Priority and Externalities in Dynamic Matching

Itai Ashlagi * Elaine Liu† Pengyu Qian‡

July 1, 2026

Abstract

Voucher programs in kidney exchange give priority to patients whose incompatible donors donate before the patient receives a kidney. Such pairs create a positive externality by initiating chains, but priority for these patients can limit this externality by terminating chains early.

We study a dynamic matching market where each donor is compatible with each other pair's patient independently with probability p, and compare priority mechanisms that vary in how aggressively voucher patients are prioritized. We first analyze a *Standard Priority* rule, which extends a chain greedily until a voucher patient can be matched. Under this rule, a small fraction of voucher pairs q is enough (provided that $q \geq p$) for both voucher and non-voucher patients to obtain near optimal waiting times; when voucher pairs are too scarce, non-voucher patients have long waiting times. We use simulations to compare Standard Priority with two alternatives. A *Weak Priority* rule, which extends the chain provided that the next pair can terminate with a voucher patient, substantially reduces waiting times for non-voucher patients with only a small increase in voucher patient waiting times. A *Strict Priority* rule, which searches for a shortest path to a voucher patient, benefits voucher patients but at a greater cost to other patients. We show that a similar separation arises in asymmetric markets, where voucher patients are easy-to-match.

1 Introduction

Marketplaces often use priorities to encourage participation or actions that generate positive externalities for others. In organ allocation, priority may be granted to individuals who have

*Management Science and Engineering, Stanford University. Email address: iashlagi@stanford.edu.
†Management Science and Engineering, Stanford University. Email address: elaineliu@stanford.edu.
‡Questrom School of Business, Boston University. Email address: pqian@bu.edu.

previously donated an organ.[1] Such policies can improve efficiency, but they also raise equity concerns, as favoring one group may come at the expense of another. Moreover, the way priority is implemented may affect the positive externalities they can generate.

A motivating application for this study arises in kidney exchange. In kidney exchange, patients with an intended but incompatible living donor are to be matched with other pairs in order to receive a compatible kidney.[2][3] Traditionally, exchange programs require an incompatible patient-donor pair to receive a kidney no later than donating a kidney. This restriction limits exchanges to cycles or chains initiated by altruistic donors.[4] Chains can include many patients (Roth et al., 2006; Rees et al., 2009) and prior research shows that chains can substantially reduce waiting times, especially in sparse pools with many hard-to-match patients (Anderson et al., 2017). Recently, voucher programs emerged to relax the timing constraint in a different way: they allow a donor to donate before the intended patient receives a kidney, in return for priority in future matches (Veale et al., 2017).[5] Early-donating pairs therefore effectively initiate chains. But because these chains terminate with voucher patients, prioritizing these patients may limit their potential benefit.

This raises a basic design question: can priority for early donation benefit both voucher and non-voucher patients, or does it mainly reallocate gains toward voucher holders? On one hand, early donation relieves congestion by generating new chains. On the other hand, the priority promised to early donors determines when chains terminate. Therefore, how aggressively voucher patients are prioritized also affects how much of the chain's benefit is preserved for all other patients.

We study these questions in a stylized dynamic matching model motivated by kidney exchange. Patient-donor pairs arrive over time and differ in whether they are willing to donate before the intended patient receives a kidney. Upon arrival, a pair is willing (and therefore receives a *voucher*) independently with probability q.[6] Each donor is compatible with each patient from every other pair independently with probability p. After each arrival, the clearinghouse forms matches through chains initiated by *free* donors (a donor from a voucher pair is free upon arrival, whereas a donor from a non-voucher pair becomes free only after the intended patient has received a kidney).[7]

[1]See also (Sönmez et al., 2020) who propose to incentivize compatible pairs to join kidney exchange by receiving priority in the deceased donor waiting list. Priority policies also arise outside organ allocation; for example, vaccination programs often prioritize essential or volunteer workers in high-contact roles.

[2]See (Roth et al., 2004, 2007) for the emergence of kidney exchange.

[3]In 2025 there were 1408 kidney transplants facilitated through kidney exchanges in the United States (`https://www.hrsa.gov/optn?from=optn.transplant.hrsa.gov`).

[4]An altruistic donor has no intended patient.

[5]For example, the National Kidney Registry introduced a voucher program; see (Veale et al., 2017, 2021).

[6]The pair type is determined exogenously and we abstract away from strategic behavior.

[7]Matched patients and donors leave the market and we assume patients do not leave without a match.

We compare priority rules that differ in how strongly they prioritize a voucher patient. Our baseline rule, Standard Priority, extends a chain greedily and terminates it once a voucher patient can be matched. We also study weaker and stronger priority rules that respectively allow chains to continue longer before serving a voucher patient, or search more aggressively for the shortest path to a voucher patient.

We find that Standard Priority can generate almost-optimal waiting times for both voucher and non-voucher patients as long as the portion of voucher pairs is large relative to the stochastic compatibility (Theorem 1 and Proposition 2), assuming $p, q \to 0$.[8] More formally, when $q > p$, both voucher and non-voucher patients have almost-optimal expected waiting times in steady state. With high probability, the steady-state waiting time for voucher patients is $O\left(\frac{1}{p}\left(\log\frac{1}{q}\right)^{1/2+\epsilon}\right)$ and for non-voucher patients it is $\Theta\left(\frac{1}{p}\log\frac{1}{q} + \frac{1}{q^{1+\epsilon}}\log\frac{1}{q}\right)$, for all $\epsilon > 0$. When the fraction of voucher pairs is low — in particular $q < p$ — the system has no steady state, with the expected steady-state waiting time for voucher patients being $o(1)$, and the steady-state waiting time for non-voucher patients being unbounded.

Some intuition for this result is the following. In a market without voucher patients, an altruistic donor-initiated chain extends indefinitely and the expected waiting time is $\tilde{O}(1/p)$ (Anderson et al., 2017; Blum and Mansour, 2020).[9] In our setting free donors initiate chains but they terminate when a voucher patient is matched. Therefore a sufficient fraction of voucher pairs is required to maintain a steady state for the queue of non-voucher pairs. One implication of the result is that forming cycles remains a complementary way to serve non-voucher pairs when there are very few voucher pairs.[10] See further discussion and simulations in Sections 3.1 and 4.

The design of priority rules and how aggressive voucher patients are prioritized matters. We identify a *Weak Priority* rule that can generate bounded waiting times for non-voucher patients even when $q < p$, while keeping waiting times for voucher patients almost optimal.[11] Weak Priority is more flexible than Standard Priority without sacrificing priority: whenever a chain can terminate with a voucher patient, it instead proceeds in a greedy manner to another non-voucher pair provided that the donor of that pair can donate to a voucher patient. This modification results in a bounded waiting time for non-voucher patients, even when Standard Priority alone leads to an unbounded waiting time. The differences are more pronounced in heterogeneous markets: if voucher patients are easy-to-match, waiting times

[8] The assumption that $q \to 0$ is imposed for technical reasons, and simulations confirm that the result is robust to larger values of q.

[9] Here $f(n) = \tilde{O}(g(n))$ means $f(n) = O(g(n)\text{polylog}(n))$.

[10] The Standard Priority analysis does not consider cycles for tractability reasons.

[11] While the result is similar to Standard Priority, the proof technique does not apply due to correlations. Instead, we measure waiting times through simulations.

of non-voucher pairs may be unbounded under Standard Priority for any $q < 0.5$ but under Weak Priority for all q (Theorems 2 and 3). The *Strict Priority* rule, which always searches for a shortest path to a voucher patient, generates very short chains and therefore harms non-voucher patients for a larger threshold q in comparison to Standard Priority.[12]

Section 4 provides a comparison of the different priority rules using numerical simulations. The findings and trade-offs are robust to variants of the Standard, Weak, and Strict Priority rules, in which non-voucher pairs can also match through short cycles.

The paper makes two technical contributions towards analyzing the steady state of the system under Standard Priority. A challenge in obtaining the equilibrium is that, queue lengths (and by Little's Law expected waiting times) of voucher and non-voucher patients are interdependent and lead to a feedback loop: A larger voucher pool helps initiate chains that remove many non-voucher patients; a larger non-voucher pool makes it easier to find a compatible donor for a voucher patient; a smaller voucher pool ensures the chain of non-voucher pairs does not end prematurely when it finds a compatible voucher patient. To resolve this feedback loop we develop an inductive technique that iteratively tightens the bound on the length of the two interdependent queues in steady state. For insights into the iterative method see Section 5.2. This method may be useful for analyzing equilibrium behavior of dynamic markets with endogenous supply and demand.

The second technical contribution establishes a high probability bound on the length of queues with unbounded departures. Unlike classical queuing settings, where departures are bounded (Bertsimas et al., 2001), the priority mechanism can remove a large number of non-voucher agents at once through long chains. This leads to an unbounded decrease in the queue. Lemma 1 formalizes this key technical contribution that allows for the analysis of systems with unbounded jumps. The argument may be useful more broadly for dynamic markets with batch clearing.

1.1 Related Literature

Our paper contributes first to the literature on kidney exchange and dynamic matching. Early work introduced kidney exchange through cycles and altruistic-donor chains (Roth et al., 2004, 2007; Ünver, 2010). Anderson et al. (2017) study a stochastic dynamic barter model and compare matching using short cycles and chains initiated by altruistic donors. They find that greedy policies are near-optimal and that chains deliver substantially shorter waiting times than short cycles in sparse markets. Blum and Mansour (2020) extend these results to greedy (maximal) chains. Ashlagi et al. (2019) identify that in a market with

[12]Similar to Weak Priority, Strict Priority introduces dependencies that our proof techniques are not equipped to analyze. We instead compute the average waiting times using simulations.

easy- and hard-to-match agents, the imbalance in the market plays an important role for the desirable matching technology. This literature establishes the value of chains. We build on these findings, but study a different design problem: when chains are initiated by early-donating pairs who receive priority in return, the priority rule itself determines when chains terminate and therefore how the benefits of early donation are distributed across voucher and non-voucher patients.

A closely related literature studies how kidney exchange can relax the timing constraint that a pair must receive a kidney no later than donating one (Ausubel and Morrill, 2014; Akbarpour et al., 2020). Akbarpour et al. (2020) analyze waiting times under the assumption that all pairs are uncoupled and compare the performance to standard approaches that use altruistic donor chains and cycles. This paper instead studies a voucher-like environment, in which only some pairs are voucher pairs, and compares rules for prioritizing such pairs. Thus, our paper does not relax the timing constraint for the whole market, but is concerned with how priority affects the chain externality created by early-donating pairs.

Our paper contributes also to the broader literature on designing priorities in matching markets (Abdulkadiroğlu et al., 2009; Ashlagi and Shi, 2016; Bloch and Cantala, 2017; Çelebi and Flynn, 2022b; Kominers and Sönmez, 2016; Agarwal et al., 2019; Shi, 2022; Çelebi and Flynn, 2022a; Akbarpour et al., 2024; Sweat, 2025). We contribute to this literature by considering markets in which agents differ based on their contribution to adding thickness or easing congestion.

The paper also relates to the literature on exchanging favors and scrip systems (Möbius, 2001; Friedman et al., 2006; Hauser and Hopenhayn, 2008; Kash et al., 2015; Johnson et al., 2014; Ashlagi et al., 2026). These papers study whether cooperation can be sustained in equilibrium, and how a marketplace should prioritize agents based on their previous contributions. Our setting has a similar flavor, since early donation creates a claim for future priority. The key difference is that the "favor" is a single kidney donation, and the priority claim affects a dynamic compatibility-based matching process in which the same donation can also initiate a chain for others.

Finally, this paper contributes to the literature on Lyapunov methods for stochastic systems. Our work expands upon the rich literature on queuing systems, e.g. Kleinrock (1975), Asmussen (2003). Yet, our model departs from conventional queuing models as agents in our model are both "jobs" and "servers." We use classical Lyapunov methods as in Bertsimas et al. (2001) to give lower and upper bounds on queue lengths. We also provide a technical lemma that bounds the expected lengths of queues with unbounded, geometric decrease. This helps us obtain the lower bound on the expected waiting time for non-voucher patients.

1.2 Notation

We write $f(x) = o(g(x))$, or equivalently, $f(x) \ll g(x)$, to denote that $\lim_{x \to 0^+} \frac{f(x)}{g(x)} = 0$. We adapt the $\omega(\cdot)$ notation analogously. We write $f(x) = O(g(x))$ if there exists $C < \infty$ such that $|f(x)| \leq C \cdot g(x)$ as $x \to 0^+$. We adapt the $\Theta(\cdot)$ and $O(\cdot)$ notations analogously.

1.3 Organization

The rest of the paper is organized as follows. We describe our model formally in Section 2. We state the main results on both homogeneous and heterogeneous markets, discuss their insights and implications, and provide relevant simulations in Section 3. We emphasize that, with as little as p-fraction of voucher pairs, all agents achieve near-optimal waiting times. We also emphasize the delicate nature of priority mechanisms and how the outcome varies by market characteristics like the concentration of voucher pairs and compatibility heterogeneity. Section 4 provides simulation results, robustness checks, and comparative statics across different compatibility probabilities and mechanisms. We also give further intuition for the threshold effect under Strict Priority. In Section 5, we provide a proof sketch to our main theorem and highlight our technical contributions. We discuss in Section 6 implications of our results and a few extensions.

2 Model

We consider a discrete-time, infinite-horizon model of a kidney exchange market in which incompatible pairs arrive and match over time.

Each incompatible pair consists of a patient and an incompatible donor. Pairs are of two types, W (voucher) and U (non-voucher), distinguished by whether the pair participates in the voucher program. We assume that any donor is compatible with any patient from another pair with probability p, independently across donor–patient pairs, following the standard online random graph model used in kidney exchanges (Anderson et al., 2017).

Time is discrete, and one pair arrives in each period. The arriving pair is of type W with probability q and type U with probability $1 - q$. Each patient and donor exits the market upon matching. For a non-voucher pair (p, d), we assume that d does not exit before p. This restriction need not hold for a voucher pair.

It is convenient to represent the pool as a directed compatibility graph G from donors to patients. The compatibility graph is stochastic and the nodes and edges update with arrivals and departures. Compatibilities are realized upon each arrival: edges from the arriving donor to all patients in the pool, and from all existing donors to the arriving patient.

Remark. Our basic model assumes homogeneous types and abstracts away from blood types and other biological details that arise in the context of kidney exchange. One interpretation is that the model focuses on a submarket with identical blood types, such as O–O pairs, where incompatibility is primarily driven by tissue type. Blood types create structured imbalances.[13] This imbalance may make it undesirable to allow pairs who face a shortage of supply to donate early (i.e., to be voucher pairs). See Section 3.2 for discussion about heterogeneous types.

2.1 Priority Mechanisms

We are interested in studying matching mechanisms that assign priority to voucher patients. We restrict attention to mechanisms that attempt to form matches upon arrival of each pair. Matched patients and donors leave the pool immediately.

Mechanisms can vary in how aggressively priority is implemented. We define three priority mechanisms: Standard, Weak, and Strict. Our analysis focuses on the Standard Priority mechanism.

To describe the mechanisms, it is helpful to decompose the pool of patients and donors at any given time as follows. Let C (for coupled) be the set of non-voucher patient-donor pairs in the pool, i.e., both the patient and the donor are still unmatched. Let P (for *priority*) be the set of patients of voucher pairs in the pool. Finally, let D (for donor) be the set of *free* donors who can donate at any time: these are donors of voucher pairs and donors of non-voucher pairs whose paired patient has already matched. The sets D, P and C partition all donors and patients in the pool and are updated after every arrival and after every match as follows. When a pair (p, d) arrives, (p, d) is added to C if it is a non-voucher pair, and otherwise p is added to P and d is added to D. When a patient of a non-voucher pair matches, her paired (incompatible) donor, if still unmatched, is added to D. Note that the sets P and D are of equal size at all times.

At times, we slightly abuse the notation by saying D can donate to patient p to mean that there exists a donor in D that is compatible with patient p. Similarly, saying that a donor d can donate to P means that d is compatible with at least one patient in P. Also, when we say a donor "donates to" or "matches" a patient, it is implied the donor is compatible with that patient.

Standard Priority. In each period, a patient–donor pair (p, d) arrives. If it is a voucher

[13] For example, in practice there are more O-A than A-O patient-donor pairs because many O donors are compatible with their intended A patient and do not join exchange while the A donors ABO incompatible with their intended patient.

pair, the mechanism simultaneously attempts to match the arriving patient p with a free donor in D and to initiate a chain with the arriving donor d. If it is a non-voucher pair, the mechanism initiates a chain with d only if p can be matched with a donor in D (in which case d becomes free). In either case, the chain proceeds greedily while always prioritizing patients in P. The formal description of the Standard Priority mechanism is as follows.

For each arriving pair (p, d), update D, P, C, and then:

(a) *Match the arriving patient*: If some free donor in D is compatible with p, then form such a match. If a match is formed and $(p, d) \in C$, move d to the set of free donors D.

(b) If $d \in D$ then, *proceed with a greedy chain from* d as follows:

  i. If d is compatible with a patient in P: form this match and end the chain.

  ii. Otherwise, if d is compatible with a different patient p' in C, form this match, add its paired donor d' to D, and reiterate this step to proceed with the chain from d' (replacing d with d').

  iii. If neither (i) nor (ii) applies, terminate the chain.

When multiple donations are possible, the mechanism resolves ties by order of arrival. The following example illustrates the Standard Priority mechanism.

Example 1. *This example starts at the end of the fourth period. The red circles are patients and donors in sets P and D, and the blue circles are pairs in C. The subscript denotes their arrival period, and dashed lines denote compatibility. We separate voucher patients from their paired donor to emphasize that, under rules of the priority mechanism, the voucher pairs can be treated as decoupled agents.*

Prior to the fifth arrival, no matches occur. This is because the only free donor, d_1, is not compatible with any patient. In the fifth period, the voucher pair (p_5, d_5) arrives and d_1 and p_5 are immediately matched. Simultaneously, d_5 is compatible with p_2 and p_3, and after breaking ties by arrival, the chain proceeds with a donation to p_2. p_2's paired donor d_2 continues the chain by donating to p_4. At this point, no further donation is possible, as d_4 is not compatible with any patient. The solid arrows show the realized chain. Matched agents $(d_1, p_5, d_5, p_2, d_2, p_4)$ leave the pool. Since the chain ended with a non-voucher patient p_4, its paired donor d_4 is now free and is moved to the set D. At the end of the fifth period, the only remaining agents are p_1, d_4, and the pair (p_3, d_3). Note that the number of agents in P and D remains the same, and even though p_1 and d_4 did not arrive as a pair, they are stochastically equivalent to the case in which they did.

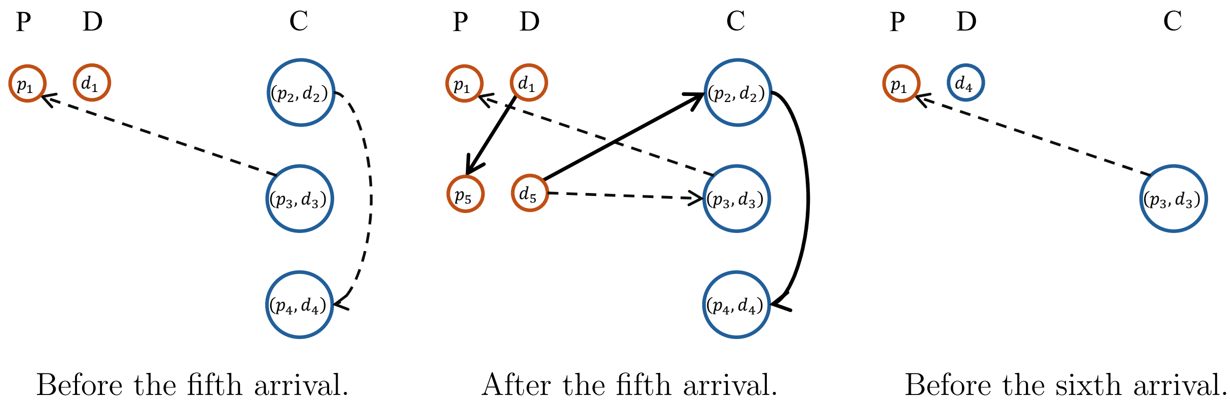


Figure 1: Before, during, and after a chain removal under Standard Priority. Each red circle is a voucher patient or donor, and each blue circle is a non-voucher pair, each labeled by their identity. The dashed arrows are compatible matches, and solid arrows are matches chosen by the mechanism. Initially, only p_1, d_1, and pairs (p_2, d_2) (p_3, d_3) (p_4, d_4) are in the market, but no matches are formed because the only voucher donor d_1 cannot start a chain. When p_5, d_5 arrive, d_1 matches with p_5, and d_5 starts a chain that ends with (p_4, d_4). When matched pairs leave the market, d_4 becomes a free donor before the sixth period starts.

Weak Priority. Weak Priority differs from Standard Priority only in step (b)i, which governs how the mechanism extends a chain from a free donor $d \in D$. When d is compatible with a voucher patient, instead of matching them immediately and ending the chain as in Standard Priority, Weak Priority first checks whether the chain can be extended by matching d to a non-voucher pair in C whose paired donor can donate to some voucher patient (thus preserving priority). If no such extension is available, the chain proceeds as in Standard Priority.

(b)i. If donor d is compatible with some voucher patient $p' \in P$: If there exists a pair $(p'', d'') \in C$ such that d is compatible with p'' and d'' can donate to P, then match d with p'' and start again (b)i with the new donor d'' (replacing d). Otherwise, match d with p'.

Strict Priority. Strict Priority follows Standard Priority except in step (b)i, where it selects the *shortest directed path* from the current donor to a voucher patient.

(b)i. If there exists a path from d to some patient in P, find a shortest directed path from d to any voucher patient in P and match donors to patients along this path.

We illustrate the three priority mechanisms with Example 2, highlighting how they differ in chain construction. We show how the three mechanisms remove different chains despite starting with the same pool of agents.

Example 2. *Figure 2 illustrates a snapshot of the compatibility graph. Red circles denote voucher pairs (patients in P and donors in D), and blue circles denote non-voucher pairs (set C). A dashed arrow from (p_i, d_i) to (p_j, d_j) indicates that d_i can donate to p_j. Suppose pairs $(p_1, d_1), \ldots, (p_7, d_7)$ arrive in that order, and that (p_1, d_1), (p_5, d_5), and (p_7, d_7) are voucher pairs. Upon the arrival of the last voucher pair (p_7, d_7), we initiate a chain with donor d_7. Before (p_7, d_7) arrives, no matches have occurred, so the set of free donors D coincides with the set of voucher donors. Figure 3 illustrates the chains removed under Standard, Weak, and Strict Priority.*

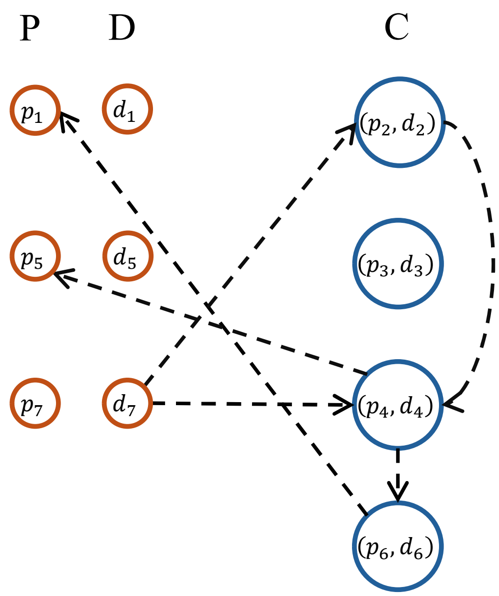
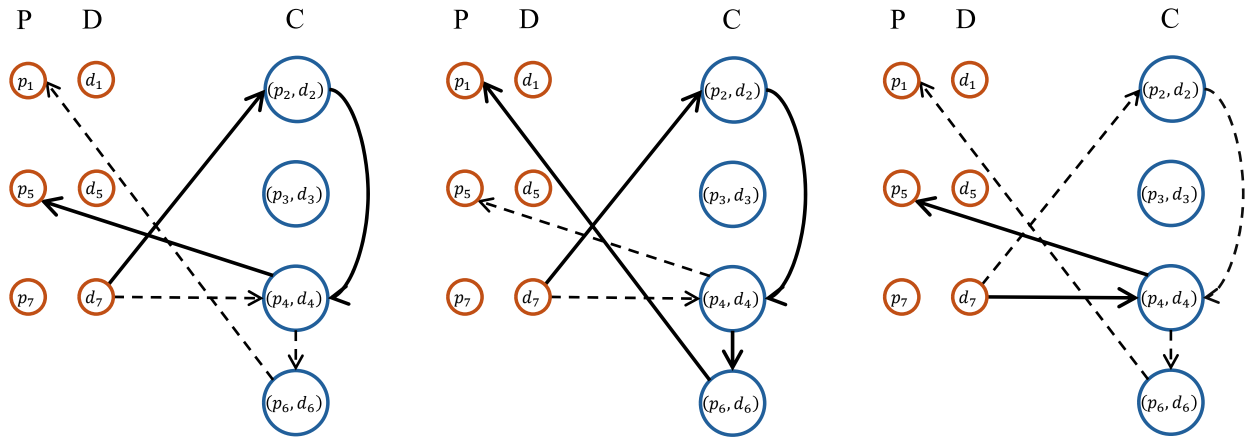


Figure 2: Underlying compatibility graph. The dashed arrows denote compatible matches.

