Congested Waiting Lists and Organ Allocation*

Itai Ashlagi† Ravi Jagadeesan‡ Pengyu Qian§

September 11, 2026

Abstract

More than 25% of the kidneys recovered from deceased donors in the U.S. and offered to patients on the national waiting list are not utilized. This paper shows that waiting-list designs can suffer from a form of congestion that can lead to valuable organs being discarded. As organs have a limited "cold-ischemia time" during which they can be offered before expiring, an organ may expire before being offered to a patient lower on the waiting list who would accept it. We develop a parsimonious framework to study equilibria of waiting lists with congestion and show that congestion creates a strong enough externality to substantively affect utilization and welfare. The delegation of organ-acceptance decisions to risk-averse transplant centers can worsen congestion and thereby increase all waiting times. We discuss how recent changes to the waiting-list design affect congestion, and policies and market designs that can mitigate congestion and thereby improve welfare.

1 Introduction

In 2024, 3,571 patients died while waiting for a kidney transplant from a deceased donor in the United States, and 4,758 patients were removed from the waiting list due to being too sick to receive transplants (Health Resources and Services Administration, 2025). During the same year, 27% of kidneys recovered from deceased donors were not utilized—a rate that has been increasing in recent years (Mohan et al., 2023; Wood et al., 2025). This paper seeks to understand how practical features of the organ allocation system can lead to excess non-utilization of recovered organs, and investigates implications for market design.

*An extended abstract of this paper appeared in the *Proceedings of the 27th ACM Conference on Economics and Computation (EC'26)*. We thank Dirk Bergemann, Navin Kartik, Al Roth, Larry Samuelson, Philipp Strack, and several seminar audiences for helpful comments.

†Department of Management Science and Engineering, Stanford University. Email address: iashlagi@stanford.edu. Ashlagi was supported by National Science Foundation grant 2312156.

‡Department of Economics, Stanford University. Email address: ravi.jagadeesan@gmail.com.

§Questrom School of Business, Boston University. Email address: pqian@bu.edu.

The key issue we focus on is that organs are perishable and therefore have a limited time to be offered before they expire. We show that this friction can cause "congestion" in the sense that some patient on the waiting list would be willing to accept some organ but the organ expires before being offered to that patient—thereby leading to excess non-utilization. We argue that understanding the welfare effects of a market design requires accounting for its impact on congestion, and evaluate recent attempts to improve utilization rates in practice.

We proceed in three steps. First, we provide descriptive evidence that suggests that organ expiry leads to congestion, and that the resulting under-utilization of organs is economically significant. Second, we investigate the welfare implications of congestion by developing a tractable theoretical model of waiting lists with heterogeneous, expiring objects. We show that congestion leads to artificially lower endogenous demand for lower-quality organs—thereby generating severe inefficiencies and leading to policies sometimes having counterintuitive welfare implications. Third, we use our model to evaluate recent market design innovations and proposals. For example, we show that attempts to increase access to organ offers, such as in a 2021 reform in the U.S., can exacerbate congestion. On the other hand, the recent practice of accelerating placement of organs at risk of non-utilization via out-of-sequence offers may alleviate congestion and hence improve welfare.

In practice, when an organ becomes available, it is typically offered sequentially to patients in priority order. Once an offer is made, the patient can refuse the offer without penalty (i.e., without forgoing any priority). Patients near the top of the list may be inclined to refuse organs of reasonable quality, as such patients stand a chance of being offered a high-quality organ in the future. By contrast, patients further from the top of the waiting list may accept lower-quality organs, as they have less of a chance of surviving (and remaining healthy enough to undergo transplantation) until being offered a high-quality organ.

Refusals of potentially useful organs come at a cost because there is a limited "cold-ischemia time" after recovery for which organs are viable for transplantation. We show that due to this limit on viability, waiting lists can enter into a counterintuitive state that we call *congestion*. Specifically, lower quality organs may expire before being offered to a patient who would be willing to accept them—despite there being such patients lower down the waiting list. Section 2 provides descriptive evidence suggesting that this form of congestion is prevalent and economically

significant in practice.

To investigate the welfare implications of congestion, we develop a parsimonious fluid framework to study steady states of waiting lists with heterogeneous, expiring objects (Section 3). Patients and organs arrive as flows of constant rates. Patients stochastically depart unmatched (i.e., due to death or becoming too sick to undergo transplantation) at rates that can depend on both their types (e.g., sicker or older patients may depart unmatched at a faster rate) and on the time they have spent on the waiting list (e.g., patients get sicker over time). To capture a key dimension of heterogeneity in organ allocation, we suppose that organs can be high-quality (safe) or low-quality (risky).[1] Patients are heterogeneous in how they value risky organs relative to safe organs. Last, to capture the limited cold-ischemia time for which organs are viable, we assume that organs expire after being offered to γ mass of patients.[2]

A consequence of organs' limited viability is that organs—including risky organs—are never offered to patients who are not among the top γ mass of patients on the waiting list. If enough patients strongly prefer safe organs, risky organs may never make it far enough down the waiting list to be offered to patients who would be willing to accept them. Hence, there may be congestion.

Congestion is relevant not only to patients who may be willing to accept risky organs. Indeed, congestion incentivizes such patients to wait for safe organs, thereby raising the waiting time for safe organs and hurting even patients who strongly prefer safe organs. When congestion occurs, all patients would strictly prefer a counterfactual with non-perishable organs, and hence no congestion. In this sense, when congestion arises, it harms all patients.

To illustrate the severity of the welfare loss caused by congestion, we consider a hypothetical policy that results in the procurement of more safe organs. Without congestion, this policy would obviously weakly benefit all patients and strictly benefit at least some. By contrast, we find that in the presence of congestion, such a policy can strictly hurt all patients. Indeed, an increase in the availability of safe organs results in more types waiting for safe organs. This can exacerbate congestion so severely that it raises waiting times for everyone.

[1] For interpretation, we focus on patients of a single blood type and abstract from other forms of incompatibility (e.g., tissue-type incompatibility).

[2] We abstract from other forms of decay in organ quality with the elapse of cold-ischemia time.

Another feature of the organ allocation system can exacerbate congestion and harm patients. In practice, patients do not decide whether to refuse an organ offer themselves—rather, transplant centers make these decisions on behalf of patients. However, transplant centers are penalized for adverse post-transplant outcomes, but not deaths on the waiting list, and so are incentivized to refuse risky organs even when patients may benefit from them (Chan and Roth, 2024; Mildebrath et al., 2024).

This misalignment of incentives inflates the relative demand for safe organs. Without congestion, the change in demand increases the waiting time for safe organs but lowers the waiting time for risky organs. However, in the presence of congestion, the change in demand also worsens congestion, which we show decreases utilization enough to raise the waiting time for risky organs as well. As a result, if the system would be congested under patient-optimal refusal decisions, the delegation of decisions to transplant centers harms all patients. Thus, impacts on congestion make the misalignment of incentives between patients and transplant centers particularly harmful.

Design choices can worsen congestion and potentially harm welfare. For example, we show that efforts to improve market thickness can worsen congestion and harm patients. This insight may shed light on some of the increase in organ non-utilization. In March 2021, the Organ Procurement and Transplantation Network moved from offering a "local" priority to patients based on an organ-procurement-organization donor service area to offer priority to patients within 250 nautical miles of the donor hospital. This change increased the number of patients that are given local priority. At the same time the non-utilization rate of kidneys from deceased donors increased from 21.2% in 2020 to 24.5% in 2021.

In response to the increasing non-utilization rate, organ procurement organizations have recently begun to *accelerate* organs by offering them to transplant centers out-of-sequence (Liyanage et al., 2025). We argue why this can help alleviate congestion, and discuss the difficulties of alleviating congestion under sequential offers.

Methodologically, the possibility of congestion means that we cannot use standard approaches to analyze queuing systems. When patients' departure rates are heterogeneous, there may be multiple steady states that suffer from congestion, and we cannot solve for the steady states in closed form. Our analysis instead relies on monotone-comparative-statics methods (Milgrom and Roberts, 1994), which had not been previously used to study queues. These methods also allow us to analyze queuing systems in which participants have time-varying departure rates.

1.1 Related Literature

This paper contributes to the literature on dynamic allocation without transfers. A large body of work studies the design of waiting lists and priority mechanisms for allocating scarce objects as in organ allocation and public housing (Su and Zenios, 2004, 2005, 2006; Bloch and Cantala, 2017; Agarwal et al., 2021; Ata et al., 2021; Leshno, 2022; Arnosti and Shi, 2020; Waldinger, 2021; Murra-Anton and Thakral, 2024; Schummer, 2021).[3] In these models, an object is not utilized only if no agent present in the allocation system is willing to accept it. By contrast, in our setting, making sequential offers takes time, and an organ may expire before the offer process reaches a patient willing to accept it. Non-utilization of organs can therefore arise despite having patients on the list who are willing to accept them.

Castro et al. (2021) considers object expiry in the context of ride-sharing queues in which perishable requests are offered sequentially to strategic drivers. They design novel mechanisms that make unattractive requests move through the queue more quickly, thereby reducing wastage due to repeated rejections. We instead take as given the waiting-list structure used in organ allocation and study its equilibrium implications. Our model also allows patients to differ in their willingness to wait for quality and their departure-rate trajectories, which are central features of organ allocation.

Our analysis also complements work explaining why organ offers are refused. Zhang (2010), De Mel et al. (2021), and Doval et al. (2024) study informational forces generated by previous acceptance and refusal decisions, while Chan and Roth (2024) and Mildebrath et al. (2024) study the incentives created by transplant-center regulation. These papers identify forces affecting individual acceptance decisions. We instead focus on studying how refusals propagate through the market when offers must be processed sequentially and organs are perishable.

Our analysis also contributes to the literature on queues with abandonment (Mandelbaum et al., 2002; Whitt, 2006; Dai and He, 2010). In that literature, agents are assumed to be non-strategic and accept service when it is offered. In our setting, objects are heterogeneous and perishable, and agents strategically refuse certain objects, which causes congestion. We obtain tractability in the presence of these richer features by assuming that unmatched agents' departures follow a version of the Cox (1972) proportional hazards model.

[3] See also Baccara et al. (2020) for an analysis of congestion in two-sided matching markets, as well as Doval (2025) and Ashlagi and Roth (2021) for surveys of dynamic matching and organ allocation.

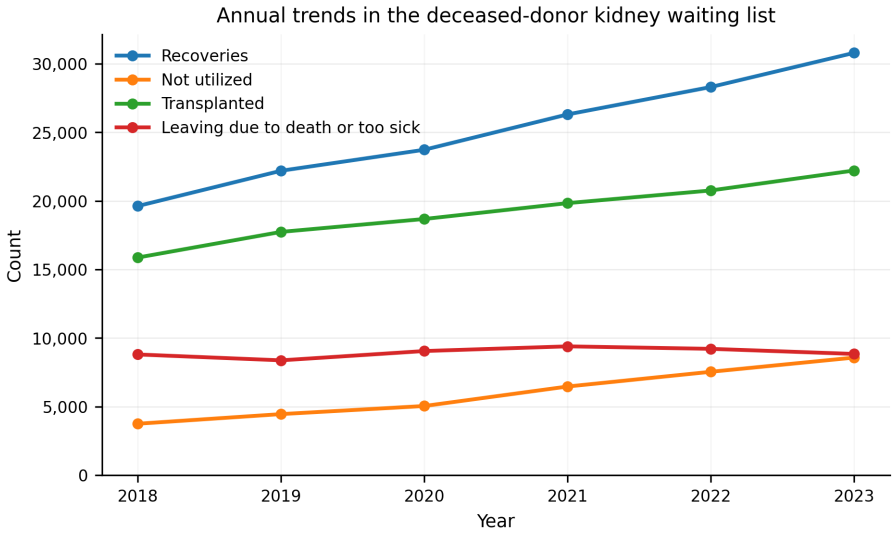

Figure 1: Annual trends in waiting list for deceased-donor kidneys.

2 Background

Kidney transplantation improves health outcomes and reduces the cost of treatment relative to remaining on dialysis (Wolfe et al., 1999; Held et al., 2016). Yet deceased-donor kidneys remain severely scarce. Figure 1 documents the corresponding growth in kidney recovery, transplantation, and non-utilization in recent years.[4]

When a deceased-donor kidney becomes available, the allocation system constructs an offering order based on compatibility and waiting time, as well as other priority criteria such as sensitization, geography, and other medical and demographic characteristics. The kidney is then offered according to this order. A patient or transplant center may refuse an offer without losing the patient's priority for future kidneys.[5]

The institutional setting has three features that are the focus of our analysis. First, deceased-donor kidneys differ substantially in quality, with lower-quality kidneys less likely to be utilized. Second, processing these offers takes time which results in degradation of organs, especially for organs of lower quality. Third, acceptance decisions

[4]Data reported here have been supplied by UNOS as the contractor for the Organ Procurement and Transplantation Network (OPTN). Data is taken from the Standard Transplant Analysis and Research (STAR) files that include de-identified information on all patients, and the Potential Transplant Recipient (PTR) files that include de-identified organ offers. The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the OPTN or the U.S. Government.

[5]Some refusals are unavoidable because information revealed during the offer process may show that a particular kidney is incompatible with a patient.

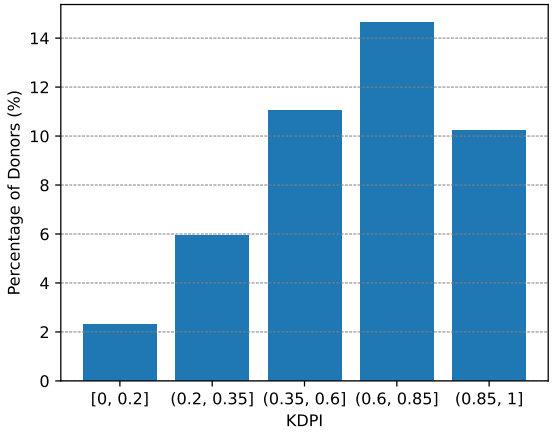

Figure 2: Percent of donors with two recovered kidneys for whom only one kidney was transplanted, by Kidney Donor Profile Index (KDPI) bin.

are often made on behalf of patients by transplant centers, which have incentives to refuse lower-quality kidneys. We describe these features in detail below.

Quality Heterogeneity and Non-Utilization. Recovered kidneys differ substantially in their expected graft longevity. The Kidney Donor Profile Index (KDPI) is a percentile index that summarizes the donor's pre-recovery medical characteristics that are associated with graft failure; higher values indicate lower expected graft longevity. High-KDPI kidneys are consequently less desirable, but they can nevertheless provide substantial benefits to suitable patients relative to remaining on dialysis. Indeed, while only around 40% of patients survive on dialysis for more than 5 years (USRDS, 2013), even organs with KDPI over 85% (i.e., the highest 15 percentiles of risk) have five-year graft-survival rates above 60% (Lentine et al., 2025).

Moreover, many kidneys that are not transplanted have observable characteristics similar to kidneys that are successfully transplanted (Mohan et al., 2018; Aubert et al., 2019). Notably, in 2023, there were 1,547 deceased donors from whom both kidneys were recovered but only one was transplanted. Figure 2 shows that the frequency of such outcomes increases with KDPI. Thus, non-utilization is concentrated among lower-quality kidneys, but is not limited to kidneys that are medically unusable.

Sequential Offers and Cold-Ischemia Time. Since patients with a high priority may be inclined to refuse offers of organs with low quality, such organs may accumulate many refusals. Figure 3 reports the median number of offers among kidneys

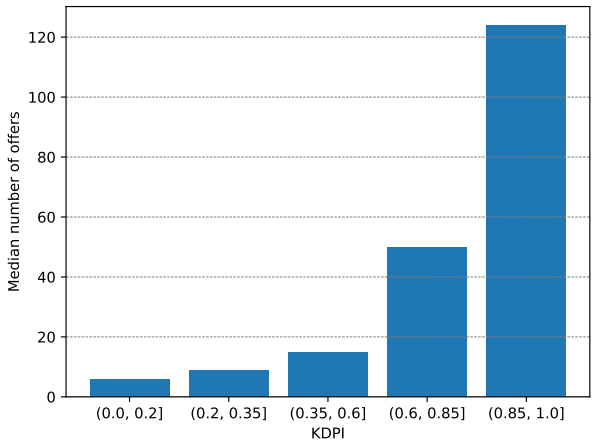

Figure 3: Median number of offers for transplanted organs by Kidney Donor Profile Index (KDPI) bin.

