+i)}(k_E^*+i)] \\
&\leq \sum_{i=k}^{\infty} \pi_{S(k_E^*+i)}(k_E^*+i) + \sum_{i=k}^{\infty} \frac{\rho^{\frac{\ln(2)}{p_H}+k_E^*+i}}{1-\rho} \\
&\leq \sum_{i=k}^{\infty} \pi_{S(k_E^*+i)}(k_E^*+i) + \sum_{i=k}^{\infty} \frac{\hat{\rho}^i}{1-\hat{\rho}} \\
&\leq \sum_{i=k}^{\infty} \hat{\rho}^i \pi_{S(k_E^*)}(k_E^*) + \hat{\rho}^{k_E^*} \sum_{i=k}^{\infty} i\hat{\rho}^i + \frac{\hat{\rho}^k}{(1-\hat{\rho})^2} \\
&\leq \sum_{i=k}^{\infty} \hat{\rho}^i + \sum_{i=k}^{\infty} i\hat{\rho}^i + \frac{\hat{\rho}^k}{(1-\hat{\rho})^2} \\
&\leq \frac{\hat{\rho}^k}{1-\hat{\rho}} \left(1 + \frac{1+k}{1-\hat{\rho}} + \frac{1}{1-\hat{\rho}}\right) \\
&\leq \frac{\hat{\rho}^k}{1-\hat{\rho}} \frac{3+k}{1-\hat{\rho}}
\end{aligned} \tag{C.12}$$

where the first inequality uses (C.10). The second uses $\hat{\rho} \geq \rho$. The third uses (C.11). The fourth uses the fact that both $\pi_{S(k_E^*)}(k_E^*)$ and $\hat{\rho}$ are smaller than one. The penultimate inequality uses the following fact:

$$\sum_{i=0}^{n} i\hat{\rho}^i = \frac{\hat{\rho} - (n+1)\hat{\rho}^{n+1} + n\hat{\rho}^{n+2}}{(1-\hat{\rho})^2} \tag{C.13}$$

and the fact that

$$\begin{aligned}\sum_{i=k}^{+\infty} i\hat{\rho}^i &= \sum_{i=0}^{+\infty} i\hat{\rho}^i - \sum_{i=0}^{k-1} i\hat{\rho}^i \\ &= \frac{\hat{\rho}}{(1-\hat{\rho})^2} - \frac{\hat{\rho} - k\hat{\rho}^k + (k-1)\hat{\rho}^{k+1}}{(1-\hat{\rho})^2} \\ &= \frac{k\hat{\rho}^k + \hat{\rho}^{k+1} - k\hat{\rho}^{k+1}}{(1-\hat{\rho})^2} \qquad \text{(C.14)} \\ &\leq \frac{k\hat{\rho}^k + \hat{\rho}^k}{(1-\hat{\rho})^2} \qquad \text{(C.15)}\end{aligned}$$

where the second equality comes from equation (C.13) and $\hat{\rho} \in (0,1)$ is used for this equality as well as for the inequality. □

C.3.2 Completing of the lower-bound result for hard-to-match patients

Proposition C.6. *Fix any $\delta > 0$, $\pi_H \left\{ k_H : \frac{k_H}{1/(-\ln(1-p_H))} \leq \ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right) - \delta \right\} \to 0$ as p_H vanishes.*

In the sequel, we fix k_E^* as defined in Lemma C.5 and define $S := \{k_E : k_E \leq k_E^* + 1/\sqrt{p_H}\}$ and, as before, $\pi_S(k_H) = \sum_{k_E \in S} \pi(k_H, k_E)$. We first prove the following lemma.

Lemma C.7. *For any $\varepsilon \in (0,1)$ and any p_H small enough, there exist constants $\hat{\rho} \in (0,1)$ such that, for any integer $z > 0$:*

$$\pi_S(k_H^*(1-\varepsilon) - z) \leq \tilde{\rho}^z \pi_S(k_H^*(1-\varepsilon)) + \phi(p_H)\tilde{\rho}\frac{1-\tilde{\rho}^z}{1-\tilde{\rho}}$$

where $\phi(p_H) := \left(3 + \frac{1}{\sqrt{p_H}}\right)\frac{\hat{\rho}^{1/\sqrt{p_H}}}{(1-\hat{\rho})^2}$.

Proof of Lemma C.7. Let us recall that using the GBE we obtained

$$\begin{aligned}&\sum_{k_E=0}^{\infty} \pi(k_H, k_E) n\lambda (1-p_H)^{k_E+k_H}(1-p_H)^{k_H} \\ = &\sum_{k_E \in S} \pi(k_H+1, k_E) \left[\begin{array}{c} n\lambda\left[1-(1-p_H)^{k_E+k_H+1}\right]\left[1-(1-p_H)^{k_H+1}\right] \\ +(\mu + n(1-\lambda))\left[1-(1-p_H)^{k_H+1}\right] \end{array} \right] \\ &+ \sum_{k_E \notin S} \pi(k_H+1, k_E) \left[\begin{array}{c} n\lambda\left[1-(1-p_H)^{k_E+k_H+1}\right]\left[1-(1-p_H)^{k_H+1}\right] \\ +(\mu + n(1-\lambda))\left[1-(1-p_H)^{k_H+1}\right] \end{array} \right].\end{aligned}$$

Note that for p_H small enough, $(1-p_H)^{k_E} \geq (1-p_H)^{k_E^*}.c^{\sqrt{p_H}}$ with $c \in (0,1)$ whenever $k_E \in S =$

$\{k_E : k_E \leq k_E^* + 1/\sqrt{p_H}\}$.[C.6] This observation allows us to lower-bound the left-hand side of the above displayed equation. Thus, by upper-bounding the right-hand side as well (simply using the facts that for $k_E \in S$, $k_E \leq k_E^* + 1/\sqrt{p_H}$ and $1-(1-p_H)^{k_E+k_H+1} \leq 1$), we get

$$\begin{aligned}
& \pi_S(k_H)\left[n\lambda(1-p_H)^{k_E^*}.c^{\sqrt{p_H}}(1-p_H)^{2k_H}\right] \\
\leq\ & \pi_S(k_H+1)\left[\begin{array}{c} n\lambda\left[1-(1-p_H)^{k_E^*+1/\sqrt{p_H}+k_H+1}\right]\left[1-(1-p_H)^{k_H+1}\right] \\ +((1-\lambda)n+\mu)\left[1-(1-p_H)^{k_H+1}\right] \end{array}\right] \\
& +(\pi_H(k_H+1)-\pi_S(k_H+1))\left[\begin{array}{c} n\lambda\left[1-(1-p_H)^{k_E^*+1/\sqrt{p_H}+k_H+1}\right]\left[1-(1-p_H)^{k_H+1}\right] \\ +((1-\lambda)n+\mu)\left[1-(1-p_H)^{k_H+1}\right] \end{array}\right].
\end{aligned}$$

In the sequel we fix $\varepsilon \in (0,1)$ and consider $k_H \leq k_H^*(1-\varepsilon)$. We will be using the following two inequalities which hold for p_H small enough : $k_E^*+1/\sqrt{p_H}+k_H+1 \leq k_H^*$ and $k_H+1 \leq k_H^*$. Since the second inequality holds if the first one holds, we just provide the argument for the first one. Since k_H^* goes to $\left[\ln\left[n(1+\lambda)+\mu\right]-\ln\left(n+\mu\right)\right]/p_H$ as p_H vanishes, we must have $\varepsilon k_H^* \geq k_E^*+1/\sqrt{p_H}+1$ for p_H small enough (recall by Lemma C.5 that k_E^* is a constant which does not depend on p_H). Since we assumed that $k_H \leq k_H^*(1-\varepsilon)$, we must have $k_E^*+1/\sqrt{p_H}+k_H+1 \leq k_E^*+1/\sqrt{p_H}+k_H^*(1-\varepsilon)+1$ which is thus smaller than k_H^* for p_H small enough, as claimed. These two inequalities allow us to further bound the right-hand side of the above displayed equation:

$$\begin{aligned}
& \pi_S(k_H)\left[n\lambda(1-p_H)^{k_E^*}.c^{\sqrt{p_H}}(1-p_H)^{2k_H}\right] \\
\leq\ & \pi_S(k_H+1)\left[n\lambda\left[1-(1-p_H)^{k_H^*}\right]^2+(n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H^*}\right]\right] \\
& +(\pi_H(k_H+1)-\pi_S(k_H+1))\left[n\lambda\left[1-(1-p_H)^{k_H^*}\right]^2+(n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H^*}\right]\right].
\end{aligned}$$

Then, using the fact that $k_H \leq k_H^*(1-\varepsilon)$, we can lower-bound the left-hand side of the above equation to get

$$\begin{aligned}
& \pi_S(k_H)\left[n\lambda(1-p_H)^{k_E^*}.c^{\sqrt{p_H}}(1-p_H)^{2k_H^*(1-\varepsilon)}\right] \\
\leq\ & \pi_S(k_H+1)\left[n\lambda\left[1-(1-p_H)^{k_H^*}\right]^2+(n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H^*}\right]\right] \\
& +(\pi_H(k_H+1)-\pi_S(k_H+1))\left[n\lambda\left[1-(1-p_H)^{k_H^*}\right]^2+(n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H^*}\right]\right].
\end{aligned}$$

[C.6] Simply note that, for any $k_E \in S$,

$$\begin{aligned}
(1-p_H)^{k_E} &\geq (1-p_H)^{k_E^*}(1-p_H)^{1/\sqrt{p_H}} \\
&= (1-p_H)^{k_E^*}\left((1-p_H)^{1/p_H}\right)^{\sqrt{p_H}}.
\end{aligned}$$

Since $(1-p_H)^{1/p_H}$ converges from below to $1/e$, we can ensure that for p_H small enough, $(1-p_H)^{k_E} \geq (1-p_H)^{k_E^*}\left(\frac{0.9}{e}\right)^{\sqrt{p_H}}$.

This may be rewritten as:

$$\pi_S(k_H) \leq \tilde{\rho}_1(p_H)\pi_S(k_H+1) + \tilde{\rho}_1(p_H)(\pi_H(k_H+1) - \pi_S(k_H+1)) \tag{C.16}$$

with

$$\tilde{\rho}_1(p_H) := \left(\frac{(1-p_H)^{2\varepsilon k_H^*}}{(1-p_H)^{k_E^* c\sqrt{p_H}}}\right) \frac{n\lambda\left[1-(1-p_H)^{k_H^*}\right]^2 + (n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H^*}\right]}{n\lambda(1-p_H)^{2k_H^*}}.$$

Now, we claim that, for p_H small enough, $\tilde{\rho}_1(p_H) \leq \tilde{\rho}_1$ where $\tilde{\rho}_1 < 1$ does not depend on p_H. Indeed,

$$\begin{aligned}\tilde{\rho}_1(p_H) &= \left(\frac{(1-p_H)^{2\varepsilon k_H^*}}{(1-p_H)^{k_E^* c\sqrt{p_H}}}\right) \frac{n\lambda\left[1-(1-p_H)^{k_H^*}\right]^2 + (n(1-\lambda)+\mu)\left[1-(1-p_H)^{k_H^*}\right]}{n\lambda(1-p_H)^{2k_H^*}} \\ &= \frac{(1-p_H)^{2\varepsilon k_H^*}}{(1-p_H)^{k_E^* c\sqrt{p_H}}} = \frac{1}{(1-p_H)^{k_E^*.c\sqrt{p_H}}}\left(\frac{n+\mu}{n(1+\lambda)+\mu}\right)^{2\varepsilon}\end{aligned}$$

where the penultimate equality holds by Equation (C.5) while the last one holds by Equation (C.6). Now, $1/\left((1-p_H)^{k_E^* c\sqrt{p_H}}\right)$ converges from above to 1 as p_H vanishes (recall by Lemma C.5 that k_E^* is a constant which does not depend on p_H), thus, the above term converges from above to $\left(\frac{n+\mu}{n(1+\lambda)+\mu}\right)^{2\varepsilon} < 1$ as p_H vanishes. So, for p_H small enough, $\tilde{\rho}_1(p_H) \leq \tilde{\rho}_1$ where $\tilde{\rho}_1 < 1$ does not depend on p_H.

Thus, from inequality (C.16) we get that for p_H small enough,

$$\pi_S(k_H) \leq \tilde{\rho}_1\pi_S(k_H+1) + \tilde{\rho}_1(\pi_H(k_H+1) - \pi_S(k_H+1))$$

Now, by Lemma C.5, for any k_H,

$$\begin{aligned}\pi_H(k_H+1) - \pi_S(k_H+1) &= \sum_{k_E \notin S} \pi(k_H+1, k_E) \\ &\leq \sum_{k_H=0}^{\infty} \sum_{k_E \notin S} \pi(k_H, k_E) \\ &\leq \frac{3+1/\sqrt{p_H}}{1-\hat{\rho}}\left(\frac{\hat{\rho}^{1/\sqrt{p_H}}}{1-\hat{\rho}}\right)\end{aligned}$$

with $\hat{\rho} \in (0,1)$. Hence, for $k_H \leq k_H^*(1-\varepsilon)$, we obtain

$$\begin{aligned}\pi_S(k_H) &\leq \tilde{\rho}_1\left[\pi_S(k_H+1) + \frac{3+1/\sqrt{p_H}}{1-\hat{\rho}}\left(\frac{\hat{\rho}^{1/\sqrt{p_H}}}{1-\hat{\rho}}\right)\right] \\ &\leq \tilde{\rho}\left[\pi_S(k_H+1) + \phi(p_H)\right]\end{aligned}$$

where $\tilde{\rho} := \max(\hat{\rho}, \tilde{\rho}_1)$ and

$$\phi(p_H) := \frac{3+1/\sqrt{p_H}}{1-\tilde{\rho}}\frac{\tilde{\rho}^{1/\sqrt{p_H}}}{1-\tilde{\rho}}.$$

An inductive argument yields

$$\pi_S(k_H^*(1-\varepsilon)-i) \leq \tilde{\rho}^i \pi_S(k_H^*(1-\varepsilon)) + \phi(p_H)\tilde{\rho}\frac{1-\tilde{\rho}^i}{1-\tilde{\rho}}. \tag{C.17}$$

□

Completion of the proof of Proposition C.6. We fix $\delta > 0$ and claim that

$$\pi_H \left\{ k_H : \frac{k_H}{1/(-\ln(1-p_H))} \leq \ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right) - \delta \right\} \to 0$$

which is equivalent to showing that

$$\pi_H \left\{ k_H : k_H \leq \frac{1}{p_H}\left[\ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right)\right] - \frac{1}{p_H}\delta \right\} \to 0.$$

Pick p_H and ε small enough so that

$$\frac{1}{p_H}\left[\ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right)\right] - \frac{1}{p_H}\delta \leq k_H^*(1-\varepsilon) - 1/\sqrt{p_H}.$$

Clearly, for our purpose, it is enough to show that

$$\pi_H\{k_H \leq k_H^*(1-\varepsilon) - 1/\sqrt{p_H}\} \to 0$$

as p_H vanishes. In order to see this, observe that

$$\begin{aligned}
\pi_H\{k_H &\leq k_H^*(1-\varepsilon) - 1/\sqrt{p_H}\} \\
&= \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} \pi_H(k_H^*(1-\varepsilon) - i) \\
&= \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} [\pi_H(k_H^*(1-\varepsilon) - i) - \pi_S(k_H^*(1-\varepsilon) - i)] + \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} \pi_S(k_H^*(1-\varepsilon) - i) \\
&\leq \sum_{k_H=0}^{\infty} \sum_{k_E \notin S} \pi(k_H, k_E) + \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} \pi_S(k_H^*(1-\varepsilon) - i) \\
&\leq \phi(p_H) + \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} \pi_S(k_H^*(1-\varepsilon) - i) \\
&\leq \phi(p_H) + \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} \tilde{\rho}^i \pi_S(k_H^*(1-\varepsilon)) + \sum_{i=0}^{k_H^*(1-\varepsilon)} \phi(p_H)\tilde{\rho}\frac{1-\tilde{\rho}^i}{1-\tilde{\rho}} \\
&\leq \phi(p_H) + \sum_{i=1/\sqrt{p_H}}^{\infty} \tilde{\rho}^i + \sum_{i=0}^{k_H^*(1-\varepsilon)} \phi(p_H)\tilde{\rho}\frac{1-\tilde{\rho}^i}{1-\tilde{\rho}} \\
&\leq \phi(p_H) + \frac{\tilde{\rho}^{1/\sqrt{p_H}}}{1-\tilde{\rho}} + k_H^*(1-\varepsilon)\phi(p_H)\frac{\tilde{\rho}}{1-\tilde{\rho}} \to 0
\end{aligned}$$

where the second inequality comes from Lemma C.5 and the third from the Lemma C.7. In order to prove the convergence result let us first note that, as p_H vanishes, $\phi(p_H) \to \frac{\tilde{\rho}^{\frac{1}{\sqrt{p_H}}}}{(1-\tilde{\rho})^2\sqrt{p_H}}$ and, since k_H^* is of order $1/p_H$, $k_H^*\phi(p_H) \to \frac{c\tilde{\rho}^{\frac{1}{\sqrt{p_H}}}}{(1-\tilde{\rho})^2 p_H\sqrt{p_H}} = \frac{c\tilde{\rho}^{\frac{1}{\sqrt{p_H}}}}{(1-\tilde{\rho})^2(\sqrt{p_H})^3}$, with c a positive constant. In order to prove that those two terms tend to zero, it is sufficient to prove that, for any $\alpha \in (0,1)$ and any finite integer $n \geq 1$, we have $x^n\alpha^x \to 0$ as $x \to +\infty$. We prove this fact by repeatedly applying L'Hospital's Rule n times:

$$\lim_{x\to+\infty} \frac{x^n}{\left(\frac{1}{\alpha}\right)^x} = \lim_{x\to+\infty} \frac{\prod_{i=0}^{n-1}(n-i)}{\left[\ln\left(\frac{1}{\alpha}\right)\right]^n \left(\frac{1}{\alpha}\right)^x} = 0 \tag{C.18}$$

since the product is a finite number and $\alpha \in (0,1)$. □

C.4 Completion of the proof of Proposition 4.2

Here, we complete the proof of Proposition 4.2 by providing the expressions of $p_H\mathbf{W}_H(Unpaired)$ and $p_H\mathbf{W}_E(Unpaired)$ when p_H vanishes.