Under Standard Priority, the mechanism starts a greedy search from d_7. Donor d_7 is incompatible with all voucher patients but compatible with p_2 and p_4. The mechanism matches d_7 to p_2, which wins the tie against p_4 by earlier arrival, and continues the chain from d_2. Donor d_2 then donates to p_4, whose paired donor is compatible with a voucher patient. The chain ends with d_4 donating to p_5. The resulting chain is highlighted by solid lines in Figure 3(a).

Under Weak Priority, the mechanism proceeds as under Standard Priority until it identifies the match from d_4 to p_5. Instead of donating to p_5 immediately, it searches for a non-voucher pair in C that can receive from d_4 and whose paired donor can donate to P. Pair (p_6, d_6) satisfies this condition. The mechanism therefore extends the chain to p_6 before donating to P. The resulting chain is $d_7 \to (p_2, d_2) \to (p_4, d_4) \to (p_6, d_6) \to p_1$, as highlighted in Figure 3(b).

Under Strict Priority, the mechanism searches for a shortest directed path from d_7 to the *set of voucher patients. In Figure 3(c), the path $d_7 \to (p_4, d_4) \to p_5$ is such a shortest path.*



(a) Standard (b) Weak (c) Strict

Figure 3: Match found under Standard, Weak, and Strict Priority mechanisms from the same market in Figure 2. Different matches are realized under different mechanisms.

In what follows, we analyze the dynamics of the Standard Priority mechanism. At any time, the state of the marketplace is represented by the compatibility graph G defined in Section 2. Each node represents a donor-patient pair, and an edge points from one pair to another if the donor in the former pair is compatible with the patient in the latter.

In the analysis, rather than realizing all compatibility edges when a pair enters the market, we find it more convenient to use an equivalent construction: an edge is revealed only when the policy first needs to determine whether the corresponding donor is compatible with the corresponding patient. Let $X(t)$ and $Y(t)$ denote the numbers of voucher and non-voucher patients in the system, respectively, immediately before the arrival of the $(t+1)$-st pair. By the *principle of deferred decisions* (Knuth, 1997), the count process $(X(t), Y(t))_{t\geq 0}$ induced by the deferred-revelation construction has the same law as the corresponding process in the original system. Under this construction, it is sufficient to track the count state $(X(t), Y(t))$, which evolves as a Markov process. Because our analysis focuses on the steady-state behavior of the system, we first establish the existence of a stationary distribution for $(X(t), Y(t))$.

Proposition 1 (Existence of steady state under Standard Priority)**.** *Fix $0 < p \leq q < 1$. Under the Standard Priority mechanism, the Markov chain*

$$Z(t) = (X(t), Y(t)) \in \mathbb{Z}_+^2$$

admits a stationary distribution.

The proof of Proposition 1 is given in Appendix B.

We measure efficiency by average waiting time. In our model, minimizing average waiting time is equivalent to minimizing the average number of agents in the system, since the two are proportional by Little's Law (Little, 1961). Accordingly, we focus on the steady-state queue lengths.[14]

3 Results

In this section, we present the main performance guarantees for Standard Priority. We first bound the steady-state queue lengths of voucher and non-voucher patients, then relate these bounds to existing results from the kidney exchange literature, and finally discuss an extension to heterogeneous patient compatibilities.

Let $\mathbf{X}^{\mathrm{SP}}$ and $\mathbf{Y}^{\mathrm{SP}}$ denote the random steady-state queue lengths of voucher and non-voucher patients, respectively, under Standard Priority.

Theorem 1. *Consider the Standard Priority mechanism with compatibility probability p and voucher-pair proportion q. Let $\mathbf{X}^{\mathrm{SP}}$ and $\mathbf{Y}^{\mathrm{SP}}$ denote the steady-state queue lengths of voucher and non-voucher patients, respectively. As $q \to 0$ and $0 < p \le q$, there is a finite constant C such that, for every fixed $\epsilon > 0$,*

$$\mathbf{X}^{\mathrm{SP}} \;\le\; C\frac{q}{p}\Big(\log\frac{1}{q}\Big)^{1/2+\epsilon}$$

holds with high probability.[15] *Moreover, for $q \to 0$ and $q \ge p^{1-\delta}$ for some $\delta > 0$,*

$$\mathbf{Y}^{\mathrm{SP}} \;\le\; \frac{1+\epsilon}{p}\log\frac{1}{q}\,.$$

holds with high probability.

Remark 1. *Theorem 1 also yields results on waiting time proxies. Define*

$$\widehat{W}_{\mathrm{W}}^{\mathrm{SP}} \triangleq \frac{\mathbf{X}^{\mathrm{SP}}}{q}\,, \qquad \widehat{W}_{\mathrm{U}}^{\mathrm{SP}} \triangleq \frac{\mathbf{Y}^{\mathrm{SP}}}{1-q}$$

where the subscripts W and U denote voucher and non-voucher patients, respectively. Under the same conditions as in Theorem 1, with high probability

$$\widehat{W}_{\mathrm{W}}^{\mathrm{SP}} \;\le\; C\frac{1}{p}\Big(\log\frac{1}{q}\Big)^{1/2+\epsilon}, \qquad \widehat{W}_{\mathrm{U}}^{\mathrm{SP}} \;\le\; \frac{1+\epsilon}{p}\log\frac{1}{q}\,.$$

[14] The within-type order in which patients are matched does not affect queue length or expected waiting time, by Little's Law.

[15] Given a sequence of events $\{\mathcal{E}_n\}$, we say this sequence occurs with high probability if $\lim_{n\to\infty}\mathbb{P}(\mathcal{E}_n) = 1$.

By Little's Law, the expectations of these proxies coincide with the corresponding steady-state waiting times. Little's Law, however, does not identify their tail distributions; for this reason we refer to them as waiting time proxies.

The proof of Theorem 1 is outlined in Section 5 and given in Appendix D. A few implications are useful before turning to the detailed comparison with prior work in the next subsection. It is standard in the kidney-exchange literature to study the regime in which the compatibility probability p tends to zero. In this regime, our bounds require only a small fraction of voucher pairs: it suffices that $q \geq p$ for the voucher-patient results, whereas the non-voucher-patient results require the slightly stronger condition $q \geq p^{1-\delta}$ for some $\delta > 0$. In the latter regime, our bound matches the lower bound in the literature, obtained under altruistic donors or universal voucher participation ($q = 1$), up to logarithmic factors, even though our model has no altruistic donors and q is vanishingly small.

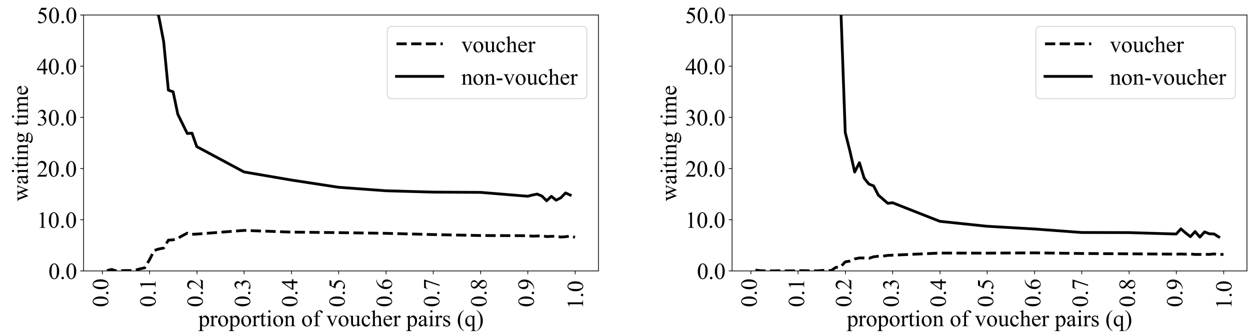
To complement the result above, the following proposition provides a converse: if voucher pairs are too rare, then non-voucher patients experience unbounded waiting times.

Proposition 2. *Suppose $q < p$ and let $\mathbf{W}_{\mathrm{U}}^{\mathrm{SP}}(k)$ be the (random) waiting time of the kth-arriving non-voucher pair under the Standard Priority mechanism. Then, for any initial state,*

$$\mathbb{E}[\mathbf{W}_{\mathrm{U}}^{\mathrm{SP}}(k)] \to \infty \quad \textit{as } k \to \infty.$$

The proof of Proposition 2 is in Appendix C. The proposition isolates the role of vouchers by prohibiting matches outside of priority chains. Specifically, it rules out two-way matches among non-voucher pairs. In practice, such cycles can also be formed and yield average waiting times of order $O(1/p^2)$ (Anderson et al., 2017). Thus, Theorem 1 and Proposition 2 can be read together as a threshold result: under Standard Priority, voucher participation substantially benefits both voucher and non-voucher patients (relative to two-/three-way matching) when voucher pairs are moderately prevalent ($q > p$), whereas the benefit for non-voucher patients vanishes when voucher pairs are too rare ($q < p$).

We next illustrate the threshold numerically by varying the voucher-pair share $q \in [0, 1]$ under Standard Priority. For each value of q, we simulate 10,000 arrivals and compute the average waiting times of voucher and non-voucher patients. Figure 4 reports the results for $p = 0.1$ and $p = 0.2$. For both values of p, the simulations display a phase transition around $q = p$. Below this threshold, voucher patients wait essentially zero time while non-voucher patients experience unbounded waiting times. Above the threshold, both types have bounded waiting times close to $1/p$, the optimal order achievable by any matching mechanism (Anderson et al., 2017).



(a) $p = 0.1$

(b) $p = 0.2$

Figure 4: Average waiting times of voucher and non-voucher patients under Standard Priority. The horizontal axis is the voucher-pair proportion q, and the vertical axis is the average waiting time. Dashed and solid curves are the voucher and non-voucher patients' waiting times, respectively.

3.1 Discussion of Main Result

Previous findings. To understand how prioritization affects different types of patients, we first place the result against waiting time bounds from the kidney exchange literature. Table 1 summarizes the comparison.

Under the same stochastic compatibility structure, $\Omega(1/p)$ is a lower bound on average waiting time (Anderson et al., 2017). Several papers study markets in which all pairs are non-voucher and match either through cycles or through chains initiated by altruistic donors.[16] Cycles-based policies achieve expected waiting time $O(1/p^2)$ with 2-way cycles and $O(1/p^{3/2})$ when both 2-way and 3-way cycles are allowed (Anderson et al., 2017). Chain-based policies achieve $O(1/p \log(1/p))$ under greedy and altruistically-initiated chains (Blum and Mansour, 2020). This greedy chain bound is within only a logarithmic factor of the lower bound.[17] Akbarpour et al. (2020) study the setting in which all pairs are voucher pairs ($q = 1$). They assume a fraction of patients are easy-to-match,[18] and find that the lower bound $O(1/p)$ can be achieved even with a vanishing fraction of easy-to-match pairs.

The effect of prioritization. Theorem 1 provides insights into both the strengths and the shortcomings of prioritization under different market characteristics. With a small but sufficiently frequent flow of voucher pairs, Standard Priority gives nearly-optimal waiting times to both voucher and non-voucher patients: voucher patients wait less than under

[16] Altruistic donors are necessary to form chains when all pairs are non-voucher.

[17] The lower bound can be attained using an altruistic donor chain that searches for the longest segment rather than being extended greedily (Anderson et al., 2017).

[18] With a constant probability for compatibility.

greedy altruistic-donor chains, while non-voucher patients wait substantially less than under cycles and close to the order achieved by greedy chains. The intuition is that voucher donors act as chain initiators. When voucher pairs are too scarce, however, too few chains are initiated, and the expected waiting time of non-voucher patients grows without bound.

These findings also have distributional implications. When voucher pairs are sufficiently prevalent, Standard Priority improves waiting times for non-voucher pairs relative to short cycles, and the additional waiting time advantage from being a voucher patient is only logarithmic. In this regime, priority rewards early donation while preserving much of the chain externality for non-voucher pairs. When voucher pairs are too scarce, the waiting time gap between voucher and non-voucher pairs becomes much larger. Thus, the distributional consequences of priority depend not only on the priority rule itself, but also on whether the market contains enough early-donating pairs.

In practice, a kidney exchange clearinghouse can also form cycles. Our findings suggest that cycles are especially important when voucher pairs are scarce ($q < p$) because in that region non-voucher pairs can still be matched through cycles with expected waiting time of $O(1/p^2)$. Section 4 provides simulation results that combine the priority rules with a simple cycle-augmentation.

The Standard Priority mechanism with contingent priorities. Theorem 1 assumes, for tractability, that Standard Priority assigns priority to all voucher patients regardless of whether their intended donors have already donated. In practice, priority is often contingent on the intended donor having already donated; one example is the standard voucher program through the NKR (National Kidney Registry, 2025). We verify the robustness of Standard Priority by simulating a modified version of Standard Priority with contingent priority.

Figure 5 compares the average waiting times for voucher and non-voucher patients under the baseline Standard Priority rule and modified contingent-priority variant, using $p = 0.1$ and 10,000 arrivals. The two mechanisms generate similar qualitative patterns.

Table 1: Expected waiting time comparison, where p is the compatibility probability and q is the proportion of voucher donors. For the Standard Priority, we further assume that $q > p$ and both $p, q \to 0$.

<table>
<tr><th></th><th>Standard Priority
(this paper)</th><th>Cycles
Anderson et al. (2017)</th><th>Altruistic Chains
Anderson et al. (2017)
Blum and Mansour (2020)</th><th>Unpaired
(Optimal)
Akbarpour et al. (2020)</th></tr>
<tr><td>Voucher</td><td>$\frac{1}{p}\left(\log\frac{1}{q}\right)^{1/2+\epsilon}$</td><td rowspan="2">$\frac{1}{p^2}$ (2-cycles)
$\frac{1}{p^{3/2}}$ (2- and 3-cycles)</td><td rowspan="2">$\frac{1}{p}\log\frac{1}{p}$ (greedy)
$\frac{1}{p}$ (longest)</td><td rowspan="2">$\frac{1}{p}$</td></tr>
<tr><td>Non-voucher</td><td>$\frac{1}{p}\log\frac{1}{q}$</td></tr>
</table>

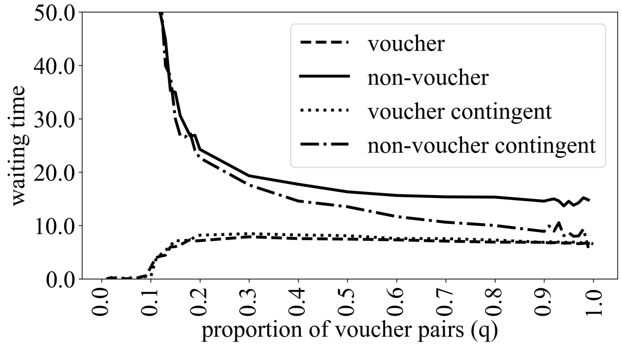


Figure 5: Standard Priority with contingent priorities ($p = 0.1$). The horizontal axis is the voucher pair proportion q, and the vertical axis is average waiting time. Dashed and solid curves report voucher and non-voucher waiting times under the baseline Standard Priority rule. Dotted and dash-dot curves report the corresponding waiting times under contingent priorities.

3.2 Heterogeneous Compatibilities

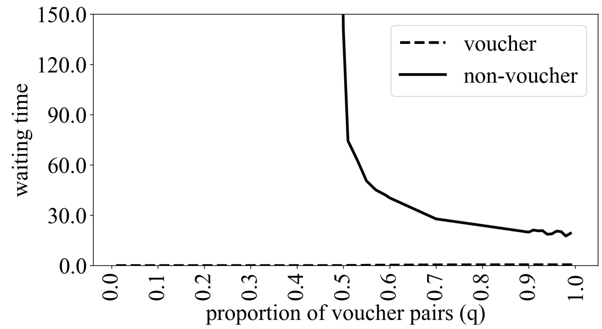
We next discuss how priority interacts with heterogeneity in patient compatibility. Suppose voucher and non-voucher patients are compatible with a random donor with probabilities p_W and p_U, respectively, and assume $p_W > p_U$. This assumption is natural since easy-to-match patients may be more willing to take the risk to donate before receiving an organ, and therefore are more likely to enter as voucher pairs.

We first illustrate the effects of different priority mechanisms in the case in which voucher patients are easy-to-match, with $p_W = 1$, and non-voucher patients are hard-to-match, with $p_U = 0.1$. Figure 6 shows a striking difference between Standard Priority and Weak Priority. Under Standard Priority, the expected waiting time for non-voucher patients is unbounded for any $q < 0.5$. Under Weak Priority, this instability disappears: non-voucher patients have finite expected waiting time even as $q \to 0$. Theorem 2 and Theorem 3 formalize these differences; proofs are deferred to Appendix E.

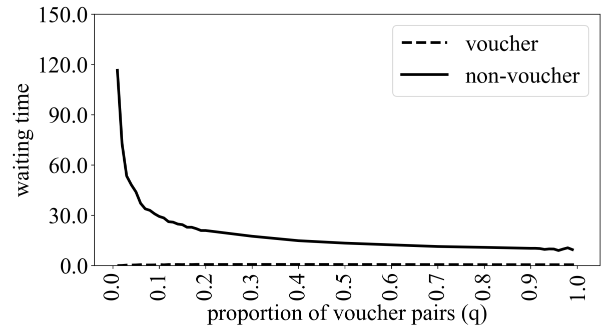
Theorem 2. *Under the Standard Priority mechanism, when $p_W = 1, p_U = p < 1, q < 0.5$, the expected waiting time for non-voucher patients is unbounded. The expected waiting time for voucher patients is $O\left(\frac{1}{q}\right)$.*

Theorem 3. *Under the Weak Priority mechanism, when $p_W = 1, p_U = p < 1$, the expected waiting time for non-voucher patients is $O\left(\frac{1}{p(1-q)}\log\frac{1}{q} + \frac{1}{q(1-q)}\right)$ for all $p, q \in (0, 1)$. For voucher patients, the expected wait is $O\left(\min\left\{\frac{1}{q}, \frac{1}{p}\right\}\right)$.*

Theorem 2 has the same spirit as Ashlagi et al. (2019): in a heterogeneous market, easy-to-match patients can only help as many other patients as there are themselves. As such,



(a) Standard Priority



(b) Weak Priority

Figure 6: Average waiting times of voucher and non-voucher patients when $p_W = 1$ and $p_U = 0.1$. Solid curves report waiting times for non-voucher patients, which are unbounded for $q < 0.5$ under Standard Priority but finite under Weak Priority.

the waiting time for hard-to-match patients (non-voucher in this work and hard-to-match in Ashlagi et al. (2019)) depends critically on whether easy-to-match (voucher) pairs constitute at least half of the market ($q > 0.5$).

When $p_W = 1$, the system can have at most one waiting voucher pair, because a second voucher pair would match immediately with the first voucher pair upon arrival. The intuition for Theorem 2 is that, given $p_W = 1$, only two types of matches are possible under the Standard Priority mechanism: a 2-way cycle between two voucher pairs, or a 2-way cycle between one voucher pair and one non-voucher pair. As such, the Standard Priority reduces to a cycle mechanism where easy-to-match pairs are prioritized, and hard-to-match pairs cannot match with each other (which only differs from Ashlagi et al. (2023) in the sense that, in their paper, hard-to-match are prioritized). Therefore, the number of non-voucher patients that can be matched is bounded by the number of voucher pairs. By contrast, Weak Priority allows the chain to continue greedily through non-voucher pairs. To prove Theorem 3, we bound the expected wait of non-voucher patients using the (M, K, ρ, β)-random walk of Blum and Mansour (2020); for completeness, we state the construction in Definition 1 and Lemma 4 in Appendix E.

The unbounded waiting time for non-voucher patients in Theorem 2 highlights the importance of how priority is implemented. A strong priority rule (Standard) exploits opportunities to match easy-to-match voucher patients quickly, but can leave hard-to-match non-voucher patients with very long waiting times. Weak Priority mitigates this effect by allowing chains to pass through additional non-voucher pairs before ending with a voucher patient.

Blood types offer another natural interpretation of this consequence due to heterogeneity. Suppose, for example, that voucher pairs are A-O patient-donor pairs, while non-voucher pairs are O-A patient-donor pairs. O donors are ABO compatible with any patient, whereas

O patients are ABO compatible only with O donors. Under Standard Priority, a free O donor may immediately donate to a voucher A patient and end the chain, using a highly flexible donor on an easy-to-match patient. In contrast, Weak Priority can first use the O donor to match a non-voucher O patient, whose paired A donor can then donate to a voucher A patient.

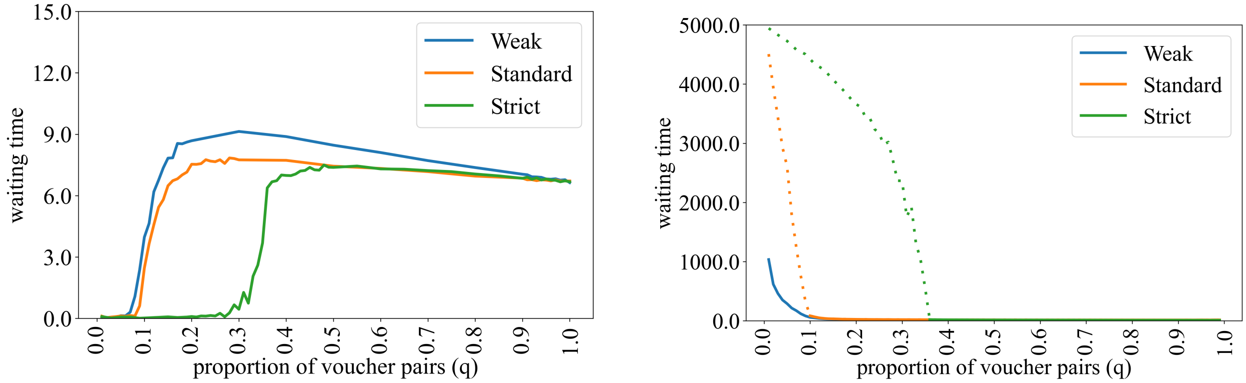
4 Simulations

We complement the theoretical analysis with simulations comparing Weak, Standard, and Strict Priority. We also consider cycle-augmented variants of these mechanisms: after each arrival, the priority rule is applied first, and then available 2-way cycles among non-voucher pairs are formed. Throughout this section, we set $p = 0.1$, vary q, and report averages over five simulations with 10,000 arrivals each.

Figure 7 reports waiting times when no non-voucher cycles are allowed. The simulations confirm the analytical threshold under Standard Priority (Proposition 2 and Theorem 1): non-voucher waiting times are large when q is below p and fall sharply once voucher pairs are more common. Under Weak Priority, a similar phase transition also happens around $q = p$. Under Strict Priority, the phase transition happens at a threshold larger than p. Importantly, the waiting times of non-voucher patients are bounded under Weak Priority for all q, but still decline sharply around $q = p$. Under Strict Priority their waiting times are longest when q is small. Voucher patients, by contrast, wait very little for all values of q under all three rules, especially in the low-q region, where the large non-voucher pool makes compatible donors easy to find.

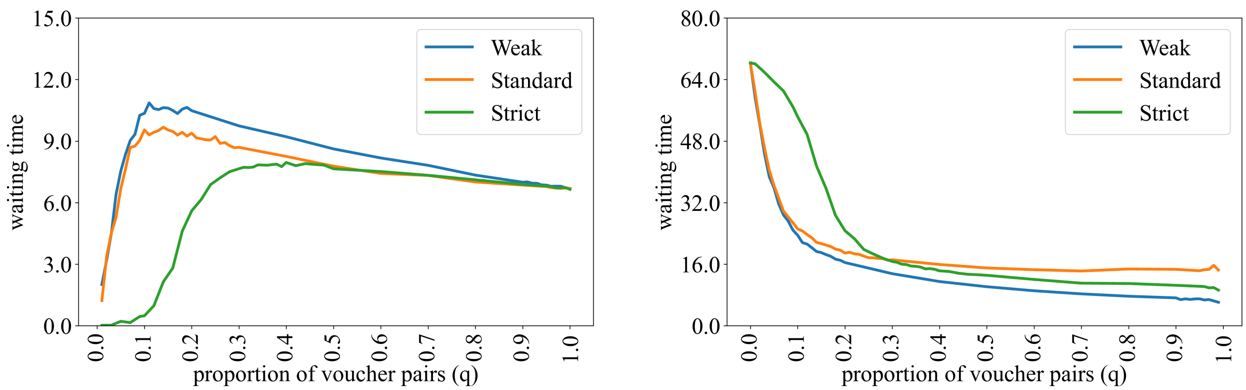
Figure 8 shows that allowing 2-way cycles among non-voucher pairs substantially reduces congestion for non-voucher patients in the low-q regimes. In particular, non-voucher waiting times remain bounded for all q under all three cycle-augmented mechanisms, and are below the benchmark $(\log 2)/p^2 + o(1/p^2)$ from Anderson et al. (2017), which corresponds to matching only through 2-cycles. This benchmark is reflected at $q = 0$, by an average waiting time of 68.32 in the simulation. The ordering across priority rules remains the same: Cycle-Augmented Weak Priority gives the shortest waiting times for non-voucher patients, while Cycle-Augmented Strict Priority gives the longest.