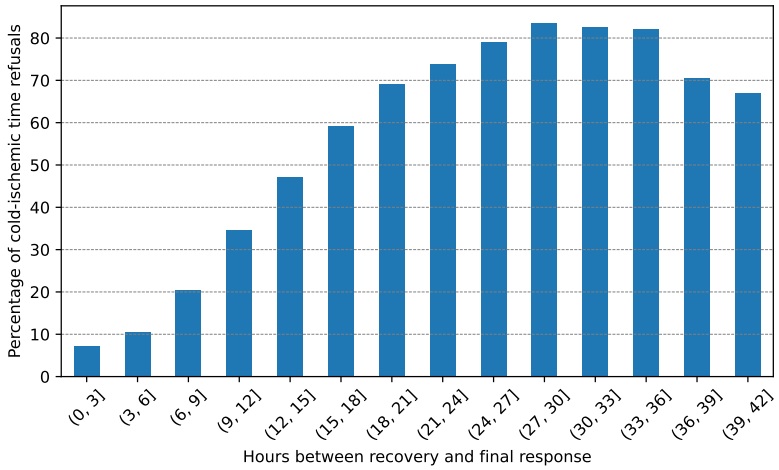

Figure 4: Percent of offer refusals that are due to "actual or projected cold-ischemia time too long," by (binned) number of hours after organ recovery.

that are eventually transplanted by KDPI bin. The number increases steeply with KDPI: even conditional on successful transplantation, a high-KDPI kidney is offered to more than one hundred patients at the median before being accepted.

Each additional offer consumes time. This is consequential because a recovered kidney remains viable for only a limited period of "cold-ischemia time," and longer cold-ischemia time reduces expected graft survival and makes subsequent acceptance less likely (Quiroga et al., 2006; Peters-Sengers et al., 2019; Debout et al., 2015). Recovered kidneys are typically medically viable for around 24–36 hours. Figure 4 shows that the share of refusals attributed to "actual or projected cold-ischemia time

too long" rises sharply with the time elapsed since recovery.[6] So a lower-quality kidney may move too slowly through the priority order and *expire* (Barah et al., 2022; Guan et al., 2025) before reaching a patient who would be willing to accept the offer.

Delegated Decisions. In practice, transplant centers often make decisions to refuse offers on behalf of their patients. Transplant centers may decline an organ separately for a particular patient or decline it simultaneously for multiple patients. Guan et al. (2025) document that more than 69% of refusals in 2022–2023 were multi-patient refusals, indicating that much of the screening occurs at the transplant-center-level rather than the patient-level. As transplant centers are penalized for adverse post-transplant outcomes but not deaths on the waiting list, they are incentivized to refuse lower-quality organs (Chan and Roth, 2024; Mildebrath et al., 2024). In addition, survey data suggest that patients are more inclined to accept low-quality organs than their transplant centers are (Mehrotra et al., 2022; Sepulveda et al., 2023). Our analysis begins with patients making acceptance decisions, but we return in Section 7 to discuss how equilibrium outcomes change when transplant centers make decisions.

3 Model

This section develops our model of waiting lists with expiring objects. We take a continuous-time fluid approach, which makes the equilibrium analysis and comparative statics tractable by avoiding fluctuations due to stochastic arrivals and departures. Thus, we suppose a continuum of patients arrives on the waitlist at a flow of a constant rate $p > 0$. The analogue of position in a standard queue is that we refer to a patient who has waited for time t as being at *position* t.

3.1 Organs

There are two types of organs: safe grafts and risky grafts. A continuum of safe organs (resp. risky organs) arrives at a flow of constant rate $sp > 0$ (resp. $rp > 0$). We assume that $s < 1$, capturing the shortage of safe grafts.

Each arriving organ is offered sequentially to patients in descending order of waiting time. However, an organ expires after $\gamma > 0$ mass of patients refuse the organ.

[6] See also Guan et al. (2025) for a detailed analysis of offer refusals.

Here, γ can be interpreted as a measure of the post-recovery viability of a graft—i.e., the limit on cold-ischemia time to remain viable. Thus, each organ goes to a highest-ranking patient on the waitlist (at the time of the organ's arrival) who is willing to accept the organ, except that patients who have more than γ mass (of patients) ahead of them on the waitlist do not receive organs. We allow for the case of $\gamma = \infty$, which captures the case in which organs do not expire.

3.2 Patients' Preferences and Decisions

Patients are heterogeneous in their individual characteristics, as well as in the characteristics of transplant centers that manage their care. We summarize these characteristics in a type $\theta \in \Theta$, where Θ is a compact domain in a finite-dimensional Euclidean space $\mathbb{R}^k$. We assume that the distribution of characteristics of arriving patients is constant over time, and admits a (Borel-measurable) probability density function $f(\theta)$. A patient of type θ has Bernoulli utility

$$u_{\text{patient}} = \begin{cases} 1 & \text{if get safe organ} \\ \nu(\theta) & \text{if get risky organ} \\ 0 & \text{if do not get an organ} \end{cases},$$

where $\nu(\theta) \in [0,1]$ is the willingness to accept a risky organ.[7]

Based on a nonlinear version of the Cox (1972) proportional hazards model, patients of type θ who remain unmatched after having been on the waiting list for t time depart at rate $\lambda(t)\beta(\theta)$, where $\lambda(t) > 0$ is a baseline hazard rate and $\beta(\theta) > 0$ is a baseline scaling factor.[8] For interpretation, $\beta(\theta)$ captures persistent differences in departure risk across patients, such as differences in baseline medical status, while $\lambda(t)$ captures how departure risk evolves with time spent waiting. A simple example is that of time-invariant departure rates—i.e., $\lambda(t) = 1$—in which case $\beta(\theta)$ gives the departure rate of unmatched patients of type θ. More naturally, departure rates may increase with time spent on the waiting list (as patients get sicker).

Letting $\Lambda(t) = \int_0^t \lambda(x)dx$ denote the baseline cumulative hazard function, the probability that a patient of type θ survives to time t absent a match is $e^{-\Lambda(t)\beta(\theta)}$. We

[7] Here, $\nu : \Theta \to [0,1]$ is a Borel measurable function.

[8] Here, $\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is a measurable function, and $\beta : \Theta \to \mathbb{R}_{>0}$ is a Borel-measurable function.

impose a technical assumption that both the baseline hazard function $\lambda(t)$ and the baseline scaling factors $\beta(\theta)$ are bounded away from 0 and bounded above.

Upon receiving an offer, a patient decides whether to accept it. We suppose that patients make these decisions to maximize their expected utility. In Section 7, we introduce an extension in which decisions are delegated to transplant centers.

3.3 Discussion of Assumptions

To isolate the impact of organ expiry due to excess cold-ischemia time in a tractable framework, our model abstracts from some features of the kidney waiting list.

In practice, there are numerous different qualities of organs, parameterized by the KDPI. For tractability, we consider two types of organs to understand the impact of congestion on the utilization of lower-quality organs.

The waiting list also involves priorities, which depend on medical and geographic characteristics of the organ and the patient, including waiting times. For example, in kidney allocation in the U.S., healthy and young patients are prioritized for organs in the top 20 percentiles of quality (i.e., KDPI less than 20%). However, a majority of organs from deceased donors are effectively allocated (within a local region) based on the amount of time on dialysis, which we interpret as waiting time in our model. Indeed, for each organ, typically few patients are assigned a high priority (as few patients have a perfect match, or are very sensitized yet still compatible with the organ). For other patients compatible with the organ, offers are mainly determined by waiting time.[9] Our analysis can be interpreted as applying within any of the majority of priority classes for which organs are offered effectively in order of waiting time.

Organs also degrade continuously with cold-ischemia time (Sampaio et al., 2018). For tractability, to focus on the substantial increase in refusals due to "actual or projected cold-ischemia time too long" at 24–36 hours post-recovery, we assume that the quality of the organ remains fixed for γ offers after which it immediately expires. The parameter γ therefore reflects both the limited cold-ischemic time for which organs are viable and the time required to process offers in sequence.

Last, we assume a proportional-hazard-rate structure for patient departures, based on a nonlinear version of Cox (1972). This specification places no restrictions on the evolution of the departure rate of a patient over time, but rather only that that

[9]See `https://optn.transplant.hrsa.gov/media/eavh5bf3/optn_policies.pdf` for more details of organ allocation policies in the U.S.

variation in the departure rate with time on the waiting list be proportional across patients.[10] Cox (1972) regression models are also widely applied in the empirical literature (Medin et al., 2000; Schaubel et al., 2009; Agarwal et al., 2021).

4 Equilibria

In this section, we introduce steady states based on waiting times. Intuitively, due to the fluid formulation, there will be two waiting times: a waiting time T_{safe} needed to obtain an offer of a safe organ, and a waiting time T_{risky} needed to obtain an offer of a risky organ. All patients accept safe organs if offered, and patients who wait at least T_{risky} continue to be offered risky organs at each moment in time (until they accept an offer or depart unmatched). In Online Appendix C, we formally define steady-state equilibria when patients can dynamically adapt their organ-acceptance decision rules, and show that steady-state equilibria precisely correspond to the waiting-time steady states introduced in this section.

Suppose a patient of type θ is offered a risky organ at time t but knows that they will receive an offer of a safe organ at time $T_{\text{safe}} \geq t$. The conditional probability that they survive to time T_{safe} is then $e^{-(\Lambda(T_{\text{safe}})-\Lambda(t))\beta(\theta)}$. Maximizing expected utility taking into account the exogenous rate of unmatched departure, they will accept the risky organ if $\Lambda(T_{\text{safe}}) - \Lambda(t) > W(\theta)$, where

$$W(\theta) = \frac{-\log \nu(\theta)}{\beta(\theta)}, \tag{1}$$

and refuse the risky organ and wait for a safe one if $\Lambda(T_{\text{safe}})-\Lambda(t) < W(\theta)$. Since these conditions compare $T_{\text{safe}} - t$ and $W(\theta)$ in the case of time-invariant departure rates, we call the quantity $W(\theta)$ the *willingness-to-wait* for a patient with type θ. (More generally, willingness-to-wait is measured in baseline cumulative-hazard units.)

If $\Lambda(T_{\text{safe}}) - \Lambda(t) = W(\theta)$, then the patient has multiple optimal strategies. To avoid technical issues with this being the case for a positive mass of patients, we make the following technical assumption throughout.

Assumption 0. The distribution of $W(\theta)$ is continuous.

The organ assignment in steady-state equilibrium is determined by an agent's

[10] For an axiomatic characterization of the proportional hazards model, see Kella (2024).

willingness-to-wait. Indeed, a patient of type θ refuses a risky organ at time T_{risky} and waits for a safe one if $W(\theta) > \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$; such patients then continue to refuse any risky organs that are offered. Conversely, a patient of type θ accepts a risky organ at time T_{risky} if $W(\theta) < \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$.

Thus, almost all participants' decisions depend only on the difference

$$\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$$

in baseline cumulative hazards between the waiting times. Under time-invariant departure rates (i.e., $\lambda(t) = 1$), this difference is simply the difference in waiting times.

We can then define waiting-time steady states in terms of the distribution of the willingness-to-wait. Such steady states must satisfy four conditions. As safe organs are in short supply and all patients accept offers of safe organs, all safe organs must be allocated. Three further conditions constrain T_{risky}. First, the demand for risky organs cannot exceed the supply of risky organs. Second, risky organs expire if refused by γ mass of patients, which can increase T_{risky}. Third, risky organs are not intentionally withheld: one of the previous two conditions must hold with equality unless everyone is offered a risky organ (i.e., unless $T_{\text{risky}} = 0$).

Definition 1. A *waiting-time steady state* consists of waiting times $0 \leq T_{\text{risky}} \leq T_{\text{safe}}$ that satisfy the following conditions, writing $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$:

1. All safe organs are allocated:

$$\text{SafeDemand}(T_{\text{risky}}, T_{\text{safe}}) := p \int_{\theta: W(\theta) > \Delta} e^{-\Lambda(T_{\text{safe}})\beta(\theta)} f(\theta) d\theta = \ sp\,. \tag{2}$$

2. Demand for risky organs cannot exceed their supply:

$$\text{RiskyDemand}(T_{\text{risky}}, T_{\text{safe}}) := p \int_{\theta: W(\theta) < \Delta} e^{-\Lambda(T_{\text{risky}})\beta(\theta)} f(\theta) d\theta \leq \ rp\,. \tag{3}$$

3. An organ expires after being refused by γ measure of agents:

$$\text{RefuseRisky}(T_{\text{risky}}, T_{\text{safe}}) := \int_{T_{\text{risky}}}^{T_{\text{safe}}} \underbrace{p \int_{\theta: W(\theta) > \Delta} e^{-\Lambda(t)\beta(\theta)} f(\theta) d\theta}_{\text{declining mass of patients at position } t} \ dt \leq \ \gamma\,. \tag{4}$$

4. Risky organs are not intentionally withheld:

$$\text{At least one of the inequalities (3), (4), and } T_{\text{risky}} \geq 0 \text{ holds with equality.} \quad (5)$$

In the case of $\gamma = \infty$, (4) cannot bind, and we obtain the more standard case of a fluid queue with non-expiring objects.

As patients make optimal decisions about organ acceptance given the waiting times, the expected utility of type θ in a waiting-time steady state $(T_{\text{risky}}, T_{\text{safe}})$ is

$$U_\theta(T_{\text{risky}}, T_{\text{safe}}) = \mathbb{E}[u_{\text{patient}}] = \max\left\{\nu(\theta)e^{-\Lambda(T_{\text{risky}})\beta(\theta)}, e^{-\Lambda(T_{\text{safe}})\beta(\theta)}\right\}.$$

4.1 Existence

We next establish the existence of a waiting-time steady state.

Proposition 1 (Existence)**.** *In each market, there exist waiting-time steady states.*

A proof with arbitrary numbers of types of organs can be given based on Kakutani's Fixed Point Theorem. We instead give a more direct argument that uses the two-type structure but helps to illustrate the logic behind equilibrium in our model.

Since almost all participants' decisions depend only on the difference $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$ in baseline cumulative hazards between the waiting times, the flow balance for safe organs (2) can be expressed as a relationship between T_{safe} and Δ. The larger the difference, the fewer patients wait for a safe organ, and hence the smaller the waiting time for safe organs. As $\Lambda(T_{\text{risky}}) = \Lambda(T_{\text{safe}}) - \Delta$, this also implies a relationship between T_{risky} and Δ, which we use in our subsequent analysis.

Formally, let $\overline{\Delta} = \operatorname{ess\,sup} W(\theta) \leq \infty$, which is the supremal willingness-to-wait for a safe organ. Furthermore, consider an arbitrary strictly increasing, continuous extension of $\Lambda(t)$ to $t \in \mathbb{R}$ that is unbounded below.

Lemma 1. *(a) There is a nonincreasing, continuous function $g_{\text{safe}} : [0, \overline{\Delta}) \to \mathbb{R}$ (dependent on the parameters s, $f(\theta)$, $\nu(\theta)$, $\lambda(t)$, and $\beta(\theta)$) with $g_{\text{safe}}(0) > 0$ and $\lim_{\Delta \to \overline{\Delta}^-} g_{\text{safe}}(\Delta) < 0$ such that $(T_{\text{risky}}, T_{\text{safe}})$ satisfies (2) if and only if $T_{\text{safe}} = g_{\text{safe}}(\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}))$. If $W(\theta)$ has full support on $[0, \overline{\Delta})$, then g_{safe} is strictly decreasing.*

(b) Letting $g_{\text{risky}}(\Delta) = \Lambda^{-1}\big(\Lambda(g_{\text{safe}}(\Delta)) - \Delta\big)$*, which is continuous and strictly decreasing on* $[0, \overline{\Delta})$*, a pair* $(T_{\text{risky}}, T_{\text{safe}})$ *satisfies* (2) *if and only if we have* $T_{\text{risky}} = g_{\text{risky}}(\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}))$.

Proof. Consider the demand for safe organs expressed as a function of T_{safe} and $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$, which is $D_{\text{safe}}(T_{\text{safe}}, \Delta) = \text{SafeDemand}(\Lambda^{-1}(\Lambda(T_{\text{safe}}) - \Delta), T_{\text{safe}})$. Note that for each $\Delta \in [0, \overline{\Delta})$, the quantity $D_{\text{safe}}(T_{\text{safe}}, \Delta)$ is strictly decreasing and continuous in T_{safe}, converges to ∞ as $T_{\text{safe}} \to -\infty$, and converges to $0 < sp$ as $T_{\text{safe}} \to \infty$. Hence, for each $\Delta \in [0, \overline{\Delta})$, there exists a unique value $T_{\text{safe}} = g_{\text{safe}}(\Delta) \in \mathbb{R}$ such that $D_{\text{safe}}(T_{\text{safe}}, \Delta) = sp$—i.e., (2) holds for $(T_{\text{risky}}, T_{\text{safe}})$ when $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$

As $D_{\text{safe}}(T_{\text{safe}}, \Delta)$ is nonincreasing in Δ and continuous in $(T_{\text{safe}}, \Delta)$, the function g_{safe} is nonincreasing and continuous. (If $W(\theta)$ has full support on $[0, \overline{\Delta})$, then $D_{\text{safe}}(T_{\text{safe}}, \Delta)$ is strictly decreasing in Δ on $[0, \overline{\Delta})$, and hence so is g_{safe}.) We also have that $D_{\text{safe}}(0, 0) = p > sp = D_{\text{safe}}(g_{\text{safe}}(0), 0)$ since $s < 1$. As $D_{\text{safe}}(T_{\text{safe}}, 0)$ is decreasing in T_{safe}, it follows that $g_{\text{safe}}(0) > 0$. Since the distribution of $W(\theta)$ is continuous, we have $\lim_{\Delta \to (\overline{\Delta})^-} D_{\text{safe}}(0, \Delta) = 0 < sp$. As $D_{\text{safe}}(T_{\text{safe}}, \Delta)$ is decreasing in T_{safe}, it follows that $g_{\text{safe}}(\Delta) < 0$ for all Δ sufficiently close to $\overline{\Delta}$.

