Preliminary and incomplete.

Rationing through Classification*

Edwin Muñoz-Rodriguez† James Schummer‡

March 6, 2024

Abstract

In various object allocation problems, including organ allocation, potential recipients are prioritized according to their actions (e.g. treatment decisions). While prioritization can improve object allocation, *strict* prioritization has been shown to distort action choices, especially in competitive environments. We examine how competitiveness impacts this welfare tradeoff when the planner more generally can *partially* prioritize recipients via *rationing through classification* (RTC): the planner rations a fractional share of the objects amongst recipients taking some critical action, rationing the rest amongst the remaining recipients.

In "competitive" environments (where RTC is without loss), optimal utilitarian welfare is obtained by maximizing this fractional share subject to eliminating distorted decisions (see also Braverman and Garg (2020), Perez-Richet and Skreta (2022)). In less competitive environments (i.e. few agents control many recipients' actions) our

*We are grateful for comments from Baris Ata, Itai Gurvich, John Duggan, Andrew Mackenzie, Romans Pancs, Vasiliki Skreta, Olivier Tercieux, William Thomson, and audiences at the University of Rochester, 13th Conference on Economic Design, and 15th Social Choice and Welfare Conference. We have received valuable research assistance from Michelle Avataneo and Isaac Lara. Muñoz-Rodriguez is grateful for funding from the Ford Foundation through a PRAEM program grant at El Colegio de Mexico, A.C. The views and conclusions expressed in this article are solely those of the authors, and do not necessarily reflect those of the Ford Foundation.

†Center for Economic Studies, El Colegio de México A.C., Picacho Ajusco 20, Mexico City, 14110, eamunoz@colmex.mx.

‡Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208, schummer@kellogg.northwestern.edu.

model becomes a (not well-behaved) atomic congestion game sometimes yielding a second form of equilibria with excessively distorted actions. We provide conditions making this second form of equilibria less plausible. In such cases we can extend the earlier welfare result and show that competitiveness decreases optimal welfare. Numerical analysis confirms these results and suggests the latter equilibria typically yield lower welfare.

1 Introduction

There are many contexts in which limited resources are allocated based on potential recipients' action choices. As one leading example, patients awaiting organ transplants are prioritized—partly or even fully—based on interim treatment choices made by their care providers. For heart transplant patients a prioritization is predominantly based on the choice of mechanical circulatory or ventilatory support; liver transplant patients historically received higher priority when placed in an intensive care unit (ICU). Other examples include some school choice settings (where priorities can depend on choice of residential location), and organizations that rank or rate agents.[1]

The prioritization of recipients based on actions leads to two consequences, both confirmed by empirical observation. Firstly, action choices become distorted. As an illustration of this again within the context of organ allocation, a policy change in October of 2018 increased the relative priority of heart patients treated with an intra-aortic balloon pump (IABP) over those treated with high-dose inotropes (HDI). Coinciding with this change was a roughly three-fold increase in IABP usage (Ran et al., 2021). Another version of this same story occurred for liver patients in 2002, when the removal of ICU status as a prioritization factor coincided with a roughly 50% drop in the admission of liver patients to ICU's (Snyder, 2010).

The second consequence is linked to a crucial feature of these organ allocation examples which motivates our main model: more than one potential recipient (patient) falls under the care of a single strategic agent in the form of a Transplant Center (TC). The possibly coordinated decision-making across multiple recipients is a factor in these kinds of allocation problems that is

[1] A related example is certification agencies, though they may not face an exogenous budget constraint when allocating certifications; see Frankel and Kartik (2021), Ball (2023), Perez-Richet and Skreta (2022), and Perez-Richet and Skreta (2023).

novel in the analysis of allocation problems (see Subsection 1.1). Intuitively, increasing priority for one patient might crowd out other patients. If this congestion effect were to be internalized by that patient's TC, one might expect the above distortions to be weakened. Indeed, Parker et al. (2018) provide compelling evidence of this in the context of heart allocation, showing that patient "over-treatment" was noticeably correlated with the number of competing Transplant Centers within a given area. Viewing the strategic agent to be the TC rather than the patient, this evidence the intuition that decentralization of action choices intensifies the above distortion.