The role of cycles is most important when voucher pairs are scarce or when priority is strict. Figure 9 shows that, for small q, a large share of non-voucher patients are matched through 2-cycles rather than through priority chains. As q increases, non-voucher patients increasingly match through priority chains, and the three mechanisms become more similar. Thus, cycles smooth the threshold effect: they do not eliminate the trade-off created by priority, but they provide an important complementary channel for matching non-voucher patients when voucher pairs are too scarce or when chains are too short.



(a) Voucher (b) Non-voucher

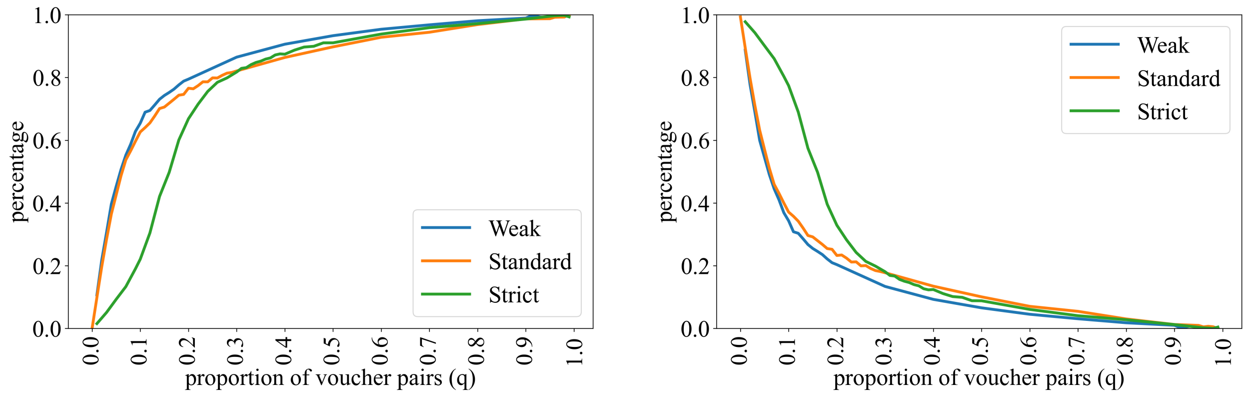
Figure 7: Average waiting times when no non-voucher cycles are allowed ($p = 0.1$). The left and right panels report voucher and non-voucher patients. Blue, orange, and green curves correspond to Weak, Standard, and Strict priorities. Solid lines denote the average waiting time in steady state; dotted lines indicate parameter values for which the corresponding infinite-horizon average waiting time is unbounded, although the plotted finite-horizon simulation average remains finite over 10,000 arrivals.



(a) Voucher (b) Non-voucher

Figure 8: Average waiting times under cycle-augmented priority mechanisms for $p = 0.1$, averaged over five independent simulations.

Further insight comes from the behavior of chains under different priority rules augmented with cycles. A *proper chain* is a concatenation of chain segments that begins with a voucher donor and ends with a voucher patient; all matches can be partitioned into such proper chains and 2-way cycles among non-voucher pairs. Table 2 reports, for $p = 0.1$, the number of voucher arrivals, non-voucher–non-voucher (U-U) cycles, chain segments, proper chains, and



(a) Priority chain

(b) Non-voucher–Non-voucher cycle

Figure 9: Percentage of non-voucher patients matched through priority chains and non-voucher–non-voucher cycles for $p = 0.1$, averaged over five independent simulations.

average proper-chain length. The table highlights the same threshold effect. When $q = 0.02$, voucher pairs are scarce and priority chains are valuable but limited. Strict Priority finds voucher patients quickly, and all 176 segments become proper chains, but chains are short: the average proper-chain length is only 2.57, compared with 10.21 under Standard Priority and 12.07 under Weak Priority. When $q = 0.4$, voucher pairs are common, chains are short under all three mechanisms, non-voucher–non-voucher (U-U) cycles are much less common, and the differences across priority rules are small.

The simulations confirm that the way priority is implemented matters: weakening priority slightly can preserve more of the chain externality at little cost to voucher patients, while Strict Priority shortens chains and thereby harms non-voucher patients. Second, short cycles are an important complementary channel for ensuring that non-voucher pairs are matched, especially when voucher pairs are scarce. Appendix F reports simulations results for other parameters, which reveal similar patterns.

	$q = 0.02$					$q = 0.40$				
	#W	#U-U cycles	#seg	#prop chain	avg chain length	#W	#U-U cycles	#seg	#prop chain	avg chain length
Standard	197	3981	243	195	10.21	4004	383	4453	3161	2.92
Weak	204	3738	278	204	12.07	3957	285	4866	3192	2.95
Strict	177	4742	176	176	2.57	4037	360	4799	3278	2.83

Table 2: Chain, segment, and cycle metrics for $p = 0.1$ and $q \in \{0.02, 0.40\}$. #W denotes the number of voucher-pair arrivals, #U-U cycles denotes the number of non-voucher–non-voucher 2-cycles, #seg denotes the number of match segments excluding U-U cycles, # prop chain denotes the number of proper chains excluding U-U cycles, and avg chain length denotes the average proper-chain length.

5 Proof Sketch of Main Results

In our model, the expected wait is proportional to the steady-state expected queue length by Little's Law (Little, 1961). We focus our analysis on the expected queue lengths. Throughout this section, we write

$$(X, Y) \sim \pi$$

for a random state drawn from the stationary distribution of the Markov chain $(X(t), Y(t))$, where X is the number of waiting voucher patients and Y is the number of waiting non-voucher patients. All probabilities are taken under π, unless explicitly stated otherwise.

The proof is nonstandard and challenging for two reasons. First, unlike in traditional queuing models, a single arrival can trigger a chain and remove a random, potentially large, number of patients. Second, the two queues are coupled: the size of the voucher queue affects how often chains start and terminate, while the size of the non-voucher queue affects how far a chain can travel before priority forces it to stop at a voucher patient. The proof therefore does not attempt to construct one global Lyapunov function for (X, Y). Instead, it uses an iterative bootstrap. A crude bound on X implies a lower bound on Y; that lower bound on Y then improves the upper tail of X; the improved upper tail of X then improves the lower tail of Y. Repeating this loop yields the final voucher-pool bound, and a separate drift argument yields the final non-voucher upper bound.

5.1 Preliminaries

We establish a useful lemma to analyze the unique queuing dynamics under the priority mechanism.

Lemma 1 provides a stationary tail bound for a (generalized) Lyapunov function $V(\cdot)$ of a discrete-time Markov chain on a countable state space. It extends standard drift-based tail bounds such as Theorem 1 of Bertsimas et al. (2001) in two respects. First, the lemma is well suited to cases in which the state evolves through infrequent but potentially large one-step changes (as in kidney exchange, where initiating a chain can trigger a substantial jump). Classical drift arguments typically summarize one-step variability through a single worst-case increment bound, which can yield overly conservative tail estimates when large jumps are rare. Second, the lemma accommodates a non-global drift condition: V is required to have a uniform negative drift only on a designated "good" set $\mathcal{E}$ (and only above the threshold B), while outside $\mathcal{E}$ the drift may be positive but is controlled through the bound γ_{out}. The resulting tail estimate therefore degrades gracefully with the stationary mass of the exceptional region, via the factor $\mathbb{P}_\pi(X \notin \mathcal{E})$.

We state the lemma for discrete-time Markov chain $\{X(t)\}$ with countable state space $\mathcal{X}$. For $x, x' \in \mathcal{X}$, we denote $\mathbb{P}(X(t+1) = x' \mid X(t) = x)$ by $p(x, x')$.

Lemma 1 (Tail bound of Lyapunov function with an exceptional set)**.** *Let $\{X(t) : t \in \mathbb{N}\}$ be a discrete-time Markov chain on a countable state space $\mathcal{X}$ with transition probabilities $p(\cdot, \cdot)$ and stationary distribution π. Let $V : \mathcal{X} \to \mathbb{R}_+$ satisfy the drift condition*

$$\mathbb{E}[\Delta V(X(t)) \mid X(t) = x] \leq -\gamma \quad \text{for all } x \in \mathcal{E} \text{ with } V(x) \geq B,$$

for some $\gamma > 0$, $B \geq 0$, and $\mathcal{E} \subseteq \mathcal{X}$; here $\Delta V(X(t)) \triangleq V(X(t+1)) - V(X(t))$. Assume $\mathbb{E}_\pi[V(X)] < \infty$ and that the one-step increments are bounded:

$$\nu_{\max} \triangleq \sup_{x, x' : \, p(x, x') > 0} |V(x') - V(x)| < \infty.$$

Define

$$\gamma_{\text{out}} \triangleq \sup_{x \notin \mathcal{E}} \mathbb{E}[\Delta V(X(t)) \mid X(t) = x] < \infty,$$

and, with $(u)_+ = \max\{u, 0\}$,

$$q_{\text{in}} \triangleq \sup_{x \in \mathcal{E}} \mathbb{E}[(V(X(t+1)) - V(x))_+ \mid X(t) = x] < \infty,$$

$$q_{\text{out}} \triangleq \sup_{x \notin \mathcal{E}} \mathbb{E}[(V(X(t+1)) - V(x))_+ \mid X(t) = x] < \infty.$$

Let

$$\xi = \frac{q_{\text{in}}}{q_{\text{in}} + \gamma}, \qquad \phi = \frac{q_{\text{out}} + \gamma_{\text{out}} + \gamma}{\gamma}.$$

Then for all $m \in \mathbb{N}$,

$$\mathbb{P}_\pi(V(X) \geq B + 2\nu_{\max} m) \leq \xi^m + \phi \, \mathbb{P}_\pi(X \notin \mathcal{E}),$$

where $X \sim \pi$.

The proof of Lemma 1 is in Appendix A. In Appendix A, we also adapt the proof technique of Lemma 1 to a random walk with geometric jumps, a setting in which Lemma 1 cannot be applied directly because $\nu_{\max} = \infty$.

5.2 Main Steps of Theorem 1

The proof relies on three recurring effects under Standard Priority. Recall that chains can start in two ways: upon the arrival of a voucher pair, or upon the arrival of a non-voucher patient who can receive from a free donor.

First, when X is small, a newly arriving non-voucher patient is unlikely to be compatible with a waiting free donor. Such arrivals therefore rarely free the arriving donor to start a chain. This suppresses departures from the non-voucher queue and gives us a lower bound on Y.

Second, when Y is large, a donor-initiated chain is unlikely to get stuck among non-voucher pairs before it encounters a compatible voucher patient. Since Standard Priority stops a chain once it can match a voucher patient, a large Y gives us an upper bound on X.

Third, once X is sufficiently small, chains tend to travel farther through the non-voucher pool before priority terminates them. This creates the negative drift used for the final upper bound on Y.

The argument proceeds through a base case (Steps 1–2) and an induction step (Steps 3). The base case provides a crude high-probability upper bound on X and then converts it into a high-probability lower bound on Y. The induction step repeatedly alternates: (i) a lower bound on Y implies a sharper upper-tail bound on X, and (ii) that sharper upper bound on X implies a sharper lower-tail bound on Y. Figure 10 summarizes the proof flow.

Base case:

$\xRightarrow{\text{Proposition 3}}$ $\boxed{\text{Step 1: } X = O(q^{1/2}/p) \text{ w.h.p.}}$ $\xRightarrow{\text{Proposition 4}}$ $\boxed{\text{Step 2: } Y \geq \frac{1}{2p}\log\left(\frac{1}{q}\right) \text{ w.h.p.}}$

Induction step:

$\boxed{Y \geq \frac{1-2^{-k}}{p}\log\left(\frac{1}{q}\right) \text{ w.h.p.}}$ $\xRightarrow{\text{Proposition 5}}$ $\boxed{\text{Step 3a: } X = O(q^{1-2^{-(k+1)}}/p) \text{ w.h.p.}}$

$\xRightarrow{\text{Proposition 6}}$ $\boxed{\text{Step 3b: } Y \geq \frac{1-2^{-(k+1)}}{p}\log\left(\frac{1}{q}\right) \text{ w.h.p.}}$

Figure 10: Proof flow for the iterative bounds. The base case gives the first lower bound on Y. The induction alternates between a lower-tail bound for Y and an upper-tail bound for X.

5.2.1 Step 1: A crude upper bound on the voucher pool

As a first step, we control the steady-state size of the voucher pool X. A key technical challenge is that the removal rate from the voucher pool depends not only on its own size but also on the non-voucher pool in a nonlinear way. Proposition 3 provides a coarse high-probability upper bound of order $\sqrt{q}/p$ for X. The bound is obtained without using the non-voucher pool Y: we restrict attention to two match mechanisms that are always available, namely (i) voucher–voucher matches triggered by a voucher arrival, and (ii) two-way exchanges of the form $D \to$ (new non-voucher pair) $\to P$ triggered by a non-voucher arrival. This crude control of X serves as an input in subsequent steps, where we explicitly couple the voucher and non-voucher dynamics.

Proposition 3 (Step 1)**.** *There exists $\epsilon > 0$ such that for any $q \in (0, \min\{\epsilon, 1/2\})$ and any $p \le q$,*

$$\mathbb{P}_\pi\left(X \ge 2\frac{\sqrt{q}}{p}\right) \le \exp\left(-\frac{\log 2}{4} \cdot \frac{\sqrt{q}}{p}\right) \le \exp\left(-0.17 \cdot \frac{1}{\sqrt{p}}\right).$$

The proof of Proposition 3 is in Appendix D.

5.2.2 Step 2: A lower bound on the non-voucher pool size using Step 1

In this step we leverage the crude upper bound on the voucher pool from Step 1 to obtain a high-probability lower bound for the non-voucher pool. Intuitively, when X is small, it is difficult for an arriving non-voucher pair to match with a voucher donor and thereby free its donor to initiate a chain; as a result, departures from the non-voucher pool are limited while non-voucher arrivals continue to accumulate. Proposition 4 formalizes this intuition by conditioning on the "small voucher-pool" event $\mathcal{E}_1 = \{X \le 2\sqrt{q}/p\}$, whose probability is close to one by Proposition 3. Combining this conditioning with a concentration bound for random walks with geometric jumps yields an explicit high-probability lower-tail estimate for Y.

Proposition 4 (Step 2)**.** *Define the "small voucher-pool" event $\mathcal{E}_1 \triangleq \left\{X \le 2\frac{\sqrt{q}}{p}\right\}$. There exists $\epsilon > 0$ such that for all $q \in (0, \min\{\epsilon, 1/2\})$, all $p \le q$, and all $m \in \mathbb{N}$,*

$$\begin{aligned}&\mathbb{P}_\pi\left(Y \le \frac{1}{2p}\left(\log\frac{1}{q} - 3\right) - 0.5\, m\, q^{-3/4}\right)\\ &\le 0.8^m + 4.4\left(\left(0.95 + \frac{0.75}{\sqrt{q}}\right)\mathbb{P}_\pi(\mathcal{E}_1^c) + 0.75e^{-q^{-1/4}}\right).\end{aligned}$$

The proof of Proposition 4 is in Appendix D.

5.2.3 Step 3: Iteratively improving the bounds on the voucher and non-voucher pool sizes

The next two propositions form the induction step. Fix an integer $k \ge 1$. Assuming a high-probability lower bound on Y of the form $Y \gtrsim (1 - 2^{-k})/p \log(1/q)$, Proposition 5 shows that the voucher pool has a uniform negative drift above a threshold of order $q^{1-2^{-(k+1)}}/p$, yielding an exponentially decaying stationary upper tail for X. Proposition 6 then uses this improved upper control on X to strengthen the lower-tail bound for Y from level k to level $k+1$. This alternation is the mechanism behind the iterative tightening in Figure 10.

Proposition 5 is the first half of the induction step. It shows that on the event $\mathcal{G}_k$ where the non-voucher pool is sufficiently large, the voucher-patient queue X exhibits a negative drift once it exceeds an explicit threshold, and consequently its stationary upper tail decays exponentially. Intuitively, when Y is large, chains triggered by arriving non-voucher pairs are likely to run long enough to encounter a compatible voucher patient and terminate there, thereby removing a voucher patient from the system. The proof formalizes this mechanism by lower bounding the probability that a chain initiated in a given period ends at a voucher patient, and then translating this bound into a drift condition for X on $\mathcal{G}_k$.

Proposition 5. *For each integer $k \geq 1$, define the event*

$$\mathcal{G}_k \triangleq \left\{(x, y) \in \mathbb{N}^2 : \; y \geq \frac{1}{p}\left(1 - 2^{-k}\right)\left(\log\frac{1}{q} - 3.41\right)\right\}.$$

There exists $\epsilon > 0$ such that for all $q \in (0, \epsilon)$, all $p \leq q$, and all $m \in \mathbb{N}$,

$$\mathbb{P}_\pi\left(X \geq 9.9\,\frac{q^{1-2^{-(k+1)}}}{p} + 2m\right) \;\leq\; 0.96^m \;+\; 46\,\mathbb{P}_\pi((X, Y) \notin \mathcal{G}_k)\,.$$

The proof of Proposition 5 is in Appendix D.

Proposition 6 is the second half of the induction step. It mirrors the logic of Step 2: conditioning on the event $\mathcal{E}_k$ that the voucher pool X is small, an arriving non-voucher pair is unlikely to find a compatible voucher donor, so its donor is rarely freed to initiate a chain. Consequently, departures from the non-voucher pool are suppressed while non-voucher arrivals continue to accumulate, pushing Y upward. Combined with the improved upper-tail control on X from Proposition 5, this closes the induction loop and yields progressively tighter bounds as k increases.

Proposition 6. *Fix an integer $k \geq 2$ and define the "small voucher-pool" event at level k by*

$$\mathcal{E}_k \triangleq \left\{\, X \leq 10\,\frac{q^{1-2^{-k}}}{p} \,\right\}.$$

There exists $\epsilon > 0$ such that for all $q \in (0, \min\{\epsilon, 1/2\})$, all $p \leq q$, and all $m \in \mathbb{N}$, letting

$$y_k \triangleq \frac{1}{p}\left(1 - 2^{-k}\right)\left(\log\frac{1}{q} - 3.4\right),$$

we have

$$\mathbb{P}_\pi\Big(Y \le y_k - 0.17\, m\, q^{-(1-2^{-(k+1)})}\Big)$$
$$\le 0.96^m + 24.3\left[\left(0.95 + \frac{0.25}{q^{1-2^{-k}}}\right)\mathbb{P}_\pi(\mathcal{E}_k^c) + 0.91\, e^{-q^{-2^{-(k+1)}}}\right].$$

The proof of Proposition 6 is in Appendix D.

5.2.4 Step 4: Final bounds on the voucher and non-voucher pool sizes via induction

This step establishes that the stationary voucher-pool size $X(\infty)$ is very unlikely to be larger than $10\frac{q}{p}\left(\log\frac{1}{q}\right)^\alpha$ for any $\alpha > \frac{1}{2}$. We introduce a sequence of "good" events $\mathcal{E}_k$ that assert progressively stronger upper bounds on $X(\infty)$, starting from a crude bound and tightening it across levels. At each level k, we also define an auxiliary "good" event $\mathcal{G}_k$ that says the companion variable $Y(\infty)$ is not too small (so the system has accumulated enough "buffer" on the Y-side).

The core mechanism of Proposition 7 is as follows. If $\mathcal{G}_k$ holds, then Proposition 5 provides a strong one-step tail bound for $X(\infty)$, which upgrades control of X from level k to level $k+1$. Meanwhile, Proposition 6 shows if $\mathcal{E}_k$ holds, then $Y(\infty)$ is very unlikely to fall below the threshold defining $\mathcal{G}_k$; in other words, the "bad" event $\mathcal{G}_k^c$ is itself controlled by the current failure probability $\mathbb{P}(\mathcal{E}_k^c)$ plus an exponentially small remainder. Combining these two propositions yields a recursive inequality for the failure probabilities $P_k = \mathbb{P}(\mathcal{E}_k^c)$. Iterating this recursion across levels, we choose the terminal level $k = O(\log\log(1/q))$. We can show that $P_{k+1} \to 0$ and hence prove the claimed upper-tail bound for $X(\infty)$.

Proposition 7. *Fix any $\alpha > \frac{1}{2}$. Assume $p \le q$. Then we have*

$$\lim_{q\to 0^+}\ \sup_{0<p\le q}\ \mathbb{P}_\pi\left(X(\infty)\ \ge\ 10\,\frac{q}{p}\left(\log\tfrac{1}{q}\right)^\alpha\right) = 0,$$

where $\mathbb{P}_\pi$ denotes probability under the stationary distribution.

The proof of Proposition 7 is in Appendix D.

5.2.5 Step 5: Final upper bound for the non-voucher queue

The final step uses the small-voucher-pool regime in a different direction. When X is at most a fixed polynomial scale below q/p, a donor-initiated chain is unlikely to terminate quickly at a voucher patient. If Y is also large, the chain can keep finding non-voucher patients, so the expected number of non-voucher removals dominates the non-voucher arrival rate. The proof formalizes this force and provides an upper bound for the non-voucher queue.

Proposition 8. *For any fixed $\alpha \in (0,1)$, $q \geq p$, as $q \to 0$ we have:*

$$\mathbb{P}\left(Y \geq (1+3\alpha)\frac{1}{p}\log\frac{1}{q} + \frac{4}{q^{1+\alpha}}\log\frac{1}{q}\right) \to 0\,.$$

If we further have that $q^{1+\alpha} = \omega(p)$ as $q \to 0$, then we have:

$$\mathbb{P}\left(Y \geq (1+4\alpha)\frac{1}{p}\log\frac{1}{q}\right) \to 0\,.$$

The proof of Proposition 8 is in Appendix D.

Proof of Theorem 1. The upper bound on voucher queue length follows from Proposition 7, and the upper bound on non-voucher queue length follows from Proposition 8.

All bounds in this iterative process are in high probability. The process is illustrated in Figure 11.

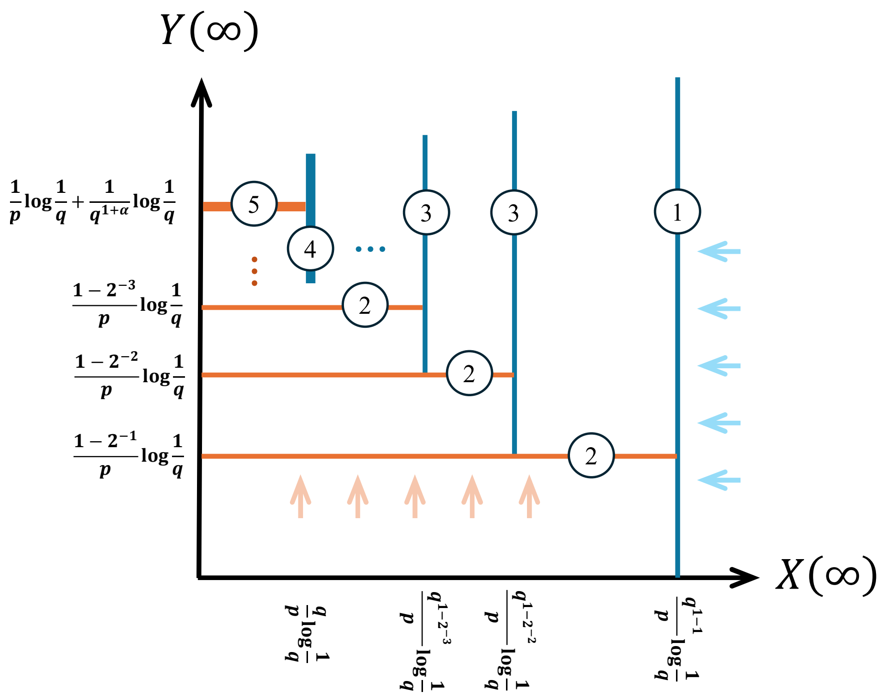


Figure 11: Iterative bounds on the queue lengths in steps 1-5. The circled numbers are bounds achieved in the corresponding step. The thickened lines denote the final upper bounds from steps 4 and 5.