The second part of the lemma is immediate from the first because we have that $\Lambda(T_{\text{risky}}) = \Lambda(T_{\text{safe}}) - \big(\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})\big)$. □

Conditioning on $\Lambda(T_{\text{safe}})$ and Δ clearing the safe organs (i.e. satisfying (2)), we can express the risky demand and the mass of agents that refuse risky organs as functions of Δ:

$$\text{RiskyDemand}_g(\Delta) := \text{RiskyDemand}(g_{\text{risky}}(\Delta), g_{\text{safe}}(\Delta))$$
$$\text{RefuseRisky}_g(\Delta) := \text{RefuseRisky}(g_{\text{risky}}(\Delta), g_{\text{safe}}(\Delta))$$

The continuity of g_{risky} and g_{safe} and the continuity of the distribution of W imply that RiskyDemand_g and RefuseRisky_g are continuous on $[0, \overline{\Delta})$. To prove Proposition 1, we show the existence of a steady-state value for Δ using a simple continuity argument.

Proof of Proposition 1. Let $g_{\text{risky}}, g_{\text{safe}}$ be as in Lemma 1. Taking $T_{\text{risky}} = g_{\text{risky}}(\Delta)$ and $T_{\text{safe}} = g_{\text{safe}}(\Delta)$, based on the conditions for steady-state, let

$$S = \big\{\Delta \in [0, \overline{\Delta}) \,\big|\, \text{RiskyDemand}_g(\Delta) \le rp,\ \text{RefuseRisky}_g(\Delta) \le \gamma, \text{ and } g_{\text{risky}}(\Delta) \ge 0\big\}.$$

Note that $\text{RiskyDemand}_g(0) = 0 < rp$ and $\text{RefuseRisky}_g(0) = 0 < \gamma$. Furthermore, since $g_{\text{safe}}(0) > 0$, we have that $g_{\text{risky}}(0) > 0$. Thus $0 \in S$, so S is non-empty.

As $\lim_{\Delta \to (\overline{\Delta})^-} g_{\text{safe}}(\Delta) < 0$ and $g_{\text{risky}}(\Delta) \leq g_{\text{safe}}(\Delta)$, the set S must be bounded away from $\overline{\Delta}$. Since the functions RiskyDemand_g, RefuseRisky_g, and g_{risky} are continuous (due to Lemma 1), S is closed. Therefore, S has a maximum $\Delta^* \in [0, \overline{\Delta})$.

As $\Delta^* + \kappa \notin S$ for all $\kappa > 0$, continuity also implies that one of the constraints in the definition of S must bind at $\Delta = \Delta^*$. By Lemma 1, it follows that $(g_{\text{risky}}(\Delta^*), g_{\text{safe}}(\Delta^*))$ gives a waiting-time steady-state. □

5 Congestion

In standard queuing systems, a waiting time for a good is positive only if the good is in excess demand—in particular, only if all units of the good are allocated. By contrast, due to the perishability of organs, the allocation system can enter a counterintuitive "congested" state in which the waiting time for risky organs is positive but nevertheless some risky organs are wasted (as will also be illustrated in Example 2).

We formally define congestion as follows.

Definition 2. A waiting-time steady state $(T_{\text{risky}}, T_{\text{safe}})$ is *congested* if $T_{\text{risky}} > 0$ and a positive mass of risky organs are discarded—i.e., (3) holds with strict inequality.

The following example exhibits a congested waiting-time steady state.

Example 1 (Congested Steady State)**.** Patients arrive at a rate of $p = 1$, safe organs arrive at a rate of $sp = 0.2$ and risky organs arrive at a rate of $rp = 0.3$. All patients depart at a time-invariant rate of $\beta = 0.2$. A patient's utility $\nu(\theta)$ for a risky organ is drawn from a standard uniform distribution $F_\nu \sim \text{Unif}[0, 1]$. Organs expire after being offered to measure $\gamma = 0.5$ of patients.

We claim that the waiting times $T_{\text{risky}} \approx 3.993$ and $T_{\text{safe}} \approx 6.020$ comprise a waiting-time steady state.[11] Given these waiting times, types whose value for risky organs satisfy $\nu(\theta) < \frac{2}{3}$ have willingness-to-wait $W(\theta) > T_{\text{safe}} - T_{\text{risky}} \approx 2.027$. As a result, the mass of allocated safe organs at each moment is $\frac{2}{3} \cdot e^{-\beta T_{\text{safe}}} = \frac{2}{3} \cdot \frac{3}{10} = sp$. The demand for risky organs at each moment is $\text{RiskyDemand}(T_{\text{risky}}, T_{\text{safe}}) = \frac{1}{3} \cdot e^{-\beta T_{\text{risky}}} = \frac{1}{3} \cdot \frac{9}{20} = 0.15 < rp$. The mass of patients that refuse a risky organ is

[11] Analytically, the waiting times are $T_{\text{risky}} = 5 \log \frac{20}{9}$ and $T_{\text{safe}} = 5 \log \frac{10}{3}$.

$\text{RefuseRisky}(T_{\text{risky}}, T_{\text{safe}}) = \int_{T_{\text{risky}}}^{T_{\text{safe}}} \frac{2}{3} e^{-\beta t} dt = \frac{2}{3} \cdot \frac{3}{4} = \gamma$. Hence, no organs are intentionally withheld.[12]

Note that despite the waiting time for risky organs being positive, half of the risky organs are not utilized because they expire before being offered to someone who would accept them. That is, not enough patients with $\nu(\theta) > \frac{2}{3}$ survive to be allocated a risky organ, and T_{risky} does not fall to clear the market due to organ expiry.

If instead organs did not expire ($\gamma = \infty$), then the waiting-time steady state would instead be given by $T_{\text{risky}} \approx 2.043$ and $T_{\text{safe}} \approx 5.045$,[13] and all organs would be utilized. Moreover, all patients would be better off, as the waiting times for both safe and risky kidneys would be lower.

To understand the impact of congestion, we first analyze the uncongested case and then compare other steady states to that case. If an uncongested steady state exists, it corresponds to a steady state of the limiting case of $\gamma = \infty$—a model we show has a unique steady state.

Proposition 2. *(a) For $\gamma = \infty$, there is a unique waiting-time steady state $(T^{\infty}_{\text{risky}}, T^{\infty}_{\text{safe}})$.*

(b) For any γ, a waiting-time steady state is uncongested if and only if it is $(T^{\infty}_{\text{risky}}, T^{\infty}_{\text{safe}})$.

It is not difficult to see that there exists $\gamma^\star < \infty$ (dependent on the other parameters of the model) such that $(T^{\infty}_{\text{risky}}, T^{\infty}_{\text{safe}})$ is a waiting-time steady state if and only if $\gamma \geq \gamma^\star$. Thus, congestion is inevitable in all steady states for γ sufficiently small—for larger γ, congestion may not arise, or may only arise if self-fulfilling.

To prove Proposition 2, we show that the difference in baseline cumulative hazards for waiting times in any steady state for $\gamma = \infty$ is an upper bound on the difference in baseline cumulative hazards for waiting times in any steady state for any γ. In the sequel, we use this property to study congested steady states.

Lemma 2. *If $(T_{\text{risky}}, T_{\text{safe}})$ is a waiting-time steady state for any γ, and $(T^{\infty}_{\text{risky}}, T^{\infty}_{\text{safe}})$ is a waiting-time steady state for $\gamma = \infty$, then $\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}) \leq \Lambda(T^{\infty}_{\text{safe}}) - \Lambda(T^{\infty}_{\text{risky}})$.*

[12] In Appendix A, we show the waiting-time steady state is unique if all patients share the same departure rate function that is nondecreasing over time (i.e., $\beta(\theta) \equiv \beta$ and $\lambda(t)$ is nondecreasing).

[13] Analytically, these waiting times are $T_{\text{risky}} = 5 \log \frac{40 - 10\sqrt{7}}{9}$ and $T_{\text{safe}} = 5 \log \frac{5\sqrt{7} - 5}{3}$. To see why this is a waiting-time steady state for $\gamma = \infty$, note that given these waiting times, types with $\nu(\theta) < \frac{\sqrt{7}-1}{3}$ have willingness-to-wait $W(\theta) > T_{\text{safe}} - T_{\text{risky}}$. As a result, at each moment, the mass of allocated safe organs is $\frac{\sqrt{7}-1}{3} \cdot e^{-\beta T_{\text{safe}}} = 0.2 = sp$, and the mass of allocated risky organs is $\left(1 - \frac{\sqrt{7}-1}{3}\right) \cdot e^{-\beta T_{\text{risky}}} = 0.3 = rp$.

The proof of Lemma 2 uses the following lemma, which establishes that demand for risky organs (conditional on waiting times clearing the market for safe organs) increases with $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$.

Lemma 3. *The function* $\text{RiskyDemand}_g(\Delta)$ *is nondecreasing in* Δ*, and strictly increasing in* Δ *on the region of* Δ *for which* $\text{RiskyDemand}_g(\Delta) > 0$*.*

Proof. Let $g_{\text{risky}}, g_{\text{safe}}$ be as in Lemma 1. Note $\text{RiskyDemand}(T_{\text{risky}}, \Lambda^{-1}(\Lambda(T_{\text{risky}}) + \Delta))$ is nonincreasing in T_{risky} and strictly decreasing in T_{risky} for $\Delta > \operatorname{ess\,inf} W(\theta)$ (as fewer agents survive at higher waiting times). Moreover, the same quantity is nondecreasing in Δ (as more types are willing to accept risky organs at higher Δ). As g_{risky} is decreasing (by Lemma 1), it follows that $\text{RiskyDemand}_g(\Delta)$ is nondecreasing in Δ, and strictly increasing in Δ for $\Delta > \operatorname{ess\,inf} W(\theta)$. The lemma follows as $\text{RiskyDemand}_g(\Delta) > 0$ if and only if $\Delta > \operatorname{ess\,inf} W(\theta)$. □

Proof of Lemma 2. Writing $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$ and $\Delta^\infty = \Lambda(T^\infty_{\text{safe}}) - \Lambda(T^\infty_{\text{risky}})$, we divide into cases based on the value of T^∞_{risky}.

- Case 1: $T^\infty_{\text{risky}} = 0$. Let g_{risky} be as in Lemma 1. By Lemma 1, we have that $g_{\text{risky}}(\Delta^\infty) = T^\infty_{\text{risky}} = 0$ and $g_{\text{risky}}(\Delta) = T_{\text{risky}} \geq 0$. Since g_{risky} is decreasing (by Lemma 1), it follows that $\Delta \leq \Delta^\infty$.

- Case 2: $T^\infty_{\text{risky}} > 0$. (3) must hold with equality at $(T^\infty_{\text{risky}}, T^\infty_{\text{safe}})$ by definition. By Lemma 1, we therefore have $\text{RiskyDemand}_g(\Delta^\infty) = pr$. Since $(T_{\text{risky}}, T_{\text{safe}})$ is a waiting-time steady state, (3) must hold at those waiting times. Lemma 1 then implies that $\text{RiskyDemand}_g(\Delta) \leq pr$. By Lemma 3, it follows that $\Delta \leq \Delta^\infty$.

As the cases exhaust all possibilities, we have proven the lemma. □

Proposition 2 follows straightforwardly from Lemma 2 by using the lemma to rank differences $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$ in baseline cumulative hazards of waiting times, which leads to rankings of waiting times $(T_{\text{risky}}, T_{\text{safe}})$ by Lemma 1.

Proof of Proposition 2. To prove Part (a), note that Proposition 1 provides existence. For uniqueness, let $(T_{\text{risky}}, T_{\text{safe}})$ and $(T^\infty_{\text{risky}}, T^\infty_{\text{safe}})$ be waiting-time steady states for $\gamma = \infty$. Lemma 2 implies that $\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}) \leq \Lambda(T^\infty_{\text{safe}}) - \Lambda(T^\infty_{\text{risky}})$. Exchanging the roles of the two steady states, Lemma 2 also implies that $\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}) \geq$

$\Lambda(T^{\infty}_{\text{safe}}) - \Lambda(T^{\infty}_{\text{risky}})$. Hence, we have $\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}) = \Lambda(T^{\infty}_{\text{safe}}) - \Lambda(T^{\infty}_{\text{risky}})$. By Lemma 1, we then have that $T_{\text{safe}} = T^{\infty}_{\text{safe}}$, and hence that $T_{\text{risky}} = T^{\infty}_{\text{risky}}$.

Now consider an arbitrary $\gamma > 0$. If $(T^{\infty}_{\text{risky}}, T^{\infty}_{\text{safe}})$ is a waiting-time steady state, since it is a waiting-time steady state for $\gamma = \infty$, (5) implies that either (3) holds with equality or $T^{\infty}_{\text{risky}} = 0$, so $(T^{\infty}_{\text{risky}}, T^{\infty}_{\text{safe}})$ is uncongested. Conversely, consider any uncongested waiting-time steady state $(T_{\text{risky}}, T_{\text{safe}})$. By construction, $(T_{\text{risky}}, T_{\text{safe}})$ is also a waiting-time steady state for $\gamma = \infty$. By Part (a), we then have that $(T_{\text{risky}}, T_{\text{safe}}) = (T^{\infty}_{\text{risky}}, T^{\infty}_{\text{safe}})$. □

We can also show that the waiting times for the $\gamma = \infty$ case are lower than the waiting times in any congested steady state. Clearly, by leading to the discard of risky organs, congestion harms those who would accept such organs. However, congestion also affects those who refuse risky organs: with a higher waiting time for risky organs, patients who are offered risky organs have a better chance of obtaining a safe organ, making congestion increase the demand for safe organs. As a result, congestion raises the waiting time for safe organs as well once patients' decisions are taken into account. Hence, congestion, if it arises, harms everyone.

The formal argument is based on showing that the gap $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$ between baseline cumulative hazards at waiting times in any congested waiting-time steady state is strictly below $\Lambda(T^{\infty}_{\text{safe}}) - \Lambda(T^{\infty}_{\text{risky}})$.

Proposition 3. *If $(T_{\text{risky}}, T_{\text{safe}})$ is a congested waiting-time steady-state, then we have that $\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}) < \Lambda(T^{\infty}_{\text{safe}}) - \Lambda(T^{\infty}_{\text{risky}})$, that $T_{\text{risky}} > T^{\infty}_{\text{risky}}$, and that $T_{\text{safe}} \geq T^{\infty}_{\text{safe}}$, with strict inequality if $W(\theta)$ has full support on $[0, \overline{\Delta})$.*

Proof. By Lemma 2, we have $\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}) \leq \Lambda(T^{\infty}_{\text{safe}}) - \Lambda(T^{\infty}_{\text{risky}})$. If equality held, then it would follow from Lemma 1 that $(T_{\text{risky}}, T_{\text{safe}}) = (T^{\infty}_{\text{risky}}, T^{\infty}_{\text{safe}})$, which is not congested (by Proposition 2(b)). Hence, the inequality must hold strictly. The remaining parts of the proposition then follow by Lemma 1. □

5.1 Multiplicity and Equilibrium Selection

The following example illustrates that the waiting-time steady state is not generally unique. In fact, we show that congestion can arise in one steady state but not another.

Example 2 (Multiple Steady States)**.** Patients are equally likely to be *healthy* (θ_h) and *unhealthy* (θ_u). Types have time-invariant departure rates $\beta(\theta_h) = 0.2$ and

$\beta(\theta_u) = 0.4$—reflecting that healthy patients have longer expected lifetimes. Patients' utility from risky organs is $\nu(\theta_h) = \nu(\theta_u) = 0.1$. Hence, types' willingness-to-wait for quality are $W(\theta_h) \approx 11.51$ and $W(\theta_u) \approx 5.76$.[14]

Patients arrive at rate $p = 1$, safe organs arrive at rate $sp = 0.15$, and risky organs arrive at rate $rp = 0.5$. Organs expire after refusal by mass $\gamma = 1.9$ of patients.