With both of these consequences in mind, we construct a stylized model to consider (i) the degree to which a planner should prioritize recipients based on action choices, and more significantly (ii) how that degree is affected by the level of competition, vis à vis the degree to which action choices are decentralized as described above. The planner trades off the potential distortion in action choices against possible improvements in allocation decisions. At least in the case of organ allocation, discussions on addressing these distortions have often centered on the binary decision of whether to give *full* or *no* priority to otherwise identical patients who take one action vs. another. E.g. in the heart example discussed above, IABP patients were given strict priority over HDI following perceived abuse of the latter treatment; for livers, the abuse of ICU status resulted in its complete elimination as a factor.

Rather than consider the tradeoff as a binary one, we allow a planner to *partially* prioritize the class of recipients that take a particular action over the class who do not. Namely, we describe a "rationing through classification" (RTC) approach that allows the planner to offer less-than-strict priority to the former class by reserving a more-than-zero share of resources for the latter. Our approach is one way to convexify the (all-or-nothing) absolute- and no-prioritization alternatives. Though it sounds reminiscent of a reserves-based mechanism design approach, a key difference is the endogenous choice of classification membership through equilibrium actions of the agents. In addition, while our approach appears to constrain a "general mechanism design" approach, it turns out that in very competitive environments our restriction to RTC is without loss of generality (details to be discussed in Subsection 2.2).

Due to the potential for a poorly behaved optimization problem, we restrict attention to a stylized model in which there are two types of (many) potential recipients ("patients"). Types differ in their (i) benefit from taking a critical action (invasive medical "treatment") and (ii) marginal benefit from

receiving an object ("transplant"). What drives the planner to use actions as a basis for allocation decisions is a natural assumption of correlation in these benefits: patients who benefit most from treatment are those who marginally benefit most from transplants.[2] To allocate objects the planner uniformly rations some fraction k of them to the class of recipients who take the critical action, and rations the rest to the remaining recipients. The stark case of $k = 1$ strictly prioritizes all agents who take the critical action over those who do not, analogous to organ allocation applications where different treatment choices strictly prioritize otherwise identical patients. Lower values of k can be thought of a as partial prioritization, in that some objects are reserved for recipients who do not take the action. In some cases (Section 3) we can even replicate *no* prioritization—where the planner ignores action choices—by choosing an appropriately low level of k. In summary, k can be thought of as the "weight" that the action choice plays in the allocation decision. By increasing this weight the planner might improve object allocation but worsen decisions, as in the heart and liver transplant examples mentioned earlier.

The primary novelty of our contribution is to incorporate the level of competition (i.e., decentralization of action choices) into an analysis of this tradeoff between improved object allocation and distorted decision making. We first consider "perfectly competitive" environments (where recipients choose their own actions), showing that allocative efficiency is maximized when the planner maximizes the ration k subject to eliminating any distortions in decision-making. As a corollary we show that utilitarian welfare is maximized in the same way, resembling conclusions in related work by Braverman and Garg (2020) and Perez-Richet and Skreta (2022).

Next we consider a general model of "imperfect competition" where patients are partitioned into n "Transplant Centers," each of which chooses actions on behalf of its own patients. In a stylized way this captures the phenomenon in our leading examples that allocation is based on individual actions but these actions are *chosen* by an agent acting on behalf of numerous individuals. This model yields an atomic congestion game that, despite a restriction to two types and two "routes," is poorly behaved relative to other such games that have been studied. When equilibria in this setting are anal-

[2]Positive correlation simply determines which group of patients the planner prefers to receive transplants; the model could be solved symmetrically in the opposite case of negative correlation with analogous results.

ogous to those in the "perfect competition" model we generalize the above optimality results and show that welfare decreases in the level of competition (n). However we also demonstrate through numerical analysis that a second form of equilibria can occur that lead to highly distorted action choices. Additional, preliminary numerical analysis suggests that these latter equilibria yield lower welfare under plausible modeling assumptions.