5.3 Intuition for Proposition 2

Proposition 2 asserts that when $q < p$, under Standard Priority, the waiting time for non-voucher patients grows without bound. We give an outline of the proof and defer the details to Appendix C. Recall that a proper chain is the concatenation of segments such that it starts with a voucher donor and ends with a voucher patient. All non-voucher patient removals can be charged to proper chains. Each proper chain removes at most a geometric number of non-voucher patients, with mean at most $(1-p)/p$. Voucher donors arrive at rate q, so when $q < p$ the non-voucher arrival rate exceeds the maximal non-voucher removal rate.

6 Final Comments

Voucher programs in kidney exchange allow incompatible pairs to donate early in return for priority in future matches. Early-donating pairs create value by initiating chains, but priority also determines where those chains stop: once a voucher patient is matched, the chain terminates. This paper studies the interaction between these forces and finds that the benefit of voucher programs depends on the priority implementation and on market characteristics, especially the share of voucher pairs and the compatibility structure.

The findings suggest that priority can preserve much of the benefit of chains when the market contains enough early-donating pairs. In this case, priority rewards the patients whose donors donate early while still allowing non-voucher pairs to benefit from the positive externality created by those early-donation chains. When early-donating pairs are too scarce, however, standard priority may direct too many chains toward voucher patients before enough non-voucher pairs are matched. In this regime, short cycles among non-voucher pairs, or weaker priority rules that allow chains to continue while preserving the ability to match a voucher patient, can play an important complementary role.

The paper is only a first step in understanding the role of priorities in dynamic matching markets. One direction is to incorporate richer heterogeneity in compatibilities and ask further when to restrict the use of vouchers. A related direction is incorporating exogenous departures to address the challenge of not being able to redeem vouchers. Finally, our model treats willingness to donate early as exogenous. It is important to explore priority rules with endogenous types. While a stronger priority may increase participation in voucher programs, it may also create pressure for some pairs to donate early.

References

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2009). Strategy-proofness versus efficiency in matching with indifferences: Redesigning the nyc high school match. *American Economic Review 99*(5), 1954–1978.

Agarwal, N., I. Ashlagi, E. Azevedo, C. R. Featherstone, and Ö. Karaduman (2019). Market failure in kidney exchange. *American Economic Review 109*(11), 4026–4070.

Akbarpour, M., E. Budish, P. Dworczak, and S. D. Kominers (2024). An economic framework for vaccine prioritization. *The Quarterly Journal of Economics 139*(1), 359–417.

Akbarpour, M., J. Combe, Y. He, V. Hiller, R. Shimer, and O. Tercieux (2020). Unpaired kidney exchange: Overcoming double coincidence of wants without money. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 465–466.

Anderson, R., I. Ashlagi, D. Gamarnik, and Y. Kanoria (2017). Efficient dynamic barter exchange. *Operations Research 65*(6), 1446–1459.

Ashlagi, I., M. Burq, P. Jaillet, and V. Manshadi (2019). On matching and thickness in heterogeneous dynamic markets. *Operations Research 67*(4), 927–949.

Ashlagi, I., S. Kerimov, O. Tamuz, and G. Zhao (2026). The power of two in token systems. *Management Science, forthcoming*.

Ashlagi, I., A. Nikzad, and P. Strack (2023). Matching in dynamic imbalanced markets. *The Review of Economic Studies 90*(3), 1084–1124.

Ashlagi, I. and P. Shi (2016). Optimal allocation without money: An engineering approach. *Management Science 62*(4), 1078–1097.

Asmussen, S. (2003). *Applied probability and queues*. Springer.

Ausubel, L. M. and T. Morrill (2014). Sequential kidney exchange. *American Economic Journal: Microeconomics 6*(3), 265–285.

Bertsimas, D., D. Gamarnik, and J. N. Tsitsiklis (2001). Performance of multiclass markovian queueing networks via piecewise linear lyapunov functions. *Annals of Applied Probability*, 1384–1428.

Bloch, F. and D. Cantala (2017). Dynamic assignment of objects to queuing agents. *American Economic Journal: Microeconomics 9*(1), 88–122.

Blum, A. and Y. Mansour (2020). Kidney exchange and endless paths: On the optimal use of an altruistic donor. *arXiv preprint arXiv:2010.01645*.

Çelebi, O. and J. Flynn (2022a). Adaptive priority mechanisms.

Çelebi, O. and J. P. Flynn (2022b). Priority design in centralized matching markets. *The Review of Economic Studies 89*(3), 1245–1277.

Friedman, E. J., J. Y. Halpern, and I. Kash (2006). Efficiency and nash equilibria in a scrip system for p2p networks. In *Proceedings of the 7th ACM conference on Electronic commerce*, pp. 140–149. ACM.

Hauser, C. and H. Hopenhayn (2008). Trading favors: Optimal exchange and forgiveness. *Collegio Carlo Alberto Carlo Alberto Notebooks 88*.

Johnson, K., D. Simchi-Levi, and P. Sun (2014). Analyzing scrip systems. *Operations Research 62*(3), 524–534.

Kash, I. A., E. J. Friedman, and J. Y. Halpern (2015). An equilibrium analysis of scrip systems. *ACM Transactions on Economics and Computation 3*(3), 13.

Kleinrock, L. (1975). *Theory, volume 1, queueing systems*. Wiley-interscience.

Knuth, D. E. (1997). *Stable marriage and its relation to other combinatorial problems: An introduction to the mathematical analysis of algorithms*, Volume 10. American Mathematical Soc.

Kominers, S. D. and T. Sönmez (2016). Matching with slot-specific priorities: Theory. *Theoretical Economics 11*(2), 683–710.

Little, J. D. C. (1961, June). A proof for the queuing formula: $l = \lambda w$. *Oper. Res. 9*(3), 383–387.

Möbius, M. (2001). Trading favors. *Mimeo, Microsoft Research*.

National Kidney Registry (2025). Voucher program.

Rees, M. A., J. E. Kopke, R. P. Pelletier, D. L. Segev, M. E. Rutter, A. J. Fabrega, J. Rogers, O. G. Pankewycz, J. Hiller, A. E. Roth, et al. (2009). A nonsimultaneous, extended, altruistic-donor chain. *New England Journal of Medicine 360*(11), 1096–1101.

Roth, A. E., T. Sönmez, and M. U. Ünver (2004). Kidney exchange. *The Quarterly journal of economics 119*(2), 457–488.

Roth, A. E., T. Sönmez, and M. U. Ünver (2007). Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. *American Economic Review 97*(3), 828–851.

Roth, A. E., T. Sönmez, M. U. Ünver, F. L. Delmonico, and S. L. Saidman (2006). Utilizing list exchange and nondirected donation through 'chain'paired kidney donations. *American Journal of transplantation 6*(11), 2694–2705.

Shi, P. (2022). Optimal priority-based allocation mechanisms. *Management Science 68*(1), 171–188.

Sönmez, T., M. U. Ünver, and M. B. Yenmez (2020). Incentivized kidney exchange. *American Economic Review 110*(7), 2198–2224.

Sweat, K. R. (2025). Endogenous priority in centralized matching markets: The design of the heart transplant waitlist. In *Proceedings of the 26th ACM Conference on Economics and Computation*, pp. 441–441.

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

Veale, J. L., A. M. Capron, N. Nassiri, G. Danovitch, H. A. Gritsch, A. Waterman, J. Del Pizzo, J. C. Hu, M. Pycia, S. McGuire, et al. (2017). Vouchers for future kidney transplants to overcome "chronological incompatibility" between living donors and recipients. *Transplantation 101*(9), 2115–2119.

Veale, J. L., N. Nassiri, A. M. Capron, G. M. Danovitch, H. A. Gritsch, M. Cooper, R. R. Redfield, P. T. Kennealey, and S. Kapur (2021). Voucher-based kidney donation and redemption for future transplant. *JAMA surgery 156*(9), 812–817.

A Proofs of Key Supporting Lemmas

A.1 Proof of Lemma 1

Fix $c \geq B - \nu_{\max}$ and define $\bar{V}(x) \triangleq \max\{V(x), c\}$. Since $X(t) \sim \pi$ for all t and $\mathbb{E}_\pi[\bar{V}(X(t))] < \infty$, stationarity implies

$$0 = \mathbb{E}_\pi[\bar{V}(X(t+1)) - \bar{V}(X(t))] = \sum_{x \in \mathcal{X}} \pi(x)\, \mathbb{E}[\Delta \bar{V}(X(t)) \mid X(t) = x].$$

Partition the state space into

$$A_1 = \{x : V(x) \leq c - \nu_{\max}\}, \quad A_2 = \{x : c - \nu_{\max} < V(x) \leq c + \nu_{\max}\}, \quad A_3 = \{x : V(x) > c + \nu_{\max}\}.$$

Row 1. If $x \in A_1$, then $\bar{V}(x) = c$. Moreover, for any x' with $p(x, x') > 0$,

$$V(x') \leq V(x) + \nu_{\max} \leq c,$$

so $\bar{V}(x') = c$ and hence $\Delta \bar{V} = 0$. Therefore the total contribution of A_1 is 0.

Row 2. For any x, $\Delta \bar{V} \leq (\Delta \bar{V})_+$. Also $(\Delta \bar{V})_+ \leq (V(X(t+1)) - V(x))_+$ because $\bar{V}(x) \geq V(x)$ and $\bar{V}(X(t+1)) \leq \max\{V(X(t+1)), c\}$. Thus for $x \in \mathcal{E}$,

$$\mathbb{E}[\Delta \bar{V} \mid X(t) = x] \leq \mathbb{E}[(\Delta \bar{V})_+ \mid X(t) = x] \leq \mathbb{E}[(V(X(t+1)) - V(x))_+ \mid X(t) = x] \leq q_{\text{in}},$$

and similarly for $x \notin \mathcal{E}$ we have $\mathbb{E}[\Delta \bar{V} \mid X(t) = x] \leq q_{\text{out}}$. Hence the contribution of A_2 is at most

$$q_{\text{in}} \mathbb{P}_\pi(X \in A_2 \cap \mathcal{E}) + q_{\text{out}} \mathbb{P}_\pi(X \in A_2 \cap \mathcal{E}^c).$$

Using $\mathbb{P}_\pi(X \in A_2 \cap \mathcal{E}) = \mathbb{P}_\pi(V(X) > c - \nu_{\max}, X \in \mathcal{E}) - \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \in \mathcal{E})$ and the analogous identity for $\mathcal{E}^c$, this row is bounded by

$$\begin{aligned} &q_{\text{in}} \Big(\mathbb{P}_\pi(V(X) > c - \nu_{\max}, X \in \mathcal{E}) - \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \in \mathcal{E}) \Big) \\ &\quad + q_{\text{out}} \Big(\mathbb{P}_\pi(V(X) > c - \nu_{\max}, X \notin \mathcal{E}) - \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \notin \mathcal{E}) \Big). \end{aligned}$$

Row 3. If $x \in A_3$, then for any x' with $p(x, x') > 0$,

$$V(x') \geq V(x) - \nu_{\max} > c,$$

so $\bar{V}(x) = V(x)$ and $\bar{V}(x') = V(x')$, hence $\Delta \bar{V} = \Delta V$. Since $c \geq B - \nu_{\max}$, $x \in A_3$ implies $V(x) \geq B$. Therefore,

$$\mathbb{E}[\Delta \bar{V} \mid X(t) = x] = \mathbb{E}[\Delta V \mid X(t) = x] \leq \begin{cases} -\gamma, & x \in \mathcal{E}, \\ \gamma_{\text{out}}, & x \notin \mathcal{E}. \end{cases}$$

Thus the A_3 contribution is at most

$$-\gamma\, \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \in \mathcal{E}) + \gamma_{\text{out}}\, \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \notin \mathcal{E}).$$

Combining the three rows yields

$$\begin{aligned}0 \le q_{\text{in}} &\Big(\mathbb{P}_\pi(V(X) > c - \nu_{\max}, X \in \mathcal{E}) - \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \in \mathcal{E})\Big) \\ &+ q_{\text{out}}\Big(\mathbb{P}_\pi(V(X) > c - \nu_{\max}, X \notin \mathcal{E}) - \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \notin \mathcal{E})\Big) \\ &- \gamma\, \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \in \mathcal{E}) + \gamma_{\text{out}}\, \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \notin \mathcal{E}).\end{aligned}$$

Rearranging the terms, we have:

$$\begin{aligned}&(q_{\text{in}} + \gamma)\, \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \in \mathcal{E}) \\ &\le q_{\text{in}}\, \mathbb{P}_\pi(V(X) > c - \nu_{\max}, X \in \mathcal{E}) + q_{\text{out}}\, \mathbb{P}_\pi(V(X) > c - \nu_{\max}, X \notin \mathcal{E}) \\ &\quad - q_{\text{out}}\, \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \notin \mathcal{E}) + \gamma_{\text{out}}\, \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \notin \mathcal{E})\,.\end{aligned}$$

Adding $(q_{\text{in}} + \gamma)\mathbb{P}_\pi\left(V(X) > c + \nu_{\max}, X \notin \mathcal{E}\right)$ on both sides of the above inequality, we have

$$\begin{aligned}&(q_{\text{in}} + \gamma)\mathbb{P}_\pi\left(V(X) > c + \nu_{\max}\right) \\ &\le q_{\text{in}}\, \mathbb{P}_\pi\left(V(X) > c - \nu_{\max}\right) + q_{\text{out}}\, \mathbb{P}_\pi\left(V(X) > c - \nu_{\max}, X \notin \mathcal{E}\right) \\ &\quad - q_{\text{out}}\, \mathbb{P}_\pi\left(V(X) > c + \nu_{\max}, X \notin \mathcal{E}\right) + \gamma_{\text{out}}\, \mathbb{P}_\pi(V(X) > c + \nu_{\max}, X \notin \mathcal{E}) \\ &\quad + \gamma\, \mathbb{P}_\pi\left(V(X) > c + \nu_{\max}, X \notin \mathcal{E}\right) \\ &\le q_{\text{in}}\, \mathbb{P}_\pi\left(V(X) > c - \nu_{\max}\right) + (q_{\text{out}} + \gamma_{\text{out}} + \gamma)\mathbb{P}_\pi\left(X \notin \mathcal{E}\right)\,.\end{aligned}$$

Let $\xi = \frac{q_{\text{in}}}{q_{\text{in}}+\gamma}$ and $\kappa = \frac{q_{\text{out}}+\gamma_{\text{out}}+\gamma}{q_{\text{in}}+\gamma}$. Then

$$\mathbb{P}_\pi(V(X) > c + \nu_{\max}) \le \xi\, \mathbb{P}_\pi(V(X) > c - \nu_{\max}) + \kappa\, \mathbb{P}_\pi(X \notin \mathcal{E}).$$

Iterating this inequality m times starting from $c = B + \nu_{\max}$ gives

$$\mathbb{P}_\pi(V(X) \ge B + 2\nu_{\max} m) \le \xi^m + \kappa \sum_{i=0}^{m-1} \xi^i\, \mathbb{P}_\pi(X \notin \mathcal{E}) \le \xi^m + \frac{\kappa}{1-\xi}\, \mathbb{P}_\pi(X \notin \mathcal{E}).$$

Finally, $\frac{\kappa}{1-\xi} = \frac{q_{\text{out}}+\gamma_{\text{out}}+\gamma}{\gamma} \triangleq \phi$, completing the proof.

A.2 A tail bound for queues with geometric jumps

Lemma 2 establishes a stationary tail bound for a queue-length process whose one-step increments can be unbounded, a feature that arises naturally in kidney exchange. This setting falls outside the scope of Lemma 1, which relies on uniformly bounded one-step changes of the Lyapunov function. The key idea in Lemma 2 is to retain the same drift-decomposition argument, but to replace the bounded-increment requirement by an explicit control of the overshoot probability induced by the geometric jump distribution. The resulting bound yields a geometric decay in the tail of the stationary queue length, up to additive terms that scale with the stationary probability of being outside the "good" set and with the (exponentially small) probability of an exceptionally large jump.

Lemma 2. *Let $\{Z(t) = (X(t), Y(t))\}_{t\geq 0}$ be a discrete-time Markov chain on $\mathbb{N}^2$ with stationary distribution π. Suppose $Y(t)$ evolves according to*

$$Y(t+1) = \big[Y(t) + B(t)A(t) - D(t)\big]^+.$$

Fix parameters $a, b, q > 0$, $\alpha > 0$, and $\rho \in (0,1)$. Assume there exists a set $\mathcal{E} \subseteq \mathbb{N}^2$ such that, conditional on $Z(t)$,

- *if $Z(t) \notin \mathcal{E}$, then $B(t) = 1$ almost surely;*
- *if $Z(t) \in \mathcal{E}$, then $B(t)$ is Bernoulli with mean aq^α (so $aq^\alpha \leq 1$);*
- *$A(t)$ is geometric on $\{1, 2, \dots\}$ with parameter $bq^\alpha \in (0,1]$, and $A(t)$ is independent of $\{Z(s), B(s), D(s) : s \leq t\}$;*
- *$D(t)$ is Bernoulli with mean ρ, and $D(t)$ may be correlated with $B(t)$ but satisfies $\mathbb{E}[D(t) \mid Z(t)] = \rho$.*

Assume $a/b < \rho$. Fix any $\beta > \alpha$. Then for any $c \geq 1$ and any $m \in \mathbb{N}$,

$$\mathbb{P}_\pi\left(Y \geq c + \frac{2m}{b} q^{-\beta}\right) \leq \left(\frac{a}{\rho b}\right)^m + \frac{b}{\rho b - a}\left[\left(\rho + \frac{3}{bq^\alpha}\right)\mathbb{P}_\pi\big(Z \notin \mathcal{E}\big) + \frac{a}{b}\, e^{-q^{\alpha-\beta}}\right].$$

Proof of Lemma 2. The proof follows the same drift-decomposition strategy as Lemma 1, but must be modified because the one-step increments of $Y(t)$ are unbounded (the geometric jump can be arbitrarily large), so a uniform bound $\nu_{\max} < \infty$ is unavailable. Instead, we introduce a step size

$$h \triangleq \frac{1}{b} q^{-\beta}, \qquad \beta > \alpha,$$

and later control the probability of overshooting the truncation level by more than h using the geometric tail of the jump distribution.

Let $Z(t) \triangleq (X(t), Y(t))$ denote the Markov chain state. We take the Lyapunov function $V(Z) \triangleq Y$ and, for a fixed $c > 0$, define the truncation $\bar{V}(Z) \triangleq \max\{Y, c\}$. Since a stationary distribution π exists and $\mathbb{E}_\pi[Y] < \infty$ by assumption, we have $\mathbb{E}_\pi[\bar{V}(Z)] < \infty$, and stationarity implies

$$0 = \mathbb{E}_\pi[\bar{V}(Z(t+1)) - \bar{V}(Z(t))] = \sum_{z \in \mathbb{N}^2} \pi(z)\, \mathbb{E}[\Delta \bar{V}(Z(t)) \mid Z(t) = z],$$

where $\Delta \bar{V}(Z(t)) \triangleq \bar{V}(Z(t+1)) - \bar{V}(Z(t))$. Partition the state space into three regions according to whether Y lies below, near, or above c:

$$\{Y \leq c - h\}, \qquad \{c - h < Y \leq c + h\}, \qquad \{Y > c + h\}.$$

Accordingly,

$$\begin{aligned}0 = &\sum_{z:\, V(z)\le c-h} \pi(z)\, \mathbb{E}[\Delta \bar{V}(Z(t)) \mid Z(t)=z] \\ &+ \sum_{z:\, c-h<V(z)\le c+h} \pi(z)\, \mathbb{E}[\Delta \bar{V}(Z(t)) \mid Z(t)=z] \\ &+ \sum_{z:\, V(z)>c+h} \pi(z)\, \mathbb{E}[\Delta \bar{V}(Z(t)) \mid Z(t)=z].\end{aligned}$$

We bound the contribution of the first region $\{V(z) \le c-h\}$, where $V(z)=y$ and $h \triangleq b^{-1}q^{-\beta}$. Fix a state $z=(x,y)$ with $y \le c-h$. Since $\bar{V}(z)=c$,

$$\Delta \bar{V}(Z(t)) = \bar{V}(Z(t+1)) - \bar{V}(Z(t)) = \big(Y(t+1)-c\big)_+.$$

Using the recursion $Y(t+1) = [Y(t)+B(t)A(t)-D(t)]^+$ and the fact that $(u^+-c)_+ = (u-c)_+$ for $c>0$, we obtain

$$\begin{aligned}\big(Y(t+1)-c\big)_+ &= \big(Y(t)+B(t)A(t)-D(t)-c\big)_+ \\ &= \big(B(t)A(t)-D(t)-(c-Y(t))\big)_+.\end{aligned}$$

On the event $\{Y(t)\le c-h\}$ we have $c-Y(t)\ge h$, and since $D(t)\ge 0$,

$$\big(B(t)A(t)-D(t)-(c-Y(t))\big)_+ \le \big(B(t)A(t)-h\big)_+ \le B(t)\,(A(t)-h)_+,$$

where we used that $B(t)\in\{0,1\}$. Taking conditional expectations given $Z(t)=z$ and using that $A(t)$ is independent of $\{Z(s),B(s),D(s): s\le t\}$, we have

$$\mathbb{E}[\Delta \bar{V}(Z(t)) \mid Z(t)=z] \le \mathbb{E}[B(t)\mid Z(t)=z]\cdot \mathbb{E}[(A(t)-h)_+].$$

By the definition of $\mathcal{E}$, $\mathbb{E}[B(t)\mid Z(t)=z]=1$ if $z\notin\mathcal{E}$ and $\mathbb{E}[B(t)\mid Z(t)=z]=aq^\alpha$ if $z\in\mathcal{E}$. Therefore,

$$\sum_{z:\, V(z)\le c-h} \pi(z)\, \mathbb{E}[\Delta \bar{V}(Z(t)) \mid Z(t)=z] \le \big(\mathbb{P}_\pi(Z\notin\mathcal{E}) + aq^\alpha\big)\, \mathbb{E}[(A-h)_+].$$

Finally, if $A\sim \mathrm{Geo}(p)$ on $\{1,2,\dots\}$, then for any $h>0$,

$$\mathbb{E}[(A-h)_+] \le \sum_{n=\lfloor h\rfloor+1}^{\infty} \mathbb{P}(A\ge n) = \sum_{n=\lfloor h\rfloor+1}^{\infty} (1-p)^{n-1} \le \frac{(1-p)^h}{p} \le \frac{e^{-ph}}{p}.$$

With $p=bq^\alpha$ and $h=b^{-1}q^{-\beta}$, this yields

$$\mathbb{E}[(A-h)_+] \le \frac{1}{bq^\alpha}\, e^{-q^{\alpha-\beta}}.$$

Now we consider states z such that $c-h<V(z)\le c+h$. As in Lemma 1, we bound

$\Delta\bar{V}$ by its positive part and then by the positive one-step increase of V:

$$\Delta\bar{V}(Z(t)) \le (\Delta\bar{V}(Z(t)))_+ \le (V(Z(t+1)) - V(z))_+ \quad \text{when } Z(t) = z.$$

Define

$$q_{\text{in}} \triangleq \sup_{z\in\mathcal{E}} \mathbb{E}[(V(Z(t+1)) - V(z))_+ \mid Z(t) = z]\,, \qquad q_{\text{out}} \triangleq \sup_{z\notin\mathcal{E}} \mathbb{E}[(V(Z(t+1)) - V(z))_+ \mid Z(t) = z]\,.$$

Then the contribution of the second region is bounded by

$$\begin{aligned}&\sum_{z:\, c-h<V(z)\le c+h} \pi(z)\,\mathbb{E}[\Delta\bar{V}(Z(t)) \mid Z(t) = z]\\ &\le q_{\text{in}}\Big(\mathbb{P}_\pi(V > c-h, Z\in\mathcal{E}) - \mathbb{P}_\pi(V > c+h, Z\in\mathcal{E})\Big)\\ &\quad + q_{\text{out}}\Big(\mathbb{P}_\pi(V > c-h, Z\notin\mathcal{E}) - \mathbb{P}_\pi(V > c+h, Z\notin\mathcal{E})\Big).\end{aligned}$$

Finally, consider states z such that $V(z) > c+h$. For the queue recursion, the one-step decrease of Y is at most 1, hence $V(Z(t+1)) \ge V(z) - 1$. If we take $c \ge 1$ and $h \ge 1$, then $V(z) > c+h$ implies $V(Z(t+1)) \ge c$, so truncation is inactive and $\Delta\bar{V}(Z(t)) = \Delta V(Z(t))$ on this region. Assume there exists $\gamma > 0$ such that

$$\mathbb{E}[\Delta V(Z(t)) \mid Z(t) = z] \le -\gamma \qquad \forall\, z\in\mathcal{E} \text{ with } V(z) \ge 1,$$

and define

$$\gamma_{\text{out}} \triangleq \sup_{z\notin\mathcal{E}} \mathbb{E}[\Delta V(Z(t)) \mid Z(t) = z] < \infty.$$