We identify two waiting-time steady states—one uncongested and one congested:

- *An Uncongested Steady State.* The waiting times are $T_{\text{risky}} = 0$ and $T_{\text{safe}} \approx 6.02$.[15] As $W(\theta_u) < T_{\text{safe}} - T_{\text{risky}} < W(\theta_h)$, healthy patients refuse risky organs at these waiting times, while unhealthy patients accept risky organs. As a result, the mass of allocated safe organs at each moment is $\frac{1}{2} \cdot e^{-\beta(\theta_h)T_{\text{safe}}} = \frac{1}{2} \cdot \frac{3}{10} = sp$. The demand for risky organs at each moment is $\text{RiskyDemand}(T_{\text{risky}}, T_{\text{safe}}) = \frac{1}{2}e^{-\beta(\theta_u)T_{\text{risky}}} = \frac{1}{2} = rp$. Moreover, the mass of patients that refuse a risky organ is $\text{RefuseRisky}(T_{\text{risky}}, T_{\text{safe}}) = \int_{T_{\text{risky}}}^{T_{\text{safe}}} \frac{1}{2}e^{-\beta(\theta_h)t}dt = \frac{1}{2} \cdot \frac{7}{2} = 1.75 < \gamma$. Hence, the cold-ischemia-time limit does not bind.

- *A Congested Steady State.* The waiting times are $T_{\text{risky}} \approx 1.44$ and $T_{\text{safe}} \approx 7.10$.[16] Since $T_{\text{safe}} - T_{\text{risky}} < W(\theta_u) < W(\theta_h)$, both patient types refuse risky organs at these waiting times. As a result, the mass of allocated safe organs at each moment is $\frac{1}{2} \cdot e^{-\beta(\theta_h)T_{\text{safe}}} + \frac{1}{2} \cdot e^{-\beta(\theta_u)T_{\text{safe}}} = sp$. Since no patients accept risky organs, risky organs are underdemanded. Moreover, the mass of patients that refuse a risky organ is $\text{RefuseRisky}(T_{\text{risky}}, T_{\text{safe}}) = 1.9 = \gamma$: the cold-ischemia-time limit binds. Thus, all risky organs expire due to excess cold-ischemia time.

The above example of multiple steady states illustrates an interesting *self-fulfilling* congestion phenomenon. Starting from the uncongested steady state, if unhealthy patients switch to choosing safe organs, T_{safe} increases by only about 1.08 because they depart much faster than healthy patients. However, the shift creates congestion and raises T_{risky} by 1.44—which is enough to make risky organs unattractive to unhealthy patients and thereby sustain a congested steady state. The effect of congestion is stark enough that all risky organs are discarded, whereas all risky organs would be utilized in the uncongested steady state.[17]

[14] While the distribution of $W(\theta)$ is not continuous here, it can be made continuous by perturbation without affecting the conclusions of our example.

[15] Analytically, the waiting times are $T_{\text{risky}} = 0$ and $T_{\text{safe}} = 5 \log \frac{10}{3}$.

[16] Analytically, the waiting times are $T_{\text{risky}} = -5 \log \frac{\sqrt{232+10\sqrt{55}}-10}{10}$ and $T_{\text{safe}} = 5 \log \frac{5+\sqrt{55}}{3}$.

[17] The example relies on the heterogeneity of patients' departure rates. Indeed, as we show in

More generally, while there is at most one uncongested waiting-time steady state (by Proposition 2(b)), there may be both congested and uncongested waiting-time steady states, as well as multiple congested waiting-time steady states. Nevertheless, we show that all waiting-time steady states can be ranked in a total order. Indeed, waiting-time steady states are clearly ordered by the difference in baseline cumulative hazards between the two waiting times—which, by Lemma 1, ranks waiting times.

Proposition 4. *All waiting-time steady states are totally ordered.*

Proof. By Lemma 1, given waiting-time steady states $(T_{\text{risky}}, T_{\text{safe}})$ and $(T'_{\text{risky}}, T'_{\text{safe}})$, if $\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}) \geq \Lambda(T'_{\text{safe}}) - \Lambda(T'_{\text{risky}})$, then $T_{\text{risky}} \leq T'_{\text{risky}}$ and $T_{\text{safe}} \leq T'_{\text{safe}}$. □

Throughout the rest of the paper, we focus on the waiting-time steady state with lowest waiting times, which exists due to Proposition 4 (since the set of waiting-time steady states is non-empty by Proposition 1 and closed). This waiting-time steady state is Pareto-optimal among all waiting-time steady states, as all types of patients prefer lower waiting times. This equilibrium-selection rule minimizes the scope of the inefficiencies that we highlight. Propositions 2 and 3 tell us that this selection rule always selects the uncongested waiting-time steady state if one exists.

5.2 Welfare and Congestion Effects of the Cold-Ischemia-Time Limit

We next show how changes in the cold-ischemia-time limit γ affect waiting times. Specifically, we show that starting from a congested lowest waiting-time steady state, raising γ lowers both waiting times—benefiting all patient types (and strictly so under a full support condition). Similarly, lowering γ raises both waiting times—hurting all patient types (and strictly so under a full support condition). This comparative static formalizes a sense in which relaxing the cold-ischemia-time limit alleviates congestion, and tightening it worsens congestion.

Proposition 5. *If the lowest waiting-time steady state is congested, then raising (resp. lowering) γ strictly decreases (resp. increases) T_{risky} and weakly decreases*

Appendix A, when all patients have the same departure-rate function that is nondecreasing over time, there is a unique waiting-time steady state.

(resp. increases) T_{safe} *under the lowest waiting-time steady state selection. If furthermore the distribution of* $W(\theta)$ *has full support on* $[0, \overline{\Delta})$*, then the decrease (resp. increase) in* T_{safe} *is strict.*

To prove Proposition 5, we characterize (in Lemma 4) the lowest waiting-time steady state in terms of the functions $g_{\text{risky}}, g_{\text{safe}}$ and the waiting-time steady state for $\gamma = \infty$.[18] We then use a monotone-comparative-statics argument based on an idea of Milgrom and Roberts (1994). (The details of our setting allow for a simple argument that does not rely on any results of Milgrom and Roberts (1994).)

Lemma 4. *The lowest waiting-time steady state* $(T_{\text{risky}}, T_{\text{safe}})$ *satisfies*

$$\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}) = \max\left\{\Delta \in \left[0, \Lambda(T^{\infty}_{\text{safe}}) - \Lambda(T^{\infty}_{\text{risky}})\right] \,\middle|\, \text{RefuseRisky}_g(\Delta) \leq \gamma\right\}. \quad (6)$$

Proof. Denote the right-hand side of (6) by Δ^*, which exists since it follows from Lemma 1 that $\text{RefuseRisky}_g(\Delta)$ is continuous in Δ. Write $\Delta^{\infty} = \Lambda(T^{\infty}_{\text{safe}}) - \Lambda(T^{\infty}_{\text{risky}})$.

If $\Delta^* = \Delta^{\infty}$, then (4) holds at $(T^{\infty}_{\text{risky}}, T^{\infty}_{\text{safe}})$ (due to Lemma 1), so $(T^{\infty}_{\text{risky}}, T^{\infty}_{\text{safe}})$ is a waiting-time steady state. By Propositions 2(b) and 3, it is the lowest one.

Hence, we can restrict attention to the case in which $\Delta^* < \Delta^{\infty}$. Letting $g_{\text{risky}}, g_{\text{safe}}$ be as in Lemma 1, we first show $(g_{\text{risky}}(\Delta^*), g_{\text{safe}}(\Delta^*))$ is a waiting-time steady state. It follows from Lemmata 1 and 3 that

$$rp \geq \text{RiskyDemand}_g(\Delta^{\infty}) \geq \text{RiskyDemand}_g(\Delta^*)$$

—i.e., that (3) holds. By Lemma 1, we also have $g_{\text{risky}}(\Delta^*) > g_{\text{risky}}(\Delta^{\infty}) \geq 0$. Last, the continuity of g_{risky} and g_{safe} (by Lemma 1) and the continuity of the distribution of $W(\theta)$ imply the continuity of $\text{RefuseRisky}_g(\Delta)$, so (4) must hold with equality.

Letting $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$, we have that $(T_{\text{risky}}, T_{\text{safe}}) = (g_{\text{risky}}(\Delta), g_{\text{safe}}(\Delta))$ by Lemma 1. (4) then implies that $\text{RefuseRisky}_g(\Delta) \leq \gamma$, and Lemma 2 implies that $\Delta \leq \Delta^{\infty}$. Hence, we have that $\Delta \leq \Delta^*$. By Lemma 1, it follows that $(T_{\text{risky}}, T_{\text{safe}}) \geq (g_{\text{risky}}(\Delta^*), g_{\text{safe}}(\Delta^*))$.

Since $(T_{\text{risky}}, T_{\text{safe}})$ is the lowest waiting-time steady state, we then have that $(T_{\text{risky}}, T_{\text{safe}}) = (g_{\text{risky}}(\Delta^*), g_{\text{safe}}(\Delta^*))$, so $\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}) = \Delta^*$. □

[18] A characterization of all waiting-time steady states is available along similar lines. Suppose that a pair of waiting times $(T_{\text{risky}}, T_{\text{safe}})$ satisfies (2). The proof of Lemma 4 shows that $(T_{\text{risky}}, T_{\text{safe}})$ is a waiting-time steady state if and only if we have that $\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}) \leq \Lambda(T^{\infty}_{\text{safe}}) - \Lambda(T^{\infty}_{\text{risky}})$ and that $\text{RefuseRisky}_g(\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})) \leq \gamma$, and one of these two inequalities holds with equality.

To complete the proof of Proposition 5, we simply note that raising (resp. lowering) γ expands (resp. shrinks) the set on the right-hand side of (6), which decreases (resp. increases) waiting times. A more careful argument establishes the strictness of the decrease (resp. increase) under the full-support condition.

Proof of Proposition 5. Let the lowest waiting-time steady state be $(T_{\text{risky}}, T_{\text{safe}})$, which we assume to be congested. Let $\gamma' \neq \gamma$, and consider the new lowest waiting-time steady state $(T'_{\text{risky}}, T'_{\text{safe}})$. Let $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$, let $\Delta' = \Lambda(T'_{\text{safe}}) - \Lambda(T'_{\text{risky}})$, and let $\Delta^\infty = \Lambda(T^\infty_{\text{safe}}) - \Lambda(T^\infty_{\text{risky}})$. As $(T_{\text{risky}}, T_{\text{safe}})$ is assumed congested, Proposition 3 implies that $\Delta < \Delta^\infty$.

First, suppose that $\gamma' > \gamma$. By Lemma 4, we have $\text{RefuseRisky}_g(\Delta) \leq \gamma < \gamma'$. By the continuity of $\text{RefuseRisky}_g(\Delta)$ (which follows from Lemma 1 and the continuity of the distribution of $W(\theta)$) and Lemma 4, we then have that $\Delta' > \Delta$.[19] The conclusion of the proposition follows by Lemma 1.

Now suppose that $\gamma' < \gamma$. By Lemma 4, we have $\text{RefuseRisky}_g(\Delta') \leq \gamma' < \gamma$. Since $\Delta < \Delta^\infty$, Lemma 4 implies that $\text{RefuseRisky}_g(\Delta^\infty) > \gamma$, and hence we have that $\Delta' < \Delta^\infty$. By the continuity of RefuseRisky_g and Lemma 4, we then have that $\Delta' < \Delta$. The conclusion of the proposition follows by Lemma 1. □

6 Constrained Inefficiency under Congestion

Intuitively, increasing the pool of safe organs should improve welfare. If not, then the equilibrium is constrained inefficient in that discarding safe organs might improve welfare even without changing the underlying queuing structure. In this section, we show that this form of constrained inefficiency can arise if the steady state is congested, but (as is intuitive) not if the steady state is uncongested.

We first show that without congestion or cold-ischemia-time limits, "artificially discarding" safe kidneys can never lower waiting times.

Proposition 6. *Lowering s strictly increases T^∞_{safe} and weakly increases T^∞_{risky}.*

Proof. Consider $s > s'$, and let $(T^\infty_{\text{risky}}, T^\infty_{\text{safe}})$ and $(T^{\infty,\prime}_{\text{risky}}, T^{\infty,\prime}_{\text{safe}})$ denote the corresponding waiting-time steady states for $\gamma = \infty$. Let $\Delta = \Lambda(T^\infty_{\text{safe}}) - \Lambda(T^\infty_{\text{risky}})$ and let $\Delta' = \Lambda(T^{\infty,\prime}_{\text{safe}}) - \Lambda(T^{\infty,\prime}_{\text{risky}})$.

[19] Milgrom and Roberts (1994, Theorem 1) provides conditions to obtain a weak inequality that apply under weaker continuity conditions that allow for jumps in one direction. We obtain a strict inequality using continuity and interiority.

We do casework based on the relative values of Δ and Δ' to complete the proof.

- Case 1: $\Delta' \leq \Delta$. Due to (2), as $s' < s$, we must then have that $T^{\infty,\prime}_{\text{safe}} > T^{\infty}_{\text{safe}}$. It follows that $\Lambda(T^{\infty,\prime}_{\text{risky}}) = \Lambda(T^{\infty,\prime}_{\text{safe}}) - \Delta' > \Lambda(T^{\infty}_{\text{safe}}) - \Delta = \Lambda(T^{\infty}_{\text{risky}})$, and hence that $T^{\infty,\prime}_{\text{risky}} > T^{\infty}_{\text{risky}}$.

- Case 2: $\Delta' > \Delta$. Due to (3) and (5), and since the arrival rate of risky organs is unchanged, we must have that $T^{\infty,\prime}_{\text{risky}} \geq T^{\infty}_{\text{risky}}$. It follows that $\Lambda(T^{\infty,\prime}_{\text{safe}}) = \Lambda(T^{\infty,\prime}_{\text{risky}}) + \Delta' > \Lambda(T^{\infty}_{\text{risky}}) + \Delta = \Lambda(T^{\infty}_{\text{safe}})$, and hence that $T^{\infty,\prime}_{\text{safe}} > T^{\infty}_{\text{safe}}$.

The cases exhaust all possibilities, which completes the proof. □

The following example shows that when the state is congested, artificially discarding safe organs may result in a Pareto improvement by decreasing both steady state waiting times. (Equivalently, the arrival of more safe organs can harm all patients by increasing both steady state waiting times.) This can happen even when there is a unique waiting-time steady state.

Example 3 (Pareto improvement when artificially discarding safe organs under congestion). Patients arrive at a rate of $p = 1$, safe organs arrive at a rate of $sp = 0.05$ and risky organs arrive at a rate of $rp = 0.8$. All patients depart at a constant rate of $\beta = 0.2$. A patient's utility $\nu(\theta)$ for a risky organ is drawn from the distribution $F_\nu \sim \text{Beta}(2, 1)$, whose cdf is $F_\nu(x) = x^2$. Organs expire after being offered to mass $\gamma = 0.25$ of patients.

We claim that $T_{\text{risky}} \approx 4.581$ and $T_{\text{safe}} \approx 8.047$ comprise a waiting-time steady state.[20] Given these waiting times, types whose values for risky organs satisfy $\nu(\theta) < \frac{1}{2}$ have willingness-to-wait $W(\theta) > T_{\text{safe}} - T_{\text{risky}} \approx 3.466$—such types comprise $\frac{1}{4}$ fraction of types according to the distribution F_ν. Hence, the mass of safe organs allocated at each moment is $\frac{1}{4} \cdot e^{-\beta T_{\text{safe}}} = \frac{1}{4} \cdot \frac{1}{5} = sp$. The demand for risky organs at each moment is $\frac{3}{4} \cdot e^{-\beta T_{\text{risky}}} = \frac{3}{4} \cdot \frac{2}{5} = 0.3 < rp$. The mass of patients that refuse a risky organ is $\text{RefuseRisky}(T_{\text{risky}}, T_{\text{safe}}) = \int_{T_{\text{risky}}}^{T_{\text{safe}}} \frac{1}{4} e^{-\beta t} dt = 0.25$. Hence, no organs are intentionally withheld.

Now suppose that the arriving mass of safe organs falls to $s'p = 0.02$. We claim that $T_{\text{risky}} \approx 0.769$ and $T_{\text{safe}} \approx 7.032$ comprise a waiting-time steady state.[21] Given these waiting times, types whose values for risky organs satisfy $\nu(\theta) < \frac{2}{7}$ have

[20] Analytically, the waiting times are $T_{\text{risky}} = 5 \log \frac{5}{2}$ and $T_{\text{safe}} = 5 \log 5$.
[21] Analytically, the waiting times are $T_{\text{risky}} = 5 \log \frac{400}{343}$ and $T_{\text{safe}} = 5 \log \frac{200}{49}$.

willingness-to-wait $W(\theta) > T_{\text{safe}} - T_{\text{risky}} \approx 6.264$—such types comprise $\frac{4}{49}$ fraction of types according to the distribution F_ν. As a result, the mass of allocated safe organs at each moment is $\frac{4}{49} \cdot e^{-\beta T_{\text{safe}}} = \frac{4}{49} \cdot \frac{49}{200} = s'p$. The demand for risky organs at each moment is $\text{RiskyDemand}(T_{\text{risky}}, T_{\text{safe}}) = \frac{45}{49} \cdot e^{-\beta T_{\text{risky}}} = \frac{45}{49} \cdot \frac{343}{400} < 0.79 < rp$. The mass of patients that refuse a risky organ is $\text{RefuseRisky}(T_{\text{risky}}, T_{\text{safe}}) = \int_{T_{\text{risky}}}^{T_{\text{safe}}} \frac{4}{49} e^{-\beta t} dt = 0.25$. Hence, no organs are intentionally withheld.