Waiting time of hard-to-match patients. Given p_H, let us denote by $K_H(p_H)$ the random variable corresponding to the number of hard-to-match patients at the invariant distribution π_H.

Proposition C.1 and C.6 imply that, as p_H vanishes, $K_H(p_H)p_H$ converges in probability to constant $\ln[n(1+\lambda)+\mu] - \ln(n+\mu)$. One can show that $\{K_H(p_H)p_H\}$ is uniformly integrable, i.e., for a given $\delta > 0$, there exists $M < \infty$ large enough such that $\mathbb{E}[K_H(p_H)p_H\mathbf{1}\{K_H(p_H)p_H \geq M\}] \leq \delta$ for all the random variables in our collection $\{K_H(p_H)\}$ where $\mathbf{1}\{K_H(p_H)p_H \geq M\}$ stands for the indicator function equal to 1 if and only if $K_H(p_H)p_H \geq M$. This implies (see for instance Williams, 1991) that

$$\lim_{p_H \to 0} \mathbb{E}[K_H(p_H)p_H] = \ln[n(1+\lambda)+\mu] - \ln(n+\mu). \tag{C.19}$$

To see that the collection $\{K_H(p_H)p_H\}$ is uniformly integrable, fix $\delta, \varepsilon > 0$ and let $M := k_H^*(1+\varepsilon)p_H + z \leq [\ln[n(1+\lambda)+\mu] - \ln(n+\mu)](1+\varepsilon) + z$ where $z \geq 1$ is an integer yet to be specified.[C.7] Note that

$$\begin{aligned}
\mathbb{E}[K_H(p_H)p_H\mathbf{1}\{K_H(p_H)p_H \geq M\}] &= p_H \sum_{i=z/p_H}^{\infty} \pi_H\left(k_H^*(1+\varepsilon)+i\right)\left[k_H^*(1+\varepsilon)+i\right] \\
&\leq p_H \sum_{i=z}^{\infty} \pi_H\left(k_H^*(1+\varepsilon)+i\right)\left[k_H^*(1+\varepsilon)+i\right] \\
&\leq \sum_{i=z}^{\infty} \pi_H\left(k_H^*(1+\varepsilon)+i\right)\left[\ln[n(1+\lambda)+\mu] - \ln(n+\mu)\right](1+\varepsilon) \\
&\quad + p_H \sum_{i=z}^{\infty} \pi_H\left(k_H^*(1+\varepsilon)+i\right) i \\
&\leq \left[\ln[n(1+\lambda)+\mu] - \ln(n+\mu)\right](1+\varepsilon) \sum_{i=z}^{\infty} \rho^i \pi_H\left(k_H^*(1+\varepsilon)\right) \\
&\quad + p_H \sum_{i=z}^{\infty} \rho^i \pi_H\left(k_H^*(1+\varepsilon)\right) i \\
&\leq \left[\ln[n(1+\lambda)+\mu] - \ln(n+\mu)\right](1+\varepsilon) \sum_{i=z}^{\infty} \rho^i + \sum_{i=z}^{\infty} \rho^i i \\
&\leq \left[\ln[n(1+\lambda)+\mu] - \ln(n+\mu)\right](1+\varepsilon)\frac{\rho^z}{1-\rho} + \frac{z\rho^z + \rho^z}{(1-\rho)^2}
\end{aligned}$$

where the third inequality uses (C.8) while the last inequality uses (C.15). Since $\rho \in (0,1)$, the above term is smaller than δ when z is large enough and so when M is large enough.

Hence, as stated in (C.19) we obtain

$$\lim_{p_H \to 0} \mathbb{E}[p_H K_H(p_H)] = \ln[n(1+\lambda)+\mu] - \ln(n+\mu).$$

[C.7] For the inequality: $k_H^*(1+\varepsilon)p_H + z \leq [\ln[n(1+\lambda)+\mu] - \ln(n+\mu)](1+\varepsilon) + z$ we simply used the fact that $p_H \leq -\ln(1-p_H)$ together with the definition of k_H^*.

Now, by Little's law we get that:

$$\begin{aligned}\lim_{p_H\to 0} p_H \mathbf{W}_H(Unpaired) &= \lim_{p_H\to 0} p_H \frac{\mathbb{E}\left[K_H(p_H)\right]}{\lambda.n} \\ &= \lim_{p_H\to 0} \frac{\mathbb{E}\left[p_H K_H(p_H)\right]}{\lambda.n} = \frac{\ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right)}{\lambda.n}.\end{aligned}$$

Waiting time of easy-to-match patients. First, we prove that, at the invariant distribution, as p_H vanishes, the number of easy-to-match patients is concentrated around the constant k_E^* as defined in Lemma C.5. This indeed comes from that lemma and parallels Proposition C.1 but now for easy-to-match patients.

Proposition C.8. *Fix any $\delta > 0$,*

$$\pi_E\left\{k_E : \frac{k_E}{1/(-\ln(1-p_H))} \geq \frac{k_E^*}{1/(-\ln(1-p_H))} + \delta\right\} \to 0$$

as p_H vanishes.

Proof. Fix any $\delta > 0$ and let $p_H > 0$ be small enough so that $\sqrt{p_H} \leq \delta$. We know that

$$\begin{aligned}& \pi_E\left\{k_E : \frac{k_E}{1/(-\ln(1-p_H))} \geq \frac{k_E^*}{1/(-\ln(1-p_H))} + \delta\right\} \\ \leq\ & \pi_E\left\{k_E : \frac{k_E}{1/(-\ln(1-p_H))} \geq \frac{k_E^*}{1/(-\ln(1-p_H))} + \sqrt{p_H}\right\} \\ =\ & \pi_E\left\{k_E : k_E \geq k_E^* + \frac{\sqrt{p_H}}{-\ln(1-p_H)}\right\} \\ \leq\ & \frac{3 + \frac{\sqrt{p_H}}{-\ln(1-p_H)}}{(1-\hat{\rho})^2}\hat{\rho}^{\frac{\sqrt{p_H}}{-\ln(1-p_H)}} \to 0 \text{ as } p_H \text{ vanishes}\end{aligned}$$

where the first inequality is ensured by our choice of p_H while the last inequality is by Lemma C.5 and the convergence result holds since $\hat{\rho} \in (0,1)$ (again by Lemma C.5) and since $-\ln(1-p_H)/p_H$ goes to 1 as p_H vanishes and so $\frac{\sqrt{p_H}}{-\ln(1-p_H)}$ explodes as p_H vanishes. □

Now, given p_H, let us denote by $K_E(p_H)$ the random variable corresponding to the number of easy-to-match patients at the invariant distribution π_E. As for the case of hard-to-match patients, we want to show that the collection $\{K_E(p_H)p_H\}$ is uniformly integrable. To see this, fix $\delta > 0$ and

let $M := k_E^* p_H + z$ where $z \geq 1$ is an integer yet to be specified. Then, note that:

$$\begin{aligned}
\mathbb{E}\left[K_E(p_H)p_H \mathbf{1}\{K_E(p_H)p_H \geq M\}\right] &= p_H \sum_{i=z/p_H}^{\infty} \pi_E\left(k_E^* + i\right)\left[k_E^* + i\right] \\
&\leq p_H \sum_{i=z}^{\infty} \pi_E\left(k_E^* + i\right)\left[k_E^* + i\right] \\
&\leq k_E^* \sum_{i=z}^{\infty} \pi_E\left(k_E^* + i\right) + \sum_{i=z}^{\infty} i\pi_E\left(k_E^* + i\right) \\
&\leq k_E^* \left(\frac{\hat{\rho}^z}{1-\hat{\rho}}\right)\frac{3+z}{1-\hat{\rho}} + \sum_{i=z}^{\infty} i\pi_E\left(k_E^* + i\right) \qquad \text{(C.20)}
\end{aligned}$$

where the last inequality comes from (C.12). Moreover, we have:

$$\begin{aligned}
\sum_{i=z}^{\infty} i\pi_E\left(k_E^* + i\right) &= \sum_{i=z}^{\infty} i\pi_{S(k_E^*+i)}(k_E^* + i) + \sum_{i=z}^{\infty} i[\pi_E(k_E^* + i) - \pi_{S(k_E^*+i)}(k_E^* + i)] \\
&\leq \sum_{i=z}^{\infty} i\pi_{S(k_E^*+i)}(k_E^* + i) + \sum_{i=z}^{\infty} i\frac{\rho^{\frac{\ln(2)}{p_H}+k_E^*+i}}{1-\rho} \\
&\leq \sum_{i=z}^{\infty} i\pi_{S(k_E^*+i)}(k_E^* + i) + \hat{\rho}^{k_E^*}\sum_{i=z}^{\infty}\frac{i\hat{\rho}^i}{1-\hat{\rho}} \\
&\leq \sum_{i=z}^{\infty} i\hat{\rho}^i\pi_{S(k_E^*)}(k_E^*) + \hat{\rho}^{k_E^*}\sum_{i=z}^{\infty} i^2\hat{\rho}^i + \hat{\rho}^{k_E^*}\sum_{i=z}^{\infty}\frac{i\hat{\rho}^i}{1-\hat{\rho}} \\
&\leq \sum_{i=z}^{\infty} i\hat{\rho}^i + \sum_{i=z}^{\infty} i^2\hat{\rho}^i + \sum_{i=z}^{\infty}\frac{i\hat{\rho}^i}{1-\hat{\rho}} \\
&\leq \frac{2}{1-\hat{\rho}}\sum_{i=z}^{\infty} i\hat{\rho}^i + \sum_{i=z}^{\infty} i^2\hat{\rho}^i \\
&\leq \frac{2(1+z)\hat{\rho}^z}{(1-\hat{\rho})^3} + \sum_{i=z}^{\infty} i^2\hat{\rho}^i \qquad \text{(C.21)}
\end{aligned}$$

where the first inequality uses (C.10). The second uses $\rho \leq \hat{\rho} \leq 1$. The third uses (C.11). The fourth and the fifth use the fact that $\pi_{S(k_E^*)}(k_E^*)$ and $\hat{\rho}$ are smaller than one. The last inequality

comes from (C.15). Finally, we have:

$$\begin{aligned}
\sum_{i=z}^{\infty} i^2\hat{\rho}^i &= \hat{\rho}\sum_{i=z}^{\infty} i^2\hat{\rho}^{i-1} \\
&= \hat{\rho}\frac{d}{d\hat{\rho}}\left[\sum_{i=z}^{\infty} i\hat{\rho}^i\right] \\
&= \hat{\rho}\frac{d}{d\hat{\rho}}\left[\frac{z\hat{\rho}^z+(1-z)\hat{\rho}^{z+1}}{(1-\hat{\rho})^2}\right] \\
&= \hat{\rho}\left[\frac{(1-\hat{\rho})\left(z^2\hat{\rho}^{z-1}+(1-z)(1+z)\hat{\rho}^z\right)+2\left(z\hat{\rho}^z+(1-z)\hat{\rho}^{z+1}\right)}{(1-\hat{\rho})^3}\right] \\
&\leq \hat{\rho}\left[\frac{z^2\hat{\rho}^{z-1}+(1+3z)\hat{\rho}^z+2\hat{\rho}^{z+1}}{(1-\hat{\rho})^3}\right] \\
&\leq \hat{\rho}\left[\frac{z^2\hat{\rho}^{z-1}+(1+3z)\hat{\rho}^{z-1}+2\hat{\rho}^{z-1}}{(1-\hat{\rho})^3}\right] = \frac{(3+3z+z^2)\hat{\rho}^z}{(1-\hat{\rho})^3} \qquad \text{(C.22)}
\end{aligned}$$

where the third equality comes from (C.14) and the last inequality uses the fact that $\hat{\rho}$ is smaller than one.

Combining (C.20), (C.21) and (C.22) we obtain:

$$\mathbb{E}\left[K_E(p_H)p_H\mathbf{1}\{K_E(p_H)p_H \geq M\}\right] \leq \left[\frac{(3+z)k_E^*}{(1-\hat{\rho})^2}+\frac{(5+5z+z^2)}{(1-\hat{\rho})^3}\right]\hat{\rho}^z.$$

Since $\hat{\rho}\in(0,1)$, the above term is smaller than δ when z is large enough and so when M is large enough. Thus, the collection $\{K_E(p_H)p_H\}$ is uniformly integrable.

Again, this together with Proposition C.8 imply (e.g., Williams, 1991) that

$$\lim_{p_H\to 0}\mathbb{E}\left[p_H K_E(p_H)\right] \leq \lim_{p_H\to 0} p_H k_E^* = 0$$

Then, by Little's law we get that:

$$\lim_{p_H\to 0} p_H \mathbf{W}_E(Unpaired) = \lim_{p_H\to 0} p_H \frac{\mathbb{E}\left[K_E(p_H)\right]}{(1-\lambda).n} = 0.$$

C.5 Existence of Invariant Distribution

We now prove that the Markov chain induced by the Unpaired with DDL algorithm has a unique invariant distribution. The very same argument can be used to prove that the Markov chain induced by Pairwise with DDL has a unique distribution as well.

Proposition C.9. *The transition matrix Q has a unique invariant distribution.*

Proof. As in Ashlagi et al. (2019) we use the following lemma from Meyn and Tweedie (1993) which is especially useful in proving our proposition.

Lemma C.10 (Meyn and Tweedie (1993)). *Suppose that X_t is an irreducible continuous time Markov chain with the transition matrix Q over states $S = \mathbb{N} \times \mathbb{N}$. If there exist a nonnegative function V on S, a function $w \geq 1$ on S, a finite set $C \subset S$, and constants $c > 0$ and $b \in \mathbb{R}$ such that, for all $i = (k_E, k_H) \in S$:*

$$\sum_{j \in S} Q(i,j)V(j) \leq -cw(i) + b.\mathbb{I}_C(i),$$

where $\mathbb{I}_C$ denotes the indicator function of the set C, then the Markov chain X is ergodic.

It is clear that our Markov chain is irreducible, so our proof will focus on finding a suitable set C, functions V and w, and constants c, and b.

Recall that for a continuous Markov chain, $\sum_{j \neq i} Q(i,j) = -Q(i,i)$. Hence, we must have

$$\sum_{j \in S} Q(i,j)V(j) = \sum_{j \neq i} Q(i,j)(V(j) - V(i)).$$

Now, let $V(k_E, k_H) = k_E + k_H$. For any state $i = (k_E, k_H) \in S$, we have:

$$\sum_{j \neq i} Q(i,j)(V(j) - V(i))$$

$$= Q([k_H, k_E], [k_H{+}1, k_E]) - Q([k_H, k_E], [k_H{-}1, k_E]) + Q([k_H, k_E], [k_H, k_E{+}1]) - Q([k_H, k_E], [k_H, k_E{-}1])$$

Let $\lambda_H = \lambda n$ and $\lambda_E = (1-\lambda)n$. Then the above is equal to:

$$\lambda_H(1-p_H)^{k_H+k_E}(1-p_E)^{k_E}(1-p_H)^{k_H} - \lambda_H(1-(1-p_H)^{k_H+k_E})(1-(1-p_H)^{k_H})$$

$$-\lambda_E(1-(1-p_E)^{k_H+k_E})(1-(1-p_H)^{k_H}) - \mu(1-(1-p_H)^{k_H}) + \lambda_E(1-p_E)^{k_H+k_E}(1-p_H)^{k_H}(1-p_E)^{k_E}$$

$$-\lambda_H(1-(1-p_H)^{k_H+k_E})(1-p_H)^{k_H}(1-(1-p_E)^{k_E}) - \lambda_E(1-(1-p_E)^{k_H+k_E})(1-p_H)^{k_H}(1-(1-p_E)^{k_E})$$

$$-\mu(1-(1-p_H)^{k_H})(1-(1-p_E)^{k_E})$$

$$= -n + n(1-p_H)^{k_H}(1-p_E)^{k_E} + \lambda_E(1-p_E)^{k_H+k_E} + \lambda_H(1-p_H)^{k_H+k_E}$$

$$-2\mu + 2\mu(1-p_H)^{k_H} + \mu(1-p_E)^{k_E} - \mu(1-p_H)^{k_H}(1-p_E)^{k_E}$$

Let $b = 2(n+3\mu)$, $w = n+3\mu$, and $c = \frac{1}{3}$. Now take M such that $(1-p_H)^M \leq \frac{1}{3}$ and $(1-p_E)^M \leq \frac{1}{3}$, and set $C = \{(k_E, k_H) | k_E \leq M, k_H \leq M\}$. Note that C is finite.