Then

$$\sum_{z:\, V(z)>c+h} \pi(z)\,\mathbb{E}[\Delta\bar{V}(Z(t)) \mid Z(t) = z] \le\ -\gamma\,\mathbb{P}_\pi(V > c+h, Z\in\mathcal{E}) + \gamma_{\text{out}}\,\mathbb{P}_\pi(V > c+h, Z\notin\mathcal{E}).$$

Combining the bounds from the three regions, let $h \triangleq b^{-1}q^{-\beta}$. From stationarity, $0 = \sum_z \pi(z)\mathbb{E}[\Delta\bar{V}(Z(t)) \mid Z(t) = z]$, and the three row bounds imply

$$\begin{aligned}0 \le R_1 &+ q_{\text{in}}\Big(\mathbb{P}_\pi(V > c-h, Z\in\mathcal{E}) - \mathbb{P}_\pi(V > c+h, Z\in\mathcal{E})\Big)\\ &+ q_{\text{out}}\Big(\mathbb{P}_\pi(V > c-h, Z\notin\mathcal{E}) - \mathbb{P}_\pi(V > c+h, Z\notin\mathcal{E})\Big)\\ &- \gamma\,\mathbb{P}_\pi(V > c+h, Z\in\mathcal{E}) + \gamma_{\text{out}}\,\mathbb{P}_\pi(V > c+h, Z\notin\mathcal{E}),\end{aligned}$$

where the Row 1 term is

$$R_1 \triangleq \big(\mathbb{P}_\pi(Z\notin\mathcal{E}) + aq^\alpha\big)\cdot\frac{1}{bq^\alpha}e^{-q^{\alpha-\beta}}.$$

Rearranging yields

$$
\begin{aligned}
(q_{\text{in}} + \gamma)\, \mathbb{P}_\pi(V > c+h, Z \in \mathcal{E}) \le q_{\text{in}}\, \mathbb{P}_\pi(V > c-h, Z \in \mathcal{E}) + q_{\text{out}}\, \mathbb{P}_\pi(V > c-h, Z \notin \mathcal{E}) \\
+ (\gamma_{\text{out}} - q_{\text{out}})\, \mathbb{P}_\pi(V > c+h, Z \notin \mathcal{E}) + R_1.
\end{aligned}
$$

Adding $(q_{\text{in}} + \gamma)\, \mathbb{P}_\pi\,(V > c+h, Z \notin \mathcal{E})$ on both sides, we have:

$$
\begin{aligned}
&(q_{\text{in}} + \gamma)\, \mathbb{P}_\pi\,(V > c+h) \\
&\le q_{\text{in}}\, \mathbb{P}_\pi\,(V > c-h) + (q_{\text{out}} + \gamma_{\text{out}} + \gamma)\, \mathbb{P}_\pi\,(Z \notin \mathcal{E}) + R_1 \\
&= q_{\text{in}}\, \mathbb{P}_\pi\,(V > c-h) + \left(q_{\text{out}} + \gamma_{\text{out}} + \gamma + \frac{e^{-q^{\alpha-\beta}}}{bq^\alpha}\right) \mathbb{P}_\pi\,(Z \notin \mathcal{E}) + \frac{a}{b} e^{-q^{\alpha-\beta}}. \qquad (1)
\end{aligned}
$$

Define

$$
\xi \triangleq \frac{q_{\text{in}}}{q_{\text{in}} + \gamma}, \qquad \kappa \triangleq \frac{1}{q_{\text{in}} + \gamma} \left[\left(q_{\text{out}} + \gamma_{\text{out}} + \gamma + \frac{e^{-q^{\alpha-\beta}}}{bq^\alpha} \right) \mathbb{P}_\pi(Z \notin \mathcal{E}) + \frac{a}{b} e^{-q^{\alpha-\beta}} \right].
$$

Then (1) can be rewritten as

$$
\mathbb{P}_\pi(V > c+h) \le \xi\, \mathbb{P}_\pi(V > c-h) + \kappa.
$$

Recursively applying the above one-step recursion with $c = c_0 + (2i-1)h$ for $i = 1, 2, \ldots, m$, yields

$$
\begin{aligned}
\mathbb{P}_\pi(V > c_0 + 2mh) &\le \xi^m\, \mathbb{P}_\pi(V > c_0) + \kappa \sum_{i=0}^{m-1} \xi^i \\
&\le \xi^m + \frac{\kappa}{1-\xi},
\end{aligned}
$$

where we used $\mathbb{P}_\pi(V > c_0) \le 1$ and $\xi \in (0, 1)$.

Next, we bound $q_{\text{in}}, q_{\text{out}}, \gamma, \gamma_{\text{out}}$ using $V(Z) = Y$. Since $B(t) \in \{0, 1\}$ and $D(t) \ge 0$,

$$
(Y(t+1) - Y(t))_+ \le (B(t)A(t))_+ = B(t)A(t).
$$

If $Z(t) \in \mathcal{E}$, then $\mathbb{E}[B(t) \mid Z(t)] = aq^\alpha$ and $A(t)$ is independent of $B(t)$, so

$$
\begin{aligned}
q_{\text{in}} &= \sup_{z \in \mathcal{E}} \mathbb{E}\big[(Y(t+1) - Y(t))_+ \mid Z(t) = z\big] \\
&\le \sup_{z \in \mathcal{E}} \mathbb{E}[B(t)A(t) \mid Z(t) = z] = aq^\alpha \cdot \mathbb{E}[A(t)] = \frac{a}{b}.
\end{aligned}
$$

If $Z(t) \notin \mathcal{E}$, then $B(t) \equiv 1$, hence

$$
q_{\text{out}} \le \mathbb{E}[A(t)] = \frac{1}{bq^\alpha}.
$$

For the drift on $\mathcal{E}$, assume that $\mathbb{E}[D(t) \mid Z(t) = z] = \rho$ for all z (or at least for $z \in \mathcal{E}$).

Then for $z \in \mathcal{E}$ with $Y \geq 1$,

$$\mathbb{E}[\Delta V(Z(t)) \mid Z(t) = z] = \mathbb{E}[B(t) \mid z]\mathbb{E}[A(t)] - \mathbb{E}[D(t) \mid z] \leq \frac{a}{b} - \rho = -\gamma, \qquad \gamma \triangleq \rho - \frac{a}{b} > 0.$$

Also,

$$\gamma_{\text{out}} \leq \sup_{z \notin \mathcal{E}} \mathbb{E}[B(t)A(t) \mid Z(t) = z] \leq \mathbb{E}[A(t)] = \frac{1}{bq^{\alpha}}.$$

With these bounds, $q_{\text{in}} + \gamma = \rho$, and hence

$$\xi = \frac{q_{\text{in}}}{q_{\text{in}} + \gamma} \leq \frac{a/b}{\rho} = \frac{a}{\rho b}.$$

Moreover, from the definition of κ in the recursion,

$$\kappa \leq \frac{1}{\rho}\left(\rho + \frac{3}{bq^{\alpha}}\right)\mathbb{P}_{\pi}(Z \notin \mathcal{E}) + \frac{a}{\rho b}\, e^{-q^{\alpha-\beta}},$$

where we used $\gamma \leq \rho$, $q_{\text{out}} + \gamma_{\text{out}} \leq \frac{2}{bq^{\alpha}}$, and $\frac{e^{-q^{\alpha-\beta}}}{bq^{\alpha}} \leq \frac{1}{bq^{\alpha}}$. □

A.3 A bound for the truncated geometric random variable

The next lemma lower bounds the expectation of a truncated geometric random variable. This will be used frequently in our analysis.

Lemma 3. *Consider a Geometric distribution X with parameter p, i.e, the probability mass function is $\mathbb{P}(X = k) = (1-p)^k p$ for $k = 0, 1, 2 \cdots$. Define the truncated geometric variable $\tilde{X}$ as*

$$\tilde{X} \triangleq X \cdot \mathbb{1}\{X \leq K\},$$

then for $\epsilon > 2p$ and $K \geq \frac{2}{p}\log\left(\frac{4}{\epsilon}\right)$, we have

$$\mathbb{E}[\tilde{X}] \geq \frac{1-\epsilon}{p}.$$

Proof of Lemma 3. Let K be a positive integer. We have:

$$\begin{aligned}
\mathbb{E}[\tilde{X}] &= \sum_{i=0}^{K} (1-p)^i \cdot p \cdot i \\
&= \sum_{i=0}^{\infty} (1-p)^i \cdot p \cdot i - \sum_{i=K+1}^{\infty} (1-p)^i \cdot p \cdot i \\
&= \frac{1-p}{p} - \sum_{j=0}^{\infty} (1-p)^{(j+K+1)} \cdot p \cdot (j+K+1) \\
&= \frac{1-p}{p} - (1-p)^{K+1} \sum_{j=0}^{\infty} (1-p)^j \cdot p \cdot (j+K+1) \\
&= \frac{1-p}{p} - (1-p)^{K+1} \left[\sum_{j=0}^{\infty} (1-p)^j \cdot p \cdot j + (K+1) \sum_{j=0}^{\infty} (1-p)^j \cdot p \right] \\
&= \frac{1-p}{p} - (1-p)^{K+1} \left[\frac{1-p}{p} + (K+1) \right] \\
&= \frac{1-p}{p} - (1-p)^{K+1} \frac{1-p}{p} - (K+1)(1-p)^{K+1} . \\
&= \frac{1-p}{p} \left(1 - (1-p)^{K+1}(1+p(K+1)) \right)
\end{aligned}$$

We then have:

$$\begin{aligned}
\frac{1-p}{p} \left(1 - (1-p)^{K+1}(1+p(K+1)) \right) &\geq \frac{1-p}{p} \left(1 - e^{-p(K+1)}(1+p(K+1)) \right) \\
&\geq \frac{1}{p} \left(1 - e^{-p(K+1)}(1+p(K+1)) \right) - 1 \\
&= \frac{1}{p} \left(1 - p - e^{-p(K+1)}(1+p(K+1)) \right)
\end{aligned}$$

Choose $K \geq \frac{2}{p} \log\left(\frac{4}{\epsilon}\right)$, we have:

$$e^{-p(K+1)}(1+p(K+1)) \leq e^{-p(K+1)} 2 e^{\frac{1}{2}p(K+1)} = 2e^{-\frac{1}{2}p(K+1)} \leq \frac{\epsilon}{2} ,$$

where in the first inequality we use the inequality $1 + x \leq 2e^{x/2}$, $\forall x \in \mathbb{R}$.

We further require that $\epsilon > 2p$, then,

$$\frac{1}{p} \left(1 - p - e^{-p(K+1)}(1+p(K+1)) \right) \geq \frac{1}{p}(1-\epsilon) .$$

This concludes the proof. □

A.4 Other Useful Facts

We state two inequalities that we use repeatedly:

$$1 - x \le e^{-x}, \quad \forall x \in \mathbb{R}, \tag{2}$$

and

$$e^{-x} \le 1 - x + \frac{x^2}{2}, \quad \forall x \ge 0. \tag{3}$$

A variant of the above inequality is:

$$1 - x \ge e^{-x-x^2}, \quad \forall x \in [0, \frac{1}{2}] \tag{4}$$

B Proof of Proposition 1

Let

$$a_x := 1 - (1-p)^x, \qquad b_x := (1-p)^x.$$

Thus a_x is the probability that a given patient is compatible with at least one of x free donors, and also the probability that a given donor is compatible with at least one of x voucher patients.

We first record two elementary drift estimates.

Step 1: Drift of the voucher pool when X is large. In one period, X can increase by at most one and decrease by at most one. Consider the event that the arriving patient is compatible with some current free donor and that the arriving donor is compatible with some current voucher patient. Conditional on $X(t) = x$, this event has probability a_x^2. On this event, the arriving patient is matched by an existing free donor, the arriving donor becomes free, and then Standard Priority immediately matches this donor to a voucher patient. Hence X decreases by one. On the complement, X increases by at most one. Therefore

$$\mathbb{E}[\Delta X(t) \mid X(t) = x, Y(t) = y] \le -a_x^2 + (1 - a_x^2) = 1 - 2a_x^2, \tag{5}$$

where $\Delta X(t) = X(t+1) - X(t)$.

In addition, X can increase by one only if the arrival is voucher pair, the arriving patient is not matched by any of the x current free donors, and the arriving donor is not compatible with any of the x current voucher patients. Hence

$$\mathbb{P}(\Delta X(t) = 1 \mid X(t) = x, Y(t) = y) \le q b_x^2. \tag{6}$$

Step 2: Drift of the non-voucher pool when X is bounded and Y is large. We claim that for every fixed integer $K < \infty$, there exist constants $\eta_K > 0$ and $Y_K < \infty$ such that

$$\mathbb{E}[\Delta Y(t) \mid X(t) = x, Y(t) = y] \le -\eta_K \qquad \text{for all } 0 \le x \le K,\ y \ge Y_K. \tag{7}$$

To prove the claim, fix x and let $y \to \infty$. Suppose first that $x \ge 1$. If a non-voucher patient arrives, then with probability b_x she is not matched by any free donor, in which case Y increases by one. With probability a_x, she is matched, her donor becomes free, and this free donor initiates a chain.

Conditional on such a chain starting with x voucher patients and y non-voucher pairs, the probability that the chain removes at least ℓ non-voucher patients is bounded below by

$$b_x^\ell \prod_{j=0}^{\ell-1} \bigl(1-(1-p)^{y-j}\bigr).$$

Indeed, for each of the first ℓ steps, it is enough that the current donor avoids all x voucher patients and is compatible with at least one remaining non-voucher patient. Therefore, for every fixed L,

$$\mathbb{E}[\text{non-voucher removals from such a chain}] \geq \sum_{\ell=1}^{L} b_x^\ell \prod_{j=0}^{\ell-1} \bigl(1-(1-p)^{y-j}\bigr).$$

Letting $y \to \infty$ and then $L \to \infty$, we obtain

$$\liminf_{y\to\infty} \mathbb{E}[\text{non-voucher removals from a matched non-voucher arrival}] \geq \sum_{\ell=1}^{\infty} b_x^\ell = \frac{b_x}{a_x}. \tag{8}$$

Now consider a voucher arrival. Its donor initiates a chain. To lower bound the number of non-voucher patients removed, require first that the arriving donor avoid the x incumbent voucher patients and match an non-voucher patient. After the first non-voucher patient is removed, the chain has at most $x+1$ voucher patients to avoid. Thus, for every fixed L,

$$\mathbb{E}[\text{non-voucher removals from a voucher arrival}] \geq \sum_{\ell=1}^{L} b_x b_{x+1}^{\ell-1} \prod_{j=0}^{\ell-1} \bigl(1-(1-p)^{y-j}\bigr).$$

Letting $y \to \infty$ and then $L \to \infty$,

$$\liminf_{y\to\infty} \mathbb{E}[\text{non-voucher removals from a voucher arrival}] \geq \sum_{\ell=1}^{\infty} b_x b_{x+1}^{\ell-1} = \frac{b_x}{a_{x+1}}. \tag{9}$$

Combining (8) and (9), for every fixed $x \geq 1$,

$$\begin{aligned}\limsup_{y\to\infty} \mathbb{E}[\Delta Y(t) \mid X(t)=x, Y(t)=y] &\leq (1-q)\left(b_x - a_x\frac{b_x}{a_x}\right) - q\frac{b_x}{a_{x+1}} \\ &= -q\frac{b_x}{a_{x+1}} < 0.\end{aligned} \tag{10}$$

It remains to handle $x=0$. If $x=0$, a non-voucher arrival cannot be matched by a free donor, so it increases Y by one. A voucher arrival creates one voucher patient and a free donor. With probability tending to one as $y \to \infty$, the voucher donor matches a non-voucher patient first. Thereafter each subsequent non-voucher donor hits the voucher patient with probability p, and otherwise can continue through another non-voucher patient with probability tending to one. Hence

$$\liminf_{y\to\infty} \mathbb{E}[\text{non-voucher removals from a voucher arrival} \mid X(t)=0, Y(t)=y] \geq \sum_{\ell=1}^{\infty}(1-p)^{\ell-1} = \frac{1}{p}.$$

Therefore

$$\limsup_{y\to\infty} \mathbb{E}[\Delta Y(t) \mid X(t)=0, Y(t)=y] \leq (1-q) - \frac{q}{p} \leq -p, \tag{11}$$

where the last inequality uses $q \geq p$.

Since the set $\{0, 1, \dots, K\}$ is finite, (10) and (11) imply the uniform claim (7).

We now construct the Lyapunov function. Choose $B = 4$. Since $a_R \to 1$ and $b_R \to 0$ as $R \to \infty$, choose an integer $R \geq 1$ large enough that

$$a_R^2 \geq \frac{3}{4} \qquad \text{and} \qquad Bb_R a_{R+1} \leq \frac{1}{4}. \tag{12}$$

Define

$$V(x,y) := y + B(x-R)^+. \tag{13}$$

This function has finite sublevel sets: if $V(x,y) \leq M$, then $y \leq M$ and $x \leq R + M/B$.

We show that V has negative drift outside a finite set.

First suppose $x \geq R+1$. Since $\Delta X(t) \in \{-1, 0, 1\}$, we have

$$(X(t+1)-R)^+ - (x-R)^+ = \Delta X(t).$$

Moreover, $\Delta Y(t) \leq 1$, because at most one non-voucher patient arrives in one period. Using (5) and (12),

$$\begin{aligned} \mathbb{E}[\Delta V(t) \mid X(t)=x, Y(t)=y] &= \mathbb{E}[\Delta Y(t) \mid x,y] + B\mathbb{E}[\Delta X(t) \mid x,y] \\ &\leq 1 + B(1-2a_x^2) \\ &\leq 1 + B\left(1 - 2\cdot\frac{3}{4}\right) = 1 - \frac{B}{2} = -1. \end{aligned} \tag{14}$$

Next suppose $x \leq R-1$. Then a one-step increase in X cannot make $(X-R)^+$ positive, so

$$(X(t+1)-R)^+ - (x-R)^+ = 0.$$

By the claim (7), applied with $K = R$, there exist $Y_R < \infty$ and $\eta_R > 0$ such that for all $y \geq Y_R$,

$$\mathbb{E}[\Delta V(t) \mid X(t)=x, Y(t)=y] = \mathbb{E}[\Delta Y(t) \mid X(t)=x, Y(t)=y] \leq -\eta_R. \tag{15}$$

It remains to consider the boundary $x = R$. At $x = R$, the kinked term can increase only if $\Delta X(t) = 1$, so by (6),

$$\mathbb{E}\big[(X(t+1) - R)^+ - (X(t) - R)^+ \mid X(t) = R, Y(t) = y\big] \leq q b_R^2. \tag{16}$$

By (10), increasing Y_R if necessary, we may also ensure that for all $y \geq Y_R$,

$$\mathbb{E}[\Delta Y(t) \mid X(t) = R, Y(t) = y] \leq -\frac{q}{2} \frac{b_R}{a_{R+1}}. \tag{17}$$

Combining (16), (17), and the choice of R in (12), we get

$$\begin{aligned} \mathbb{E}[\Delta V(t) \mid X(t) = R, Y(t) = y] &\leq -\frac{q}{2} \frac{b_R}{a_{R+1}} + Bqb_R^2 \\ &= -q \frac{b_R}{a_{R+1}} \left(\frac{1}{2} - Bb_R a_{R+1} \right) \\ &\leq -\frac{q}{4} \frac{b_R}{a_{R+1}} < 0. \end{aligned} \tag{18}$$

Let

$$C := \{0, 1, \dots, R\} \times \{0, 1, \dots, Y_R - 1\}.$$

This set is finite. Equations (14), (15), and (18) imply that there exists $\varepsilon > 0$ such that

$$\mathbb{E}[V(Z(t+1)) - V(Z(t)) \mid Z(t) = (x, y)] \leq -\varepsilon \qquad \text{for all } (x, y) \notin C. \tag{19}$$

Finally, the chain is irreducible on $\mathbb{Z}_+^2$. From any state, there is positive probability of reaching $(0, 0)$: voucher patients can be removed one at a time by arrivals whose patients match existing free donors and whose donors then match voucher patients, and once $X = 0$, a voucher arrival can initiate a finite chain that removes all remaining non-voucher patients and terminates at the new voucher patient. Conversely, from $(0, 0)$, any state (x, y) can be reached by x voucher arrivals and y non-voucher arrivals with no compatible edges realized among the finite set of relevant donors and patients. Each of these finite compatibility patterns has strictly positive probability.

Thus $Z(t)$ is an irreducible countable-state Markov chain with a Lyapunov function V having finite sublevel sets and uniformly negative drift outside a finite set. By the Foster-Lyapunov criterion, $Z(t)$ is positive recurrent. Consequently it admits a stationary distribution.

C Proof of Proposition 2

We calculate the rate at which non-voucher patients are removed. We charge every non-voucher removal to a *proper chain*: recall that a proper chain is the concatenation of chain segments, each of which except the last one ending with a non-voucher patient.

We upper bound the length of proper chains. Since all chains must be initiated by a free donor, there must be at least one voucher patient every time the chain extends. So the probability that the chain ends is at least p, independent of the past. Additionally, since matching is done upon *every* arrival, only one proper chain can extend at a time. Hence the number of non-voucher removals in a proper chain is stochastically dominated by a geometric random variable on $\{0, 1, 2, \cdots\}$ with success probability p, and has expectation at most $(1-p)/p$.[19] Each proper chain starts with a voucher donor, who arrives at a rate of q. Therefore, the removal rate of non-voucher patients is at most

$$q(1-p)/p\,.$$

On the other hand, the arrival rate of non-voucher pairs is $1-q$. Therefore, when $q < p$, we have

$$\frac{q}{p}(1-p) < 1-q\,,$$

thus the number of non-voucher patients grows at least linearly in expectation with time and cannot have a stationary distribution with finite mean. Little's Law then rules out a finite steady-state expected waiting time for non-voucher patients.

D Proofs in Section 3

D.1 Proof of Proposition 3

Note that $X(t)$ denotes the number of *waiting voucher patients*. Let $\Delta X(t) \triangleq X(t+1)-X(t)$. Let $\mathcal{F}_t$ denote the natural filtration up to time t. We will upper bound $\mathbb{E}[\Delta X(t) \mid \mathcal{F}_t]$ by accounting for arrivals and removals during period $t \to t+1$.

In period $t \to t+1$, with probability q a voucher pair arrives. In that case, at most one new voucher patient can be added to the waiting pool (the arriving patient), and only if the arrival is not immediately matched. Thus, the arrival contributes at most $+1$ to $X(t)$.

Next we bound the number of voucher patients removed in period $t \to t+1$. Since each donor–patient compatibility edge is present independently with probability p, conditional on the state at time t, the probability that an arriving donor is compatible with at least one of the $X(t)$ waiting voucher patients equals $1-(1-p)^{X(t)}$. Likewise, the probability that an arriving patient is compatible with at least one waiting voucher donor equals $1-(1-p)^{X(t)}$.

- *Voucher-pair arrival.* A voucher patient is removed if the arriving donor can donate to some waiting voucher patient; this occurs with probability $1-(1-p)^{X(t)}$. In addition, the arriving voucher patient may be removed (matched immediately) if they are compatible with some waiting voucher donor; this also occurs with probability $1-(1-p)^{X(t)}$.

- *Non-voucher-pair arrival.* A waiting voucher patient can be removed if the arriving pair enables a two-way exchange between the voucher pool and the arrival. A sufficient condition is that (i) the arriving patient is compatible with at least one waiting voucher donor and (ii) the arriving donor is compatible with at least one waiting voucher patient. We ignore other possible removal mechanisms (which can only increase the

[19] Statistics on the average proper chain lengths corroborate this result in Table 3.

removal probability), so focusing on this sufficient condition yields an upper bound of $X(t+1)$. Since the donor–patient compatibilities are independent and each edge is present with probability p, the two compatibility events occur with probability $1-(1-p)^{X(t)}$ each, and hence the joint event has probability $\big(1-(1-p)^{X(t)}\big)^2$.

Therefore, we have

$$\begin{aligned}\mathbb{E}[\Delta X(t) \mid \mathcal{F}_t] &\leq q - 2q\Big[1-(1-p)^{X(t)}\Big] - (1-q)\Big[1-(1-p)^{X(t)}\Big]^2 \\ &\leq q - (1-q)\Big[1-(1-p)^{X(t)}\Big]^2.\end{aligned}$$

The first inequality uses that, under a voucher-pair arrival, at most one voucher patient arrives (contributing $+1$ to X), while two independent compatibility events can each remove one voucher patient, and each event occurs with probability $1-(1-p)^{X(t)}$. Under a non-voucher-pair arrival, the sufficient two-way exchange event occurs with probability $\big[1-(1-p)^{X(t)}\big]^2$ and removes one waiting voucher patient.