As we demonstrate in Appendix A, the homogeneity and time-invariance of departure rates across patients imply that the waiting-time steady state is unique. Therefore, lowering the safe arrival rate from s to s' lowers the (unique) steady-state waiting times for both risky and safe organs under congestion.

For intuition, when the system is congested, an increase in the supply of safe organs results in more types waiting for safe organs. This worsens congestion, and the example shows it can exacerbate congestion so severely as to increase waiting times for both safe and risky organs.

While our example is stylized, it illustrates the severity of the congestion externality. The externality can be so severe that a benevolent social planner who is forced to use a queuing system might, in principle, prefer to destroy safe organs. This result shows that the queuing system is constrained inefficient under congestion, in that the inefficiency could be mitigated by taking a feasible action without changing the underlying structure of the queuing system. In the remainder of the paper, we show how the congestion externality is relevant to other features of the market and to understanding the welfare effects of alternative waiting list policies.

7 Delegated Decisions: A Congestion Externality

In practice, organ-acceptance decisions are delegated to the patient's transplant center, which may have less of an incentive to transplant risky organs. To allow for misaligned incentives in this delegation problem, we suppose that transplant centers make organ-acceptance decisions based instead on the (Bernoulli) utility function

$$u_{\text{transplant center}} = \begin{cases} 1 & \text{if get safe organ} \\ \widehat{\nu}(\theta) & \text{if get risky organ} \\ 0 & \text{if do not get an organ} \end{cases},$$

where $\widehat{\nu}(\theta) \in [0,1]$ is the transplant center's effective willingness to accept a risky organ on behalf of a patient of type θ.

A similar definition to (1) shows that the transplant center's willingness-to-wait for quality for a patient of type θ is

$$\widehat{W}(\theta) = \frac{-\log \widehat{\nu}(\theta)}{\beta(\theta)}.$$

Throughout, we assume that the distribution of $\widehat{W}(\theta)$ is also continuous.

As transplant centers are penalized for adverse post-transplant outcomes but not deaths on the waiting list (Chan and Roth, 2024; Mildebrath et al., 2024), we assume that transplant centers are overall less willing to accept risky organs than patients would be themselves (Mehrotra et al., 2022; Sepulveda et al., 2023): i.e., that $\widehat{W}$ first-order stochastically dominates $W(\theta)$. We then show that if the steady state with patient decisions is congested, decisions being made by transplant centers weakly raises both steady-state waiting times—and strictly raises them if there exists $\varepsilon > 0$ such that $\widehat{W}$ first-order stochastically dominates $W(\theta) + \varepsilon$. Intuitively, transplant centers' greater reluctance to accept risky organs worsens congestion.

We introduce an additional assumption on each of ν and $\widehat{\nu}$ under which we can demonstrate the comparative static.

Definition 3. We say that W satisfies *the healthy-more-willing-to-wait condition* if for all $\theta, \theta' \in \Theta$, if $\beta(\theta) < \beta(\theta')$, then $W(\theta) \geq W(\theta')$.

Consider for example a market with older patients, as well as younger and healthier patients who have lower departure rates at each waiting time—and hence lower baseline scaling factors β—than the older patients. The healthy-more-willing-to-wait condition would then require that the younger and healthier patients all have a weakly higher willingness-to-wait than all the older patients. We impose this condition on both patients' and transplant centers' preferences.

Proposition 7. *Suppose that $\widehat{W}(\theta)$ first-order stochastically dominates $W(\theta)$, and that both W and $\widehat{W}$ satisfy the healthy-more-willing-to-wait condition. Suppose furthermore that $W(\theta)$ has full support on $[0, \overline{\Delta})$.*

(a) Moving from patient to transplant-center decisions weakly increases T_{safe} under the lowest waiting-time steady state selection.

(b) *If the lowest waiting-time steady state is congested under patient decisions, then moving from patient to transplant-center decisions weakly increases* T_{risky} *under the lowest waiting-time steady state selection and the lowest waiting-time steady state must also be congested under transplant-center decisions.*

If there exists $\varepsilon > 0$ *such that* $\widehat{W}(\theta)$ *first-order stochastically dominates* $W(\theta) + \varepsilon$*, then both increases in waiting times are strict.*

The proof is in Appendix B and also uses monotone-comparative-statics ideas.

For intuition, note that transplant centers' more conservative acceptance decisions shift demand toward safe organs, and therefore increase T_{safe}. Without congestion, the corresponding decline in demand for risky organs would lower T_{risky}.

By contrast, under congestion, the additional refusals consume the organ's limited placement window and prevent risky organs from reaching willing recipients. This congestion externality reverses the standard demand effect and raises T_{risky} as well. Thus, it is through the congestion externality that misalignment between patients and transplant centers exacerbates the inefficiencies of the waiting list design.

In particular, if there would be congestion under patient decisions, then decisions being made by transplant centers harms all patients—including patients whose transplant centers make patient-aligned decisions. Indeed, in addition to transplant centers making decisions that may be suboptimal given the waiting times, their greater reluctance to accept risky organs worsens congestion enough to raise waiting times.

8 Market Design Implications

Our framework and findings can help in understanding how market design innovations in kidney allocation can affect congestion and, in turn, welfare.

8.1 Market Thickness

In March 2021, the Organ Procurement and Transplantation Network increased the size of the regions in which patients are given a (local) geographic priority.[22] The non-utilization rate increased by approximately 3 percentage points. Motivated by this change, we derive comparative statics with respect to the market thickness p.

[22]Instead of a small fixed service area for each organ procurement organization under previous policy, patients within 250 NM from the donor hospital are given priority.

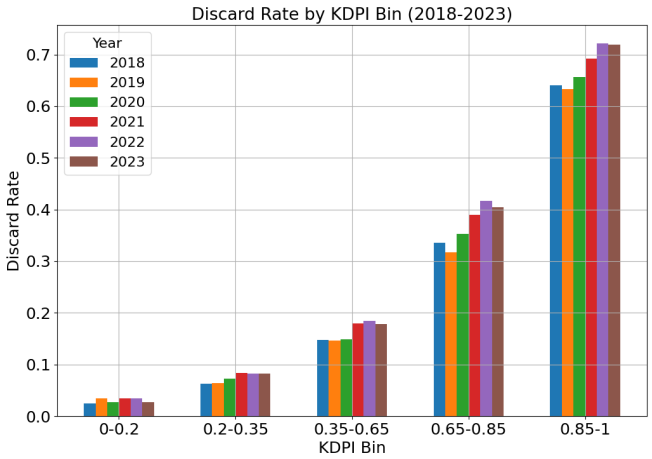

Figure 5: Annual trends of non-utilization rate by KDPI bin.

Increasing p scales up the flows of patients and organs but leaves the mass γ of offers that can be processed before expiry unchanged. Mathematically, this is equivalent to changing γ in the opposite direction, so Proposition 5 provides comparative statics. For simplicity, we focus on the practically relevant case of increasing thickness.

Proposition 8. *If the lowest waiting-time steady state is congested, then raising p strictly increases T_{risky} and weakly increases T_{safe} under the lowest waiting-time steady state selection. If furthermore the distribution of $W(\theta)$ has full support on $[0, \overline{\Delta})$, then the increase in T_{safe} is strict.*

Proof. The steady-state conditions (2) and (3) are homogeneous of degree 1 in p on both sides. As the left-hand side of (4) is homogeneous of degree 1 in p and the right-hand side is γ, changing p to p' is equivalent to changing γ to $\gamma' = p\gamma/p'$. The result then follows from Proposition 5. □

Consistent with Proposition 8, following the modification of local prioritization in March 2021, the non-utilization rate increased especially for low-quality organs (i.e., organs with high KDPI) as depicted in Figure 5.

Note that market thickness may still benefit patients when preferences exhibit substantial horizontal heterogeneity, for example due to compatibility differences (Agarwal et al., 2021). Our fluid formulation abstracts from these forms of heterogeneity to highlight the role of congestion on waiting lists.

8.2 Offering Processes

In the U.S., a deceased-donor kidney is offered sequentially to patients based on a fixed priority order. While high-quality kidneys are accepted quickly, kidneys with marginal quality are refused by many patients (see Figure 3). Guan et al. (2025) find that refusals due to long cold-ischemia time increase substantially 4 hours after recovery, and "actual or projected cold-ischemia time too long" becomes the main refusal reason 10 hours after recovery (see also Figure 4).[23] The large non-utilization rate has led organ procurement organizations to expedite the offering process: in 2023, more than 20% of accepted organs were placed out-of-sequence (Liyanage et al., 2025).

It is straightforward to see that in partial equilibrium (i.e., not taking into account effects on patients' decisions and equilibrium waiting times), expediting raises welfare by allocating risky organs to patients who would otherwise not receive an organ. However, expediting many organs will also have general equilibrium effects due to patients' decisions and equilibrium waiting times. Our model allows us to analyze these broader impacts.

Consider the following modification of the first-come-first-serve offering policy, denoted by Π: Safe organs are offered based on first-come-first-serve, and risky organs are offered in sequence only to patients who have not yet refused a risky organ. Under this policy, risky organs bypass patients whose earlier decisions reveal that they are unlikely to accept, while priority is preserved among the remaining patients.[24] Observe that this policy achieves the same waiting times as in the case of $\gamma = \infty$. Since there is no uncertainty about when a patient will receive an offer of safe or risky kidney in a fluid model, their decisions will be based solely on (steady-state) waiting times. As a result, the following result is immediate.

Proposition 9. *For any $\gamma > 0$, the offering policy Π achieves the same welfare as the first-come-first-serve policy achieves for $\gamma = \infty$.*

Evidence from other markets supports this intuition.[25] In the U.K., organs that are refused more than three times are offered simultaneously to all transplant centers, and the organ is given to the accepting patient whose priority is highest (Callaghan et al., 2017). In the U.S., organ procurement organizations have increasingly adopted

[23] Guan et al. (2025) also find that 4 hours after recovery, an organ has been offered to patients from only 11 transplant centers (out of 220).

[24] This is similar to a buffer queue as studied in Leshno (2022).

[25] Expediting unattractive options increases utilization in ride-sharing (Castro et al., 2021).

ad-hoc expedited placement practices by making open offers to transplant centers (Jadlowiec et al., 2023). These practices, however, have raised equity concerns (such as "list skipping"), prompting the Health Resources and Services Administration to direct the Organ Procurement and Transplantation Network to develop a standardized policy for expedited placements.

A complementary approach to reducing congestion is to elicit more refined preferences from physicians and patients, and thereby avoid unnecessarily making offers that are likely to be refused. Although current systems allow transplant centers to filter certain types of offers, refusal rates remain high (Guan et al., 2025).

9 Conclusion

This paper shows that the perishability of organs—driven by the undesirability of organs with long cold-ischemia times—can generate congestion on deceased-donor waiting lists and thereby lead to substantial inefficiencies and externalities. When the waiting list is congested, many risky organs are not utilized because they fail to reach patients who are willing to accept them quickly enough.

Congestion is also exacerbated by misaligned incentives between patients and transplant centers in acceptance decisions. One way to mitigate these inefficiencies is to involve patients more directly in refusal decisions (Mohan and Husain, 2022). Another way is to avoid offers that are likely to be refused. Patients who recently joined the waiting list may be more willing to accept lower-quality organs (Mehrotra et al., 2022; Sepulveda et al., 2023). This opens the door to mechanisms such as dedicated lists for risky organs—e.g., requiring explicit prospective consent from the transplant center to receive offers with KDPI above 85%. Another option is to relax transplant centers' disincentives to accept higher-risk organs.

Our findings have broader implications for organ allocation policy. When the waitlist is congested, thickening the market by expanding service regions may exacerbate congestion. A natural approach to alleviate congestion and increase utilization of low-quality organs is to expedite the offering processes for them.[26]

[26]See Ashlagi and Roth (2026) for a discussion of expediting organ placements.

A Uniqueness with Homogeneous Departure Rates

When all patient types share the same departure-rate function that is nondecreasing over time, we can establish the uniqueness of waiting-time steady state. This holds, for example, if all types share the same constant departure rate (as in Examples 1 and 3).

Proposition 10. *If $\beta(\theta) \equiv \beta$ and the baseline hazard rate $\lambda(t)$ is nondecreasing in t, then the waiting-time steady state is unique.*

The key to the proof of Proposition 10 is the following lemma.

Lemma 5. *Under the hypotheses of Proposition 10,* $\text{RefuseRisky}_g(\Delta)$ *is strictly increasing on* $\{\Delta \mid g_{\text{risky}}(\Delta) \geq 0\}$.

Proof. When $\beta(\theta) \equiv \beta$, we have that

$$\begin{aligned}\text{RefuseRisky}_g(\Delta) &= \int_{g_{\text{risky}}(\Delta)}^{g_{\text{safe}}(\Delta)} pe^{-\Lambda(t)\beta} \Pr{}_\theta[W(\theta) > \Delta] dt \\ &= pe^{-\Lambda(g_{\text{safe}}(\Delta))\beta} \Pr{}_\theta[W(\theta) > \Delta] \int_{g_{\text{risky}}(\Delta)}^{g_{\text{safe}}(\Delta)} e^{(\Lambda(g_{\text{safe}}(\Delta))-\Lambda(t))\beta} dt.\end{aligned}$$

As $\text{SafeDemand}(T_{\text{risky}}, T_{\text{safe}}) = pe^{-\Lambda(T_{\text{safe}})\beta} \Pr_\theta[W(\theta) > \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})]$ when $\beta(\theta) \equiv \beta$, we have that $pe^{-\Lambda(g_{\text{safe}}(\Delta))\beta} \Pr_\theta[W(\theta) > \Delta] = sp$ by the definition of g_{safe}. By the definition of g_{risky}, it follows that

$$\begin{aligned}\text{RefuseRisky}_g(\Delta) &= sp \int_{\Lambda^{-1}(\Lambda(g_{\text{safe}}(\Delta))-\Delta)}^{g_{\text{safe}}(\Delta)} e^{(\Lambda(g_{\text{safe}}(\Delta))-\Lambda(t))\beta} dt \\ &= sp \int_{-\Delta}^{0} \frac{e^{-\beta x}}{\lambda\big(\Lambda^{-1}\big(x + \Lambda(g_{\text{safe}}(\Delta))\big)\big)} dx,\end{aligned}$$

where the second equality is by substituting $x = \Lambda(t) - \Lambda(g_{\text{safe}}(\Delta))$. The integrand is positive, and nondecreasing in Δ for each x since $\lambda(t)$ is nondecreasing (by hypothesis) and g_{safe} is nonincreasing (by Lemma 1). The lemma follows. □

Proof of Proposition 10. Suppose for the sake of deriving a contradiction that there are two distinct waiting-time steady states $(T_{\text{risky}}, T_{\text{safe}})$ and $(T'_{\text{risky}}, T'_{\text{safe}})$. Letting $\Delta = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$ and $\Delta' = \Lambda(T'_{\text{safe}}) - \Lambda(T'_{\text{risky}})$. By Lemma 1, we have that $(T_{\text{risky}}, T_{\text{safe}}) = (g_{\text{risky}}(\Delta), g_{\text{safe}}(\Delta))$ and $(T'_{\text{risky}}, T'_{\text{safe}}) = (g_{\text{risky}}(\Delta'), g_{\text{safe}}(\Delta'))$. In particular, we have that $\Delta \neq \Delta'$. Without loss of generality, assume that $\Delta' < \Delta$.

(3) and (4) for $(T_{\text{risky}}, T_{\text{safe}}) = (g_{\text{risky}}(\Delta), g_{\text{safe}}(\Delta))$ and $T_{\text{risky}} \geq 0$ imply that $\text{RiskyDemand}_g(\Delta) \leq rp$, that $\text{RefuseRisky}_g(\Delta) \leq \gamma$, and that $g_{\text{risky}}(\Delta) = T_{\text{risky}} \geq 0$, respectively. By Lemmata 3 and 5, we then have that $\text{RiskyDemand}_g(\Delta') < rp$ and that $\text{RefuseRisky}_g(\Delta') < \gamma$. By Lemma 1, we have that $g_{\text{risky}}(\Delta') > 0$. These inequalities contradict (5) for $(T'_{\text{risky}}, T'_{\text{safe}}) = (g_{\text{risky}}(\Delta'), g_{\text{safe}}(\Delta'))$. □

B Proof of Proposition 7

Let F_W and $F_{\widehat{W}}$ denote the cdfs of $W(\theta)$ and $\widehat{W}(\theta)$, respectively. For $\alpha \in [0, 1)$, define their lower quantile functions by

$$q_-(\alpha) = \inf\{\Delta \geq 0 \mid F_W(\Delta) \geq \alpha\},$$
$$\widehat{q}_-(\alpha) = \inf\{\Delta \geq 0 \mid F_{\widehat{W}}(\Delta) \geq \alpha\}.$$

We first characterize first-order stochastic dominance in terms of quantiles.