For any $i = (k_E, k_H) \notin C$, we must have:

$$\sum_{j \neq i} Q(i,j)(V(j) - V(i)) \leq -n + \frac{1}{3}n + \frac{1}{3}\lambda_E + \frac{1}{3}\lambda_H - 2\mu + \frac{2}{3}\mu + \frac{1}{3}\mu = -\frac{n+3\mu}{3} = -cw(i) + b.\mathbb{I}_C(i)$$

For any $i = (k_E, k_H) \in C$, we have:

$$\sum_{j \neq i} Q(i,j)(V(j) - V(i)) \leq -n + n + \lambda_E + \lambda_H - 2\mu + 2\mu + \mu = n + \mu \leq -\frac{n + 3\mu}{3} + 2(n + 3\mu) = -cw(i) + b.\mathbb{I}_C(i)$$

Thus, the Markov chain is ergodic, which means that it has a unique invariant distribution. □

D Unpaired with DDL versus Optimal with DDL

In this section, we prove the following result which generalizes Theorem 2.8 to the case where $\mu \geq 0$. First, we define the average waiting time achieved by the Optimal algorithm, $\mathbf{W}(Optimal)$, as

$$\inf \mathbf{W}(ALG)$$

where the infimum is taken over all matching algorithms. Similar to the definition in Section 2 (without DDL), a matching algorithm selects a matching M_t in the compatibility graph $\mathcal{G}_t$ (which now potentially includes a deceased donors among the nodes).[D.8] For matching algorithms, we impose that each time a patient is matched to a deceased donor, one donor exits the system with the interpretation that he gives his kidney right away to a compatible patient waiting on the national waiting list for deceased donors. (Recall footnote 46 for the motivation of this modeling choice)

Theorem D.1. *Fix a matching algorithm ALG. We must have*

$$\lim_{p_H \to 0} \sup \frac{\mathbf{W}(Unpaired)}{\mathbf{W}(ALG)} \leq \frac{1}{\lambda}\left(2+\frac{\mu}{n}\right)\ln\left(\frac{1+\lambda+\mu/n}{1+\mu/n}\right) \leq 2\frac{\ln(1+\lambda)}{\lambda}.$$

Let us denote the size of the pool by $\tilde{k}$.[D.9] In the sequel, $\tilde{\mathbf{W}}(\text{ALG})$ is the random variable describing the average waiting time of an arriving patient. Note that a necessary condition for a patient to be matched is that (i) he is compatible with a donor in the pool upon arriving or (ii) he is compatible with a donor in the future or he is compatible with a DD kidney in the future. In the former case, his waiting time is simply 0 while in the latter case, by the Poisson thinning property, the expected waiting time is lower bounded by $\frac{1}{np_T+\mu p_T}$ for a patient of type $T \in \{E,H\}$.

$$\begin{aligned}
\mathbb{E}\left[\tilde{\mathbf{W}}(ALG)\,\middle|\,\tilde{k}=k\right] &\geq \lambda(1-p_H)^k\frac{1}{np_H+\mu p_H}+(1-\lambda)(1-p_E)^k\frac{1}{np_E+\mu p_E}\\
&\geq \lambda(1-kp_H)\frac{1}{p_H(n+\mu)}+(1-\lambda)(1-kp_E)\frac{1}{p_E\,(n+\mu)}\\
&= \frac{\lambda}{p_H(n+\mu)}+\frac{1-\lambda}{p_E\,(n+\mu)}-k\frac{1}{n+\mu}.
\end{aligned}$$

Thus, using the fact that, by Little's law, $\mathbf{W}(\text{ALG}) = \frac{\mathbb{E}[\tilde{k}]}{n}$, we have

$$\begin{aligned}
\mathbf{W}(ALG) &= \mathbb{E}\left[\mathbb{E}\left[\tilde{\mathbf{W}}(ALG)\,\middle|\,\tilde{k}=k\right]\right]\\
&\geq \frac{\lambda}{p_H(n+\mu)}+\frac{1-\lambda}{p_E\,(n+\mu)}-\mathbb{E}\left[\tilde{k}\right]\frac{1}{n+\mu}\\
&= \frac{\lambda}{p_H(n+\mu)}+\frac{1-\lambda}{p_E\,(n+\mu)}-\frac{n}{n+\mu}\mathbf{W}(ALG).
\end{aligned}$$

[D.8] We maintain the restriction to algorithms inducing stochastic processes having an invariant distribution.

[D.9] Note that, by assumption, the number of patients remaining in the pool equals the number of donors remaining. This is the case because we impose the constraint that, each time a patient receive from a deceased donor, a living donor is removed from the system.

This gives us

$$\mathbf{W}(ALG) \geq \frac{\lambda}{(2n+\mu)\,p_H} + \frac{1-\lambda}{(2n+\mu)\,p_E}. \tag{D.23}$$

Now, we are in a position to prove the proposition. Indeed,

$$\begin{aligned}
\lim_{p_H \to 0} \sup \frac{\mathbf{W}(Unpaired)}{\mathbf{W}(ALG)} &= \frac{\lim_{p_H \to 0} p_H \mathbf{W}(Unpaired)}{\lim_{p_H \to 0} \inf p_H \mathbf{W}(ALG)} \\
&\leq \frac{\left(\ln\left[n(1+\lambda)+\mu\right] - \ln\left(n+\mu\right)\right)/n}{\lambda/\left(2n+\mu\right)} \\
&= \frac{1}{\lambda}\left(2+\frac{\mu}{n}\right)\ln\left(\frac{1+\lambda+\mu/n}{1+\mu/n}\right) \\
&\leq 2\frac{\ln(1+\lambda)}{\lambda}.
\end{aligned}$$

where the first inequality comes from Propositon 2.6 together with Equation (D.23). To prove the last inequality, let us define $f(x) = \frac{1}{\lambda}(2+x)\ln\left(\frac{1+\lambda+x}{1+x}\right)$ and remark that:

$$\begin{aligned}
f'(x) &= \frac{1}{\lambda}\left[\ln\left(1+\frac{\lambda}{1+x}\right) - \lambda\frac{1}{1+x}\frac{2+x}{1+\lambda+x}\right] \\
&\leq \frac{1}{\lambda}\left[\ln\left(1+\frac{\lambda}{1+x}\right) - \lambda\frac{1}{1+x}\right]
\end{aligned}$$

which is negative since, for all $y \in (0,1)$, $\ln(1+y) < y$. Hence, $f(\frac{\mu}{n})$ is maximized when there is no deceased donors ($\mu = 0$).

E Results for Pairwise with DDL

In this section, we prove the following proposition (corresponding to Proposition 4.4 in the main text) which generalizes Ashlagi et al. (2019) to the case where one may have an inflow of deceased donors in the system.

Proposition E.1. *Under the Pairwise algorithm, the waiting time of easy-to-match patients,* $\mathbf{W}_E(Paired)$*, satisfies*

$$\lim_{p_H \to 0} p_H \mathbf{W}_E(Paired) = 0$$

while the average waiting time of hard-to-match patient, $\mathbf{W}_H(Paired)$ *satisfies*

- *If* $\mu > n(2\lambda - 1)$*:*

$$\lim_{p_H \to 0} p_H \mathbf{W}_H(Paired) = \frac{c}{\lambda \cdot n}.$$

  where c *solves*[E.10]

$$n(1-\lambda)e^{-cp_E} + \mu e^{-c} = n(1-2\lambda) + \mu.$$

- *If* $\mu < n(2\lambda - 1)$*:*

$$\lim_{p_H \to 0} p_H^2 \mathbf{W}_H(Paired) = \frac{\ln(2\lambda n) - \ln(n+\mu)}{\lambda \cdot n}.$$

E.1 Preliminaries

We denote by Q the transition rate matrix over states $\mathbb{N} \times \mathbb{N}$. We will mainly focus on the following transition rates:

$$\begin{aligned}
Q([k_H,k_E],[k_H+1,k_E]) &= n\lambda(1-p_H^2)^{k_H}(1-p_Ep_H)^{k_E} \\
Q([k_H,k_E],[k_H-1,k_E]) &= n\left\{\lambda\left(1-(1-p_H^2)^{k_H}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H}\right)\right\} \\
&\quad +\mu[1-(1-p_H)^{k_H}] \\
Q([k_H,k_E],[k_H,k_E+1]) &= n(1-\lambda)(1-p_Ep_H)^{k_H}(1-p_E^2)^{k_E} \\
Q([k_H,k_E],[k_H,k_E-1]) &= n\left\{\lambda(1-p_H^2)^{k_H}\left(1-(1-p_Hp_E)^{k_E}\right)\right. \\
&\quad \left.+(1-\lambda)(1-p_Ep_H)^{k_H}\left(1-(1-p_E^2)^{k_E}\right)\right\}+\mu(1-p_H)^{k_H}\left(1-(1-p_E)^{k_E}\right)
\end{aligned}$$

Let also first recall that the Global Balance Equations (GBE) are a set of equations that characterize the invariant distribution of a Markov chain, when such a distribution exists. The above stochastic process is a Markov chain which has an invariant distribution as proved in Appendix C.5. In the sequel, we let π be this invariant distribution. The GBE can be stated as follows: for any subset $S \subset \mathbb{N} \times \mathbb{N}$, we must have:

$$\sum_{j \in S} \pi(j) \sum_{i \notin S} Q(j,i) = \sum_{i \notin S} \pi(i) \sum_{j \in S} Q(i,j) \tag{E.24}$$

[E.10] It is easily checked that $c \in \left[\ln(n(1-\lambda)+\mu) - \ln(n(1-2\lambda)+\mu), \frac{1}{p_E}\left(\ln(n(1-\lambda)+\mu) - \ln(n(1-2\lambda)+\mu)\right)\right]$.

Let us define the following function $f : \mathbb{R}_+ \to \mathbb{R}$:

$$f(k) := n\lambda(1-p_H^2)^k - n\left\{\lambda\left(1-(1-p_H^2)^k\right) + (1-\lambda)\left(1-(1-p_E p_H)^k\right)\right\} - \mu[1-(1-p_H)^k]$$

Observe that $f(k)$ is decreasing with $f(0) = n\lambda$ and $\lim_{k\to+\infty} f(k) = -(n+\mu)$ such that there is a unique positive value k_H^* such that $f(k_H^*) = 0$. We show the following intermediary result:

Lemma E.2. *As p_H vanishes, we have that*

- *if $\mu > n(2\lambda-1)$,*
$$-k_H^* \ln(1-p_H) \to c$$
where c solves
$$n(1-\lambda)e^{-cp_E} + \mu e^{-c} = n(1-2\lambda) + \mu.$$
- *if $\mu < n(2\lambda-1)$,*
$$-k_H^* \ln(1-p_H^2) \to \ln(2n\lambda) - \ln(n+\mu).$$

Proof. For each p_H, we fix a solution $k_H^*(p_H)$ of the equation $f(k) = 0$ (we make explicit the dependence of k_H^* to p_H). We claim that, when $\mu > n(2\lambda-1)$, $(1-p_H^2)^{k_H^*(p_H)} \to 1$. Indeed, assume not, i.e., $(1-p_H^2)^{k_H^*(p_H)} \nrightarrow 1$. Since the sequence $(1-p_H^2)^{k_H^*(p_H)}$ lies in the compact set $[0,1]$, there must exist a subsequence of $(1-p_H^2)^{k_H^*(p_H)}$ which converges to a positive number $c < 1$.[E.11] From now on, we focus on this subsequence, we have $(1-p_H^2)^{k_H^*(p_H)} = \left((1-p_H^2)^{1/p_H^2}\right)^{k_H^*(p_H)p_H^2} \to c$ where $c \in [0,1)$. Hence, $k_H^*(p_H)p_H^2$ does not go to 0 as p_H vanishes. Put differently, $k_H^*(p_H)$ grows at the order at least $1/p_H^2$, in particular, $k_H^*(p_H)p_H \to \infty$. In such a case, $(1-p_H)^{k_H^*(p_H)} = \left((1-p_H)^{1/p_H}\right)^{k_H^*(p_H)p_H} \to 0$ and, similarly, $(1-p_E p_H)^{k_H^*(p_H)} \to 0$. Hence

$$f(k_H^*(p_H)) \to n\lambda c - n\lambda(1-c) - (1-\lambda)n - \mu < n(2\lambda-1) - \mu \leq 0,$$

such that $k_H^*(p_H)$ cannot satisfy $f(k_H^*(p_H)) = 0$ for a p_H small enough, which yields a contradiction with the definition of $k_H^*(p_H)$. We conclude that $(1-p_H^2)^{k_H^*(p_H)} \to 1$. Thus, in this case, we must have that, as p_H vanishes,

$$n(1-\lambda)(1-p_E p_H)^{k_H^*(p_H)} + \mu(1-p_H)^{k_H^*(p_H)} = n(1-2\lambda) - \mu.$$

From this, it is easy to see that $-k_H^*(p_H)\ln(1-p_H)$ must converge to a constant c as p_H vanishes. Thus, as p_H vanishes, the above equation can be rewritten as

$$n(1-\lambda)(1-p_E p_H)^{\frac{1}{p_E p_H} p_E c} + \mu(1-p_H)^{\frac{1}{p_H} c} = n(1-2\lambda) - \mu$$

[E.11] Recall that for a given sequence if each possible converging subsequence extracted from that sequence converges to the same limit, say 1, then the original sequence also converges to that 1.

which, as p_H vanishes, yields

$$n(1-\lambda)e^{-p_E c} + \mu e^{-c} = n(1-2\lambda) - \mu.$$

We conclude that $-k_H^*(p_H)\ln(1-p_H)$ converges to c as p_H vanishes where c solves the above equation, as claimed.

Let us now consider the case $\mu < n(2\lambda - 1)$. Again, we fix a solution $k_H^*(p_H)$ of the equation $f(k) = 0$ and we show that $(1-p_H^2)^{k_H^*(p_H)}$ converge to a constant $c \in (0,1)$. Assume not, there must exist a subsequence of $(1-p^2)^{k_H^*(p_H)}$ which converges to 1. Focusing on this subsequence we must have

$$\begin{aligned} f(k_H^*(p_H)) &\to (1-\lambda)n(1-p_E p_H)^{k_H^*(p_H)} + \mu(1-p_H)^{k_H^*(p_H)} + n(2\lambda-1) - \mu \\ &> (1-\lambda)n(1-p_E p_H)^{k_H^*(p_H)} + \mu(1-p_H)^{k_H^*(p_H)} \geq 0 \end{aligned}$$

where the first inequality arises since $\mu < n(2\lambda - 1)$. The above inequality contradicts the definition of $k_H^*(p_H)$ so that we must have $(1-p_H^2)^{k_H^*(p_H)} \to c \in (0,1)$. It implies that $k_H^*(p_H)$ grows at the order at least $1/p_H^2$, such that $(1-p_H)^{k_H^*(p_H)} \to 0$ and $(1-p_E p_H)^{k_H^*(p_H)} \to 0$. Hence, as p_H vanishes, $k_H^*(p_H)$ must satisfies:

$$2n\lambda(1-p_H^2)^{k_H^*(p_H)} = n + \mu$$

Hence, $(1-p_H^2)^{k_H^*(p_H)} \to \frac{n+\mu}{2n\lambda}$. This proves that, as p_H vanishes,$-k_H^*(p_H)\ln(1-p_H^2)$ converges to $\ln(2n\lambda) - \ln(n+\mu)$, as claimed. □

E.2 Upper-bound result

In the sequel, we first prove the following result providing an upper-bound on the number of hard-to-match patients.

Proposition E.3. *Assume $\mu > n(2\lambda - 1)$: for any $\delta > 0$,*

$$\pi_H\left\{k_H : \frac{k_H}{1/(-\ln(1-p_E p_H))} \geq c + \delta\right\} \to 0$$

as p_H vanishes. Assume $\mu < n(2\lambda - 1)$: for any $\delta > 0$,

$$\pi_H\left\{k_H : \frac{k_H}{1/(-\ln(1-p_H^2))} \geq \ln(2n\lambda) - \ln(n+\mu) + \delta\right\} \to 0$$

as p_H vanishes.

In order to show this, we need to prove the following lemma.

Lemma E.4. *For any $\varepsilon > 0$, there exists a constant $\rho \in (0,1)$ such that, for any $p_H > 0$ and for*

any integer $k_H \geq k_H^*(1+\varepsilon)$

$$\frac{\pi_H(k_H+1)}{\pi_H(k_H)} \leq \rho.$$

Proof of Lemma E.4. The GBE (Equation (E.24)) gives us

$$\sum_{k_E=0}^{\infty} \pi(k_H,k_E)Q([k_H,k_E],[k_H+1,k_E]) = \sum_{k_E=0}^{\infty} \pi(k_H+1,k_E)Q([k_H+1,k_E],[k_H,k_E])$$

Using the expressions of the transition rates, this can be rewritten as:

$$\begin{aligned} & \sum_{k_E=0}^{\infty} \pi(k_H,k_E)\left[n\lambda(1-p_H^2)^{k_H}(1-p_Ep_H)^{k_E}\right] \\ = & \sum_{k_E=0}^{\infty} \pi(k_H+1,k_E)\left[\begin{array}{c} n\left\{\lambda\left(1-(1-p_H^2)^{k_H+1}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H+1}\right)\right\} \\ +\mu[1-(1-p_H)^{k_H+1}] \end{array}\right]. \end{aligned}$$

Observing that the term in brackets in left-hand side of the above equality is maximized at $k_E = 0$, we get:

$$\begin{aligned} & \textstyle\sum_{k_E=0}^{\infty} \pi(k_H,k_E)\left[n\lambda(1-p_H^2)^{k_H}\right] \\ \geq & \textstyle\sum_{k_E=0}^{\infty} \pi(k_H+1,k_E)\left[\begin{array}{c} n\left\{\lambda\left(1-(1-p_H^2)^{k_H+1}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H+1}\right)\right\} \\ +\mu[1-(1-p_H)^{k_H+1}] \end{array}\right]. \end{aligned}$$

It implies that

$$\frac{\pi_H(k_H)}{\pi_H(k_H+1)} \geq \frac{n\left\{\lambda\left(1-(1-p_H^2)^{k_H+1}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H+1}\right)\right\}+\mu[1-(1-p_H)^{k_H+1}]}{n\lambda(1-p_H^2)^{k_H}}. \tag{E.25}$$

where we recall that $\pi_H(k_H) = \sum_{k_E=0}^{\infty} \pi(k_H,k_E)$.