We will apply Lemma 1 to the Markov chain $Z(t) = (X(t), Y(t))$ with Lyapunov function $V(Z) = X$. We take $\mathcal{E} = \mathcal{X}$ (the full state space), so $\mathbb{P}_\pi(Z \notin \mathcal{E}) = 0$. Set

$$B \triangleq \frac{3}{2}\frac{\sqrt{q}}{p}.$$

We have:

$$\begin{aligned}\mathbb{E}[\Delta X(t) \mid \mathcal{F}_t,\, X(t) \geq B] &\leq q - (1-q)\Big[1-(1-p)^{X(t)}\Big]^2 \\ &\leq q - (1-q)\Big[1-e^{-pX(t)}\Big]^2 \\ &\leq q - (1-q)\Big[1-e^{-pB}\Big]^2 = q - (1-q)\Big[1-e^{-\frac{3}{2}\sqrt{q}}\Big]^2. \end{aligned} \tag{20}$$

The second inequality uses (2), the third inequality uses $X(t) \geq B$.

Using (3) with $x = \frac{3}{2}\sqrt{q}$, for all $q \in (0,1)$,

$$e^{-\frac{3}{2}\sqrt{q}} \leq 1 - \frac{3}{2}\sqrt{q} + \frac{9}{8}q. \tag{21}$$

Combining (21) with (20) yields, for $X(t) \geq B$,

$$\begin{aligned}\mathbb{E}[\Delta X(t) \mid \mathcal{F}_t,\, X(t) \geq B] &\leq q - (1-q)\left(1-e^{-\frac{3}{2}\sqrt{q}}\right)^2 \\ &\leq q - (1-q)\left(\frac{3}{2}\sqrt{q} - \frac{9}{8}q\right)^2.\end{aligned}$$

Expanding the right-hand side gives

$$q - (1-q)\left(\frac{3}{2}\sqrt{q} - \frac{9}{8}q\right)^2 = -\frac{5}{4}q + O(q^{3/2}).$$

Therefore, there exists $\epsilon > 0$ such that for all $q < \epsilon$,

$$\mathbb{E}[\Delta X(t) \mid \mathcal{F}_t,\, X(t) \geq B] \leq -q.$$

Therefore, Lemma 1 applies with drift parameter $\gamma = q$ and threshold $B = \frac{3}{2}\frac{\sqrt{q}}{p}$. Moreover, since at most one voucher patient can arrive per period, $(\Delta X(t))_+ \leq 1$ and $\mathbb{E}[(\Delta X(t))_+ \mid Z(t) = z] \leq q$, so $q_{\text{in}}, q_{\text{out}} \leq q$. Since $\mid \Delta X(t) \mid \leq 1$, then $\nu_{\max} = 1$ and hence

$$\xi = \frac{q_{\text{in}}}{q_{\text{in}} + \gamma} \leq \frac{q}{q+q} = \frac{1}{2}.$$

Therefore, for all $m \in \mathbb{N}$,

$$\mathbb{P}_\pi(X \geq B + 2m) \leq \left(\frac{1}{2}\right)^m = e^{-m \log 2}.$$

In particular, for the level $2\frac{\sqrt{q}}{p}$, note that

$$2\frac{\sqrt{q}}{p} - B = \frac{1}{2}\frac{\sqrt{q}}{p}.$$

Choosing $m = \left\lceil \frac{1}{4}\frac{\sqrt{q}}{p} \right\rceil$ ensures $B + 2m \geq 2\frac{\sqrt{q}}{p}$, hence

$$\mathbb{P}_\pi\left(X \geq 2\frac{\sqrt{q}}{p}\right) \leq \exp\left(-\log 2 \cdot \left\lceil \frac{1}{4}\frac{\sqrt{q}}{p} \right\rceil\right) \leq \exp\left(-\frac{\log 2}{4}\frac{\sqrt{q}}{p}\right).$$

Given that $q \geq p$, then $\frac{\sqrt{q}}{p} \geq \frac{1}{\sqrt{p}}$, giving

$$\mathbb{P}_\pi\left(X \geq 2\frac{\sqrt{q}}{p}\right) \leq \exp\left(-\frac{\log 2}{4} \cdot \frac{1}{\sqrt{p}}\right) \leq \exp\left(-0.17\frac{1}{\sqrt{p}}\right).$$

This concludes the proof.

D.2 Proof of Proposition 4

In this step, we bound Y from below conditioned on X is small. Let

$$\Delta Y(t) \triangleq Y(t+1) - Y(t).$$

For notational convenience, we work with $-\Delta Y(t)$, which captures the net number of non-voucher patients removed in period $t \to t+1$ after accounting for any new non-voucher arrival.

Fix time t and let $\mathcal{F}_t$ denote the natural filtration up to time t. Let $W(t)$ indicate the arrival type in period $t \to t+1$, where $W(t) = 1$ corresponds to a voucher arrival and $W(t) = 0$ to a non-voucher arrival; conditional on $\mathcal{F}_t$, we have $W(t) \sim \text{Bernoulli}(q)$.

When $W(t) = 0$ (a non-voucher arrival), define $U(t)$ as the indicator that the arriving non-voucher patient is compatible with at least one voucher donor present at time t. Under the independent-compatibility model,

$$\mathbb{P}\big(U(t) = 1 \mid \mathcal{F}_t\big) = 1 - (1-p)^{X(t)}.$$

Finally, let $Z(t)$ denote the number of non-voucher patients removed by the chain initiated by a "free donor" in that period, whenever such a chain is triggered.

The one-step change admits the representation

$$-\Delta Y(t) = W(t)\, Z(t) + (1 - W(t))\big(U(t)(1 + Z(t)) - 1\big).$$

To see this, note that if a voucher pair arrives ($W(t) = 1$), no new non-voucher patient enters Y, and the arriving donor may initiate a chain that removes $Z(t)$ non-voucher patients. If a non-voucher pair arrives ($W(t) = 0$), then Y increases by 1; moreover, if the arriving patient is matched to a voucher donor ($U(t) = 1$), in addition to the arriving non-voucher patient, $Z(t)$ non-voucher patients will be removed by a chain initiated by the arriving donor who becomes free.

The distribution of the chain length $Z(t)$ is complicated. We therefore upper bound $Z(t)$ in stochastic order by a geometric random variable whose parameter depends only on $Y(t)$.

Claim. Conditional on $\mathcal{F}_t$, the number $Z(t)$ of non-voucher patients removed by a donor-initiated chain is stochastically dominated by a geometric random variable $\tilde{Z}(t)$ on $\{0, 1, 2, \dots\}$ with success (termination) probability $(1-p)^{Y(t)}$; that is, $Z(t) \preccurlyeq \tilde{Z}(t)$ and

$$\mathbb{P}(\tilde{Z}(t) = k \mid \mathcal{F}_t) = \big(1 - (1-p)^{Y(t)}\big)^k (1-p)^{Y(t)}, \qquad k \geq 0.$$

Justification. Consider any step of an active chain. In order for the chain to remove one additional non-voucher patient, the current donor must (i) have no compatible voucher patient among the $X(t)$ voucher patients, and (ii) have at least one compatible non-voucher patient among the (at most) $Y(t)$ non-voucher patients. Under independent edge formation, these events have probabilities at most $(1-p)^{X(t)}$ and $1 - (1-p)^{Y(t)}$, respectively. Hence the probability of extending the chain by one more non-voucher patient is at most $\big[1 - (1 - p)^{Y(t)}\big](1-p)^{X(t)}$, and therefore the probability that the chain terminates before removing another non-voucher patient is at least

$$1 - \big[1 - (1-p)^{Y(t)}\big](1-p)^{X(t)} \;\geq\; (1-p)^{Y(t)}.$$

Moreover, as the chain removes non-voucher patients, Y decreases, which only increases the termination probability. Thus, treating the termination probability as constantly $(1-p)^{Y(t)}$ yields a worst-case upper bound on the chain length, implying $Z(t) \preccurlyeq \tilde{Z}(t)$.

Let

$$A_0(t) \triangleq W(t) + (1 - W(t))U(t), \qquad D(t) \triangleq 1 - W(t).$$

Then $A_0(t) \in \{0, 1\}$ indicates whether a free donor is available after the arrival-stage matching, and

$$-\Delta Y(t) \leq A_0(t)(Z(t) + 1) - D(t).$$

The extra $+1$ only weakens the upper bound and incorporates the immediate removal of a non-voucher arrival when $U(t) = 1$. Conditional on $\mathcal{F}_t$,

$$\mathbb{E}[A_0(t) \mid \mathcal{F}_t] = q + (1-q)\big(1 - (1-p)^{X(t)}\big), \qquad \mathbb{E}[D(t) \mid \mathcal{F}_t] = 1 - q.$$

Using the stochastic domination $Z(t) \preccurlyeq \tilde{Z}(t)$ and setting $G(t) = \tilde{Z}(t) + 1$, where $G(t)$ is geometric on $\{1, 2, \dots\}$ with parameter $(1-p)^{Y(t)}$, we may apply Lemma 2 to the upper bound $A_0(t)G(t) - D(t)$.

We now upper bound the jump under the event

$$\mathcal{E}_1 \triangleq \left\{ X(t) \le B\frac{\sqrt{q}}{p} \right\}.$$

Bounding $\mathbb{E}[U(t) \mid \mathcal{F}_t]$ on $\mathcal{E}_1$. Recall that, conditional on $\mathcal{F}_t$, $U(t)$ is Bernoulli with mean $1 - (1-p)^{X(t)}$. For $p \le \frac{1}{2}$, using (4) and (2), we have

$$1 - (1-p)^{X(t)} \le 1 - e^{-p(1+p)X(t)} \le (1+p)pX(t). \tag{22}$$

Therefore, on $\mathcal{E}_1$,

$$\mathbb{E}[U(t) \mid \mathcal{F}_t] = 1 - (1-p)^{X(t)} \le (1+p)pX(t) \le (1+p)B\sqrt{q}.$$

Lower bounding the geometric parameter $(1-p)^{Y(t)}$. Recall that the success probability (termination probability) of $\tilde{Z}(t)$ is $(1-p)^{Y(t)}$. For $p \le \frac{1}{2}$, again using (4), we have

$$(1-p)^{Y(t)} \ge e^{-p(1+p)Y(t)}.$$

If in addition

$$Y(t) \le \frac{1}{2p}\left(\log\frac{1}{q} - C\right),$$

then

$$(1-p)^{Y(t)} \ge \exp\!\left(-\frac{1+p}{2}\left(\log\frac{1}{q} - C\right)\right) = e^{\frac{1+p}{2}C}\, q^{\frac{1+p}{2}} \ge e^{\frac{C}{2}}\, q^{1/2}\, q^{p} \ge e^{\frac{C}{2}}\, q^{1/2}\, q^{q}. \tag{23}$$

The last inequality follows from $p \le q$ and $q \in (0,1)$. Since $\lim_{q \downarrow 0} q^q = 1$, there exists $\epsilon_0 > 0$ such that for all $q < \epsilon_0$, $q^q \ge 0.95$. Combining this with (23), we obtain the uniform lower bound

$$(1-p)^{Y(t)} \;\ge\; 0.95\, e^{C/2}\, q^{1/2} \qquad \text{whenever} \qquad Y(t) \le \frac{1}{2p}\left(\log\frac{1}{q} - C\right).$$

Therefore, in Lemma 2 we set

$$\alpha = \frac{1}{2}, \qquad b = 0.95\, e^{C/2}.$$

Next, on the event $\mathcal{E}_1 = \{X(t) \le B\frac{\sqrt{q}}{p}\}$, inequality (22) gives

$$\mathbb{E}[U(t) \mid \mathcal{E}_1] \le (1+p)B\sqrt{q}.$$

Hence the probability that a "free donor" is available to start a chain in period $t \to t+1$ is at most

$$q + (1-q)\mathbb{E}[U(t) \mid \mathcal{E}_1] \le q + (1+p)B\sqrt{q} \le \big((1+p)B + 1\big)\sqrt{q},$$

using $q \le \sqrt{q}$ for $q \in (0,1)$. For p sufficiently small we may further simplify $(1+p)B \le B + 0.05$, and thus take

$$a = B + 1.05 \qquad \text{so that} \qquad q + (1-q)\mathbb{E}[U(t) \mid \mathcal{E}_1] \le a\, q^{1/2}.$$

We now apply Lemma 2 to the Lyapunov function $V(x,y) = -y$, which has increment $\Delta V(t) = -\Delta Y(t)$.

$$y_0 \triangleq \frac{1}{2p}\Big(\log\frac{1}{q} - C\Big), \qquad c \triangleq -y_0, \qquad \beta = \frac{3}{4}, \qquad \rho = 0.95.$$

Then, for all $m \in \mathbb{N}$,

$$\mathbb{P}\bigg(Y(\infty) \le y_0 - \frac{2}{b}\, m\, q^{-3/4}\bigg) \le \Big(\frac{a}{0.95\, b}\Big)^m + \frac{b}{0.95\, b - a}\left[\bigg(0.95 + \frac{3}{b\, q^{1/2}}\bigg)\mathbb{P}(\mathcal{E}_1^c) + \frac{a}{b}e^{-q^{-1/4}}\right],$$

where we used $\alpha - \beta = \frac{1}{2} - \frac{3}{4} = -\frac{1}{4}$.

Finally, choosing $B = 2$ and $C = 3$ yields $a = 3.05$ and $b = 0.95e^{3/2} \approx 4.26$. Substituting these values and conservatively rounding the constants, we have

$$\begin{aligned}
&\mathbb{P}\left(Y(\infty) \le \frac{1}{2p}\left(\log\frac{1}{q} - 3\right) - 0.5m \cdot q^{-3/4}\right) \\
&\le 0.8^m + 4.4 \cdot \left(\left(0.95 + \frac{0.75}{\sqrt{q}}\right)\mathbb{P}(\mathcal{E}_1^c) + 0.75e^{-q^{-1/4}}\right).
\end{aligned}$$

This concludes the proof.

D.3 Proof of Proposition 5

We express the one-step change of the number of voucher patients $\Delta X(t) \triangleq X(t+1) - X(t)$ in terms of arrival-type and chain-removal variables. Let $W(t) \in \{0,1\}$ indicate the arrival type in period $t \to t+1$, where $W(t) = 1$ denotes a voucher arrival and $W(t) = 0$ denotes a non-voucher arrival; conditional on the natural filtration $\mathcal{F}_t$, we have $W(t) \sim \text{Bernoulli}(q)$.

When the arrival is non-voucher, let $U(t) \in \{0,1\}$ be the indicator that the arriving non-voucher patient is compatible with at least one waiting voucher donor at time t. Under the independent-compatibility model,

$$\mathbb{P}\big(U(t) = 1 \mid \mathcal{F}_t\big) = 1 - (1-p)^{X(t)}.$$

Finally, let $Z'(t)$ denote the number of voucher patients removed by the chain initiated by a free donor in period $t \to t+1$, under the convention that the chain "sees" only the $X(t)$ voucher patients present at the start of the period (so $Z'(t)$ is defined as a state-dependent lower bound on the actual number removed).

With these definitions,

$$\Delta X(t) \le W(t)\big(1 - Z'(t)\big) - (1 - W(t))\, U(t)\, Z'(t). \tag{24}$$

The term $W(t)(1 - Z'(t))$ is an upper bound for the voucher-arrival case: the arriving voucher patient may also be matched immediately, which can only reduce X. The term $(1 - W(t))U(t)Z'(t)$ captures the non-voucher-arrival case, in which a voucher patient is removed only if the arriving patient matches a free donor and the resulting chain terminates at a voucher patient.

Note that when $W(t) = 1$ and the arriving voucher patient is not immediately matched by an existing voucher donor, the voucher pool size during chain extension is $X(t) + 1$. In that case, the actual chain can only remove weakly more voucher patients than a chain that is restricted to the initial pool size $X(t)$. Thus $Z'(t)$ is a lower bound on voucher-patient removals, and (24) provides an upper bound on $\Delta X(t)$, which is sufficient for our drift analysis.

The distribution of $Z'(t)$ is complicated because it depends on the state and the chain evolution. We therefore lower bound $Z'(t)$ by a simpler Bernoulli random variable $\tilde{Z}'(t)$ (conditional on the current state), so that $\tilde{Z}'(t) \preccurlyeq Z'(t)$.

Recall that $Z'(t) \in \{0, 1\}$ indicates whether the donor-initiated chain terminates by matching a voucher (priority) patient, and hence $\mathbb{E}[Z'(t) \mid \mathcal{F}_t] = \mathbb{P}(Z'(t) = 1 \mid \mathcal{F}_t)$.

Claim. Fix an integer $L \in \{0, 1, \ldots, Y(t)\}$. Define

$$s_X \triangleq 1 - (1-p)^{X(t)}, \qquad r_L \triangleq (1-p)^{X(t)}\Big(1 - (1-p)^{Y(t)-L+1}\Big).$$

Let $\tilde{Z}'(t)$ be Bernoulli with mean

$$\mathbb{P}(\tilde{Z}'(t) = 1 \mid \mathcal{F}_t) \;\triangleq\; s_X \sum_{\ell=0}^{L} r_L^{\ell} = s_X \frac{1 - r_L^{L+1}}{1 - r_L}.$$

Then $\tilde{Z}'(t) \preccurlyeq Z'(t)$.

Proof of claim. For $\ell = 0, 1, \ldots, L$, let E_ℓ be the event that the chain extends through ℓ non-voucher patients and then terminates with a voucher patient at the next step. These events are disjoint, and on E_ℓ the chain ends with a voucher patient, so $Z'(t) = 1$.

At any step $i \le L - 1$, the chain can be extended by another non-voucher patient if (i) the current donor is compatible with no voucher patient among the $X(t)$ voucher patients, and (ii) it is compatible with at least one of the remaining $Y(t) - i \ge Y(t) - L + 1$ non-voucher patients. Under the independent-compatibility model, this continuation probability is at least

$$(1-p)^{X(t)}\Big(1 - (1-p)^{Y(t)-L+1}\Big) = r_L.$$

Moreover, at any step, the chain terminates with a voucher patient if the current donor is compatible with at least one voucher patient, an event of probability s_X. Therefore,

$$\mathbb{P}(E_\ell \mid \mathcal{F}_t) \;\geq\; r_L^\ell \, s_X, \qquad \ell = 0, 1, \ldots, L.$$

Summing over the disjoint events yields

$$\mathbb{P}(Z'(t) = 1 \mid \mathcal{F}_t) \;\geq\; \sum_{\ell=0}^{L} \mathbb{P}(E_\ell \mid \mathcal{F}_t) \;\geq\; s_X \sum_{\ell=0}^{L} r_L^\ell = \mathbb{P}(\tilde{Z}'(t) = 1 \mid \mathcal{F}_t),$$

which implies $\tilde{Z}'(t) \preccurlyeq Z'(t)$. $\square$

This simplification allows us to compute $\mathbb{E}[\tilde{Z}' \mid X, Y]$ in closed form. By construction, $\tilde{Z}'$ is Bernoulli with mean

$$\mathbb{E}[\tilde{Z}' \mid X, Y] = \big(1 - (1-p)^X\big) \sum_{\ell=0}^{L} \Big((1-p)^X \big(1 - (1-p)^{Y-L+1}\big)\Big)^\ell,$$

and therefore

$$\begin{aligned}
\mathbb{E}[\tilde{Z}' \mid X, Y] &= \big(1 - (1-p)^X\big) \cdot \frac{1 - \Big((1-p)^X \big(1 - (1-p)^{Y-L+1}\big)\Big)^{L+1}}{1 - (1-p)^X \big(1 - (1-p)^{Y-L+1}\big)} \\
&= \big(1 - (1-p)^X\big) \cdot \frac{1 - \big(1 - (1-p)^{Y-L+1}\big)^{L+1} (1-p)^{X(L+1)}}{1 - \big(1 - (1-p)^{Y-L+1}\big) (1-p)^X}.
\end{aligned}$$

Next, a chain can be initiated by a non-voucher arrival precisely when the arrival is non-voucher and its patient is compatible with at least one waiting voucher donor. Conditional on X, this occurs with probability

$$(1-q)\big(1 - (1-p)^X\big).$$

Moreover, since $\tilde{Z}' \preccurlyeq Z'$ and $\tilde{Z}'$ is independent of U conditional on the state, we have the conditional lower bound

$$\mathbb{E}[(1-W)UZ' \mid X, Y] \;\geq\; (1-q)\big(1 - (1-p)^X\big)\, \mathbb{E}[\tilde{Z}' \mid X, Y].$$

Accordingly, define

$$\mathcal{T} \triangleq (1-q)\big(1 - (1-p)^X\big)\, \mathbb{E}[\tilde{Z}' \mid X, Y],$$

so that $\mathcal{T}$ is a convenient explicit lower bound on the conditional expectation of $(1-W)UZ'$ in (24). Substituting the expression above yields

$$\begin{aligned}
\mathcal{T} &= (1-q)\big(1 - (1-p)^X\big)^2 \cdot \frac{1 - \Big((1-p)^X \big(1 - (1-p)^{Y-L+1}\big)\Big)^{L+1}}{1 - (1-p)^X \big(1 - (1-p)^{Y-L+1}\big)} \\
&= (1-q)\big(1 - (1-p)^X\big)^2 \cdot \frac{1 - \big(1 - (1-p)^{Y-L+1}\big)^{L+1} (1-p)^{X(L+1)}}{1 - \big(1 - (1-p)^{Y-L+1}\big) (1-p)^X}.
\end{aligned}$$

Lower bounding $\mathcal{T}$ is algebraically involved, so we break it into several steps.

- *Step 1. Simplifying the numerator of $\mathcal{T}$.* Recall

$$\mathcal{T} = (1-q)\big(1-(1-p)^X\big)^2 \cdot \frac{1-r^{L+1}}{1-r}, \qquad r \triangleq (1-p)^X\big(1-(1-p)^{Y-L+1}\big).$$

Since $r \in (0,1)$, we have $1-r^{L+1} \geq \frac{1}{2}$ whenever $r^{L+1} \leq \frac{1}{2}$, i.e.,

$$L+1 \;\geq\; \frac{\log 2}{-\log r}. \tag{25}$$

Moreover,

$$-\log r = -\log\!\big(1-(1-p)^{Y-L+1}\big) - \log(1-p)^X.$$

Using $-\log(1-x) \geq x$ for $x \in [0,1)$, we obtain

$$-\log r \;\geq\; (1-p)^{Y-L+1} + pX, \quad \text{and hence} \quad \frac{\log 2}{-\log r} \;\leq\; \frac{\log 2}{pX}.$$

Therefore, whenever $L \geq \frac{\log 2}{pX}$ we have (25), and thus $1-r^{L+1} \geq \frac{1}{2}$.

In particular, if $X \geq Bq^{1-2^{-(k+1)}}/p$, then $pX \geq Bq^{1-2^{-(k+1)}}$. We take

$$L = \left\lceil \frac{\log 2}{Bq^{1-2^{-(k+1)}}} \right\rceil.$$

For all sufficiently small q, the good-set lower bound on Y implies $Y-L+1 \geq 0$ because Y is of order $p^{-1}\log(1/q)$ whereas $L = O(q^{-(1-2^{-(k+1)})})$ and $p \leq q$. This integer choice ensures $L \geq \log(2)/(pX)$ and hence $1-r^{L+1} \geq 1/2$. With this choice,

$$\mathcal{T} \;\geq\; \frac{1}{2}(1-q) \cdot \frac{\big(1-(1-p)^X\big)^2}{1-r}.$$

- *Step 2. Finding the minimizing X.* In the lower bound from Step 1, the dependence on X appears only through $(1-p)^X$. To obtain a uniform lower bound over all

$$X \;\geq\; B\frac{q^{1-2^{-(k+1)}}}{p},$$

it suffices to study the monotonicity of

$$f(x) \triangleq \frac{(1-x)^2}{1-ax}, \qquad x \in (0,1),$$

where we identify $x = (1-p)^X \in (0,1)$ and $a = 1-(1-p)^{Y-L+1} \in (0,1)$.

A direct differentiation gives

$$\begin{aligned} f'(x) &= \frac{-2(1-x)(1-ax) + a(1-x)^2}{(1-ax)^2} \\ &= \frac{(1-x)\big[-2(1-ax) + a(1-x)\big]}{(1-ax)^2} = \frac{(1-x)\big((2-a)x - 2 + a\big)}{(1-ax)^2}. \end{aligned}$$

Since $a \in (0,1)$, we have $2-a > 0$ and, for all $x \in (0,1)$,

$$(2-a)x - 2 + a \ \le\ (2-a)\cdot 1 - 2 + a \ =\ 0,$$

with strict inequality when $x < 1$. Therefore $f'(x) < 0$ for all $x \in (0,1)$, and f is strictly decreasing on $(0,1)$.