Lemma 6. *The random variable $\widehat{W}(\theta)$ first-order stochastically dominates $W(\theta)$ if and only if*

$$\widehat{q}_-(\alpha) \geq q_-(\alpha) \qquad \text{for every } \alpha \in (0, 1).$$

Moreover, if there exists $\varepsilon > 0$ such that $\widehat{W}(\theta)$ first-order stochastically dominates $W(\theta) + \varepsilon$, then

$$\widehat{q}_-(\alpha) \geq q_-(\alpha) + \varepsilon \qquad \text{for every } \alpha \in (0, 1).$$

Proof. Suppose first that $F_{\widehat{W}}(x) \leq F_W(x)$ for all x. Fix $\alpha \in (0, 1)$ and $x < q_-(\alpha)$. Then $F_W(x) < \alpha$, so $F_{\widehat{W}}(x) < \alpha$ as well. Hence $x < \widehat{q}_-(\alpha)$. Taking the supremum over all $x < q_-(\alpha)$ gives that $\widehat{q}_-(\alpha) \geq q_-(\alpha)$.

Conversely, suppose that $\widehat{q}_-(\alpha) \geq q_-(\alpha)$ for all quantiles $\alpha \in (0, 1)$. Fix $x \geq 0$ and $\alpha < F_{\widehat{W}}(x)$. By the definition of the lower quantile, we have that $\widehat{q}_-(\alpha) \leq x$. Hence we have that $q_-(\alpha) \leq x$, which implies that $F_W(x) \geq \alpha$. Letting $\alpha \to F_{\widehat{W}}(x)^-$ gives that $F_W(x) \geq F_{\widehat{W}}(x)$. Thus, $\widehat{W}(\theta)$ first-order stochastically dominates $W(\theta)$.

For the final claim, apply the first part to $\widehat{W}(\theta)$ and $W(\theta) + \varepsilon$, and use the fact that the lower quantile function for $W(\theta) + \varepsilon$ equals $q_-(\alpha) + \varepsilon$. □

To prove Proposition 7, we apply monotone-comparative-statics ideas based on a parameterization that is invariant across the change in decision makers. Rather than

considering the difference in baseline cumulative hazards across waiting times, we consider the fraction α of patients who would accept a risky organ if offered at time T_{risky}. Let F_β denote the cdf of baseline scaling factors $\beta(\theta)$. For $t \in \mathbb{R}$ and $\alpha \in [0, 1)$, define

$$\tilde{D}_{\text{safe}}(t, \alpha) := p \int_0^\infty e^{-\Lambda(t)x} \, d \min\{F_\beta(x), 1 - \alpha\}, \tag{B.1}$$

$$\tilde{D}_{\text{risky}}(t, \alpha) := p \int_0^\infty e^{-\Lambda(t)x} \, d\Big(F_\beta(x) - \min\{F_\beta(x), 1 - \alpha\}\Big). \tag{B.2}$$

Thus, $\tilde{D}_{\text{safe}}(t, \alpha)$ is the mass at position t among the fraction $1 - \alpha$ of patients with lowest baseline scaling factors, while $\tilde{D}_{\text{risky}}(t, \alpha)$ is the mass at position t among the α fraction of patients with highest baseline scaling factors. Note that $\tilde{D}_{\text{safe}}$ and $\tilde{D}_{\text{risky}}$ are continuous in (t, α).

We next show the following more technical claim. Intuitively, this claim shows that $\tilde{D}_{\text{safe}}(T_{\text{safe}}, \alpha)$ (resp. $\tilde{D}_{\text{risky}}(T_{\text{risky}}, \alpha)$) gives the demand for safe (resp. risky) organs when the types with the lowest $1 - \alpha$ fraction of baseline scaling factors hold out for safe organs.

Claim 1. *Suppose that W satisfies the healthy-more-willing-to-wait condition. Fix $c \geq 0$ with $F_W(c) < 1$, and let $F_W(c) = \alpha$. Then, for every $t \geq 0$, we have that*

$$p \int_{\theta : W(\theta) > c} e^{-\Lambda(t)\beta(\theta)} f(\theta) d\theta = \tilde{D}_{\text{safe}}(t, \alpha),$$
$$p \int_{\theta : W(\theta) < c} e^{-\Lambda(t)\beta(\theta)} f(\theta) d\theta = \tilde{D}_{\text{risky}}(t, \alpha),$$

where $\tilde{D}_{\text{safe}}(t, \alpha)$ and $\tilde{D}_{\text{risky}}(t, \alpha)$ are as defined in (B.1) *and* (B.2)*, respectively.*

Proof. Since $W(\theta)$ has a continuous distribution, the domains of integration have masses $1 - \alpha$ and α, respectively, under the measure $f(\theta)d\theta$. Moreover, for each x, the sets $\{\theta \mid W(\theta) > c\}$ and $\{\theta \mid \beta(\theta) \leq x\}$ are nested: if they were not, there would exist $\theta \in \Theta$ with $W(\theta) > c$ and $\beta(\theta) > x$ and $\theta' \in \Theta$ with $W(\theta') \leq c$ and $\beta(\theta') \leq x$, which would contradict the healthy-more-willing-to-wait condition. Consequently, for every $x \geq 0$, we have that

$$\Pr\big[W(\theta) > c \text{ and } \beta(\theta) \leq x\big] = \min\{F_\beta(x), 1 - \alpha\}.$$

Hence, the cdf of departure rates restricted to the Borel set $\{\theta \mid W(\theta) > c\}$ is $\min\{F_\beta, 1-\alpha\}$, while the cdf on the complementary Borel set $\{\theta \mid W(\theta) < c\}$ is $F_\beta - \min\{F_\beta, 1-\alpha\}$. The claim follows by change of variables. □

The following lemma provides an analogue of Lemma 1.

Lemma 7. *(a) There is a strictly decreasing, continuous function $\tilde{g}_{\text{safe}} : [0,1) \to \mathbb{R}$ independent of $\nu(\theta)$ (but dependent on s, $f(\theta)$, $\lambda(t)$, and $\beta(\theta)$) with $\tilde{g}_{\text{safe}}(1-s) = 0$ such that if W satisfies the healthy-more-willing-to-wait condition, then a pair $(T_{\text{risky}}, T_{\text{safe}})$ of waiting times satisfies (2) if and only if, writing*

$$\alpha = F_W\big(\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})\big),$$

we have that $T_{\text{safe}} = \tilde{g}_{\text{safe}}(\alpha)$.

(b) Suppose that W satisfies the healthy-more-willing-to-wait condition. If $(T_{\text{risky}}, T_{\text{safe}})$ satisfies (2), then writing $\alpha = F_W(\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}}))$, we have that $T_{\text{risky}} \leq \Lambda^{-1}\big(\Lambda(\tilde{g}_{\text{safe}}(\alpha)) - q_-(\alpha)\big)$, with equality if W has full support on $[0, \overline{\Delta})$.

Proof. Let $\tilde{D}_{\text{safe}}$ be defined as in (B.1). For each $\alpha \in [0,1)$, the quantity $\tilde{D}_{\text{safe}}(T_{\text{safe}}, \alpha)$ is continuous and strictly decreasing in T_{safe}, with $\lim_{T_{\text{safe}} \to -\infty} \tilde{D}_{\text{safe}}(T_{\text{safe}}, \alpha) = \infty$ and $\lim_{T_{\text{safe}} \to \infty} \tilde{D}_{\text{safe}}(T_{\text{safe}}, \alpha) = 0$. Hence, for each $\alpha \in [0,1)$, there exists a unique value $T_{\text{safe}} = \tilde{g}_{\text{safe}}(\alpha) \in \mathbb{R}$ such that $\tilde{D}_{\text{safe}}(T_{\text{safe}}, \alpha) = sp$. Moreover, as $\tilde{D}_{\text{safe}}(0, 1-s) = sp$, we have that $\tilde{g}_{\text{safe}}(1-s) = 0$. Since $\tilde{D}_{\text{safe}}(t, \alpha)$ is strictly decreasing in α and continuous in (t, α), the function $\tilde{g}_{\text{safe}}$ is strictly decreasing and continuous.

To prove Part (a), suppose that W satisfies the healthy-more-willing-to-wait condition. Consider a pair $(T_{\text{risky}}, T_{\text{safe}})$ of waiting times, and let α be defined as in the statement of the lemma. Since the distribution of $W(\theta)$ is continuous, Claim 1 (taking $c = \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$) implies that $\text{SafeDemand}(T_{\text{risky}}, T_{\text{safe}}) = \tilde{D}_{\text{safe}}(T_{\text{safe}}, \alpha)$. By the definition of $\tilde{g}_{\text{safe}}$, (2) then holds if and only if $T_{\text{safe}} = \tilde{g}_{\text{safe}}(\alpha)$.

To prove Part (b), note that the definition of the lower quantile gives that $q_-(\alpha) \leq \Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})$, with equality if the distribution of $W(\theta)$ has full support on $[0, \overline{\Delta})$. As $\Lambda(T_{\text{risky}}) = \Lambda(T_{\text{safe}}) - \big(\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})\big)$, Part (b) then follows from Part (a). □

To complete the proof of Proposition 7, for simplicity, we refer to the hypothesis that there exists $\varepsilon > 0$ such that $\widehat{W}(\theta)$ first-order stochastically dominates $W(\theta) + \varepsilon$ as "strict dominance."

Let $(T_{\text{risky}}, T_{\text{safe}})$ be the lowest waiting-time steady state under patient decisions, and let $\alpha^* = F_W\big(\Lambda(T_{\text{safe}}) - \Lambda(T_{\text{risky}})\big)$. Let $(\widehat{T}_{\text{risky}}, \widehat{T}_{\text{safe}})$ be the lowest waiting-time steady state under transplant-center decisions, and let $\widehat{\alpha}^* = F_{\widehat{W}}\big(\Lambda(\widehat{T}_{\text{safe}}) - \Lambda(\widehat{T}_{\text{risky}})\big)$. (2) and Lemma 7 (applied to both patient and transplant-center decisions) imply that

$$T_{\text{safe}} = \tilde{g}_{\text{safe}}(\alpha^*), \qquad T_{\text{risky}} = \Lambda^{-1}\big(\Lambda(\tilde{g}_{\text{safe}}(\alpha^*)) - q_-(\alpha^*)\big), \tag{B.3}$$

$$\widehat{T}_{\text{safe}} = \tilde{g}_{\text{safe}}(\widehat{\alpha}^*), \qquad \widehat{T}_{\text{risky}} \leq \Lambda^{-1}\big(\Lambda(\tilde{g}_{\text{safe}}(\widehat{\alpha}^*)) - \widehat{q}_-(\widehat{\alpha}^*)\big). \tag{B.4}$$

By Lemma 7, we have that $\widehat{T}_{\text{safe}} = \tilde{g}_{\text{safe}}(\widehat{\alpha}^*) > 0$, and hence that $\widehat{\alpha}^* < 1 - s$.

To prove the proposition, we first show that

$$\alpha^* \geq \widehat{\alpha}^*, \text{ with strict inequality under strict dominance.} \tag{B.5}$$

To show this claim, we divide into cases based on whether $\widehat{\alpha}^* = 0$.

- Case 1: $\widehat{\alpha}^* = 0$. Note (5) requires that $\text{RiskyDemand}(T_{\text{risky}}, T_{\text{safe}}) = rp > 0$ or $\text{RefuseRisky}(T_{\text{risky}}, T_{\text{safe}}) = \gamma > 0$ or $T_{\text{risky}} = 0 < T_{\text{safe}}$. In any of these cases, we must have that $T_{\text{risky}} < T_{\text{safe}}$, and hence that $\alpha^* > 0 = \widehat{\alpha}^*$ as W has full support on $[0, \overline{\Delta})$.

- Case 2: $\widehat{\alpha}^* > 0$. Define functions

$$\begin{aligned}
\tilde{g}_{\text{risky}}(\alpha) &= \Lambda^{-1}\big(\Lambda(\tilde{g}_{\text{safe}}(\alpha)) - q_-(\alpha)\big) \\
\text{RiskyDemand}_{\tilde{g}}(\alpha) &= \text{RiskyDemand}(\tilde{g}_{\text{risky}}(\alpha), \tilde{g}_{\text{safe}}(\alpha)) \\
\text{RefuseRisky}_{\tilde{g}}(\alpha) &= \text{RefuseRisky}(\tilde{g}_{\text{risky}}(\alpha), \tilde{g}_{\text{safe}}(\alpha)).
\end{aligned}$$

  The definition of $\tilde{g}_{\text{risky}}$, Lemma 6, and (B.4) together give that

$$\tilde{g}_{\text{risky}}(\widehat{\alpha}^*) = \Lambda^{-1}\big(\Lambda(\tilde{g}_{\text{safe}}(\widehat{\alpha}^*)) - q_-(\widehat{\alpha}^*)\big) \geq \Lambda^{-1}\big(\Lambda(\tilde{g}_{\text{safe}}(\widehat{\alpha}^*)) - \widehat{q}_-(\widehat{\alpha}^*)\big) \geq \widehat{T}_{\text{risky}}. \tag{B.6}$$

It follows that

$$\begin{aligned}\text{RiskyDemand}_{\tilde{g}}(\widehat{\alpha}^*) &= p\int_{\theta:W(\theta)<q_-(\widehat{\alpha}^*)} e^{-\Lambda(\tilde{g}_{\text{risky}}(\widehat{\alpha}^*))\beta(\theta)} f(\theta)d\theta \\ &= \tilde{D}_{\text{risky}}(\tilde{g}_{\text{risky}}(\widehat{\alpha}^*), \widehat{\alpha}^*) \\ &= p\int_{\theta:\widehat{W}(\theta)<\Lambda(\widehat{T}_{\text{safe}})-\Lambda(\widehat{T}_{\text{risky}})} e^{-\Lambda(\tilde{g}_{\text{risky}}(\widehat{\alpha}^*))\beta(\theta)} f(\theta)d\theta \\ &\leq p\int_{\theta:\widehat{W}(\theta)<\Lambda(\widehat{T}_{\text{safe}})-\Lambda(\widehat{T}_{\text{risky}})} e^{-\Lambda(\widehat{T}_{\text{risky}})\beta(\theta)} f(\theta)d\theta \leq rp, \quad \text{(B.7)}\end{aligned}$$

where the second and third equalities follow from Claim 1 (under patient and transplant-center decisions, respectively), the first inequality uses (B.6), and the second inequality is (3). Similarly, we have that

$$\begin{aligned}\text{RefuseRisky}_{\tilde{g}}(\widehat{\alpha}^*) &= \int_{\tilde{g}_{\text{risky}}(\widehat{\alpha}^*)}^{\widehat{T}_{\text{safe}}} p\int_{\theta:W(\theta)>q_-(\widehat{\alpha}^*)} e^{-\Lambda(t)\beta(\theta)} f(\theta)d\theta dt \\ &= \int_{\tilde{g}_{\text{risky}}(\widehat{\alpha}^*)}^{\widehat{T}_{\text{safe}}} \tilde{D}_{\text{safe}}(t, \widehat{\alpha}^*)\, dt \\ &= \int_{\tilde{g}_{\text{risky}}(\widehat{\alpha}^*)}^{\widehat{T}_{\text{safe}}} p\int_{\theta:\widehat{W}(\theta)>\Lambda(\widehat{T}_{\text{safe}})-\Lambda(\widehat{T}_{\text{risky}})} e^{-\Lambda(t)\beta(\theta)} f(\theta)d\theta dt \\ &\leq \int_{\widehat{T}_{\text{risky}}}^{\widehat{T}_{\text{safe}}} p\int_{\theta:\widehat{W}(\theta)>\Lambda(\widehat{T}_{\text{safe}})-\Lambda(\widehat{T}_{\text{risky}})} e^{-\Lambda(t)\beta(\theta)} f(\theta)d\theta dt \leq \gamma, \quad \text{(B.8)}\end{aligned}$$

where the first equality is due to (B.4), the second and third equalities follow from Claim 1, the first inequality uses (B.6), and the second inequality is (4).