Fix any $\varepsilon > 0$ and an arbitrary $k_H \geq k_H^*(1+\varepsilon)$. From the inequality above we deduce:

$$\begin{aligned} \frac{\pi_H(k_H+1)}{\pi_H(k_H)} & \leq \frac{n\lambda(1-p_H^2)^{k_H}}{n\left\{\lambda\left(1-(1-p_H^2)^{k_H+1}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H+1}\right)\right\}+\mu[1-(1-p_H)^{k_H+1}]} \\ & \leq \frac{n\lambda(1-p_H^2)^{k_H}}{n\left\{\lambda\left(1-(1-p_H^2)^{k_H}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H}\right)\right\}+\mu[1-(1-p_H)^{k_H}]} \\ & \leq \frac{n\lambda(1-p_H^2)^{k_H^*(1+\varepsilon)}}{n\left\{\lambda\left(1-(1-p_H^2)^{k_H^*(1+\varepsilon)}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H^*(1+\varepsilon)}\right)\right\}+\mu[1-(1-p_H)^{k_H^*(1+\varepsilon)}]} \\ & \leq \frac{n\lambda(1-p_H^2)^{k_H^*}}{n\left\{\lambda\left(1-(1-p_H^2)^{k_H^*(1+\varepsilon)}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H^*(1+\varepsilon)}\right)\right\}+\mu[1-(1-p_H)^{k_H^*(1+\varepsilon)}]} \end{aligned}$$

As p_H vanishes, both the denominator and the numerator converge to a constant. We denote the ratio of these constant by ρ. Clearly, by definition of k_H^*, at $\varepsilon = 0$, this ratio is equal to 1. Hence, given that $\varepsilon > 0$, we must have $\rho < 1$. Hence we obtain a positive constant $\rho < 1$ independent of

p_H, such that for all $k_H \geq (1+\varepsilon)k_H^*$: $\frac{\pi_H(k_H+1)}{\pi_H(k_H)} \leq \rho$, as claimed. □

Using the result stated in Lemma E.4 we can show the following:

Lemma E.5. *For any $\varepsilon > 0$, there exists a constant $\rho \in (0,1)$ such that, for any $p_H > 0$ and for any integer $z > 0$:*

$$\pi_H \{k_H : k_H \geq k_H^*(1+\varepsilon) + z\} \leq \frac{\rho^z}{1-\rho}.$$

Proof of Lemma E.5. The proof is the same as for Unpaired with DDL (see Lemma C.4). □

Completion of the proof of Proposition C.1. Assume that $\mu < n(2\lambda - 1)$. Fix any $\delta > 0$. We want to show that

$$\pi_H \left\{ k_H : \frac{k_H}{1/(-\ln(1-p_H^2))} \geq \ln(2n\lambda) - \ln(n+\mu) + \delta \right\} \to 0$$

as p_H vanishes. Fix $\varepsilon > 0$ and $p_H > 0$ small enough so that

$$[\ln(2n\lambda) - \ln(n+\mu)](1+2\varepsilon) + p_H \leq \ln(2n\lambda) - \ln(n+\mu) + \delta.$$

Hence, we obtain

$$\begin{aligned}
& \pi_H \left\{ k_H : \frac{k_H}{1/(-\ln(1-p_H^2))} \geq \ln(2n\lambda) - \ln(n+\mu) + \delta \right\} \\
\leq\ & \pi_H \left\{ k_H : \frac{k_H}{1/(-\ln(1-p_H^2))} \geq [\ln(2n\lambda) - \ln(n+\mu)](1+2\varepsilon) + p_H \right\} \\
\leq\ & \pi_H \left\{ k_H : k_H \geq k_H^*(1+\varepsilon) + \frac{p_H}{-\ln(1-p_H^2)} \right\} \\
\leq\ & \frac{\rho^{\frac{p_H}{-\ln\left(1-p_H^2\right)}}}{1-\rho} \to 0
\end{aligned}$$

where the first inequality is ensured by our choice of ε and p_H. The second inequality uses the fact that $[\ln(2n\lambda) - \ln(n+\mu)](1+2\varepsilon) \geq -\ln(1-p_H^2))k_H^*(1+\varepsilon)$ for p_H small enough since, by Lemma E.2, $-\ln(1-p_H^2))k_H^* \to \ln(2n\lambda) - \ln(n+\mu)$ as p_H vanishes. In turn, the last inequality follows from Lemma C.4 and the convergence result holds since $\rho \in (0,1)$ (still by Lemma E.5) and since $\frac{p_H}{-\ln(1-p_H^2)}$ explodes as p_H vanishes.

Assume now that $\mu > n(2\lambda - 1)$ and fix any $\delta > 0$. We want to show that

$$\pi_H \left\{ k_H : \frac{k_H}{1/(-\ln(1-p_H))} \geq c + \delta \right\} \to 0$$

as p_H vanishes. Fix $\varepsilon > 0$ and $p_H > 0$ small enough so that

$$c(1+2\varepsilon) + \sqrt{p_H} \leq c + \delta.$$

Hence, we obtain

$$\begin{aligned}
& \pi_H\left\{k_H : \frac{k_H}{1/(-\ln(1-p_H))} \geq c+\delta\right\} \\
\leq\ & \pi_H\left\{k_H : \frac{k_H}{1/(-\ln(1-p_H))} \geq c(1+2\varepsilon)+\sqrt{p_H}\right\} \\
\leq\ & \pi_H\left\{k_H : k_H \geq k_H^*(1+\varepsilon)+\frac{\sqrt{p_H}}{-\ln(1-p_H)}\right\} \\
\leq\ & \frac{\rho^{\frac{\sqrt{p_H}}{-\ln(1-p_H)}}}{1-\rho} \to 0
\end{aligned}$$

where the first inequality is ensured by our choice of ε and p_H. The second inequality uses the fact that $c(1+2\varepsilon) \geq -k_H^*(1+\varepsilon)\ln(1-p_H)$ for p_H small enough since, by Lemma E.2, $-k_H^*\ln(1-p_H) \to c$ as p_H vanishes. In turn, the last inequality follows from Lemma E.5 and the convergence result holds since $\rho \in (0,1)$ (still by Lemma E.5) and since $\frac{\sqrt{p_H}}{-\ln(1-p_H)}$ explodes as p_H vanishes. □

E.3 Lower-bound result

Before providing a lower-bound on the number of hard-to-match patients (Section E.3.2) we need to establish an upper-bound on the number of easy-to-match patients (Section C.3.1).

E.3.1 An upper-bound on the number of easy-to-match patients

Lemma E.6. *For any p_H small enough,*

$$\pi_E\left\{k_E : k_E \geq \frac{1}{\sqrt{p_H}}+k\right\} \leq \frac{3+k}{(1-\hat{\rho})^2}\hat{\rho}^k$$

where $\hat{\rho} < 1$.[E.12]

Proof of Lemma E.6. Fix an arbitrary $k_E \geq 0$ and let us consider the set $S = \mathbb{N}\times\{0,1,...,k_E\}$. Then, the GBE (Equation (E.24)) writes as:

$$\sum_{k_H=0}^{\infty}\pi(k_H,k_E)Q([k_H,k_E],[k_H,k_E+1]) = \sum_{k_H=0}^{\infty}\pi(k_H,k_E+1)Q([k_H,k_E+1],[k_H,k_E]) \quad \text{(E.26)}$$

[E.12] While we do not formally prove it, one can show that when $\mu \geq n(2\lambda-1)$, there is a constant k_E^* such that, for any k and for any p_H small enough,

$$\pi_E\{k_E : k_E \geq k_E^*+k\} \leq \frac{3+k}{(1-\hat{\rho})^2}\hat{\rho}^k$$

where $\hat{\rho} < 1$, very much as Lemma C.5 that we obtained for the Unpaired algorithm.

Using the expressions of the transition rates, this can be rewritten as:

$$\begin{aligned}
&\sum_{k_H=0}^{\infty} \pi(k_H,k_E) n(1-\lambda)(1-p_E p_H)^{k_H}(1-p_E^2)^{k_E} \\
= &\sum_{k_H=0}^{\infty} \pi(k_H,k_E+1) \left[\begin{array}{c} n\lambda(1-p_H^2)^{k_H}\left(1-(1-p_H p_E)^{k_E+1}\right) \\ +n(1-\lambda)(1-p_E p_H)^{k_H}\left(1-(1-p_E^2)^{k_E+1}\right) \\ +\mu(1-p_H)^{k_H}\left(1-(1-p_E)^{k_E+1}\right) \end{array} \right]
\end{aligned}$$

Observing that the expression in brackets in the left hand side is maximized at $k_H = 0$ and that the expression in brackets in right hand-side may be bounded below by disregarding the third term, we get that:

$$\begin{aligned}
&\sum_{k_H=0}^{\infty} \pi(k_H,k_E)(1-\lambda)(1-p_E^2)^{k_E} \\
\geq &\sum_{k_H=0}^{\infty} \pi(k_H,k_E+1)\left\{\lambda(1-p_H^2)^{k_H}\left(1-(1-p_H p_E)^{k_E+1}\right)+(1-\lambda)(1-p_E p_H)^{k_H}\left(1-(1-p_E^2)^{k_E+1}\right)\right\} \\
\geq &\sum_{k_H=0}^{\infty} \pi(k_H,k_E+1)\left\{\lambda(1-p_H^2)^{k_H}\left(1-(1-p_H p_E)^{k_E}\right)+(1-\lambda)(1-p_E p_H)^{k_H}\left(1-(1-p_E^2)^{k_E}\right)\right\}
\end{aligned}$$

Let us disregard the second term of the right hand side of the above inequality such that:

$$\pi_E(k_E)(1-\lambda)(1-p_E^2)^{k_E} \geq \sum_{k_H=0}^{\infty} \pi(k_H,k_E+1)\lambda(1-p_H^2)^{k_H}\left(1-(1-p_H p_E)^{k_E}\right) \tag{E.27}$$

where π_E denotes the marginal of π on the number of easy-to-match patients in the pool, i.e., $\pi_E(k_E) = \sum_{k_H=0}^{\infty} \pi(k_H,k_E)$.

In the sequel, for each k_E and for any $\varepsilon > 0$, we define $S(k_E) := \{k_H : k_H \leq (1+\varepsilon)k_H^* + k_E\}$ as well as $\pi_{S(k_E)}(k_E) := \sum_{k_H \in S(k_E)} \pi(k_H,k_E)$. We must have:

$$\begin{aligned}
\pi_E(k_E) - \pi_{S(k_E)}(k_E) &= \sum_{k_H \notin S(k_E)} \pi(k_H,k_E) \\
&\leq \sum_{k_E=0}^{\infty} \sum_{k_H \notin S(k_E)} \pi(k_H,k_E) = \sum_{k_H \notin S(k_E)} \pi_H(k_H).
\end{aligned}$$

Hence, by Lemma E.5, for any k_E,

$$\pi_E(k_E) - \pi_{S(k_E)}(k_E) \leq \frac{\rho^{k_E}}{1-\rho} \tag{E.28}$$

where $\rho \in (0,1)$. Note that for p_H small enough and for each $k_H \in S(k_E)$, we have

$$(1-p_H^2)^{k_H} \geq (1-p_H^2)^{(1+\varepsilon)k_H^*+k_E} \geq d^{1+\varepsilon}(1-p_H^2)^{k_E}$$

where d is a constant smaller than 1. The second inequality comes from Lemma E.2 stating that as p_H vanishes, $(1-p_H^2)^{k_H^*}$ converges to a constant strictly smaller than 1 when $\mu < n(2\lambda-1)$ and converges to 1 when $\mu \geq n(2\lambda-1)$.

Thus, we can rewrite Equation (E.27) as

$$\begin{aligned} & (1-\lambda)(1-p_E^2)^{k_E}\pi_{S(k_E)}(k_E) + (1-\lambda)(1-p_E^2)^{k_E}\left(\pi_E(k_E) - \pi_{S(k_E)}(k_E)\right) \\ \geq \quad & \lambda d^{1+\varepsilon}(1-p_H^2)^{k_E+1}[1-(1-p_H p_E)^{k_E}]\pi_{S(k_E+1)}(k_E+1) \end{aligned}$$

where we used the fact that $\pi_E(k_E+1) \geq \pi_{S(k_E+1)}(k_E+1)$.

Let us denote

$$\bar{\rho}(k_E) := \frac{1-\lambda}{\lambda d^{1+\varepsilon}(1-p_H^2)}\left(\frac{1-p_E^2}{1-p_H^2}\right)^{k_E}\frac{1}{1-(1-p_H p_E)^{k_E}}.$$

The above inequality can be written as

$$\bar{\rho}(k_E)\pi_{S(k_E)}(k_E) + \bar{\rho}(k_E)\left(\pi_E(k_E) - \pi_{S(k_E)}(k_E)\right) \geq \pi_{S(k_E+1)}(k_E+1).$$

Now, observing that $\bar{\rho}(k_E)$ is decreasing in k_E, we get that for $k_E \geq \frac{1}{\sqrt{p_H}}$:

$$\bar{\rho}(k_E) \leq \frac{1-\lambda}{\lambda d^{1+\varepsilon}(1-p_H^2)}\left(\frac{1-p_E^2}{1-p_H^2}\right)^{\frac{1}{\sqrt{p_H}}}\frac{1}{1-(1-p_H p_E)^{\frac{1}{\sqrt{p_H}}}}. \tag{E.29}$$

We can show that the above upper bound goes to 0 as p_H vanishes. To see this, it is enough to show that

$$\left(\frac{1-p_E^2}{1-p_H^2}\right)^{\frac{1}{\sqrt{p_H}}}\frac{1}{1-(1-p_H p_E)^{\frac{1}{\sqrt{p_H}}}} \to 0$$

as p_H vanishes. This is equivalent to showing that

$$\lim_{x\to 0}\frac{\alpha^{\frac{1}{x}}}{1-\beta^x} = \lim_{x\to 0}\frac{\alpha^{\frac{1}{x}}}{x^2}\left(\frac{\log(\alpha)}{\beta^x\log(\beta)}\right) = 0$$

with $\alpha := (1-p_E^2) < 1$ and $\beta := \exp(-p_E) < 1$. The equivalence comes from the fact that $(1-p_H p_E)^{\frac{1}{\sqrt{p_H}}} = \left((1-p_H p_E)^{\frac{1}{p_H}}\right)^{\sqrt{p_H}} \to (\exp(-p_E))^{\sqrt{p_H}}$ as p_H vanished and the first equality comes from the L'Hospital's Rule. As the term in brackets tends to the positive constant $\frac{\log(\alpha)}{\log(\beta)}$ when x vanishes, for our purpose, we need to show that $\frac{\alpha^{\frac{1}{x}}}{x^2} \to 0$ as $x \to 0$. Note that this is equivalent to showing that $x^2\alpha^x \to 0$ as $x \to \infty$ which is true as shown in equation (C.18).