Because $x = (1-p)^X$ decreases as X increases, the minimum of $f((1-p)^X)$ over $X \ge B\frac{q^{1-2^{-(k+1)}}}{p}$ is attained at the largest feasible x, i.e., at the smallest feasible X. Hence the minimum of $\mathcal{T}$ over this range is achieved by setting

$$X = B\frac{q^{1-2^{-(k+1)}}}{p}.$$

- *Step 3. Bounding the numerator.* In Steps 1–2, the numerator term is $\big(1-(1-p)^X\big)^2$, and Step 2 shows the worst case over $X \ge B\frac{q^{1-2^{-(k+1)}}}{p}$ occurs at $X = B\frac{q^{1-2^{-(k+1)}}}{p}$. For this value,

$$pX = Bq^{1-2^{-(k+1)}}.$$

Using $(1-p)^X \le e^{-pX}$, we obtain

$$\big(1-(1-p)^X\big)^2 \ \ge\ \big(1-e^{-pX}\big)^2 = \Big(1-e^{-Bq^{1-2^{-(k+1)}}}\Big)^2.$$

By (3), $e^{-x} \le 1 - x + \frac{x^2}{2}$ for $x \ge 0$, hence

$$1 - e^{-x} \ \ge\ x - \frac{x^2}{2} = x\Big(1-\frac{x}{2}\Big).$$

Applying this with $x = Bq^{1-2^{-(k+1)}}$ and squaring yields

$$\Big(1-e^{-Bq^{1-2^{-(k+1)}}}\Big)^2 \ \ge\ B^2\Big(1-\frac{B}{2}q^{1-2^{-(k+1)}}\Big)^2 q^{2(1-2^{-(k+1)})}.$$

Since $q^{1-2^{-(k+1)}} \to 0$ as $q \to 0$, there exists $\epsilon > 0$ such that for all $q < \epsilon$,

$$\Big(1-\frac{B}{2}q^{1-2^{-(k+1)}}\Big)^2 \ \ge\ \frac{1}{1.05}.$$

Consequently, for all $q < \epsilon$,

$$\big(1-(1-p)^X\big)^2 \;\geq\; \frac{1}{1.05}\, B^2\, q^{\,2(1-2^{-(k+1)})}.$$

- *Step 4. Bounding the denominator.* We upper bound

$$1-(1-p)^X\big(1-(1-p)^{Y-L+1}\big)$$

for the choices of X and L from Steps 1–2.

Bounding $(1-p)^X$ from below. Assume $p \leq \frac{1}{2}$. By (4) with $x = p$,

$$1-p \;\geq\; e^{-p-p^2}, \quad \text{so} \quad (1-p)^X \;\geq\; e^{-(p+p^2)X} = e^{-p(1+p)X}.$$

Using $e^{-u} \geq 1-u$ for $u \geq 0$, we further obtain

$$(1-p)^X \;\geq\; 1-p(1+p)X.$$

With $X = B\frac{q^{1-2^{-(k+1)}}}{p}$, this becomes

$$(1-p)^X \;\geq\; 1-B(1+p)q^{\,1-2^{-(k+1)}}.$$

In particular, there exists $\epsilon > 0$ (depending on B) such that whenever $p \leq \epsilon$, $B(1+p) \leq B+0.05$, and hence

$$(1-p)^X \;\geq\; 1-(B+0.05)\,q^{\,1-2^{-(k+1)}}.$$

Bounding $(1-p)^{Y-L+1}$ from above. By (2), $(1-p)^n \leq e^{-pn}$ for $n \geq 0$, so

$$(1-p)^{Y-L+1} \;\leq\; e^{-p(Y-L+1)} \;\leq\; e^{-pY}e^{pL}.$$

Using $Y \geq \frac{1}{p}(1-2^{-k})(\log\frac{1}{q} - C)$ yields

$$e^{-pY} \;\leq\; \exp\!\left(-(1-2^{-k})\left(\log\frac{1}{q} - C\right)\right) = (qe^C)^{\,1-2^{-k}}.$$

With $L = \frac{\log 2}{Bq^{\,1-2^{-(k+1)}}}$, we have $e^{pL} = 2^{\overline{\frac{p}{Bq^{\,1-2^{-(k+1)}}}}}$. Since $p \leq q$, then $2^{\overline{\frac{p}{Bq^{\,1-2^{-(k+1)}}}}} \leq 2^{1/B}$, and therefore

$$(1-p)^{Y-L+1} \;\leq\; 2^{1/B}(qe^C)^{\,1-2^{-k}}.$$

Combining the bounds. Let $u \triangleq (B+0.05)q^{1-2^{-(k+1)}}$ and $v \triangleq 2^{1/B}(qe^C)^{1-2^{-k}}$. Then $(1-p)^X \geq 1-u$ and $(1-p)^{Y-L+1} \leq v$, so $1-(1-p)^{Y-L+1} \geq 1-v$. Hence

$$\begin{aligned}
1-(1-p)^X\big(1-(1-p)^{Y-L+1}\big) &\leq 1-(1-u)(1-v) \\
&= u+v-uv \\
&\leq u+v \\
&= (B+0.05)\, q^{1-2^{-(k+1)}} + 2^{1/B}e^C\, q^{1-2^{-k}}.
\end{aligned}$$

Combining Steps 1–4. From Steps 1–4, for $X \geq B\frac{q^{1-2^{-(k+1)}}}{p}$ and with $L = \frac{\log 2}{Bq^{1-2^{-(k+1)}}}$, we obtained

$$\mathcal{T} \;\geq\; \frac{1-q}{2}\cdot\frac{1}{1.05}B^2q^{2(1-2^{-(k+1)})}\cdot\frac{1}{(B+0.05)q^{1-2^{-(k+1)}}+2^{1/B}e^C\,q^{1-2^{-k}}}.$$

Factoring out $q^{1-2^{-k}}$ from the denominator gives

$$\mathcal{T} \;\geq\; \frac{B^2}{2.1}(1-q)\cdot\frac{q}{(B+0.05)q^{2^{-(k+1)}}+2^{1/B}e^C}.$$

Since $q^{2^{-(k+1)}} \leq 1$ for $q \in (0,1)$, we further have

$$\mathcal{T} \;\geq\; \frac{B^2}{2.1}(1-q)\cdot\frac{q}{B+0.05+2^{1/B}e^C}.$$

Now fix $q < 0.05$ and take $C = 3.41$, $B = 9.9$. Then the right-hand side is at least $1.045\,q$, so

$$\mathcal{T} \;\geq\; 1.045\,q.$$

Applying Lemma 1. Let $V(x,y) = x$ and let the good set be $\mathcal{E} = \mathcal{G}_k$. Take the drift threshold

$$B_X \triangleq 9.9\,\frac{q^{1-2^{-(k+1)}}}{p}.$$

On $\mathcal{G}_k$ and whenever $X \geq B_X$, the expected net change satisfies

$$\mathbb{E}[\Delta X \mid X, Y] \;\leq\; q-\mathcal{T} \;\leq\; -0.045\,q,$$

so we may set $\gamma = 0.045q$. Moreover, $\Delta X \in \{-1,0,1\}$, hence $\nu_{\max} = 1$. The one-step positive increment is at most 1 and occurs with probability at most q, so $q_{\text{in}} \leq q$ and $q_{\text{out}} \leq q$. Finally, outside $\mathcal{G}_k$ the drift is at most q, so $\gamma_{\text{out}} \leq q$.

Therefore Lemma 1 yields

$$\begin{aligned}
&\mathbb{P}(X \geq B_X + 2m) \leq \xi^m + \phi\,\mathbb{P}((X,Y)\notin\mathcal{G}_k)\,, \\
&\xi = \frac{q}{q+0.045q} = \frac{1}{1.045}, \quad \phi = \frac{q+q+0.045q}{0.045q} = \frac{2.045}{0.045}.
\end{aligned}$$

In particular, using the looser numerical bounds $\xi \le 0.96$ and $\phi \le 46$,

$$\mathbb{P}\left(X \ge 9.9\frac{q^{1-2^{-(k+1)}}}{p} + 2m\right) \le 0.96^m + 46\,\mathbb{P}((X,Y) \notin \mathcal{G}_k)\,.$$

This concludes the proof.

D.4 Proof of Proposition 6

The proof follows the $k = 1$ argument verbatim up to (and including) the stochastic domination claim $Z(t) \preccurlyeq \tilde{Z}(t)$ with $\tilde{Z}(t) \sim \mathrm{Geo}((1-p)^{Y(t)})$; see the claim and justification in the proof of the $k = 1$ case. We only record the modifications needed for $k \ge 2$.

Bounding $\mathbb{E}[U(t) \mid \mathcal{F}_t]$ *on* $\mathcal{E}_k$. Using the same inequalities as in (22), for $p \le \frac{1}{2}$,

$$1 - (1-p)^{X(t)} \le (1+p)pX(t).$$

On $\mathcal{E}_k$, this yields

$$\mathbb{E}[U(t) \mid \mathcal{F}_t] = 1 - (1-p)^{X(t)} \le (1+p)B\,q^{1-2^{-k}}.$$

Lower bounding the geometric parameter. If $Y(t) \le \frac{1}{p}(1-2^{-k})(\log(1/q) - C)$, then the same steps that lead to (23) give

$$(1-p)^{Y(t)} \ge e^{C(1-2^{-k})}\,q^{1-2^{-k}}\,q^q.$$

For q sufficiently small, $q^q \ge 0.95$, and since $k \ge 2$ implies $1 - 2^{-k} \ge 3/4$, we obtain

$$(1-p)^{Y(t)} \ge 0.95\,e^{3C/4}\,q^{1-2^{-k}}.$$

Hence in Lemma 2 we may take

$$\alpha = 1 - 2^{-k}, \qquad b = 0.95\,e^{3C/4}.$$

Bounding the "free-donor" probability. As in the $k = 1$ proof,

$$\mathbb{E}[\tilde{A}(t) \mid \mathcal{F}_t] = q + \big(1 - (1-p)^{X(t)}\big) \le q + (1+p)B\,q^{1-2^{-k}} \le \big((1+p)B + 1\big)q^{1-2^{-k}},$$

using $q \le q^{1-2^{-k}}$ for $q \in (0,1)$. For p sufficiently small, $(1+p)B \le B + 0.05$, so we may set $a = B + 1.05$.

Apply the geometric-jump tail bound. Apply Lemma 2 to the Lyapunov function $V(x,y) = -y$ with parameters

$$\rho = 0.95, \quad \alpha = 1 - 2^{-k}, \quad \beta = 1 - 2^{-(k+1)}, \quad a = B + 1.05, \quad b = 0.95e^{3C/4},$$

Let

$$y_k \triangleq \frac{1}{p}\Big(1 - 2^{-k}\Big)\Big(\log\frac{1}{q} - C\Big), \qquad c \triangleq -y_k.$$

With the bounds established above, on the good set $\mathcal{E}_k$ the Bernoulli parameter is at most aq^α with $a = B + 1.05$, and whenever $Y(t) \le y_k$ the geometric termination probability is at least bq^α with $b = 0.95e^{3C/4}$. Taking $\rho = 0.95$ in Lemma 2 therefore gives, for all $m \in \mathbb{N}$,

$$\mathbb{P}\Big(Y(\infty) \le y_k - \frac{2}{b} m q^{-\beta}\Big) = \mathbb{P}\Big(V(X(\infty), Y(\infty)) \ge c + \frac{2}{b} m q^{-\beta}\Big)$$
$$\le \Big(\frac{a}{0.95\, b}\Big)^m + \frac{b}{0.95\, b - a}\left[\Big(0.95 + \frac{3}{b\, q^\alpha}\Big)\mathbb{P}(\mathcal{E}_k^c) + \frac{a}{b} e^{-q^{\alpha-\beta}}\right].$$

When $k \ge 2$, choose $B = 10$ and $C = 3.4$, so that $a = 11.05$ and $b = 0.95e^{3C/4} \approx 12.16$. Substituting these values and conservatively rounding constants yields

$$\mathbb{P}\left(Y(\infty) \le \frac{1}{p}\left(1 - 2^{-k}\right)\left(\log\frac{1}{q} - 3.4\right) - 0.17\, m\, q^{-(1-2^{-(k+1)})}\right)$$
$$\le 0.96^m + 24.3\left(\left(0.95 + \frac{0.25}{q^{1-2^{-k}}}\right)\mathbb{P}\left(\mathcal{E}_k^c\right) + 0.91\, e^{-q^{-2^{-(k+1)}}}\right).$$

This concludes the proof.

D.5 Proof of Proposition 7

Fix $\alpha > \frac{1}{2}$. Recall the events

$$\mathcal{E}_1 := \left\{X \le 2\frac{\sqrt{q}}{p}\right\}, \qquad \mathcal{E}_k := \left\{X \le 10\frac{q^{1-2^{-k}}}{p}\right\} \ (k \ge 2),$$

and

$$\mathcal{G}_k := \left\{Y \ge \frac{1}{p}(1 - 2^{-k})\Big(\log\frac{1}{q} - 3.41\Big)\right\} \ (k \ge 1).$$

Write $P_k := \mathbb{P}_\pi(\mathcal{E}_k^c)$.

Step 1: Base bound on P_2. By Proposition 3, there exists $\epsilon_1 > 0$ such that for all $q \in (0, \epsilon_1)$,

$$\mathbb{P}_\pi(\mathcal{E}_1^c) \le \exp\Big(-0.17\,\frac{1}{\sqrt{p}}\Big). \tag{26}$$

Next,

$$\mathcal{G}_1^c = \left\{Y < \frac{1}{2p}\Big(\log\frac{1}{q} - 3.41\Big)\right\}.$$

Apply Proposition 4 with

$$m_1 := \left\lceil \frac{0.41}{p} q^{3/4} \right\rceil \in \mathbb{N}.$$

Then

$$\frac{1}{2p}\Big(\log\frac{1}{q} - 3\Big) - 0.5\, m_1\, q^{-3/4} \le \frac{1}{2p}\Big(\log\frac{1}{q} - 3.41\Big),$$

so Proposition 4 yields

$$\mathbb{P}_\pi(\mathcal{G}_1^c) \le 0.8^{m_1} + 4.4\left[\left(0.95 + \frac{0.75}{\sqrt{q}}\right)\mathbb{P}_\pi(\mathcal{E}_1^c) + 0.75e^{-q^{-1/4}}\right]. \tag{27}$$

Since $p \le q$, we have $m_1 \ge 0.41\, q^{-1/4}$, hence $0.8^{m_1} \le e^{-cq^{-1/4}}$ for some $c > 0$. Together with (26), this implies that for some $c_1 > 0$ and all sufficiently small q,

$$\mathbb{P}_\pi(\mathcal{G}_1^c) \le e^{-c_1 q^{-1/4}}. \tag{28}$$

Now apply Proposition 5 with $k = 1$ and

$$m_1' := \left\lceil 0.05\, \frac{q^{3/4}}{p} \right\rceil \in \mathbb{N}.$$

Then

$$P_2 = \mathbb{P}_\pi\left(X > 10\frac{q^{3/4}}{p}\right) \le 0.96^{m_1'} + 46\,\mathbb{P}_\pi(\mathcal{G}_1^c).$$

Using $0.96^{m_1'} \le e^{-c_2 q^{3/4}/p} \le e^{-c_2 q^{-1/4}}$ (since $p \le q$) and (28), we obtain for some $c_3 > 0$ and all small q,

$$P_2 \le e^{-c_3 q^{-1/4}}. \tag{29}$$

Step 2: A recursion for $k \ge 2$. For notational simplicity, from this point we treat $k \ge 2$ as a real parameter.

Fix $k \ge 2$. Proposition 6 gives, for any $m \in \mathbb{N}$,

$$\mathbb{P}_\pi\left(Y \le y_k - 0.17\, m\, q^{-(1-2^{-(k+1)})}\right) \le 0.96^m + 24.3\left[\left(0.95 + \frac{0.25}{q^{1-2^{-k}}}\right) P_k + 0.91e^{-q^{-2^{-(k+1)}}}\right],$$

where $y_k = \frac{1}{p}(1-2^{-k})\left(\log\frac{1}{q} - 3.4\right)$. Choose

$$m_k := \left\lceil \frac{0.01}{0.17}(1-2^{-k})\frac{q^{1-2^{-(k+1)}}}{p} \right\rceil \in \mathbb{N}.$$

Then $y_k - 0.17 m_k q^{-(1-2^{-(k+1)})} \le \frac{1}{p}(1-2^{-k})\left(\log\frac{1}{q} - 3.41\right)$, so the left-hand event contains $\mathcal{G}_k^c$, and hence

$$\mathbb{P}_\pi(\mathcal{G}_k^c) \le 0.96^{m_k} + 24.3\left[\left(0.95 + \frac{0.25}{q^{1-2^{-k}}}\right) P_k + 0.91e^{-q^{-2^{-(k+1)}}}\right]. \tag{30}$$

Using $0.95 + \frac{0.25}{q^{1-2^{-k}}} \le \frac{1.2}{q^{1-2^{-k}}}$ for $q \in (0,1)$, and $m_k \gtrsim q^{1-2^{-(k+1)}}/p \ge q^{-2^{-(k+1)}}$ (since $p \le q$), we may rewrite (30) as: there exist absolute constants $A_1, B_1, a_1 > 0$ such that

$$\mathbb{P}_\pi(\mathcal{G}_k^c) \le A_1 e^{-a_1 q^{-2^{-(k+1)}}} + B_1 q^{-(1-2^{-k})} P_k. \tag{31}$$

Next, Proposition 5 gives for any $m \in \mathbb{N}$,

$$\mathbb{P}_\pi\left(X \geq 9.9\frac{q^{1-2^{-(k+1)}}}{p} + 2m\right) \leq 0.96^m + 46\,\mathbb{P}_\pi(\mathcal{G}_k^c).$$

Choose

$$m'_k := \left\lceil 0.05\,\frac{q^{1-2^{-(k+1)}}}{p}\right\rceil \in \mathbb{N}.$$

Then

$$P_{k+1} = \mathbb{P}_\pi\left(X > 10\frac{q^{1-2^{-(k+1)}}}{p}\right) \leq 0.96^{m'_k} + 46\,\mathbb{P}_\pi(\mathcal{G}_k^c).$$

Using $m'_k \geq 0.05\,q^{1-2^{-(k+1)}}/p \geq 0.05\,q^{-2^{-(k+1)}}$ (since $p \leq q$) gives $0.96^{m'_k} \leq e^{-a_2 q^{-2^{-(k+1)}}}$ for some $a_2 > 0$. Combining with (31), there exist absolute constants $A, B, a > 0$ such that for all $k \geq 2$ and all small q,

$$P_{k+1} \leq Ae^{-aq^{-2^{-(k+1)}}} + Bq^{-(1-2^{-k})}P_k. \tag{32}$$

Step 3: Unrolling. Define $c_{k+1} := Ae^{-aq^{-2^{-(k+1)}}}$ and $b_k := Bq^{-(1-2^{-k})}$. Then (32) reads $P_{k+1} \leq c_{k+1} + b_k P_k$, and iterating from 2 to k gives

$$P_{k+1} \leq \Big(\prod_{j=2}^{k} b_j\Big)P_2 + \sum_{t=3}^{k+1}\Big(\prod_{j=t}^{k} b_j\Big)c_t, \tag{33}$$

with the convention that empty products equal 1. Using $1 - 2^{-j} \leq 1$ yields the crude bound

$$\prod_{j=t}^{k} b_j = B^{k-t+1}q^{-\sum_{j=t}^{k}(1-2^{-j})} \leq \left(\frac{B}{q}\right)^{k-t+1}. \tag{34}$$

Step 4: Choose $k = k(q)$ continuously and conclude. Let $x := \log(1/q) \to \infty$. Choose the integer $K = K(q) \geq 2$ such that

$$q^{-2^{-(K+1)}} \leq x^\alpha < q^{-2^{-K}}. \tag{35}$$

Such a K exists for all sufficiently small q, and $K = O(\log x)$. Moreover, $q^{-2^{-K}} \leq x^{2\alpha}$ up to an immaterial absolute factor caused by integer rounding. Therefore the threshold defining $\mathcal{E}_{K+1}$ is at most a constant multiple of

$$\frac{q}{p}x^\alpha = \frac{q}{p}\left(\log\frac{1}{q}\right)^\alpha.$$

It remains to show $P_{K+1} \to 0$.

(i) Initial term. By (34) with $t = 2$,

$$\prod_{j=2}^{k} b_j \leq \left(\frac{B}{q}\right)^{k-1} = \exp\big((k-1)(x + \log B)\big) = \exp(O(x \log x)).$$

By (29), $P_2 \leq \exp(-c_3 q^{-1/4}) = \exp(-c_3 e^{x/4})$. Since $e^{x/4} \gg x \log x$, it follows that $\left(\prod_{j=2}^{k} b_j\right) P_2 \to 0$.

(ii) Sum term. The term $t = k+1$ equals $c_{k+1} = Ae^{-aq^{-2^{-(k+1)}}} = Ae^{-ax^{\alpha}} \to 0$. Now fix $t \in \{3, \dots, k\}$ and let $i := k - t + 1 \in \{1, \dots, k-2\}$. Using (35),

$$q^{-2^{-t}} = q^{-2^{-(k+1-i)}} = \left(q^{-2^{-(k+1)}}\right)^{2^i} = (x^{\alpha})^{2^i} = x^{\alpha 2^i},$$

so $c_t = Ae^{-ax^{\alpha 2^i}}$. Also by (34), $\prod_{j=t}^{k} b_j \leq (B/q)^i = \exp(i(x + \log B))$. Hence each summand satisfies

$$\left(\prod_{j=t}^{k} b_j\right) c_t \leq A \exp\Big(-ax^{\alpha 2^i} + i(x + \log B)\Big) \to 0,$$

because $\alpha > \frac{1}{2}$ implies $x^{2\alpha} \gg x$ (the case $i = 1$), and for $i \geq 2$ the domination is even stronger. Since there are only $k - 2 = O(\log x)$ summands, the whole sum in (33) tends to 0.

Therefore $P_{k+1} \to 0$. Finally,

$$\mathbb{P}_{\pi}\left(X \geq 10\frac{q}{p}\left(\log \frac{1}{q}\right)^{\alpha}\right) \leq P_{k+1} \to 0,$$

which proves Proposition 7.

D.6 Proof of Proposition 8

Similar to the proof of Proposition 4, we can represent the one-step change of the size of non-voucher patients as:

$$\Delta Y(t) = -W(t)\, Z(t) + (1 - W(t))\big(1 - U(t)(1 + Z(t))\big). \tag{36}$$

Recall that $W(t)$ indicates the arrival type in period $t \to t+1$, where $W(t) = 1$ corresponds to a voucher arrival and $W(t) = 0$ to a non-voucher arrival; conditional on $\mathcal{F}_t$, we have $W(t) \sim \text{Bernoulli}(q)$. When $W(t) = 0$, $U(t)$ is defined as the indicator that the arriving non-voucher patient is compatible with at least one voucher donor present at time t. We have

$$\mathbb{P}\big(U(t) = 1 \mid \mathcal{F}_t\big) = 1 - (1-p)^{X(t)}.$$

And $Z(t)$ denote the number of non-voucher patients removed by the chain initiated by a "free donor" in that period, whenever such a chain is triggered.

Step 1: A truncated-geometric surrogate for Z. The distribution of the chain length $Z(t)$ is complicated. We therefore replace Z with $\tilde{Z}$ which is stochastically dominated by Z.

Fix an integer $L \geq 1$ and assume $Y \geq L$. For each $i \in \{0, 1, \dots, L-1\}$, after i non-voucher patients have been removed, the probability that the chain can be extended by one more removal is at least

$$(1-p)^X \big(1 - (1-p)^{Y-i}\big) \;\geq\; (1-p)^X \big(1 - (1-p)^{Y-L+1}\big)$$

since $Y - i \geq Y - L + 1$.

Let Q be geometric with parameter $g \triangleq 1 - (1-p)^X \big(1 - (1-p)^{Y-L+1}\big)$ (support $\{0, 1, 2, \dots\}$) and define the truncated geometric variable

$$\tilde{Z} := Q \, \mathbf{1}\{Q \leq L\}.$$

By the preceding extension bound, the true chain length Z stochastically dominates $\tilde{Z}$, so $\mathbb{E}[Z \mid X, Y] \geq \mathbb{E}[\tilde{Z} \mid X, Y]$. Since the right-hand side of (36) is non-increasing in Z, we may replace Z by $\tilde{Z}$ to obtain the upper bound

$$\Delta Y \prec_{\mathrm{FSD}} -W\tilde{Z} + (1-W)\big(1 - U(1+\tilde{Z})\big). \tag{37}$$

Step 2: Two X-regions. Let $\alpha \in (0,1)$ and $\beta = \alpha/2$. Define

$$\mathcal{E}_A := \left\{ X \leq \frac{q^{1+\alpha}}{p} \right\}, \qquad \mathcal{E}_B := \left\{ \frac{q^{1+\alpha}}{p} \leq X \leq \frac{q^{1-\beta}}{p} \right\},$$

so that $\mathcal{E} = \mathcal{E}_A \cup \mathcal{E}_B$.