Similar to the proof of Proposition 1, based on (3), (4), and the nonnegativity of waiting times (in waiting-time steady states), let

$$\tilde{S} := \big\{\alpha \in (0, 1-s] \,\big|\, \text{RiskyDemand}_{\tilde{g}}(\alpha) \leq rp,\ \text{RefuseRisky}_{\tilde{g}}(\alpha) \leq \gamma,\ \text{and } \tilde{g}_{\text{risky}}(\alpha) \geq 0\big\}.$$

The inequalities in the previous paragraph show that $\widehat{\alpha}^* \in \tilde{S}$, so, in particular, $\tilde{S}$ is non-empty. The full support and the continuity of the distribution of $W(\theta)$ make $q_-(\alpha)$ continuous on $(0, 1)$. Lemma 7 therefore implies that $\tilde{g}_{\text{risky}}$ is continuous on $(0, 1)$, and hence that $\text{RiskyDemand}_{\tilde{g}}$ and $\text{RefuseRisky}_{\tilde{g}}$ are as well. In particular, $\tilde{S}$ is closed in $(0, 1-s]$. Moreover, the full support of the

distribution of $W(\theta)$ gives that $q_-(1-s) > 0$; since $\tilde{g}_{\text{safe}}(1-s) = 0$, we have that $\tilde{g}_{\text{risky}}(1-s) < 0$ and therefore that $1-s \notin \tilde{S}$.

Because $\tilde{S}$ is a nonempty closed subset of $(0, 1-s]$, it has a maximum α^{**}; since $\widehat{\alpha}^* \in \tilde{S}$ and $1-s \notin \tilde{S}$, we then have that $\widehat{\alpha}^* \leq \alpha^{**} < 1-s$. Since the full support and the continuity of the distribution of $W(\theta)$ give $F_W(q_-(\alpha^{**})) = \alpha^{**}$, the pair $(\tilde{g}_{\text{risky}}(\alpha^{**}), \tilde{g}_{\text{safe}}(\alpha^{**}))$ satisfies (2) by Lemma 7, satisfies (3) and (4) since $\alpha^{**} \in \tilde{S}$, and satisfies (5) because at least one of the inequalities defining $\tilde{S}$ must bind at α^{**} (otherwise, continuity and $\alpha^{**} < 1-s$ would provide $\eta > 0$ such that $\alpha^{**} + \eta \in \tilde{S}$, contradicting the maximality of α^{**}). Since $\alpha^{**} > 0$, the full support of the distribution of $W(\theta)$ gives $q_-(\alpha^{**}) > 0$, so we have that $0 \leq \tilde{g}_{\text{risky}}(\alpha^{**}) < \tilde{g}_{\text{safe}}(\alpha^{**})$. Thus, $(\tilde{g}_{\text{risky}}(\alpha^{**}), \tilde{g}_{\text{safe}}(\alpha^{**}))$ is a waiting-time steady state under patient decisions. But since $(T_{\text{risky}}, T_{\text{safe}})$ is the lowest waiting-time steady state under patient decisions, we then have that $T_{\text{safe}} \leq \tilde{g}_{\text{safe}}(\alpha^{**})$. Since $T_{\text{safe}} = \tilde{g}_{\text{safe}}(\alpha^*)$ and $\tilde{g}_{\text{safe}}$ is strictly decreasing, it follows that $\alpha^* \geq \alpha^{**} \geq \widehat{\alpha}^*$, proving the weak inequality in (B.5).

Now, assuming strict dominance, let $\varepsilon > 0$ be such that $\widehat{W}$ first-order stochastically dominates $W + \varepsilon$. Since $\widehat{\alpha}^* \in (0, 1-s)$, the first inequality in (B.6) is then strict by Lemma 6. In particular, we have that $\tilde{g}_{\text{risky}}(\widehat{\alpha}^*) > 0$. Moreover, since $F_{\widehat{W}}\big(\Lambda(\widehat{T}_{\text{safe}}) - \Lambda(\widehat{T}_{\text{risky}})\big) = \widehat{\alpha}^* > 0$, it follows that the first inequality in (B.7) holds strictly, so we have that $\text{RiskyDemand}_{\tilde{g}}(\widehat{\alpha}^*) < rp$. Since $F_{\widehat{W}}\big(\Lambda(\widehat{T}_{\text{safe}}) - \Lambda(\widehat{T}_{\text{risky}})\big) = \widehat{\alpha}^* < 1-s < 1$, it also follows that the first inequality in (B.8) holds strictly, so we have that $\text{RefuseRisky}_{\tilde{g}}(\widehat{\alpha}^*) < \gamma$. Thus, all the inequalities defining $\tilde{S}$ hold strictly at $\widehat{\alpha}^*$. By continuity, there then exists $\alpha \in \tilde{S}$ with $\alpha > \widehat{\alpha}^*$, so we have that $\alpha^{**} \geq \alpha > \widehat{\alpha}^*$, and hence that $\alpha^* > \widehat{\alpha}^*$.

Since $\tilde{g}_{\text{safe}}$ is strictly decreasing (by Lemma 7), (B.4), and (B.5) together imply that $T_{\text{safe}} \leq \widehat{T}_{\text{safe}}$ with strict inequality under strict dominance. This proves Part (a).

Suppose now that under patient decisions, the lowest waiting-time steady state is congested. Then, (4) must bind under patient decisions. We then have that

$$\gamma = \text{RefuseRisky}(T_{\text{risky}}, T_{\text{safe}}) = \int_{T_{\text{risky}}}^{T_{\text{safe}}} \tilde{D}_{\text{safe}}(t, \alpha^*)\, dt \leq \int_{T_{\text{risky}}}^{T_{\text{safe}}} \tilde{D}_{\text{safe}}(t, \widehat{\alpha}^*)\, dt, \quad \text{(B.9)}$$

where the second equality is due to Claim 1, and the inequality is due to (B.5) because $\tilde{D}_{\text{safe}}(t, \alpha)$ is decreasing in α. Moreover, (4) under transplant-center decisions and

Claim 1 give that $\gamma \geq \int_{\widehat{T}_{\text{risky}}}^{\widehat{T}_{\text{safe}}} \tilde{D}_{\text{safe}}(t, \widehat{\alpha}^*)\, dt$. We therefore have that

$$\int_{\widehat{T}_{\text{risky}}}^{\widehat{T}_{\text{safe}}} \tilde{D}_{\text{safe}}(t, \widehat{\alpha}^*)\, dt \leq \gamma \leq \int_{T_{\text{risky}}}^{T_{\text{safe}}} \tilde{D}_{\text{safe}}(t, \widehat{\alpha}^*)\, dt. \tag{B.10}$$

As $\widehat{T}_{\text{safe}} \geq T_{\text{safe}}$ (Part (a)) and $\tilde{D}_{\text{safe}}(t, \widehat{\alpha}^*) > 0$ (since $\widehat{\alpha}^* < 1 - s < 1$), it follows that $\widehat{T}_{\text{risky}} \geq T_{\text{risky}}$ (otherwise, the domain of integration on the left-hand side would strictly contain the domain of integration on the right-hand side). Under strict dominance, due to (B.5), the inequality in (B.9) is strict, which implies that the inequality in (B.10) is also strict, and hence that $\widehat{T}_{\text{risky}} > T_{\text{risky}}$.

It remains to verify that the lowest waiting-time steady state is congested under transplant-center decisions. It follows from Claim 1 that the demand for risky organs at $(\widehat{T}_{\text{risky}}, \widehat{T}_{\text{safe}})$ under transplant-center decisions is $\tilde{D}_{\text{risky}}(\widehat{T}_{\text{risky}}, \widehat{\alpha}^*)$. Moreover, we have that

$$\tilde{D}_{\text{risky}}(\widehat{T}_{\text{risky}}, \widehat{\alpha}^*) \leq \tilde{D}_{\text{risky}}(T_{\text{risky}}, \alpha^*) = \text{RiskyDemand}(T_{\text{risky}}, T_{\text{safe}}) < rp,$$

where the first inequality is because $\tilde{D}_{\text{risky}}(t, \alpha)$ is nonincreasing in t and nondecreasing in α, the equality is due to Claim 1, and the second inequality is due to congestion at the lowest waiting-time steady state under patient decisions. Last, we have that $\widehat{T}_{\text{risky}} \geq T_{\text{risky}} > 0$, where the second inequality is also due to congestion at the lowest waiting-time steady state under patient decisions.

References

Agarwal, N., I. Ashlagi, M. A. Rees, P. Somaini, and D. Waldinger (2021). Equilibrium allocations under alternative waitlist designs: Evidence from deceased donor kidneys. *Econometrica 89*(1), 37–76.

Arnosti, N. and P. Shi (2020). Design of lotteries and wait-lists for affordable housing allocation. *Management Science 66*(6), 2291–2307.

Ashlagi, I. and A. E. Roth (2021). Kidney exchange: An operations perspective. *Management Science 67*(9), 5455–5478.

Ashlagi, I. and A. E. Roth (2026). Out of sequence offers: Towards efficient, equitable organ allocation. *American Journal of Bioethics 26*(1), 5–8.

Ata, B., Y. Ding, and S. A. Zenios (2021). An achievable-region-based approach for kidney allocation policy design with endogenous patient choice. *Manufacturing & Service Operations Management 23*(1), 36–54.

Aubert, O., P. P. Reese, B. Audry, Y. Bouatou, M. Raynaud, D. Viglietti, C. Legendre, D. Glotz, J.-P. Empana, X. Jouven, et al. (2019). Disparities in acceptance of deceased donor kidneys between the United States and France and estimated effects of increased US acceptance. *JAMA Internal Medicine 179*(10), 1365–1374.

Baccara, M., S. Lee, and L. Yariv (2020). Optimal dynamic matching. *Theoretical Economics 15*(3), 1221–1278.

Barah, M., V. Kilambi, J. J. Friedewald, and S. Mehrotra (2022). Implications of accumulated cold time for US kidney transplantation offer acceptance. *Clinical Journal of the American Society of Nephrology 17*(9), 1353–1362.

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

Callaghan, C. J., L. Mumford, L. Pankhurst, R. J. Baker, J. A. Bradley, and C. J. Watson (2017). Early outcomes of the new UK deceased donor kidney fast-track offering scheme. *Transplantation 101*(12), 2888–2897.

Castro, F., H. Ma, H. Nazerzadeh, and C. Yan (2021). Randomized FIFO mechanisms. Working paper.

Chan, A. and A. E. Roth (2024). Regulation of organ transplantation and procurement: A market design lab experiment. *Journal of Political Economy 132*(11), 3827–3866.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological) 34*(2), 187–202.

Dai, J. G. and S. He (2010). Customer abandonment in many-server queues. *Mathematics of Operations Research 35*(2), 347–362.

De Mel, S. C., K. Munshi, S. Reiche, and H. Sabourian (2021). Herding with heterogeneous ability: An application to organ transplantation. Working paper.

Debout, A., Y. Foucher, K. Trébern-Launay, C. Legendre, H. Kreis, G. Mourad, V. Garrigue, E. Morelon, F. Buron, L. Rostaing, N. Kamar, M. Kessler, M. Ladrière, A. Poignas, A. Blidi, J.-P. Soulillou, M. Giral, and E. Dantan (2015). Each additional hour of cold ischemia time significantly increases the risk of graft failure and mortality following renal transplantation. *Kidney International 87*(2), 343–349.

Doval, L. (2025). Dynamic matching. *Handbook of the Economics of Matching 2*, 489–577.

Doval, L., F. Echenique, W. Huang, and Y. Xin (2024). Social learning in lung transplant decision. Working paper.

Guan, G., S. Neelam, J. Studnia, X. S. Cheng, M. L. Melcher, M. A. Rees, A. E. Roth, P. Somaini, and I. Ashlagi (2025). Insights from refusal patterns for deceased donor kidney offers. *Transplantation 109*(11), 1774–1782.

Health Resources and Services Administration (2025). Organ procurement & transplantation network: Data. Available at `http://optn.transplant.hrsa.gov/data/`.

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

Jadlowiec, C. C., S. Y. Ohara, R. Punukollu, J. Wagler, B. Ruch, K. Kumm, P. Budhiraja, H. M. Me, A. K. Mathur, K. S. Reddy, et al. (2023). Outcomes with transplanting kidneys offered through expedited allocation. *Clinical Transplantation 37*(11), e15094.

Kella, O. (2024). On independence of time and cause. *Statistics & Probability Letters 204*, 109944.

Lentine, K. L., J. M. Smith, G. R. Lyden, J. M. Miller, S. E. Booker, T. G. Dolan, K. R. Temple, S. Weiss, D. Handarova, A. K. Israni, and J. J. Snyder (2025). OPTN/SRTR 2023 annual data report: Kidney. *American Journal of Transplantation 25*(2S1), S22–S137.

Leshno, J. D. (2022). Dynamic matching in overloaded waiting lists. *American Economic Review 112*(12), 3876–3910.

Liyanage, L. N., D. Akizhanov, S. S. Patel, D. L. Segev, A. B. Massie, D. E. Stewart, and S. E. Gentry (2025). Contemporary prevalence and practice patterns of out-of-sequence kidney allocation. *American Journal of Transplantation 25*(2), 343–354.

Mandelbaum, A., W. A. Massey, M. I. Reiman, A. Stolyar, and B. Rider (2002). Queue lengths and waiting times for multiserver queues with abandonment and retrials. *Telecommunication Systems 21*(2), 149–171.

Medin, C., C.-G. Elinder, B. Hylander, B. Blom, and H. Wilczek (2000). Survival of patients who have been on a waiting list for renal transplantation. *Nephrology Dialysis Transplantation 15*(5), 701–704.

Mehrotra, S., J. M. Gonzalez, K. Schantz, J.-C. Yang, J. J. Friedewald, and R. Knight (2022). Patient preferences for waiting time and kidney quality. *Clinical Journal of the American Society of Nephrology 17*(9), 1363–1371.

Mildebrath, D., T. Lee, S. Sinha, A. J. Schaefer, and A. O. Gaber (2024). Characterizing rational transplant program response to outcome-based regulation. *Operations Research 72*(4), 1421–1437.

Milgrom, P. and J. Roberts (1994). Comparing equilibria. *American Economic Review 84*(3), 441–459.

Mohan, S., M. C. Chiles, R. E. Patzer, S. O. Pastan, S. A. Husain, D. J. Carpenter, G. K. Dube, R. J. Crew, L. E. Ratner, and D. J. Cohen (2018). Factors leading to the discard of deceased donor kidneys in the United States. *Kidney International 94*(1), 187–198.

Mohan, S. and S. A. Husain (2022). Improving the utilization of deceased donor kidneys by prioritizing patient preferences. *Clinical Journal of the American Society of Nephrology 17*(9), 1278–1280.

Mohan, S., M. Yu, K. L. King, and S. A. Husain (2023). Increasing discards as an unintended consequence of recent changes in United States kidney allocation policy. *Kidney International Reports 8*(5), 1109.

Murra-Anton, Z. and N. Thakral (2024). The public housing allocation problem. Working paper.

Peters-Sengers, H., J. H. Houtzager, M. M. Idu, M. B. Heemskerk, E. L. van Heurn, J. J. H. van der Heide, J. Kers, S. P. Berger, T. M. van Gulik, and F. J. Bemelman (2019). Impact of cold ischemia time on outcomes of deceased donor kidney transplantation: An analysis of a national registry. *Transplantation Direct 5*(5), e448.

Quiroga, I., P. McShane, D. D. Koo, D. Gray, P. J. Friend, S. Fuggle, and C. Darby (2006). Major effects of delayed graft function and cold ischaemia time on renal allograft survival. *Nephrology Dialysis Transplantation 21*(6), 1689–1696.

Sampaio, M. S., B. Chopra, A. Tang, and K. K. Sureshkumar (2018). Impact of cold ischemia time on the outcomes of kidneys with Kidney Donor Profile Index $\geq$ 85%: Mate kidney analysis—a retrospective study. *Transplant International 31*(7), 729–738.

Schaubel, D. E., M. K. Guidinger, S. Biggins, J. D. Kalbfleisch, E. A. Pomfret, P. Sharma, and R. M. Merion (2009). Survival benefit-based deceased-donor liver allocation. *American Journal of Transplantation 9*(4), 970–981.

Schummer, J. (2021). Influencing waiting lists. *Journal of Economic Theory 195*, 105263.

Sepulveda, J. M. G., S. Mehrotra, J.-C. Yang, K. J. Schantz, Y. Becker, R. Formica, D. P. Ladner, D. Kaufman, and J. Friedewald (2023). Physician preferences when selecting candidates for lower-quality kidney offers. *Clinical Journal of the American Society of Nephrology 18*(12), 1599–1609.

Su, X. and S. A. Zenios (2004). Patient choice in kidney allocation: The role of the queueing discipline. *Manufacturing & Service Operations Management 6*(4), 280–301.

Su, X. and S. A. Zenios (2005). Patient choice in kidney allocation: A sequential stochastic assignment model. *Operations Research 53*(3), 443–455.

Su, X. and S. A. Zenios (2006). Recipient choice can address the efficiency-equity trade-off in kidney transplantation: A mechanism design model. *Management Science 52*(11), 1647–1660.

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

Waldinger, D. (2021). Targeting in-kind transfers through market design: A revealed preference analysis of public housing allocation. *American Economic Review 111*(8), 2660–2696.

Whitt, W. (2006). Fluid models for multiserver queues with abandonments. *Operations Research 54*(1), 37–54.