Hence, the RHS of inequality (E.29) vanishes as p_H tends to 0. Thus, we can fix a constant $\rho_* < 1-\rho$ (< 1) such that, for p_H small enough, $\bar{\rho}(k_E) < \rho_*$ for any $k_E \geq \frac{1}{\sqrt{p_H}}$ Thus, from Equation

(E.28), for any p_H small enough and for any $k_E \geq \frac{1}{\sqrt{p_H}}$, we must have

$$\rho_* \pi_{S(k_E)}(k_E) + \rho_* \frac{\rho^{k_E}}{1-\rho} \geq \pi_{S(k_E+1)}(k_E+1)$$

Clearly, since $\frac{\rho_*}{1-\rho} < 1$, we have that

$$\rho_* \pi_{S(k_E)}(k_E) + \rho^{k_E} \geq \pi_{S(k_E+1)}(k_E+1)$$

holds for any p_H small enough and for any $k_E \geq \frac{1}{\sqrt{p_H}}$. Now, setting $\hat{\rho} := \max(\rho_*, \rho) < 1$, we obtain

$$\hat{\rho} \pi_{S(k_E)}(k_E) + \hat{\rho}^{k_E} \geq \pi_{S(k_E+1)}(k_E+1).$$

Now, proceeding inductively, for any p_H small enough, we must have

$$\hat{\rho}^i \pi_{S(k_E^*(p_H))}(k_E^*(p_H)) + i\hat{\rho}^{k_E^*(p_H)+i} \geq \pi_{S(k_E^*(p_H)+i)}(k_E^*(p_H)+i) \tag{E.30}$$

where $k_E^*(p_H) := \lceil \frac{1}{\sqrt{p_H}} \rceil$. Now, we have that for any p_H small enough:

$$\begin{aligned}
\sum_{k_E \geq k_E^*(p_H)+k}^{\infty} \pi_E(k_E) &= \sum_{i=k}^{\infty} \pi_E(k_E^*(p_H)+i) \\
&= \sum_{i=k}^{\infty} \pi_{S(k_E^*(p_H)+i)}(k_E^*(p_H)+i) + \sum_{i=k}^{\infty} [\pi_E(k_E^*(p_H)+i) - \pi_{S(k_E^*(p_H)+i)}(k_E^*(p_H)+i)] \\
&\leq \sum_{i=k}^{\infty} \pi_{S(k_E^*(p_H)+i)}(k_E^*(p_H)+i) + \sum_{i=k}^{\infty} \frac{\rho^{k_E^*(p_H)+i}}{1-\rho} \\
&\leq \sum_{i=k}^{\infty} \pi_{S(k_E^*(p_H)+i)}(k_E^*(p_H)+i) + \sum_{i=k}^{\infty} \frac{\hat{\rho}^i}{1-\hat{\rho}} \\
&\leq \sum_{i=k}^{\infty} \hat{\rho}^i \pi_{S(k_E^*(p_H))}(k_E^*(p_H)) + \hat{\rho}^{k_E^*(p_H)} \sum_{i=k}^{\infty} i\hat{\rho}^i + \frac{\hat{\rho}^k}{(1-\hat{\rho})^2} \\
&\leq \sum_{i=k}^{\infty} \hat{\rho}^i + \sum_{i=k}^{\infty} i\hat{\rho}^i + \frac{\hat{\rho}^k}{(1-\hat{\rho})^2} \\
&\leq \frac{\hat{\rho}^k}{1-\hat{\rho}} \left(1 + \frac{1+k}{1-\hat{\rho}} + \frac{1}{1-\hat{\rho}}\right) \\
&\leq \frac{\hat{\rho}^k}{1-\hat{\rho}} \frac{3+k}{1-\hat{\rho}}
\end{aligned} \tag{E.31}$$

where the first inequality uses (E.28). The second uses $\hat{\rho} \geq \rho$. The third uses (E.30). The fourth uses the fact that both $\pi_{S(k_E^*(p_H))}(k_E^*(p_H))$ and $\hat{\rho}$ are smaller than one. The penultimate inequality uses the following fact:

$$\sum_{i=0}^{n} i\hat{\rho}^i = \frac{\hat{\rho} - (n+1)\hat{\rho}^{n+1} + n\hat{\rho}^{n+2}}{(1-\hat{\rho})^2} \tag{E.32}$$

and the fact that

$$\begin{aligned}\sum_{i=k}^{+\infty} i\hat{\rho}^i &= \sum_{i=0}^{+\infty} i\hat{\rho}^i - \sum_{i=0}^{k-1} i\hat{\rho}^i \\ &= \frac{\hat{\rho}}{(1-\hat{\rho})^2} - \frac{\hat{\rho} - k\hat{\rho}^k + (k-1)\hat{\rho}^{k+1}}{(1-\hat{\rho})^2} \\ &= \frac{k\hat{\rho}^k + \hat{\rho}^{k+1} - k\hat{\rho}^{k+1}}{(1-\hat{\rho})^2} \qquad \text{(E.33)} \\ &\leq \frac{k\hat{\rho}^k + \hat{\rho}^k}{(1-\hat{\rho})^2} \qquad \text{(E.34)}\end{aligned}$$

where the second equality comes from equation (E.32) and $\hat{\rho} \in (0,1)$ is used for this equality as well as for the inequality. □

E.3.2 Completing of the lower-bound result for hard-to-match patients

Proposition E.7. *Assume $\mu > n(2\lambda - 1)$: for any $\delta > 0$,*

$$\pi_H \left\{ k_H : \frac{k_H}{1/(-\ln(1 - p_E p_H))} \geq c - \delta \right\} \to 0$$

as p_H vanishes. Assume $\mu < n(2\lambda - 1)$: for any $\delta > 0$,

$$\pi_H \left\{ k_H : \frac{k_H}{1/(-\ln(1 - p_H^2))} \geq \ln(2n\lambda) - \ln(n + \mu) - \delta \right\} \to 0$$

as p_H vanishes.

We set $k_E^* := \frac{\alpha}{p_H}$ where $\alpha > 0$ is yet to be fixed. In the sequel we define $S := \{k_E : k_E \leq k_E^* + 1/\sqrt{p_H}\}$ and, as before, $\pi_S(k_H) = \sum_{k_E \in S} \pi(k_H, k_E)$. We first prove the following lemma.

Lemma E.8. *For any $\varepsilon \in (0,1)$ and any p_H small enough, there exists $\tilde{\rho} \in (0,1)$ such that, for any integer $z > 0$:*

$$\pi_S(k_H^*(1-\varepsilon) - z) \leq \tilde{\rho}^z \pi_S(k_H^*(1-\varepsilon)) + \phi(p_H)\tilde{\rho}\frac{1-\tilde{\rho}^z}{1-\tilde{\rho}}$$

where $\phi(p_H) := \left(3 + \frac{1}{\sqrt{p_H}}\right) \frac{\tilde{\rho}^{1/\sqrt{p_H}}}{(1-\tilde{\rho})^2}$.

Proof of Lemma E.8. Let us recall that using the GBE we obtained

$$\begin{aligned}&\sum_{k_E=0}^{\infty} \pi(k_H, k_E) \left[n\lambda(1-p_H^2)^{k_H}(1-p_E p_H)^{k_E} \right] \\ = &\sum_{k_E \in S} \pi(k_H+1, k_E) \left[\begin{array}{c} n\left\{\lambda\left(1-(1-p_H^2)^{k_H+1}\right) + (1-\lambda)\left(1-(1-p_E p_H)^{k_H+1}\right)\right\} \\ +\mu[1-(1-p_H)^{k_H+1}] \end{array} \right] \\ &+ \sum_{k_E \notin S} \pi(k_H+1, k_E) \left[\begin{array}{c} n\left\{\lambda\left(1-(1-p_H^2)^{k_H+1}\right) + (1-\lambda)\left(1-(1-p_E p_H)^{k_H+1}\right)\right\} \\ +\mu[1-(1-p_H)^{k_H+1}] \end{array} \right].\end{aligned}$$

Note that for p_H small enough, $(1-p_H)^{k_E} \geq (1-p_H)^{k_E^*}.c^{\sqrt{p_H}}$ with $c \in (0,1)$ whenever $k_E \in S = \{k_E : k_E \leq k_E^* + 1/\sqrt{p_H}\}$.[E.13] This observation allows us to lower-bound the left-hand side of the above displayed equation. Thus, by upper-bounding the right-hand side as well, we get

$$\begin{aligned} & \pi_S(k_H)\left[n\lambda(1-p_H)^{k_E^*}.c^{\sqrt{p_H}}(1-p_H^2)^{k_H}\right] \\ \leq\ & \pi_S(k_H+1)\left[\begin{array}{c} n\left\{\lambda\left(1-(1-p_H^2)^{k_H+1}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H+1}\right)\right\} \\ +\mu[1-(1-p_H)^{k_H+1}] \end{array}\right] \\ & +(\pi_H(k_H+1)-\pi_S(k_H+1))\left[n+\mu\right]. \end{aligned}$$

In the sequel we fix $\varepsilon \in (0,1)$ and consider $k_H \leq k_H^*(1-\varepsilon)$. We note two obvious facts. First, $k_H \leq k_H^*(1-\frac{\varepsilon}{2})$. Second, $k_H+1 \leq k_H^*(1-\frac{\varepsilon}{2})$ for p_H small enough. To see why the last inequality holds, notice that $k_H^*(1-\frac{\varepsilon}{2}) = k_H^*(1-\varepsilon) + k_H^*\frac{\varepsilon}{2}$. Since we know that $k_H \leq k_H^*(1-\varepsilon)$ and that $1 \leq k_H^*\frac{\varepsilon}{2}$ for p_H small enough (recall that, by Lemma E.2, k_H^* explodes as p_H vanishes), we get the aforementioned inequality. These two simple inequalities allow us to further bound the right-hand side and the left-hand side of the above displayed equation to get:

$$\begin{aligned} & \pi_S(k_H)\left[n\lambda(1-p_H)^{k_E^*}.c^{\sqrt{p_H}}(1-p_H^2)^{k_H^*(1-\varepsilon/2)}\right] \\ \leq\ & \pi_S(k_H+1)\left[\begin{array}{c} n\left\{\lambda\left(1-(1-p_H^2)^{k_H^*(1-\varepsilon/2)}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H^*(1-\varepsilon/2)}\right)\right\} \\ +\mu[1-(1-p_H)^{k_H^*(1-\varepsilon/2)}] \end{array}\right] \\ & +(\pi_H(k_H+1)-\pi_S(k_H+1))\left[\begin{array}{c} n\left\{\lambda\left(1-(1-p_H^2)^{k_H^*(1-\varepsilon/2)}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H^*(1-\varepsilon/2)}\right)\right\} \\ +\mu[1-(1-p_H)^{k_H^*(1-\varepsilon/2)}] \end{array}\right]. \end{aligned}$$

This may be rewritten as:

$$\pi_S(k_H) \leq \tilde{\rho}_1(p_H)\pi_S(k_H+1) + \tilde{\rho}_1(p_H)(\pi_H(k_H+1)-\pi_S(k_H+1)) \tag{E.35}$$

with

$$\tilde{\rho}_1(p_H) := \left(\frac{1}{(1-p_H)^{k_E^*}c^{\sqrt{p_H}}}\right)\frac{\begin{array}{c} n\left\{\lambda\left(1-(1-p_H^2)^{k_H^*(1-\varepsilon/2)}\right)+(1-\lambda)\left(1-(1-p_Ep_H)^{k_H^*(1-\varepsilon/2)}\right)\right\} \\ +\mu[1-(1-p_H)^{k_H^*(1-\varepsilon/2)}] \end{array}}{n\lambda(1-p_H^2)^{k_H^*(1-\varepsilon/2)}}.$$

Now, we claim that, for p_H small enough, $\tilde{\rho}_1(p_H) \leq \tilde{\rho}_1$ where $\tilde{\rho}_1 < 1$ does not depend on p_H. Indeed, $1/\left((1-p_H)^{k_E^*}c^{\sqrt{p_H}}\right)$ converges from above to 1 as p_H vanishes (recall that $k_E^* = \frac{1}{\sqrt{p_H}}$). Further the second term in the expression of $\tilde{\rho}_1(p_H)$ is equal to 1 when $\varepsilon = 0$. Thus, it is strictly smaller than

[E.13] Simply note that, for any $k_E \in S$,

$$\begin{aligned} (1-p_H)^{k_E} &\geq (1-p_H)^{k_E^*}(1-p_H)^{1/\sqrt{p_H}} \\ &= (1-p_H)^{k_E^*}\left((1-p_H)^{1/p_H}\right)^{\sqrt{p_H}}. \end{aligned}$$

Since $(1-p_H)^{1/p_H}$ converges from below to $1/e$, we can ensure that for p_H small enough, $(1-p_H)^{k_E} \geq (1-p_H)^{k_E^*}\left(\frac{0.9}{e}\right)^{\sqrt{p_H}}$.

1 when $\varepsilon > 0$. We get that $\tilde{\rho}_1(p_H)$ is smaller than $\tilde{\rho}_1 < 1$ for p_H small enough where $\tilde{\rho}_1$ does not depend on p_H.

Thus, from inequality (E.35) we get that for p_H small enough,

$$\pi_S(k_H) \leq \tilde{\rho}_1 \pi_S(k_H + 1) + \tilde{\rho}_1(\pi_H(k_H + 1) - \pi_S(k_H + 1))$$

Now, by Lemma E.6, for any k_H,

$$\begin{aligned} \pi_H(k_H + 1) - \pi_S(k_H + 1) &= \sum_{k_E \notin S} \pi(k_H + 1, k_E) \\ &\leq \sum_{k_H=0}^{\infty} \sum_{k_E \notin S} \pi(k_H, k_E) \\ &\leq \frac{3 + 1/\sqrt{p_H}}{1 - \hat{\rho}} \left(\frac{\hat{\rho}^{1/\sqrt{p_H}}}{1 - \hat{\rho}} \right) \end{aligned}$$

with $\hat{\rho} \in (0, 1)$. Hence, for $k_H \leq k_H^*(1 - \varepsilon)$, we obtain

$$\begin{aligned} \pi_S(k_H) &\leq \tilde{\rho}_1 \left[\pi_S(k_H + 1) + \frac{3 + 1/\sqrt{p_H}}{1 - \hat{\rho}} \left(\frac{\hat{\rho}^{1/\sqrt{p_H}}}{1 - \hat{\rho}} \right) \right] \\ &\leq \tilde{\rho} \left[\pi_S(k_H + 1) + \phi(p_H) \right] \end{aligned}$$

with $\tilde{\rho} := \max(\hat{\rho}, \tilde{\rho}_1)$ and

$$\phi(p_H) := \frac{3 + 1/\sqrt{p_H}}{1 - \tilde{\rho}} \left(\frac{\tilde{\rho}^{1/\sqrt{p_H}}}{1 - \tilde{\rho}} \right).$$

An inductive argument yields

$$\pi_S(k_H^*(1 - \varepsilon) - i) \leq \tilde{\rho}^i \pi_S(k_H^*(1 - \varepsilon)) + \phi(p_H) \tilde{\rho} \frac{1 - \tilde{\rho}^i}{1 - \tilde{\rho}}. \tag{E.36}$$

□

Completion of the proof of Proposition C.6. We provide a proof only for the case $\mu < n(2\lambda - 1)$. The proof for the other case is the same. We fix $\delta > 0$ and claim that

$$\pi_H \left\{ k_H : \frac{k_H}{1/(-\ln(1 - p_H^2))} \leq \ln(2n\lambda) - \ln(n + \mu) - \delta \right\} \to 0$$

which is equivalent to showing that

$$\pi_H \left\{ k_H : k_H \leq \frac{1}{p_H^2} \left[\ln(2n\lambda) - \ln(n + \mu) \right] - \frac{1}{p_H^2} \delta \right\} \to 0.$$

Pick p_H and ε small enough so that

$$\frac{1}{p_H^2}\left[\ln(2n\lambda) - \ln(n+\mu)\right] - \frac{1}{p_H^2}\delta \leq k_H^*(1-\varepsilon) - 1/\sqrt{p_H}.$$

Clearly, for our purpose, it is enough to show that

$$\pi_H\{k_H \leq k_H^*(1-\varepsilon) - 1/\sqrt{p_H}\} \to 0$$

as p_H vanishes. In order to see this, observe that

$$\begin{aligned}
\pi_H\{k_H &\leq k_H^*(1-\varepsilon) - 1/\sqrt{p_H}\} \\
&= \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} \pi_H(k_H^*(1-\varepsilon) - i) \\
&= \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} [\pi_H(k_H^*(1-\varepsilon) - i) - \pi_S(k_H^*(1-\varepsilon) - i)] + \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} \pi_S(k_H^*(1-\varepsilon) - i) \\
&\leq \sum_{k_H=0}^{\infty} \sum_{k_E \notin S} \pi(k_H, k_E) + \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} \pi_S(k_H^*(1-\varepsilon) - i) \\
&\leq \phi(p_H) + \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} \pi_S(k_H^*(1-\varepsilon) - i) \\
&\leq \phi(p_H) + \sum_{i=1/\sqrt{p_H}}^{k_H^*(1-\varepsilon)} \tilde{\rho}^i \pi_S(k_H^*(1-\varepsilon)) + \sum_{i=0}^{k_H^*(1-\varepsilon)} \phi(p_H)\tilde{\rho}\frac{1-\tilde{\rho}^i}{1-\tilde{\rho}} \\
&\leq \phi(p_H) + \sum_{i=1/\sqrt{p_H}}^{\infty} \tilde{\rho}^i + \sum_{i=0}^{k_H^*(1-\varepsilon)} \phi(p_H)\tilde{\rho}\frac{1-\tilde{\rho}^i}{1-\tilde{\rho}} \\
&\leq \phi(p_H) + \frac{\tilde{\rho}^{1/\sqrt{p_H}}}{1-\tilde{\rho}} + k_H^*(1-\varepsilon)\phi(p_H)\frac{\tilde{\rho}}{1-\tilde{\rho}} \to 0
\end{aligned}$$

where the second inequality comes from Lemma E.6, the third from the Lemma E.8. In order to prove the convergence result let us first note that, as p_H vanishes, $\phi(p_H) \to \frac{\tilde{\rho}^{\frac{1}{\sqrt{p_H}}}}{(1-\tilde{\rho})^2\sqrt{p_H}}$ and, since k_H^* is of order $1/p_H$ when $\mu \geq n(2\lambda - 1)$ and of order $1/p_H^2$ when $\mu < n(2\lambda - 1)$, $k_H^*\phi(p_H)$ converges either to $\frac{c\tilde{\rho}^{\frac{1}{\sqrt{p_H}}}}{(1-\tilde{\rho})^2 p_H\sqrt{p_H}} = \frac{c\tilde{\rho}^{\frac{1}{\sqrt{p_H}}}}{(1-\tilde{\rho})^2(\sqrt{p_H})^3}$ or $\frac{c'\tilde{\rho}^{\frac{1}{\sqrt{p_H}}}}{(1-\tilde{\rho})^2 p_H^2\sqrt{p_H}} = \frac{c'\tilde{\rho}^{\frac{1}{\sqrt{p_H}}}}{(1-\tilde{\rho})^2(\sqrt{p_H})^5}$, with c and c' some positive constants. In both cases all those terms converge to zero if for any $\alpha \in (0,1)$ and any finite integer $n \geq 1$, we have $x^n\alpha^x \to 0$ as $x \to +\infty$. We know that it is true by equation (C.18). □

Following the same arguments as in Section C.4, we can provide the expressions of $p_H\mathbf{W}_H(Pairwise)$ and $p_H\mathbf{W}_E(Pairwise)$ when p_H vanishes.

F Comparative statics for Pairwise with DDL

In this section, we analyze how the asymptotic waiting time of hard to match patients evolves with n when λ is large and $\mu > n(2\lambda - 1)$. Let us denote by $\Omega(n;\lambda)$ this asymptotic (normalized) waiting time (i.e., $p_H W_H$ as p_H vanishes). We know from Proposition E.1 that

$$\Omega(n;\lambda) = \frac{c(n;\lambda)}{\lambda n}$$

with $c(n;\lambda)$ solution of the equation

$$n(1-\lambda)e^{-cp_E} + \mu e^{-c} = n(1-2\lambda) + \mu. \tag{F.37}$$

Hence, we must have

$$\Omega'(n;\lambda) = \frac{c'(n;\lambda)n - c(n;\lambda)}{\lambda n^2}.$$

We want to show the following.