Step 3: Drift in Region B via Lemma 3. Set

$$\varepsilon := q^\beta, \qquad L := \frac{4}{q^{1+\alpha}} \log\left(\frac{4}{\varepsilon}\right), \qquad B := (1+2\alpha)\frac{1}{p} \log \frac{1}{q} + L.$$

Fix $(X, Y) \in \mathcal{E}_B$ with $Y \geq B$. Then $Y - L + 1 \geq (1+2\alpha)\frac{1}{p}\log(1/q)$, so

$$(1-p)^{Y-L+1} \leq e^{-p(Y-L+1)} \leq e^{-(1+2\alpha)\log(1/q)} = q^{1+2\alpha}.$$

Hence

$$\begin{aligned} g &= 1 - (1-p)^X \big(1 - (1-p)^{Y-L+1}\big) \\ &\leq \big(1 - (1-p)^X\big) + q^{1+2\alpha} \leq pX + q^{1+2\alpha}. \end{aligned} \tag{38}$$

On $\mathcal{E}_B$, we have $pX \leq q^{1-\beta}$ and also $pX \geq q^{1+\alpha}$. Since $\alpha > 0$ and $\beta = \alpha/2$, for q sufficiently small we have $q^{1+2\alpha} \leq \frac{1}{2}q^\beta \cdot q^{1+\alpha} \leq \frac{1}{2}q^\beta \cdot pX$ and thus $g \leq pX + q^{1+2\alpha} \leq 2pX \leq \varepsilon/2$ (for q small). In particular $\varepsilon > 2g$.

Moreover, on $\mathcal{E}_B$ we have $g \geq 1-(1-p)^X \geq \frac{1}{2}pX \geq \frac{1}{2}q^{1+\alpha}$ for q small, and hence

$$L = \frac{4}{q^{1+\alpha}} \log\left(\frac{4}{\varepsilon}\right) \geq \frac{2}{g} \log\left(\frac{4}{\varepsilon}\right).$$

Therefore Lemma 3 applied with parameter g yields

$$\mathbb{E}[\tilde{Z} \mid X, Y] \geq \frac{1-\varepsilon}{g} \qquad \text{for all } (X,Y) \in \mathcal{E}_B \text{ with } Y \geq B. \tag{39}$$

Hence

$$\begin{aligned}
&1 - \mathbb{E}[U \mid X] \cdot \mathbb{E}[\tilde{Z} \mid X, Y] \\
&\leq 1 - \left(1-(1-p)^X\right)\left(1-q^\beta\right)\frac{1}{pX + q^{1+2\alpha}} \\
&\leq 1 - \left(1-e^{-pX}\right)\left(1-q^\beta\right)\frac{1}{pX + q^{1+2\alpha}} \\
&\leq 1 - pX(1-\frac{1}{2}pX) \cdot \left(1-q^\beta\right)\frac{1}{pX + q^{1+2\alpha}} \\
&= 1 - (1-\frac{1}{2}pX) \cdot \left(1-q^\beta\right)\frac{1}{1+\frac{q^{1+2\alpha}}{pX}}.
\end{aligned}$$

Note that the product of the last three terms can be written as $1-o(1)$, where the $o(1)$ term comes from $\frac{1}{2}pX$, q^β and $\frac{q^{1+2\alpha}}{pX}$. The dominating term is q^β, hence for small enough p, q, we have:

$$1 - \mathbb{E}[U \mid X] \cdot \mathbb{E}[\tilde{Z} \mid X, Y] \leq \frac{1}{2}q^\beta.$$

Therefore,

$$\begin{aligned}
\mathbb{E}[\Delta Y] &\leq\ -q \cdot \left(1-q^\beta\right)\frac{1}{pX + q^{1+2\alpha}} + (1-q) \cdot \frac{1}{2}q^\beta \\
&\leq\ -\frac{q(1-q^\beta)}{q^{1-\beta} + q^{1+2\alpha}} + \frac{1}{2}q^\beta \\
&=\ -\frac{1-q^\beta}{q^{-\beta} + q^{2\alpha}} + \frac{1}{2}q^\beta.
\end{aligned}$$

Here the second inequality uses the upper bound of X on region $\mathcal{E}_B$. Note that the first term has dominant term q^β for small q, therefore, for small enough q, we have:

$$\mathbb{E}\left[\Delta Y \,|X \in \mathcal{E}_B\,, Y \geq B + L\right] \leq\ -\frac{1}{4}q^\beta.$$

Step 4: Drift in Region A. In this region we only focus on the voucher-initiated chain. The probability that the chain is longer than $L = \frac{2}{q}$ is: (We set $Y = \frac{1}{p}\log\frac{5}{q} + L$.)

$$\begin{aligned}
&\left[(1-p)^X(1-(1-p)^{Y-L+1})\right]^L \\
\geq\ &\left[(1-p)^{\frac{q^{1+\alpha}}{p}}(1-(1-p)^{\frac{1}{p}\log\frac{5}{q}})\right]^L \\
\geq\ &\exp\left(-L\cdot p(1+p)\frac{q^{1+\alpha}}{p}\right)\cdot\left(1-\frac{q}{5}\right)^L \\
=\ &\exp\left(-2(1+p)q^\alpha\right)\cdot\left(1-\frac{q}{5}\right)^L \\
\geq\ &\exp\left(-2(1+p)q^\alpha - \frac{2}{5}\left(1+\frac{q}{5}\right)\right) \\
\geq\ &1-2(1+p)q^\alpha - \frac{2}{5}\left(1+\frac{q}{5}\right)
\end{aligned}$$

Here the first inequality follows from the definition of region A, and L. In the second inequality, we use (4) and (2). To see the third inequality: Setting $x := \frac{q}{5} \in [0, \frac{1}{2}]$, inequality (4) gives $1 - x \geq e^{-x-x^2}$; raising both sides to the power $L \geq 0$ yields

$$\left(1-\frac{q}{5}\right)^L \geq \exp\left(-L\left(\frac{q}{5}+\frac{q^2}{25}\right)\right).$$

Multiplying by $\exp(-2(1+p)q^\alpha)$ and plug in $L = 2/q$ gives

$$\exp(-2(1+p)q^\alpha)\left(1-\frac{q}{5}\right)^L \geq \exp\left(-2(1+p)q^\alpha - \left(\frac{2}{5}+\frac{2q}{25}\right)\right).$$

The last inequality follows from (2).

For small enough q, we have

$$2(1+p)q^\alpha < 0.04.$$

For the last term, for small enough q, the last term is smaller than 0.41. As a result, the probability that the chain is longer than $L = \frac{2}{q}$ is at least 0.55. The expected chain length is at least $0.55 \times \frac{2}{q} = \frac{1.1}{q}$, hence the negative drift is at least $q \cdot \frac{1.1}{q} - 1 = 0.1$. We have:

$$\mathbb{E}\left[\Delta Y \,\middle|\, X \leq \frac{q^{1+\alpha}}{p}, Y \geq \frac{1}{p}\log\frac{5}{q} + \frac{2}{q}\right] \leq -0.1\,.$$

Step 5: Putting everything together. Combining the above, we have that for any $\alpha > 0$, if

$$X \leq \frac{q^{1-\frac{\alpha}{2}}}{p},$$

and

$$Y \geq \frac{1+2\alpha}{p}\log\frac{1}{q} + \frac{4}{q^{1+\alpha}}\log\left(\frac{4}{q^{\alpha/2}}\right) + \frac{1}{p}\log 5 + \frac{2}{q},$$

we have

$$\mathbb{E}[\Delta Y] \leq -\frac{1}{4}q^{\frac{\alpha}{2}}\,.$$

For other X, we always have

$$\mathbb{E}[\Delta Y] \leq 1\,.$$

Therefore, applying the Lemma 1, we have:

$$\begin{aligned}&\mathbb{P}\left(Y \geq (1+2\alpha)\frac{1}{p}\log\frac{1}{q} + \frac{4}{q^{1+\alpha}}\log\frac{1}{q} + \frac{1}{p}\log 5 + \frac{2}{q} + 2m\right)\\ \leq\ &\left(\frac{1}{\frac{1}{4}q^{\alpha/2}+1}\right)^m + 9q^{-\alpha/2}\cdot\mathbb{P}\left(X \geq \frac{q^{1-\alpha/2}}{p}\right)\end{aligned} \tag{40}$$

We simplify the above inequality by properly choose m and suppress the dominated terms.

For fixed $\alpha \in (0,1)$, we first bound the order of

$$\mathbb{P}\left(X \geq \frac{q^{1-\alpha/2}}{p}\right)$$

for small q. Note that this term can be bounded by P_k introduced in the proof of Proposition 7 where $k = \log_2(1/\alpha)+2$, and

$$P_k \triangleq \mathbb{P}_\pi\left(X > 10\frac{q^{1-2^{-k}}}{p}\right)$$

We can repeat the analysis steps in the proof of Proposition 7, and we have that as $q \to 0$,

$$P_{\log_2(1/\alpha)+2} = O(e^{-c_1\cdot q^{-1/4}}) + k\cdot O(e^{-c_2\cdot q^{-\alpha}})$$

for some universal constant c_1, c_2. As a result, for fixed $\alpha > 0$, the second term on the RHS of (40) goes to zero as $q \to 0$. The first term satisfied:

$$\left(\frac{1}{\frac{1}{4}q^{\alpha/2}+1}\right)^m \leq\ \exp\left(-\frac{1}{2}q^{\alpha/2}m\right)\,.$$

We choose $m = \frac{1}{q}$, then this term also converges to zero as $q \to 0$. To sum up: for fixed $\alpha \in (0,1)$, $p \leq q$, as $q \to 0$, we have

$$\mathbb{P}\left(Y \geq (1+3\alpha)\frac{1}{p}\log\frac{1}{q} + \frac{4}{q^{1+\alpha}}\log\frac{1}{q}\right) \to 0\,.$$

If we further have that $\lim_{q\to 0}\frac{p}{q^{1+\alpha}} = 0$, then we have that as $q \to 0$,

$$\mathbb{P}\left(Y \geq (1+4\alpha)\frac{1}{p}\log\frac{1}{q}\right) \to 0\,.$$

This concludes the proof.

E Heterogeneous systems: proof of Theorem 2 and Theorem 3

Proof of Theorem 2. When $p_W = 1$, every donor from another pair is compatible with every voucher patient. Therefore, under Standard Priority, whenever a free donor faces at least one waiting voucher patient, the chain immediately matches a voucher patient and terminates. Consequently, any chain removes at most one non-voucher patient before it removes a voucher patient.

First consider the voucher queue. If $X(t) \geq 1$, then a voucher arrival decreases X by one: an incumbent free donor matches the arriving voucher patient, and the arriving donor then matches an incumbent voucher patient. A non-voucher arrival decreases X by one if its patient is compatible with at least one of the $X(t)$ free donors, and otherwise leaves X unchanged. Thus X cannot increase when $X \geq 1$. If $X(t) = 0$, then a non-voucher arrival leaves X at zero, while a voucher arrival either forms a two-cycle through a non-voucher pair and returns to $X = 0$, or leaves exactly one voucher patient waiting. Hence, from every initial state, after at most the initial value of X many voucher arrivals the process enters the set $\{0, 1\} \times \mathbb{Z}_+$ and never leaves it. In particular,

$$\limsup_{t\to\infty} \mathbb{E}[X(t)] \leq 1.$$

The long-run average voucher queue length is therefore at most one, and Little's Law gives a voucher waiting-time bound of order $1/q$.

For non-voucher patients, every non-voucher removal consumes a voucher patient in the same two-cycle. Since voucher patients enter only through voucher arrivals, when $q < 1/2$, non-voucher patients arrive at a rate of $1 - q > 1/2$ and are removed at a rate of at most $q < 1/2$. Over time, the number of non-voucher patients grows to be unbounded. The expected waiting time for non-voucher patients is infinite.

□

For the proof of Theorem 3 we begin with a definition and a simple lemma.

Definition 1. *Let Q_t be a sequence of non-negative random variables where initially $Q_1 = 0$. The queue has the dynamics that, at time $t+1$, either $Q_{t+1} = Q_t + 1$, or $Q_{t+1} = Q_t - Z_t$, where $Z_t \in [0, Q_t]$. We call such a sequence a "(M, K, ρ, β)-random walk" if whenever $Q_t \geq M + 1$, then with probability at least ρ, we have $Z_t > k$. In addition, we impose that $\beta = \rho K - 1 > 0$, which implies that the expected drift is negative when $Q_t \geq M + 1$.*

Lemma 4. *Blum and Mansour (2020) Let Q_t be a (M, K, ρ, β)-random walk, where $\beta \leq 3/5$. Then,*

$$\mathbb{E}[Q_t] \leq M + K(1 + \beta)/\beta.$$

In addition,

$$\mathbb{P}\left(Q_t \leq M + \frac{K(1+\beta)}{\beta} \log \frac{2}{\delta}\right) \geq 1 - \delta.$$

We prove Theorem 3 with Lemma 4.

Proof of Theorem 3. First we compute the upper bound of the expected wait for non-voucher. We argue that the non-voucher queue Y_t is a (M, K, ρ, β)-random walk. We only focus on voucher arrivals and chains of lengths at least K started by those voucher arrivals. For $Y_t > \frac{1}{p}\log\frac{4}{q} + \frac{4}{q}$, conditioned on a voucher arrival who is ready to start a chain, the probability of extending a chain to a non-voucher when it is already at length i is $1-(1-p)^{M-i} \geq 1-(1-p)^{M-K}$ for the first K steps. Hence, the probability that the chain has length at least $K = 4/q$ is at least

$$\begin{aligned}\left(1-(1-p)^{M-K}\right)^K &\geq \left(1-(1-p)^{\frac{1}{p}\log\frac{4}{q}}\right)^{4/q}\\ &\geq \left(1-e^{-\log\frac{4}{q}}\right)^{4/q}\\ &\geq (1-q/4)^{4/q}\\ &\geq e^{-(1+q/4)}.\end{aligned}$$

The probability that a voucher arrives is q, so we can start a chain with probability at least q. Y_t is a (M, K, ρ, β)-random walk with $M = \frac{1}{q}\log\frac{4}{q} + \frac{4}{q}, K = \frac{4}{q}, \rho = qe^{-(1+q/4)}, \beta = 4e^{-(1+q/4)} - 1 \in (0.1, 0.5)$. From Lemma 4 we have

$$\mathbb{E}[Y_t] \leq \frac{1}{p}\log\frac{4}{q} + \frac{4}{q}\frac{1+\beta}{\beta}$$

and

$$\mathbb{P}\left(Y_t \leq \frac{1}{p}\log\frac{1}{q} + \frac{4}{q} + \frac{4}{q}\frac{1+\beta}{\beta}\log\frac{2}{\delta}\right) \geq 1-\delta$$

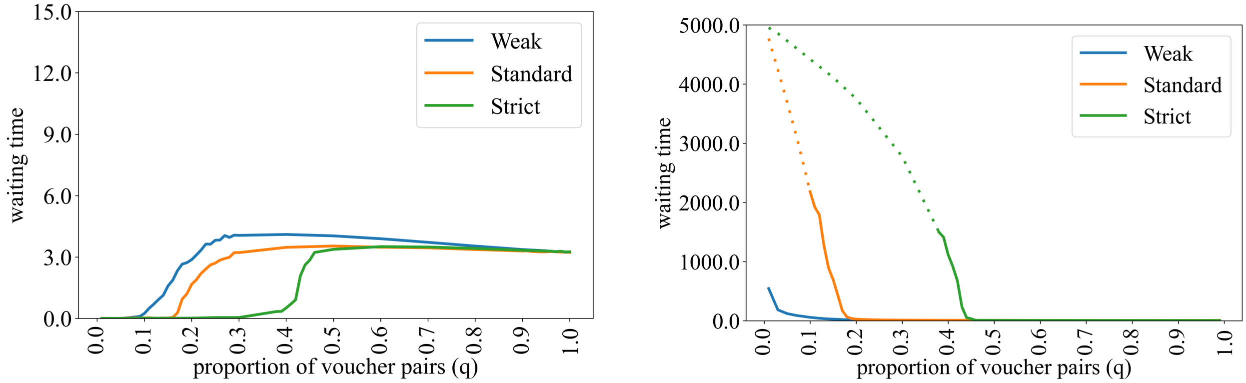
for all q.

The desired expected wait can be achieved by dividing the expected queue length by the non-voucher arrival rate $1-q$ using Little's Law (Little, 1961).

For voucher patients, the expected wait is $O\left(\min\{\frac{1}{2q}, \frac{1}{p}\}\right)$. Since any two voucher pairs can match with each other, the expected wait is at most $\frac{1}{2q}$. Additionally, the voucher pair can match with a non-voucher pair, in which case its expected wait is $\frac{1}{p}$. Together, voucher patients' expected wait is $\min\{\frac{1}{2q}, \frac{1}{p}\}$.

$\square$

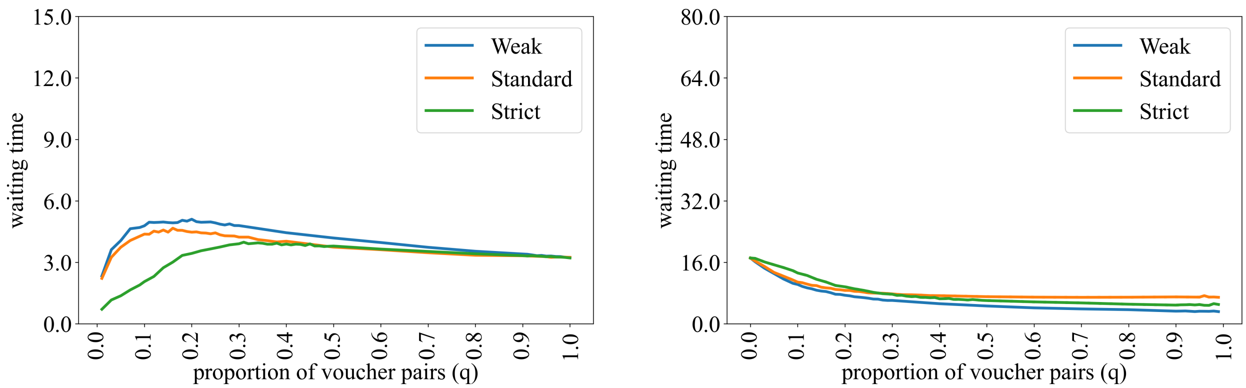
F Additional simulations



(a) $p = 0.2$, voucher

(b) $p = 0.2$, non-voucher

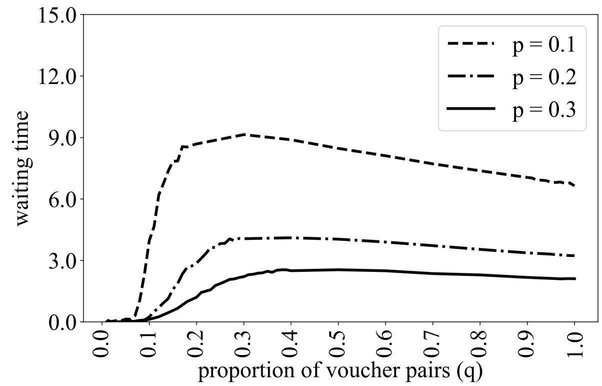
Figure 12: Average waiting times when no non-voucher cycles are allowed. The left and right panels are for patients of voucher and non-voucher pairs. Blue, orange, and green correspond to Weak, Standard, and Strict priorities. Solid lines denote the average waiting time in steady state; dotted lines denote the average waiting time for the patients when it is unbounded (but remains finite in our finite simulation with 10,000 periods).



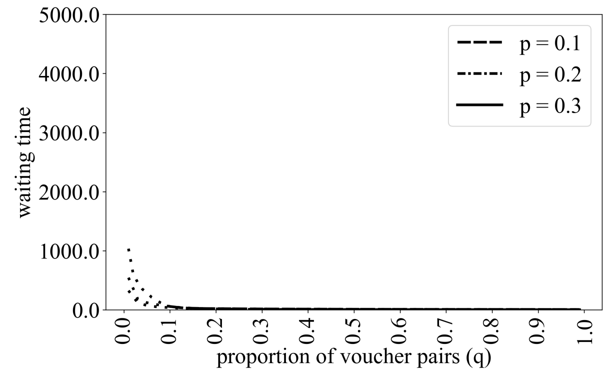
(a) $p = 0.2$, voucher

(b) $p = 0.2$, non-voucher

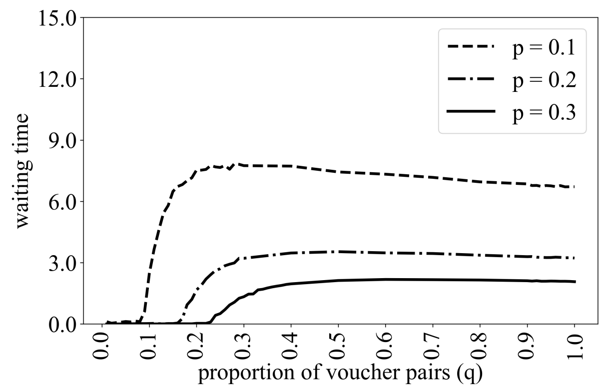
Figure 13: Average waiting times with 2-cycles for $p = 0.2$ averaged over 5 iterations.



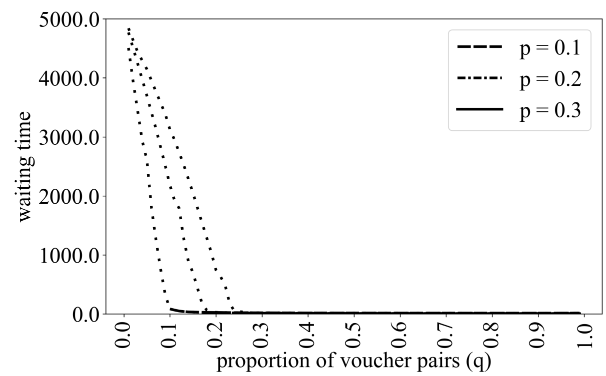
(a) Weak Priority, voucher



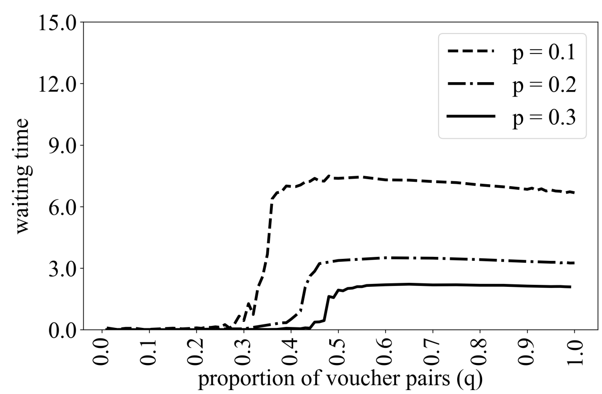
(b) Weak Priority, non-voucher



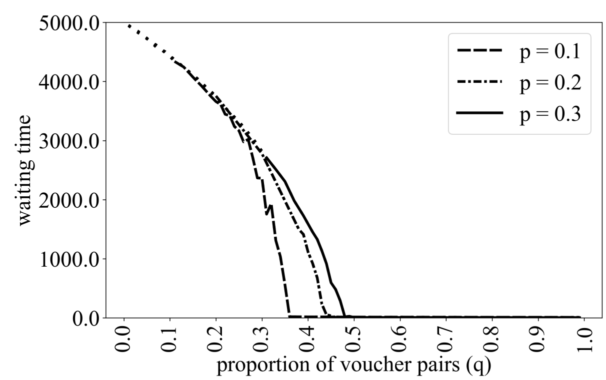
(c) Standard Priority, voucher



(d) Standard Priority, non-voucher



(e) Strict Priority, voucher



(f) Strict Priority, non-voucher

Figure 14: Average wait times when no cycles are allowed, organized by mechanism. The left panels show waiting times for voucher patients, and the right panels for non-voucher patients. Dashed, dash-dot, and solid lines correspond to $p = 0.1, p = 0.2$, and $p = 0.3$ respectively.

	$q = 0.02$				$q = 0.25$				$q = 0.40$			
	#W	#seg	#prop chain	avg chain length	#W	#seg	#prop chain	avg chain length	#W	#seg	#prop chain	avg chain length
Standard	191	191	191	11.42	2463	3198	2113	4.73	3958	4436	3189	3.13
Weak	219	219	219	42.58	2549	3695	2189	2.70	4062	3289	3935	2.04
Strict	226	228	226	2.02	2564	2569	2556	2.08	3818	4543	3150	3.17

Table 3: Chain and segment metrics for $p = 0.1$ and $q \in \{0.02, 0.25, 0.40\}$, without cycle-augmentation. #W denotes the number of voucher pairs, #seg denotes the number of match segments, #prop chain denotes the number of proper chains, and avg chain length denotes the average proper chain length.