Wolfe, R. A., V. B. Ashby, E. L. Milford, A. O. Ojo, R. E. Ettenger, L. Y. Agodoa, P. J. Held, and F. K. Port (1999). Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *New England Journal of Medicine 341*(23), 1725–1730.

Wood, N. L., D. N. VanDerwerken, A. B. Massie, D. L. Segev, J. J. Snyder, and S. E. Gentry (2025). Diagnosing the recent decrease in utilization of deceased donor kidneys. *Transplantation 109*(3), 496–503.

Zhang, J. (2010). The sound of silence: Observational learning in the US kidney market. *Marketing Science 29*(2), 315–335.

Online Appendix

C Foundation for Waiting-Time Steady States

This appendix gives a formal stationary version of the first-come-first-serve (FCFS) fluid model and shows that its equilibrium outcomes coincide with waiting-time steady states as defined in Definition 1.

C.1 The Stationary FCFS Game

For $x \in \{\text{risky}, \text{safe}\}$, let $b_x(t)$ denote the flow of type-x organs remaining after all patients with waiting times greater than t have been considered. A *residual offer flow* is a pair $b = (b_{\text{risky}}, b_{\text{safe}})$ such that

- $b_{\text{risky}} : \mathbb{R}_+ \to [0, rp]$ and $b_{\text{safe}} : \mathbb{R}_+ \to [0, sp]$ are nondecreasing and right-continuous;
- $\lim_{t\to\infty} b_{\text{risky}}(t) = rp$ and $\lim_{t\to\infty} b_{\text{safe}}(t) = sp$.

For $x \in \{\text{risky}, \text{safe}\}$, define

$$T_x^*(b) := \inf\{t \geq 0 : b_x(t) > 0\}.$$

We restrict attention to residual offer flows satisfying

$$b_x\big(T_x^*(b)\big) > 0 \qquad \text{for each } x \in \{\text{risky}, \text{safe}\}. \tag{C.1}$$

Under (C.1), monotonicity gives that $\{t \geq 0 : b_x(t) > 0\} = [T_x^*(b), \infty)$, so $T_x^*(b)$ is the first waiting time at which type-x offers are made. For notational simplicity, given a residual offer flow b, we write $\tilde{T}_{\text{risky}} := T_{\text{risky}}^*(b)$ and $\tilde{T}_{\text{safe}} := T_{\text{safe}}^*(b)$.

A *stationary pure strategy* for type θ is a Borel-measurable function $\sigma_\theta : \mathbb{R}_+ \to \{0, 1\}$, where 1 denotes acceptance of a risky offer. Safe offers are always accepted. We restrict to strategies for which, for each t

$$\big\{t' \in [t, \tilde{T}_{\text{safe}}) \,\big|\, \sigma_\theta(t') = 1\big\} \text{ has a least element whenever it is nonempty.} \tag{C.2}$$

The first risky offer accepted after the patient reaches a waiting time of t is given by

$$\tau_\theta(t) := \begin{cases} \min\left\{t' \in [t, \tilde{T}_{\text{safe}}) \,\middle|\, \sigma_\theta(t') = 1\right\} & \text{if this set is nonempty} \\ \tilde{T}_{\text{safe}} & \text{otherwise} \end{cases}.$$

Define the patient's planned allocation time

$$\tau_\theta := \begin{cases} \tau_\theta(\tilde{T}_{\text{risky}}) & \text{if } \tilde{T}_{\text{risky}} < \tilde{T}_{\text{safe}} \\ \tilde{T}_{\text{safe}} & \text{if } \tilde{T}_{\text{risky}} \geq \tilde{T}_{\text{safe}} \end{cases}.$$

For $t \in [\tilde{T}_{\text{risky}}, \tilde{T}_{\text{safe}})$, the continuation payoff conditional on surviving t units of time is

$$v_{\theta,t}(\sigma_\theta; b) := \begin{cases} \nu(\theta) e^{-(\Lambda(\tau_\theta(t)) - \Lambda(t))\beta(\theta)} & \text{if } \tau_\theta(t) < \tilde{T}_{\text{safe}} \\ e^{-(\Lambda(\tilde{T}_{\text{safe}}) - \Lambda(t))\beta(\theta)} & \text{if } \tau_\theta(t) = \tilde{T}_{\text{safe}} \end{cases}.$$

The strategy σ_θ is *sequentially optimal given* b if

$$v_{\theta,t}(\sigma_\theta; b) = \sup_{\widetilde{\sigma}_\theta} v_{\theta,t}(\widetilde{\sigma}_\theta; b) \qquad \text{for every } t \in [\tilde{T}_{\text{risky}}, \tilde{T}_{\text{safe}}),$$

where the supremum is over continuation strategies.

We require the maps $(\theta, t) \mapsto \sigma_\theta(t)$ and $\theta \mapsto \tau_\theta$ to be Borel-measurable.

C.2 Flow Balance and Stationary Equilibrium

Fix a residual offer flow b and a strategy profile σ. For every Borel set $B \subseteq \mathbb{R}_+$, define the risky- and safe-allocation masses by

$$M_{\text{risky}}(B) := p \int_\Theta \mathbf{1}\{\tau_\theta < \tilde{T}_{\text{safe}}\} \mathbf{1}\{\tau_\theta \in B\} e^{-\Lambda(\tau_\theta)\beta(\theta)} \, f(\theta) d\theta$$
$$M_{\text{safe}}(B) := p \int_\Theta \mathbf{1}\{\tau_\theta = \tilde{T}_{\text{safe}}\} \mathbf{1}\{\tilde{T}_{\text{safe}} \in B\} e^{-\Lambda(\tilde{T}_{\text{safe}})\beta(\theta)} \, f(\theta) d\theta.$$

If $\tilde{T}_{\text{risky}} < \tilde{T}_{\text{safe}}$, then the mass of refusals encountered by a risky organ while being offered to patients with waiting times between $\tilde{T}_{\text{risky}}$ and $\tilde{T}_{\text{safe}}$ is

$$\mathcal{R} := \int_{\tilde{T}_{\text{risky}}}^{\tilde{T}_{\text{safe}}} p \int_\Theta e^{-\Lambda(t)\beta(\theta)} \mathbf{1}\{\tau_\theta > t\} \, f(\theta) d\theta \, dt.$$

Set $\mathcal{R} = 0$ when $\tilde{T}_{\text{risky}} \geq \tilde{T}_{\text{safe}}$. We then define steady-state equilibrium to consist of residual offer flows, strategy profiles, and an unallocated risky-organ flow that satisfy sequential-optimality, flow-balance, refusal-limit, and no-withholding constraints.

Definition 4. A *steady-state FCFS equilibrium* is a tuple $(b, \sigma, e_{\text{risky}})$, where b is a residual offer flow, σ is a strategy profile, and $e_{\text{risky}} \geq 0$ is the flow of risky organs that remain unallocated, such that:

(i) σ_θ is sequentially optimal given b for almost every type θ;

(ii) safe-organ flow is conserved and no safe organ is discarded:

$$\begin{aligned} M_{\text{safe}}(\mathbb{R}_{\geq 0}) &= sp, \\ b_{\text{safe}}(t) &= sp - M_{\text{safe}}((t, \infty)) \qquad \text{for all } t \geq 0; \end{aligned} \tag{C.3}$$

(iii) risky-organ flow is conserved:

$$\begin{aligned} M_{\text{risky}}(\mathbb{R}_{\geq 0}) + e_{\text{risky}} &= rp, \\ b_{\text{risky}}(t) &= rp - M_{\text{risky}}((t, \infty)) - e_{\text{risky}} \mathbf{1}\{t < \tilde{T}_{\text{risky}}\} \quad \text{for all } t \geq 0; \end{aligned} \tag{C.4}$$

(iv) the refusal limit is respected:

$$\mathcal{R} \leq \gamma; \tag{C.5}$$

(v) risky organs are not intentionally withheld:

$$\text{if } e_{\text{risky}} > 0, \text{ then } \tilde{T}_{\text{risky}} = 0 \text{ or } \mathcal{R} = \gamma. \tag{C.6}$$

To analyze steady-state FCFS equilibrium, we first show that waiting times are nonnegative, with the waiting time for risky organs being strictly lower.

Lemma 8. *Every steady-state FCFS equilibrium satisfies* $0 \leq \tilde{T}_{\text{risky}} < \tilde{T}_{\text{safe}}$.

Proof. $\tilde{T}_{\text{risky}}$ and $\tilde{T}_{\text{safe}}$ are finite because we have that $\lim_{t\to\infty} b_{\text{risky}}(t) = rp > 0$ and that $\lim_{t\to\infty} b_{\text{safe}}(t) = sp > 0$. Now suppose for the sake of deriving a contradiction that $\tilde{T}_{\text{risky}} \geq \tilde{T}_{\text{safe}}$. Patients then receive safe offers before any risky offer, so $\tau_\theta = \tilde{T}_{\text{safe}}$ for every type and $M_{\text{risky}}(\mathbb{R}_+) = 0$. Due to (ii) in Definition 4, we have that

$$s = \int_\Theta e^{-\Lambda(\tilde{T}_{\text{safe}})\beta(\theta)} f(\theta) d\theta,$$

and therefore that $\tilde{T}_{\text{safe}} > 0$ since $s < 1$. (iii) in Definition 4 gives that $e_{\text{risky}} = rp > 0$, while $\mathcal{R} = 0$. Since $\tilde{T}_{\text{risky}} \geq \tilde{T}_{\text{safe}} > 0$ and $\gamma > 0$, this contradicts (C.6). □

We then have a simple cutoff characterization of sequential optimality.

Lemma 9 (Sequential optimality)**.** *A strategy is sequentially optimal if and only if for every $t \in [\tilde{T}_{\text{risky}}, \tilde{T}_{\text{safe}})$, we have that*

$$\sigma_\theta(t) = \begin{cases} 1 & \text{if } W(\theta) < \Lambda(\tilde{T}_{\text{safe}}) - \Lambda(t), \\ 0 & \text{if } W(\theta) > \Lambda(\tilde{T}_{\text{safe}}) - \Lambda(t) \end{cases}, \tag{C.7}$$

with either action allowed at equality. In particular, we have that

$$\tau_\theta = \begin{cases} \tilde{T}_{\text{risky}} & W(\theta) < \Lambda(\tilde{T}_{\text{safe}}) - \Lambda(\tilde{T}_{\text{risky}}), \\ \tilde{T}_{\text{safe}} & W(\theta) > \Lambda(\tilde{T}_{\text{safe}}) - \Lambda(\tilde{T}_{\text{risky}}). \end{cases} \tag{C.8}$$

Proof. Conditional on surviving t units of time, accepting the current risky offer gives utility $\nu(\theta)$. Accepting a risky offer at a later waiting time $t' > t$ gives expected utility $\nu(\theta)e^{-(\Lambda(t')-\Lambda(t))\beta(\theta)} \leq \nu(\theta)$ with equality only if $\nu(\theta) = 0$, while refusing all remaining risky offers and waiting for a safe organ gives expected utility $e^{-(\Lambda(\tilde{T}_{\text{safe}})-\Lambda(t))\beta(\theta)}$. Hence, sequential optimality implies (C.7).

If (C.7) does not require accepting a risky organ at time t, then, for every $t' > t$, we have that $W(\theta) \geq \Lambda(\tilde{T}_{\text{safe}}) - \Lambda(t) > \Lambda(\tilde{T}_{\text{safe}}) - \Lambda(t')$, so (C.7) requires rejecting a risky organ thereafter. It follows that the conditions in (C.7) are also sufficient for sequential optimality.

Applying (C.7) at $t = \tilde{T}_{\text{risky}}$ gives (C.8). □

C.3 Equivalence with Waiting-Time Steady States

We can now derive an equivalence between steady-state FCFS equilibria and waiting-time steady states.

Proposition 11. *Under the assumptions in Section 3 and Assumption 0, a pair $(T_{\text{risky}}, T_{\text{safe}})$ is a waiting-time steady state if and only if there exists a steady-state FCFS equilibrium with $\tilde{T}_{\text{risky}} = T_{\text{risky}}$ and $\tilde{T}_{\text{safe}} = T_{\text{safe}}$. Every such equilibrium satisfies*

$$b_{\text{risky}}(t) = rp\,\mathbf{1}\{t \geq T_{\text{risky}}\} \quad \text{and} \quad b_{\text{safe}}(t) = sp\,\mathbf{1}\{t \geq T_{\text{safe}}\}. \tag{C.9}$$

Proof. First, let $(b, \sigma, e_{\text{risky}})$ be a steady-state FCFS equilibrium and let $\tilde{\Delta} = \Lambda(\tilde{T}_{\text{safe}}) - \Lambda(\tilde{T}_{\text{risky}})$. Lemmata 8 and 9 imply that for a full measure of types θ, we have that $\tau_\theta = \tilde{T}_{\text{risky}}$ if $W(\theta) < \tilde{\Delta}$ and $\tau_\theta = \tilde{T}_{\text{safe}}$ if $W(\theta) > \tilde{\Delta}$. Since $W(\theta)$ has a continuous distribution, it follows that

$$M_{\text{safe}}(\mathbb{R}_+) = p \int_{\theta: W(\theta) > \tilde{\Delta}} e^{-\Lambda(\tilde{T}_{\text{safe}})\beta(\theta)} f(\theta) d\theta = \text{SafeDemand}(\tilde{T}_{\text{risky}}, \tilde{T}_{\text{safe}}),$$
$$M_{\text{risky}}(\mathbb{R}_+) = p \int_{\theta: W(\theta) < \tilde{\Delta}} e^{-\Lambda(\tilde{T}_{\text{risky}})\beta(\theta)} f(\theta) d\theta = \text{RiskyDemand}(\tilde{T}_{\text{risky}}, \tilde{T}_{\text{safe}}),$$
$$\mathcal{R} = \int_{\tilde{T}_{\text{risky}}}^{\tilde{T}_{\text{safe}}} p \int_{\theta: W(\theta) > \tilde{\Delta}} e^{-\Lambda(t)\beta(\theta)} f(\theta) d\theta \, dt = \text{RefuseRisky}(\tilde{T}_{\text{risky}}, \tilde{T}_{\text{safe}}).$$

Hence, (ii), (iii), and (iv) in Definition 4 give (2), (3), and (4), respectively. If $e_{\text{risky}} = 0$, risky demand equals rp. If $e_{\text{risky}} > 0$, (C.6) implies that $\tilde{T}_{\text{risky}} = 0$ or (C.5) binds. Hence (5) also holds, so $(\tilde{T}_{\text{risky}}, \tilde{T}_{\text{safe}})$ is a waiting-time steady state. As M_{safe} and M_{risky} are concentrated at $\tilde{T}_{\text{safe}}$ and $\tilde{T}_{\text{risky}}$, respectively, (C.3) and (C.4) imply (C.9).

Conversely, let $(T_{\text{risky}}, T_{\text{safe}})$ be a waiting-time steady state. We first show that $T_{\text{risky}} < T_{\text{safe}}$. Suppose for sake of deriving a contradiction that $T_{\text{risky}} = T_{\text{safe}}$. Since $W(\theta)$ has a continuous distribution, we have that $\text{RiskyDemand}(T_{\text{risky}}, T_{\text{safe}}) = 0 < rp$ and $\text{RefuseRisky}(T_{\text{risky}}, T_{\text{safe}}) = 0 < \gamma$. Since $s < 1$, (2) implies that $T_{\text{safe}} > 0$, so we also have that $T_{\text{risky}} > 0$, contradicting (5). Hence, we must have that $T_{\text{risky}} < T_{\text{safe}}$.

Defining b by (C.9), we have that $(\tilde{T}_{\text{risky}}, \tilde{T}_{\text{safe}}) = (T_{\text{risky}}, T_{\text{safe}})$. For each $\theta \in \Theta$, set

$$\sigma^*_\theta(t) := \mathbf{1}\{T_{\text{risky}} \leq t < T_{\text{safe}},\ W(\theta) < \Lambda(T_{\text{safe}}) - \Lambda(t)\}.$$

This strategy is Borel-measurable, satisfies (C.2), and is sequentially optimal by Lemma 9. It induces allocation measures M_{risky} and M_{safe} concentrated at T_{risky} and T_{safe}, respectively, with respective total masses $\text{RiskyDemand}(T_{\text{risky}}, T_{\text{safe}})$ and $\text{SafeDemand}(T_{\text{risky}}, T_{\text{safe}}) = sp$; and refusal mass $\text{RefuseRisky}(T_{\text{risky}}, T_{\text{safe}}) \leq \gamma$. Setting $e^*_{\text{risky}} := rp - \text{RiskyDemand}(T_{\text{risky}}, T_{\text{safe}}) \geq 0$, we have (ii), (iii), and (iv) in Definition 4. If $e^*_{\text{risky}} > 0$, then (5) implies that either $T_{\text{risky}} = 0$ or (C.5) binds. Hence, (C.6) also holds, so $(b, \sigma^*, e^*_{\text{risky}})$ is a steady-state FCFS equilibrium. □