Proposition F.1. *There exists a $\bar{\lambda} \in (1/2, 1)$ such that, for all $\lambda > \bar{\lambda}$, $\Omega'(n;\lambda) > 0$ for every $n < \mu/(2\lambda - 1)$.*

Proof. Note that the function $\Omega'(n;\lambda)$ is well defined for $n \in (0, \mu/(2\lambda-1))$ and $\lambda \in (0,1]$. Let us extend the domain of $\Omega'(n;\lambda)$ to (n,λ) such that $n \in [0, \mu/(2\lambda-1)]$ and $\lambda \in (0,1]$ by defining

$$\Omega'(\frac{\mu}{2\lambda-1};\lambda) := \lim_{n \to \frac{\mu}{2\lambda-1}} \Omega'(n;\lambda) = \infty$$

and

$$\Omega'(0;\lambda) := \lim_{n\to 0} \Omega'(n;\lambda) = \frac{\lambda - 2(1-\lambda)p_E}{2\mu^2}$$

for all $\lambda \in (0,1]$. Where the first limit can be directly deduced from (F.37) and the second limit uses Lemma F.2 below. Clearly, $\Omega'(n;\lambda)$ is continuous in both arguments on the extended domain.

Fix an arbitrary small $\varepsilon > 0$.[F.14] We consider the restriction of $\Omega'(n;\lambda)$ to the set of (n,λ) s.t. $n \in [0, \mu/(2\lambda-1))$ and $\lambda \in [1/2+\varepsilon, 1]$. Clearly, this restriction is continuous in both arguments over its domain. Now define $V : [1/2+\varepsilon, 1] \to \mathbb{R}$ as follows. For all $\lambda \in [1/2+\varepsilon, 1]$:

$$V(\lambda) := \inf_{n \in [0, \mu/(2\lambda-1)]} \Omega'(n;\lambda).$$

We first show that V is continuous at $\lambda = 1$. To show this, let us first remark that $V(\lambda) < 1/\mu^2$. Indeed, for all $\lambda \in [1/2+\varepsilon, 1]$

$$V(\lambda) \leq \Omega'(0;\lambda) = \frac{\lambda(1+2p_E) - 2p_E}{2\mu^2} \leq \frac{1 + 2p_E - 2p_E}{2\mu^2} < \frac{1}{\mu^2}$$

[F.14] So that $1/2 + \varepsilon < 1$. This ensures that $[1/2+\varepsilon, 1] \neq \emptyset$.

where the first inequality is by definition of $V(\lambda)$, the equality is by definition of $\Omega'(0;\lambda)$ and the second inequality is because $\Omega'(0;\lambda)$ is increasing in λ and $\lambda \in [1/2+\varepsilon, 1]$.

Now, define $\hat{\Omega}'(n;\lambda) := \min\{\Omega'(n;\lambda), 1/\mu^2\}$ and $\hat{V}(\lambda) := \inf_{n\in[0,\mu/(2\lambda-1)]} \hat{\Omega}'(n;\lambda)$. Observe that $\hat{\Omega}'(n;\lambda)$takes it values in $\mathbb{R}$ and is continuous on its domain and, in addition, the correspondence $\lambda \rightrightarrows [0, \mu/(2\lambda-1)]$ is compact-valued, non-empty valued and continuous, by Berge's Maximum Theorem, $\hat{V}$ is continuous on its domain. Now, it is clear that $\hat{V}(\lambda)$ exactly corresponds to $V(\lambda)$. Hence, V is also continuous on its, domain V is continuous at $\lambda = 1$.

To complete the proof, it remains to show that $V(1) > 0$. To do so we show that $\Omega'(n;1)$ is positive for all n. We get from (F.37), that $c(n;1) = \ln\left(\frac{\mu}{\mu-n}\right)$ such that $\Omega(n;1) = \frac{\ln\mu - \ln(\mu-n)}{n}$ and

$$\Omega'(n;1) = \frac{1}{n^2}\left[\frac{n/\mu}{1-n/\mu} - \ln\left(\frac{1}{1-n/\mu}\right)\right]$$

which is positive since $n < \mu$.

To sum-up, $V(\lambda)$ is continuous and $V(1) > 0$. Hence, there is $\bar{\lambda} > 0$ such that for all $\lambda > \bar{\lambda}$, $\Omega'(n;\lambda) > 0$ for all $n \in [0, \mu/(2\lambda-1))$. □

Below, we prove Lemma F.2 referred to in the proof of Propostion F.1 above.

Lemma F.2. $\lim_{n\to 0} \Omega'(n;\lambda) = \frac{\lambda - 2(1-\lambda)p_E}{2\mu^2}$.

Proof. Denoting

$$\begin{aligned} f(n) &:= c'(n;\lambda)n - c(n;\lambda) \\ g(n) &:= \lambda n^2 \end{aligned}$$

we have

$$\lim_{n\to 0} \Omega'(n;\lambda) = \lim_{n\to 0} \frac{f(n)}{g(n)}$$

Now, as n vanishes, we have $g(n) \to 0$ and $c(n;\lambda) \to 0$. Moreover, the differentiation of (F.37) with respect to c and n yields

$$c'(n;\lambda) = \frac{(1-\lambda)e^{-cp_E} + 2\lambda - 1}{p_E n(1-\lambda)e^{-cp_E} + \mu e^{-c}} \tag{F.38}$$

such that $c'(n) \to \lambda/\mu$ and $f(n) \to 0$. Hence we can apply l'Hospital's rule and we get that:

$$\lim_{n\to 0} \Omega'(n;\lambda) = \lim_{n\to 0} \frac{f'(n)}{g'(n)} = \lim_{n\to 0} \frac{c''(n)}{2\lambda}.$$

Using equation (F.38), we get that:

$$c''(n;\lambda) = \frac{u'(n)v(n) - v'(n)u(n)}{v(n)^2}$$

with

$$\begin{aligned}
u(n) &= (1-\lambda)e^{-c(n;\lambda)p_E} + 2\lambda - 1 \underset{n\to 0}{\longrightarrow} \lambda \\
v(n) &= p_E n(1-\lambda)e^{-c(n;\lambda)p_E} + \mu e^{-c(n)} \underset{n\to 0}{\longrightarrow} \mu \\
u'(n) &= -(1-\lambda)p_E c'^{-c(n;\lambda)p_E} \underset{n\to 0}{\longrightarrow} \frac{-\lambda(1-\lambda)p_E}{\mu} \\
v'(n) &= -n(1-\lambda)p_E^2 c'(n;\lambda)e^{-c(n;\lambda)p_E} + p_E(1-\lambda)e^{-c(n;\lambda)p_E} - \mu c'(n;\lambda)e^{-c(n;\lambda)} \underset{n\to 0}{\longrightarrow} p_E(1-\lambda) - \lambda
\end{aligned}$$

Hence we have

$$\lim_{n\to 0} c''(n;\lambda) = \frac{\lambda[\lambda - 2(1-\lambda)p_E]}{\mu^2}.$$

Finally we get that

$$\lim_{n\to 0} \Omega'(n;\lambda) = \frac{\lambda - 2(1-\lambda)p_E}{2\mu^2}$$

□

G The Omniscient Algorithm in Simulation Analyses

This appendix explains how to formulate the Omniscient algorithm in the simulations in Sections 3 and 4.2. As we shall see, the formulation depends on whether it uses DDL kidneys or not and whether incompatible pairs can exogenously exit. The objective of the algorithm is to minimize the sum of the (censored) waiting times of all patients, regardless of transplant status.

We need some additional notations. Let $\mathcal{P} = \mathcal{P}_\ell \cup \mathcal{P}_d$ where $\mathcal{P}_\ell$ is the set of patients in an incompatible pair and $\mathcal{P}_d$ is the set of "dummy patients" attached to each DDL kidney. A generic patient will be indexed by j as the patient in pair j. Let $\mathcal{D} = \mathcal{D}_\ell \cup \mathcal{D}_d$ denote the set of all kidneys to be donated, where $\mathcal{D}_\ell$ is the set of living donor kidneys and $\mathcal{D}_d$ is the set of DDL kidneys. By definition, $|\mathcal{D}_\ell| = |\mathcal{P}_\ell| := n_\ell$ and $|\mathcal{D}_d| = |\mathcal{P}_d|$. A generic donor, living or deceased, will be indexed by i. When $i = j$, donor i and patient j are in the same incompatible pair. We will also use d to denote a generic deceased donor. Below, we sometimes refer to a kidney in $\mathcal{D}$ as a donor, although a deceased donor may donate two kidneys.

For any donor kidney or patient k, we denote by $a(k)$ and $e(k)$ the arrival and the exit dates of k, respectively. For any two dates t and t', $t < t'$ means that date t is prior to date t', and $t' - t$ is the number of days between any two dates (which can be negative if date t' is after date t). For $d \in \mathcal{D}_d$, $a(d)$ is the date of arrival of the DDL kidney and the exit date is $e(d) = a(d)$ since a DDL kidney is only available on its arrival day.

The waiting time are calculated as follows. Let T^i_{ij} be the number of days donor i waits if she gives to patient j and T^j_{ij} be the number of days patient j waits if he receives from donor i. Note that $T^j_{ij} = 0$ if donor i arrives before patient j. For a patient or donor k, let $T^e_k = e(k) - a(k)$ be the number of days between its arrival and exit dates. Under the assumption of no exit, $T^e_k = T - a(k)$.

Finally, let G be the compatibility matrix between donors and patients so that $G_{ij} = 1$ if donor $i \in \mathcal{D}$ is compatible with patient $j \in \mathcal{P}$. A dummy patient associated with a DDL kidney is compatible with every donor. We also define $L = |\mathcal{P}| \times |\mathcal{D}_\ell|$.

Let $X_{ij} \in \{0, 1\}$, which is equal to 1 if donor i donates a kidney to patient j. The Omniscient algorithm, solves the following problem:[G.15]

$$\min_{(X_{ij})_{ij} \in \{0,1\}^L} \sum_{i \in \mathcal{D}, j \in \mathcal{P}} X_{ij} \times (T_{ij} - T^e_j) \tag{G.39}$$

s.t.

$$\forall i \in \mathcal{D}, j \in \mathcal{P} : (C_{ij}) \qquad X_{ij} \leq G_{ij} \times 1_{\{e(j) \geq a(i)\}} \times 1_{\{(i \in D_\ell) \vee (e(i) \geq a(j))\}}$$

$$\forall j \in \mathcal{P} : (F^p_j) \qquad \sum_{i \in \mathcal{D}} X_{ij} \leq 1$$

$$\forall i \in \mathcal{D} : (F^d_i) \qquad \sum_{j \in \mathcal{P}} X_{ij} \leq 1$$

[G.15] The total waiting time of all patients is $\sum_{i \in \mathcal{D}, j \in \mathcal{P}} X_{ij} T_{ij} + \sum_{j \in \mathcal{P}} (1 - \sum_{i \in \mathcal{D}} X_{ij}) T^e_j = \sum_{i \in \mathcal{D}, j \in \mathcal{P}} X_{ij} \times (T_{ij} - T^e_j)$.

$$\forall i \in \mathcal{D}_\ell : (E_i) \qquad \sum_{j \in \mathcal{P}: e(j) < e(i)} X_{ij} \leq \sum_{i' \in \mathcal{D}} X_{i'i}$$

$$\forall d \in \mathcal{D}_d : (DD_d) \qquad \sum_{\substack{i \in \mathcal{D}_\ell: \\ a(i) \leq a(d)}} \sum_{\substack{j \in \mathcal{P}_d: \\ a(j) \leq a(d)}} X_{ij} = \sum_{\substack{d' \in \mathcal{D}_d: \\ a(d') \leq a(d)}} \sum_{\substack{j \in \mathcal{P}_\ell: \\ a(j) \leq a(d)}} X_{d'j}$$

The constraints are:

- C_{ij}: *Compatibility constraints.* It takes into account that if a DDL kidney $i \in \mathcal{D}_d$ gives to a patient $j \in \mathcal{P}$, then this patient must arrive before the departure of the DDL kidney.
- F_j^p: *Feasibility constraints for patients.* Each patient can receive at most one kidney.
- F_i^d: *Feasibility constraints for donors.* Each donor kidney can be donated to at most one patient.
- E_i: *Exit constraints.* On the exit date, a living donor leaves with her patient but stays if her patient has received a kidney. Hence, if patient j does not receive a kidney, his associated donor $i = j$ leaves with patient j and thus cannot donate to any patient who arrives after their exit date. When we impose the no-exit assumption, this constraint is never binding.
- DD_d: *DDL kidney constraints.* We need to ensure that, at all the arrival dates of deceased kidneys, the number of living donors donating to patients on the DDL equals the number of DDL kidneys that have been used by the Omniscient.

When DDL kidneys are not used in the Omniscient algorithm, the constraints in DD_d are ignored. When pairs do not exit exogenously and stay until the end of the time period, constraints E_i are ignored.

We are interested in solving the Omniscient algorithm for different market sizes. The above problem is stated as an Integer Linear Programming (ILP). It is well known that ILP can be computationally challenging. Fortunately, in most of the simulations, the algorithm can use a polynomial-time solver. Below, we discuss each of the cases that we consider.

Implementation without exit or DDL kidneys. In Section 3, we consider the baseline version of the Omniscient algorithm that assumes no exit and does not use DDL kidneys. In this case, we formulate the Omniscient algorithm as finding a minimum-weight perfect matching.

The formulation goes as follows. We build a weighted bipartite graph. Let the set of nodes be $N := \mathcal{P}_\ell \cup \mathcal{D}_\ell$. To construct the set of edges E, we consider the complete graph such that $(j, i) \in E$ $\forall j \in \mathcal{P}_\ell$ and $\forall i \in \mathcal{D}_\ell$. For each edge (j, i), define the weight w_{ji} as follows:

$$w_{ji} = \begin{cases} T_{ij}^j & \text{if } G_{ij} = 1 \text{ and } a(j) \leq a(i) \\ T_j^e & \text{otherwise} \end{cases}$$

That is, the weight of an edge between patient j and donor i is equal to the waiting time of j on the date when i donates to j if they are compatible and if their arrival dates make it feasible to realize

a transplant. Recall that there is no exit here. If i and j are not compatible, the weight is equal to the maximum waiting time of patient j, i.e., T_j^e or $T - a(j)$. The tuple (E, N, w) defines a weighted bipartite graph.

A matching M is a set of edges $M \subset E$ such that any two distinct edges $(j,i) \in M$ and $(j',i') \in M \Rightarrow i \neq i'$ and $j \neq j'$. We say that a node is matched under M if it belongs to an edge of M. Note that donor i and patient j being matched here does not imply that there are a donation and a transplant because i and j can be incompatible. A matching M is a perfect matching if all the nodes in N are matched under M. Because our graph is complete, there exists a perfect matching. Given a perfect matching M, we let $w_M := \sum_{(j,i)\in M} w_{ji}$ be the weight of matching M. Note that, for a perfect matching M, given the definition of the weights, w_M corresponds to the overall waiting time of the patients. If $(j,i) \in M$ and $w_{ji} = T_j^e$, we interpret the edge as "patient j is not transplanted and donor i does not donate." If $(j,i) \in M$ and $w_{ji} = T_{ij}^j$, we interpret the edge as "donor j donates to patient i."

We say that M^* is a minimal weight perfect matching if it is a perfect matching and if there exists no other perfect matching M with $w_M < w_{M^*}$. Hence, a minimal weight perfect matching can identify the transplants that minimize the overall average waiting time of patients. The problem of finding a minimal weight perfect matching in a bipartite graph is standard and can be easily implemented with a polynomial algorithm that solves the linear assignment problem (Martello and Toth, 1987).

Implementation with DDL and no exit. This implementation is to obtain some of the results of the Omniscient algorithm in Section 4.2. For this case, we formulate the Omniscient algorithm as solving a minimum-cost flow problem.

Recall that Unpaired with DDL uses DDL kidneys under the constraint that if we use a DDL kidney, we need to immediately give back a living donor kidney to the DDL, which are constraints DD_d in optimization problem (G.39). In the same optimization problem, without pair exits, there are no constraints E_i.

We implement the Omniscient algorithm as follows. We start with building a weighted digraph.[G.16] Let $\mathcal{D}_{dd}$ be a set of dummy deceased donors which duplicates each deceased donor. For $d \in \mathcal{D}_d$, we denote d_c his copy in $\mathcal{D}_{dd}$. The set of nodes is $N := \mathcal{P}_\ell \cup \mathcal{D}_d \cup \mathcal{D}_{dd} \cup \mathcal{D}_\ell \cup \{s,t\}$ where s and t will be called "the source" and "the sink," respectively. The construction of the edges, E, and the weights, w, has several step. We first build edges between all patients in $\mathcal{P}_\ell$ and all living donors in $\mathcal{D}_\ell$, with their weights being the same as in the implementation without exists or DDL kidneys. Then, for each patient $j \in \mathcal{P}_\ell$ and deceased donor $d \in \mathcal{D}_d$, we add an edge (j,d) (i) if they are compatible, i.e., $G_{dj} = 1$ and (ii) if patient j arrives before donor d is available, i.e. $a(j) \leq a(d)$. The corresponding weight is the time that patient j waits in case of a donation from donor d, i.e., $w_{jd} = T_{dj}^j$. Each deceased donor $d \in \mathcal{D}_d$ is connected to his copy in $\mathcal{D}_{dd}$ by an edge $(d, dd) \in E$.[G.17]

[G.16] Contrary to the previous implementation of the Omniscient without DDL, the current graph is directed but not bipartite.

[G.17] As we shall see, these edges are used to guarantee that each deceased donor will only be used once.

For each dummy deceased donor $d_c \in \mathcal{D}_{dd}$ and living donor i, add an edge (d_c, i) if living donor i arrives before deceased donor d, of whom d_c is the copy, is available, i.e., $a(i) \leq a(d)$. These edges represent the constraint that "living donor i gives back to a patient on the DDL whenever deceased donor d is used." Lastly, $\forall j \in \mathcal{P}_\ell$, add an edge (s, j) with $w_{sj} = 0$ and $\forall i \in \mathcal{D}_\ell$, add an edge (i, t) with $w_{it} = 0$.

In the above weighted directed graph, the weights are interpreted as costs. In addition, each edge $(a, b) \in E$ is associated with a unit capacity $c_{ab} = 1$.

In sum, the source points to all patients at a zero cost. Patients are connected with living and deceased donors at costs equal to their waiting time if there is a transplantation or at costs being their maximum waiting time if there is no transplantation. Via their copies, all deceased donors point to their respective available living donors at no cost, and all living donors point to the sink at no cost.

A flow f is defined over the edges E where $f_{ab} \geq 0$ is the flow associated with edge $(a, b) \in E$. A flow is feasible if it respects the capacity of each edge, i.e., $f_{ab} \leq 1 \ \forall (a, b) \in E$. A flow is integral if f_{ab} is an integer for each $(a, b) \in E$.[G.18] For each $a \in N \backslash \{s, t\}$, a flow must be balanced, i.e. $\sum_{b \in N:(b,a) \in E} f_{ba} = \sum_{b \in N:(a,b) \in E} f_{ab}$. A flow f transports a quantity k from s to t if $\sum_{b \in N:(s,b) \in E} f_{sb} = \sum_{a \in N:(a,t) \in E} f_{at} = k$. The cost of flow f is $\sum_{(a,b) \in E} f_{ab} \times w_{ab}$. By the construction of the graph, because each edge has a unit capacity and because any living patient has an outgoing edge to any living donor, there always exists a feasible, balanced flow transporting quantity n_ℓ from s to t. Recall that n_ℓ is the number of patients in an incompatible pairs which is equal to the number of living donors.

The Omniscient algorithm amounts to finding a feasible, integral, balanced flow that transports quantity n_ℓ from s to t with the minimum cost because the cost of a flow is the sum of the waiting times of all the patients. Such a flow also respects constraints DD_d in problem (G.39): if $f_{jd} = 1$ for $j \in \mathcal{P}_\ell$ and $d \in \mathcal{D}_d$, then $f_{dd_c} = 1$ and it must be that $f_{d_c i} = 1$ for some $i \in \mathcal{D}_\ell$ and that $f_{ai} = 0$ for all $a \in (\mathcal{P}_\ell \backslash \{j\} \cup (\mathcal{D}_d \backslash \{d\})$. In other words, whenever patient j is matched with a deceased donor d at an integral feasible balanced flow, donor d (through his copy d_c) must be "matched" with an available living donor i who, in turn, cannot be matched to any other patient or deceased donor.

Hence, the omniscient algorithm has now been formulated as a minimum-cost flow problem that can be solved in polynomial time (see Korte et al., 2011). Moreover, the integrability theorem states that whenever all the capacities, costs, and flow to be transported are integers, there exists an integral flow solving the problem. The market sizes we consider allow us to simply solve the problem by linear programming.[G.19]

Implementation without DDL and exit. Appendix Table J.4 present results of the Omniscient algorithm without DDL but with exits. Because of constraints E_i, the Omniscient algorithm is implemented using ILP techniques by directly solving the optimization problem (G.39).

[G.18] In our specific case, integrability is equivalent to $f_{ab} \in \{0, 1\}$ for a feasible flow.

[G.19] In our implementation in Matlab, the linear programming solver does not guarante an integral solution. However, one can easily transform any non-integral solution into an integral one. The technique is similar to what is discussed in Galichon (2018). In our simulations, we have never encountered any non-integral solution.

H Data on Patients and Donors

We use administrative data from the *Agence de la Biomedecine* (ABM) who is in charge of organ allocation in France. Our sample period is December 2013 to February 2018. The dataset contains information on all the donors (deceased and living) who have been retrieved, all the patients who have been transplanted, and all the patient-donor pairs having participated in the Kidney Exchange Program (KEP) in France. We focus on the patient-donor pairs having ever participated in the KEP, the patient-donor pairs who have gone through a desensitization procedure, and the deceased donors. For our simulations, we use three types of information:

(i) the compatibility between any patient and any donor in our data;

(ii) the quality of a kidney from any donor in the data for any patient who has participated in the KEP or for a patient who has gone through desensitization;

(iii) the arrival date of each deceased donor kidney, the transplant date of each patient-donor pair who has gone through a desensitization procedure, and the registration date and the exit date of each patient-donor pair who have ever participated in the KEP.

In the following, we detail how we construct the relevant variables. Section H.1 explains how compatibility is calculated and how desensitization pairs and hypersensitized patients are defined, Sections H.2 and H.3 present how quality indices are constructed for kidneys from deceased and living donors, respectively, and, lastly, Section H.4 shows how these quality indices are used to select deceased-donor kidneys to be proposed to unpaired patients in the Unpaired with DDL algorithm.

H.1 Patient-Donor Compatibility and Some Definitions

Patient p_i and donor d_j are incompatible if they are either blood type incompatible or HLA (human leukocyte antigen) incompatible. We compare p_i and d_j's blood types to determine the blood type compatibility between them. To evaluate the HLA compatibility between p_i and d_j, we check if donor d_j has any antigens that is unacceptable to patient p_i: if d_j has at least one antigen that is unacceptable to p_i, p_i is HLA incompatible with d_j.

We define (p_i, d_i) as a desensitization pair if, in our data, p_i has obtained a kidney from d_i while p_i is incompatible with d_i. p_i is a *hypersensitized* patient if, in our data, p_i is HLA incompatible with at least 85 percent of all the donor kidneys (living and deceased) that have been retrieved or who have ever participated in the KEP between December 2013 and February 2018. Similarly, we define a patient to be *sensitized* if she is HLA incompatible with strictly more than 0 percent and less than 85 percent of all the donor kidneys. Finally, patients who are HLA compatible with all donor kidneys are called *non-sensitized* patients.

H.2 Quality Measure for Deceased-Donor Kidneys

We use the Kidney Donor Profile Index (KDPI) as a quality measure for deceased-donor kidneys. The KDPI, which lies in $[0, 100]$, is a relative measure. A kidney with a KDPI of $x\%$ is expected to

have higher risk of graft failure than $x\%$ of recovered deceased-donor kidneys (i.e., longer function than $(100-x)\%$ of recovered deceased-donor kidneys) in the U.S. in a reference year (for which we choose 2017). The expected risk of graft failure is measured by the Kidney Donor Risk Index (KDRI) of the donor. Below we review the calculation of the KDRI; we then explain how it is applied to our French data; in particular, how we deal with missing values; and, finally, we describe how the KDPI is obtained from the KDRI.

Calculation of Kidney Donor Risk Index (KDRI). The KDRI, developed by Rao et al. (2009), provides an estimated risk of graft failure after a transplant of a deceased donor kidney. Similar to common practice, we use a scaled, donor-only version of the KDRI (Organ Procurement and Transplant Network, 2019). It relies on 10 donor factors, including the donor's age, height, weight, ethnicity, serum creatinine, comorbidities (diabetes, hepatitis C virus—HCV, and hypertension), cause of death being cebrovascular accident (CVA) or not, and DCD status (being 1 for a donation after cardiac death or 0 otherwise).

The association between these variables and graft survival is estimated by a multivariate Cox proportional hazard model with graft outcomes from nearly 70,000 adults in the U.S. from 1995 to 2005. The estimated coefficients and the functional-form assumption lead to the following formula for the KDRI:

$$KDRI = \exp\left\{\begin{array}{l} 0.0128\times(age-40) - 0.0194\times(age-18)\times\mathbb{1}(age<18) \\ +0.0107\times(age-50)\times\mathbb{1}(age>50) - 0.0464\times\left(\frac{height-170}{10}\right) \\ -0.0199\times\left(\frac{weight-80}{5}\right)\times\mathbb{1}(weight<80) \\ +0.1790\times\mathbb{1}(\text{African American}) + 0.1260\times\mathbb{1}(\text{History of Hypertension}) \\ +0.1300\times\mathbb{1}(\text{History of Diabetes}) \\ +0.0881\times\mathbb{1}(\text{Cause of Death} = \text{CVA}) + 0.2200\times(Creatinine-1) \\ -0.2090\times(Creatinine-1.5)\times\mathbb{1}(Creatinine>1.5mg/dL) \\ +0.2400\times\mathbb{1}(HCV\ positive) + 0.1330\times\mathbb{1}(DCD) \end{array}\right\}$$

We apply this formula to our data and obtain a value of the KDRI for each deceased-donor kidney recovered in France during our sample period. Lehner et al. (2018) and Calvillo-Arbizu et al. (2018) follow the same methodology, using German and Spanish data, respectively. They confirm that the above formula, despite being estimated from a U.S. dataset, provides an accurate prediction of graft failure in these two populations.

Missing values. Donor ethnicity is not recorded in our data. We assume that all deceased donors are *Caucasian* as in Lehner et al. (2018) and Calvillo-Arbizu et al. (2018). Therefore, $\mathbb{1}(\text{African American}) = 0$ for all donors. For other variables, if there is a missing value, we set the variable at its sample mean in the calculation of the KDRI, as recommended by Organ Procurement and Transplant Network (2019).[H.20] For instance, if *History of Hypertension* is missing

[H.20] The same imputation is also used by Lehner et al. (2018) and Calvillo-Arbizu et al. (2018).

for a donor, we replace the dummy variable $\mathbb{1}$(History of Hypertension) by the proportion of donors having a history of hypertension; if the information on a donor's *serum creatinine* is missing, we assume that this donor has a serum creatinine at the mean level of serum creatinine among all the deceased donors in our sample.

Calculation of Kidney Donor Profile Index (KDPI) from KDRI The KDPI is a mapping of the KDRI from a relative risk scale to a cumulative percentage scale (Organ Procurement and Transplant Network, 2019). The reference population used for this mapping is all deceased donors in the U.S. with a kidney recovered for the purpose of transplantation in the prior calendar year. We use the year 2013 as the reference year.[H.21]

Table H.1 presents summary statistics. The median value for the KDPI is 65.73 (column 3), meaning that 65.73 percent of the deceased-donor kidneys proposed to patients in the U.S. are of a better quality than the median deceased-donor kidney in France. It is well-known that more kidneys, and thus some kidneys of a lower quality, are proposed to patients in France than in the U.S. This is also the case in other European countries, although at various degrees.[H.22] Columns (4)–(6) consider different selections among DDL kidneys to ensure that the quality of a DDL kidney is sufficiently high.

Table H.1: LKDPI and KDPI: Summary Statistics

	LKDPI within a Pair		KDPI for DDL Kidneys			
			All	Qualified for at least one patient among pairs in:		
	KEP	Desensitization		KEP	KEP & Desentitization	KEP (more selective)
	(1)	(2)	(3)	(4)	(5)	(6)
# of observations	78	508	13,036	6,142	13,019	4,230
Mean	104.45	89.33	65.73	39.43	65.74	29.33
Std Dev.	(32.39)	(41.15)	(29.41)	(19.27)	(29.40)	(15.02)
Median	125.42	125.42	71	42	71	31
Min	11.29	-4.66	0	0	0	0
Max	125.42	125.42	100	68	100	52

Notes: This table presents the summary statistics on the quality measure of a living donor kidney (LKDPI) or a DDL kidney (KDPI). LKDPI is calculated within a KEP pair or a desensitization pair. Columns (4)–(6) consider different selections among DDL kidneys in column (3) to ensure that the quality of a DDL kidney is sufficiently high. See Table 1 for more details on the sample in each column.

H.3 Quality Measure for Living-donor Kidneys

We use the Living Kidney Donor Profile Index (LKDPI) as a quality measure for living-donor kidneys. The LKDPI is developed by Massie et al. (2016) who use U.S. data to identify living donor characteristics associated with the risk of post-transplant graft failure. Importantly, the LKDPI is graft-specific, since it depends on characteristics of both the donor and the patient; by contrast,

[H.21] In other words, we use the OPTN mapping table produced in 2014 to map a KDRI to a KDPI. It is available at https://optn.transplant.hrsa.gov/media/3416/kdri-to-kdpi-mapping-table-2013.pdf; retrieved on August 2, 2021.

[H.22] Using German data, Lehner et al. (2018) conclude that 66 percent of the deceased-donor kidneys proposed in the U.S. are of a better quality than the median deceased-donor kidney in their data.

the KDPI is only donor-specific. We calculate an LKDPI for each incompatible patient-donor pair having ever participated in the KEP and for all desensitization pairs. The LKDPI is expressed on the same scale as the KDPI such that the two indices can be directly compared: if, for a given patient, the KDPI associated with a deceased-donor kidney is lower than the LKDPI associated with a living-donor kidney, the living-donor kidney has a higher expected graft failure risk.

Table H.1 presents summary statistics on the calculated LKDPI within each KEP or desensitization pair. Below we review the calculation of the LKDPI, explain how we deal with missing values, and describe how we assign an LKDPI to HLA incompatible pairs.

Calculation of the Living Kidney Donor Profile Index (LKDPI). The LKDPI, developed by Massie et al. (2016), combines 12 donor and patient factors to provide an index of graft failure risk after a living donation. The factors include the donor's age, estimated glomerular filtration (eGFR), Body Mass Index (BMI), ethnicity, history of cigarette use, and systolic blood pressure (SBP). There are also pair-specific factors: donation from a male donor to a male recipient, number of HLA-B mismatches, number of HLA-DR mismatches, donor/recipient weight ratio (D/R WR), and blood-type incompatibility.

The association between these factors and graft survival is estimated in a multivariate Cox proportional hazard model with graft outcomes from 36,025 living-donor kidney recipients in the U.S. during 2005–2013. The estimated coefficients and the functional-form assumption lead to the following formula for the LKDPI:

$$LKDPI = \exp\left\{\begin{array}{l} -11.3 + 1.85 \times (age - 50) \times \mathbb{1}(age > 50) - 0.381 \times eGFR + 1.17 \times BMI \\ +22.34 \times \mathbb{1}(\text{African American}) + 14.33 \times \mathbb{1}(\text{history of cigarette use}) \\ +0.44 \times SBP - 21.68 \times \mathbb{1}(\text{donor recipient both males}) \\ +27.30 \times \mathbb{1}(\text{donor-recipient blood-type incompatible}) \\ -10.61 \times \mathbb{1}(\text{donor recipient unrelated}) + 8.57 \times (\#\text{HLA-B mismatches}) \\ +8.26 \times (\#\text{HLA-DR mismatches}) - 50.87 \times \min\{\text{D/R WR}, 0.9\} \end{array}\right\}$$

We apply this formula to our data to calculate the LKDPI for each blood-type incompatible patient-donor pair who have ever participated in the KEP or went through a desensitization. Rehse et al. (2018) apply the same methodology to a German dataset and validate the results in Massie et al. (2016) in the German population. However, the above formula cannot be used to assess the quality of an HLA incompatible graft, which we discuss below.

Missing values. Similar to the calculation of the KDRI, missing values are replaced by the corresponding sample mean (see Section H.2 for more details).[H.23] Note that, even if we are only interested in the pairs that have participated in the KEP or desensitization, the reference population we use to calculate the sample means is all the pairs in our dataset, including 2,737 pairs that have ever participated in the KEP or whose patient has received a donated kidney from the paired

[H.23] Rehse et al. (2018) use the same imputation in their study using German data.

donor.

The *history of cigarette use* is missing for the vast majority of patients in our data. For this variable, similar to Rehse et al. (2018), we assume that this variable is zero when cigarette use is not mentioned in the patient's medical record. For the 8 pairs from Switzerland, our dataset contains a lot of missing values. However, these pairs are HLA incompatible and hence must be considered as special cases, as discussed below. Finally, for three blood-type incompatible pairs, most of the donor information is missing. These missing values are replaced by the corresponding sample means.

LKDPI for HLA incompatible pairs. HLA incompatibility is not taken into account in the LKDPI formula estimated by Massie et al. (2016). However, HLA incompatibility has a significantly negative association with graft survival (see, e.g., Bentall et al., 2013). In particular, an HLA incompatible graft is considered as being of a poorer quality than a blood-type incompatible graft. For this reason, we assign HLA incompatible pairs an LKDPI equal to the highest value of the LKDPI among HLA compatible pairs that have ever participated in the KEP (i.e., 125.42, columns 1 and 2 in Table H.1).

H.4 Selection of Deceased-Donor Kidneys

In Section 3.4, we allow DDL kidneys to be proposed to unpaired patients. We require such DDL kidneys to be of a sufficiently high quality and acceptable to unpaired patients. We consider two alternative ways for the selection:

- In the baseline simulations (Section 3.3), we consider that a DDL kidney d_j is acceptable to patient p_i if d_j's KDPI is lower than the LKDPI of the pair (p_i, d_i). In other words, a DDL kidney is considered as acceptable to p_i if the risk of graft failure associated with this kidney is lower than that associated with an incompatible graft between p_i and her paired living donor d_i. Column (4) of Table H.1 shows the summary statistics of the selected DDL kidneys for KEP pairs only. Relative to column (3), we keep less than a half of the DDL kidneys. Similarly, column (5) describes the kidneys selected for KEP and desensitization pairs. As there are 508 desensitization pairs, almost all DDL kidneys are selected for at least one pair.

- As a robustness check (Section J), we consider a more selective rule for DDL kidneys. Specifically, we first calculate the LKDPI among all HLA compatible KEP and desensitization pairs (only the existing pairs without re-pairing). This median is 52.76 in our sample, and we then consider a DDL kidney acceptable to patient p_i if the DDL kidney's KDPI is lower than 52.76. Column (6) of Table H.1 shows the summary statistics of the selected DDL kidneys for KEP pairs only with this selection rule. Relative to column (4), we keep fewer DDL kidneys (4,230) as a result of being more selective.

I Additional Simulation Results

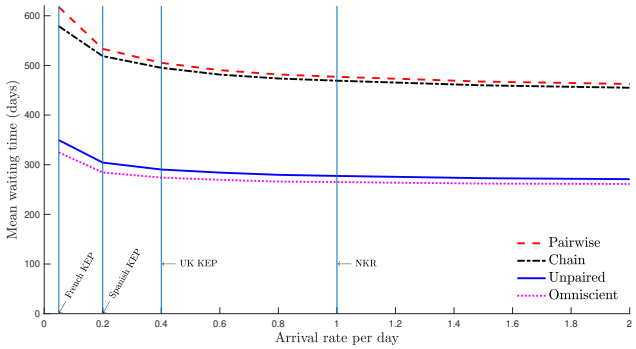

Figure I.1: Mean Waiting Time of Patients for Various Market Sizes

Notes: This figure shows the performance of the four algorithms in markets of eight different sizes, $n \in \{0.05, 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2\}$. The vertical lines indicate the size of some real-life KEPs.

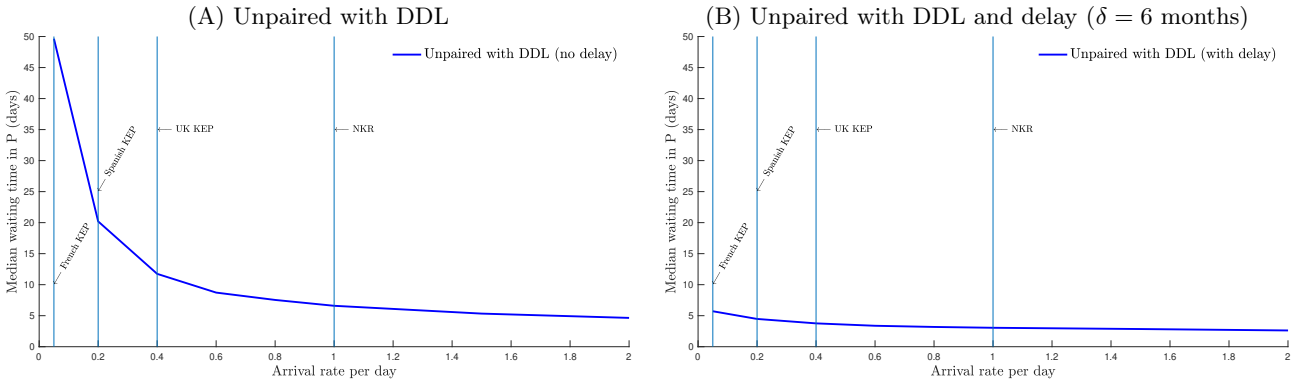

Figure I.2: Median Waiting Time of Unpaired Patients in P: Various Market Sizes

Notes: This figure shows the performance of the three algorithms with DDL in markets of eight different sizes, $n \in \{0.05, 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2\}$. The vertical lines indicate the size of some real-life KEPs.

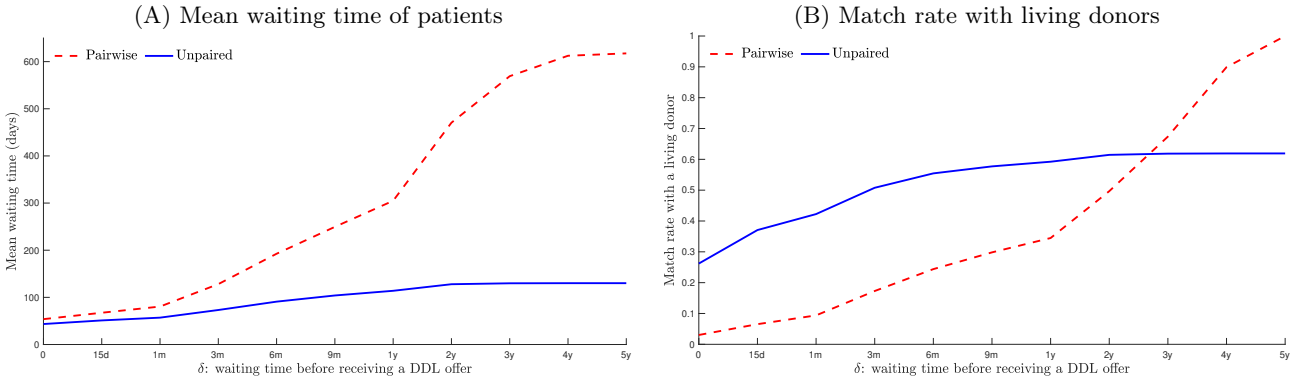

Figure I.3: Performance of Pairwise/Unpaired with DDL for Different Values of δ

Notes: This figure shows the performance of the two algorithms with DDL and different values of δ in markets of size $n = 0.05$ similar to that of the French KEP.

Table I.2: Pairwise/Unpaired with DDL, with or without Delay

	With DDL & alpha = 0 months		With DDL & alpha = 6 months	
	Pairwise	Unpaired	Pairwise	Unpaired
	(1)	(2)	(3)	(4)
Transplants				
% patients receiving transplant	94%	96%	85%	93%
% transplants from living donors	3%	26%	24%	55%
Reduction in the supply of O kidneys to the DDL[a]	41	37	34	30
O patients (in the KEP) receiving transplants	44	45	40	44
Total # of O patients benefited (KEP & DDL)[b]	3	8	6	14
Average waiting time (days)				
All patients (censored)	54	43	193	91
patients with an O donor	57	56	167	65
patients with an A donor	55	40	204	108
patients with an B donor	49	39	196	93
patients with an AB donor	44	23	209	62

Notes: The statistics are from the 1000 sets of simulations, each of which contains independent draws of pairs with a daily arrival rate of 0.05 (roughly the size of France's KEP). There are on average 83 incompatible pairs, among which 48 have an O patient.

[a] For a given algorithm, the reduction in the supply of O kidneys to the DDL is the number of O kidneys that the algorithm takes from the DDL minus the number of O living donor kidneys that it gives to the DDL.

[b] For a given algorithm, the total number of O patients (in the KEP and on the DDL) benefited is the number of O patients in the KEP who receive a transplant minus the reduction in the supply of O kidneys to the DDL.

J Robustness of the Unpaired Algorithm

We now present a set of robustness checks on the performance of our results. In most cases, we introduce a unique change into the baseline (columns 4 and 7 of Table 2) for the market size of $n = 0.05$.

Pairs opting out of donation-before-receipt. When a patient is hard-to-match, she may decide not to let her paired donor donate before she receives a transplant, while she is still willing to accept a transplant before her paired donor donates. We therefore consider the possibility of such patients opting out of donation-before-receipt (or entering P). That is, an opting-out pair remains in the KEP, and the donor cannot donate before the patient receives a transplant. We investigate the effects of two types of opt-out policies on the performance of the algorithms.

First, we let hypersensitive patients opt out. In the pool of 586 pairs, they account for 24 percent. As shown in Appendix Table J.3, relative to the baseline of no opting out, the overall waiting time significantly increases under Unpaired from 350 (column 1) to 455 days (column 2), while doubling under Unpaired with DDL, from 91 (column 4) to 184 days (column 5). The drop in transplant rate is also noticeable—from 63 to 50 percent under Unpaired, and 93 to 82 percent under Unpaired with DDL. Importantly, the reduction is larger among hypersensitized patients—15 percentage points under Unpaired and 39 percentage points under Unpaired with DDL.

Second, we let the patients with a PRA above 0.98, who are the hardest to match, opt out of donation-before-receipt. They account for 13 percent of our pool of 586 pairs. Columns (3) and (6) shows that the performance is worse than the baseline (columns 1 and 4) but not as bad as the first opt-out policy; again, the negative effects are mostly on hypersensitized or O patients.

One may alternatively let the pairs with a hypersensitized patient and an AB donor opt out of entering P. The rationale is that an AB donor is hard-to-match too, and an exchange may encourage them to opt out to reduce the overall donor waiting time in D. In our pool of 586 pairs, there are only 4 such pairs, less than 0.7 percent. As expected, the results with this opt-out policy, although not reported, are almost identical to the baseline.

In sum, the performance of the Unpaired algorithms decreases with the number of patients opting out of donation-before-receipt. However, the worst-case results, which are from the first opt-out policy, still dominate Pairwise and Chain in the baseline (Table 2) on every dimension. Importantly, these opt-out policies harm those who choose to opt out, reducing their incentive to opt out.

Higher-quality DDL kidneys in Unpaired with DDL. In the above simulations of Unpaired with DDL, we assume that a patient always accepts a DDL kidney when receives the offer. In practice, the algorithm should allow a patient to decide. Therefore, when a DDL kidney is not considered comparable to a living donor kidney, a patient may reject the DDL kidney. We test how sensitive our results are to this concern by screening DDL kidneys with a higher standard. To qualify for the algorithm, we now require a DDL kidney to have a KDPI below the median LKDPI among

Table J.3: Unpaired with DDL & $\delta = 6$ Months: Opting out and an Alternative Selection of DDL Kidneys

	Unpaired			Unpaired with DDL & $\delta = 6$ months			
		Opting out of P			Opting out of P		
	Baseline	Hypersensitized patients	Patients with PRA > 0.98	Baseline	Hypersensitized patients	Patients with PRA > 0.98	Higher-quality DDL kidneys
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Average waiting time (days)							
All patients (censored)	350	455	405	91	184	133	95
hypersensitized patients	*574*	*683*	*610*	*203*	*538*	*348*	*228*
O patients	*486*	*623*	*562*	*106*	*201*	*154*	*108*
Patients receiving transplant	176	205	192	72	89	77	74
hypersensitized patients	*281*	*266*	*280*	*145*	*239*	*154*	*166*
O patients	*311*	*443*	*373*	*83*	*99*	*87*	*84*
Transplants							
% patients receiving transplant	63%	50%	56%	93%	82%	88%	92%
hypersensitized patients	*40%*	*25%*	*35%*	*82%*	*43%*	*65%*	*80%*
O patients	*46%*	*30%*	*37%*	*91%*	*81%*	*86%*	*91%*
% transplants from living donors	100%	100%	100%	55%	46%	49%	56%
Patients going through P							
Total number	29	22	26	29	24	28	29
hypersensitized patients	*11*	*0*	*5*	*13*	*0*	*6*	*13*
O patients	*24*	*20*	*22*	*21*	*19*	*21*	*21*
Waiting time of patients in P							
Median	245	289	249	6	2	3	7
hypersensitized patients	*517*	*0*	*346*	*54*	*0*	*9*	*75*
O patients	*237*	*310*	*259*	*6*	*3*	*3*	*6*
Donors going through D							
Total number	26	20	22	36	24	29	36
AB donors	*4*	*4*	*4*	*4*	*4*	*4*	*4*
Waiting time of donors in D							
Median	339	297	314	39	3	5	46
AB donors	*618*	*589*	*604*	*48*	*4*	*7*	*55*

Notes: There are in total 83 incompatible pairs, among which 20 pairs have a hypersensitized patient and 48 have an O patient. The statistics reported are from the 1000 sets of simulations. The baseline, columns (1) and (4), is copied from columns (4) and (7) of Table 2. Columns (2)–(3) and (5)–(7) have the same simulations as in Table 2 but each has the following modification: columns (2) and (5) have hypersensitive patients opting out of entering P; columns (3) and (6) have patients with a PRA above 0.98 opting out of entering P; in column (7), a qualified DDL kidney for any patient must have a KDPI below the median LKDPI among all living donors (in both the KEP sample and the desensitization sample).

all *compatible* living donations in France in our sample period. Recall that the baseline only requires that the DDL kidney be compatible with the patient and have a KDPI below the LKDPI of the patient's paired incompatible donor. As expected, this higher quality standard reduces the supply of DDL kidneys, and thus the match rate with living donors increases, as shown in Column (8) of Table J.3, but very mildly from 55 to 56 percent. The results also show that it only slightly worsens the algorithm's performance relative to the baseline (column 4).

Allowing pairs to exit. We now relax the no-exit assumption. For each arriving pair as simulated in Section 3.2, they have a fixed daily probability of exiting. In our simulations, we fix this daily exit probability to 1% which corresponds to a monthly probability of 26%. Recall that $e(i)$ is the exit date for an incompatible pair. Furthermore, we make the following assumption: if (p_i, d_i) are still waiting at $e(i)$, both of them exit at $e(i)$; if d_i donates before $e(i)$ and if p_i has not received a kidney by $e(i)$, p_i leaves at $e(i)$; if p_i receives a kidney from someone else before $e(i)$, d_i stays until the end of our simulation, T. For pairs who do not exit in our data, we set $e(i) = T$.

With these assumptions, we simulate the first three algorithms (i.e., Pairwise, Chain, Unpaired) in the same way as before, while taking into account some pairs and patients may exit. We update the Omniscient algorithm as described in Appendix G to take into account exit constraints.

Appendix Table J.4 shows the same pattern across algorithms as in the baseline (Table 2), which implies that pair exits do not affect the performance ranking of the algorithms. Unpaired (column 4) remains similar to Omniscient (Column 6) in terms of transplant rate (57 percent vs. 58 percent), although less so in terms of waiting time (141 vs. 95 days).

Table J.4: Performance of the Matching Algorithms when Pairs May Exit

	Pairwise (2-way) Exchange (1)	2-way & 3-way Exchanges (2)	Chain & Pairwise Exchange (3)	Unpaired Exchange (4)	Omniscient (5)	With DDL & alpha = 6 months	
						Pairwise (6)	Unpaired (7)
Average waiting time (days)							
All patients (censored)	274	256	258	141	95	189	85
hypersensitized patients	*319*	*310*	*317*	*257*	*222*	*236*	*174*
O patients	*300*	*283*	*294*	*185*	*120*	*204*	*104*
Patients receiving transplant	156	155	142	98	34	157	63
hypersensitized patients	*140*	*170*	*142*	*165*	*88*	*206*	*125*
O patients	*233*	*210*	*232*	*171*	*44*	*180*	*84*
Transplants							
% patients receiving transplant	25%	29%	28%	57%	58%	46%	70%
hypersensitized patients	*12%*	*14%*	*12%*	*28%*	*31%*	*34%*	*48%*
O patients	*18%*	*22%*	*19%*	*44%*	*46%*	*42%*	*63%*
% transplants from living donors	100%	100%	95%	100%	100%	44%	72%
Patients going through P							
Total number	-	-	-	25	21	-	24
hypersensitized patients	-	-	-	*10*	*10*	-	*10*
O patients	-	-	-	*19*	*16*	-	*18*
Waiting time of patients in P							
Median	-	-	-	106	94	-	16
hypersensitized patients	-	-	-	*187*	*156*	-	*69*
O patients	-	-	-	*103*	*94*	-	*18*
Donors going through D							
Total number	-	-	4	30	40	-	32
AB donors	-	-	*1*	*4*	*4*	-	*4*
Waiting time of donors in D							
Median	-	-	195	334	297	-	119
AB donors	-	-	*344*	*610*	*581*	-	*163*

Notes: There are in total 83 incompatible pairs, among which 20 pairs have a hypersensitized patient and 48 have an O patient. The statistics reported are from the 1000 sets of simulations. The simulations are the same as in Table 2, except that pairs may exit before their patient receives a kidney. For more details, please see the notes of Table 2.

Compared with the baseline (column 4 of Table 2), the transplant rate of Unpaired is only slightly worsened when pairs may exit, 57 percent vs. 63 percent. The average (censored) waiting time is in fact lower when pairs may exit (141 vs. 350 days), because the waiting time is censored at exit date.

Regarding Unpaired with DDL and $\delta = 6$ months, compared with the baseline (column 7 of Table 2), the transplant rate is worsened significantly, 93 percent vs. 70 percent, although remaining

much higher than Unpaired and Unpaired with DDL. On the other hand, the average (censored) waiting time decrease slightly, 85 vs. 91 days.

Multiple chains. In practice, multiple altruistic donors may arrive, making multiple chains possible. To see how this can improve the performance of Chain, we draw a certain number of high-quality DDL kidneys as altruistic donors and assume that they arrive on the date they become a DDL kidney. A high-quality DDL kidney must have a KDPI below the LKDPI of one of the living donors in the KEP pairs. Similar to the previous simulations of Chain, we assume that after arrival, every DDL kidney remains available until either the end of our simulation or when it is transplanted. This may result in multiple chains. At the same time, 2-way pairwise exchanges are still allowed.

We take 58 DDL kidneys as the number of altruistic donors, corresponding to the numbers of DDL kidneys that Unpaired with DDL (and $\delta = 0$) uses in the baseline . Note that for every DDL kidney that Unpaired with DDL uses, a living donor donates a kidney to the DDL. By contrast, Chain does not require any living donor to donate to the DDL.

As expected, the results show that allowing for multiple chains improves upon the single-chain exchange. By using on average 41 altruistic donors, multiple chains reach a transplant rate of 83 percent, still lower than the rate of 93 percent under Unpaired with DDL and $\delta = 6$ months. Moreover, the average (censored) waiting time among all patients is 200 days, and that among transplanted patients is 113 days; both are higher than those under Unpaired with DDL and $\delta = 6$ months (91 and 72 days, respectively).