1.1 Related Literature

The closest work to ours addresses questions on strategic classification within both computer science and economics. This literature has focused on the case in which individual recipients strategically decide their own actions, which corresponds to our baseline "perfect competition" model in Section 3. Our results in that section mirror some results in this literature as described below. We distinguish ourselves from this work when we allow for "imperfect competition" (Section 4), reducing the number of strategic agents that control potential recipients' actions.[3] Our contribution is to show the extent and limits to which results in the baseline model extend to the general one, and how this extension is impacted by the *level* of competition.

Generally speaking, the models in this literature have the following characteristics. Agents can, at some cost, misrepresent their privately known type. The planner wishes to correctly classify an agent's type as being above/below some threshold ("high/low"), while all agents desire a high classification. In a setup like this Hardt et al. (2016) consider a planner who maximizes classification accuracy against strategically misrepresenting agents, setting aside manipulation costs. Embedding this problem into a (machine) learning context, they provide efficient algorithms that are near-optimal with high probability in each of two cases: when the classification objective is known and when it first must be learned by the algorithm through existing data. In a related model, Milli et al. (2019) analyze the tradeoff between accuracy and the resulting manipulation costs imposed on (true) "high type" agents.

In a continuous-type version of our baseline model (Section 3), Braverman and Garg (2020) maximize accuracy net of agents' manipulation costs. Under some assumptions they show that optimal classifiers (i) typically re-

[3] An additional minor difference from some of this work is that we impose a "classification budget" representing the fixed supply of resources.

quire randomization, and (ii) induce no manipulation. Our setup necessarily induces randomness by nature of our budgeted rationing problem (see footnote 3) but its degree is determined endogenously: equilibrium behavior in our model determines the rate at which high-type agents receive favorable classification. Nevertheless we obtain an analogous result (Theorem 2) in the baseline model.

Perez-Richet and Skreta (2022) allow the planner to commit to a probabilistic testing function that maps (misrepresented) types into distributions over test outcomes. The planner then (optimally) uses equilibrium-induced test outcomes to make classification decisions. Under an increasing-returns assumption on misrepresentation costs accuracy-maximizing mechanisms have two characteristics. First they "raise the bar" by offering the greatest chance of high classification only to observed types above some artificially high threshold. Second, the only agents who achieve this threshold in equilibrium are those whose true type is above the true desired threshold. Other types engage in no misrepresentation, being compensated with just enough probability of high classification to offset the net benefit of misrepresenting.[4] As a next step, Perez-Richet and Skreta (2023) impose this no-misrepresentation condition as a constraint under which they find optimal mechanisms in the presence of an allocation budget constraint.

Other work further removed from ours examines variations on the question of mechanism, scoring, or ratings design under costly misrepresentation. Frankel and Kartik (2021) consider agents who vary both in their type as above and in their misrepresentation costs. This dual heterogeneity leads the planner to commit to under-weight information, thereby improving its accuracy in equilibrium. When the types are multidimensional Ball (2023) shows that the planner can benefit by under-weighting only some dimensions while over-weighting others. With a model and objectives that diverge from ours, Lee and Suen (2023) consider allocating university seats to students with high entrance exam scores obtained either naturally (high types) or through costly tutoring (low types). They show that an increase in seat scarcity can decrease the level of inefficient tutoring.

While the above papers (and ours) consider static settings, Munoz-Rodriguez (2024) considers how a planner might leverage inter-temporal incentives to

[4]Our Theorem 1 reaches the same conclusion in our baseline model where individual recipients are the agents. That result is also consistent with the fact that when there are only two types in their setting, optimal mechanisms eliminate manipulation entirely.

effectively assign transplant organs. Roughly speaking, his model can be interpreted as an overlapping generations version of our "perfect competition" model, but where types have either zero or infinite differences in values across actions. Deriving optimal dynamic mechanisms (i.e. using report *histories*), an implication is that biasing organ allocation towards high types is achieved by leveraging the low types' option value inherent in this dynamic setup. As in our perfect competition setting, it is not optimal to fully bias allocation towards high types.

Finally, our general model of (Section 4) can be viewed as a generalization of congestion games as pioneered by Wardrop (1952). Budgeted allocation through classification induces congestion: increasing one's allocation probability through misrepresentation necessarily reduces someone else's. Indeed, fixing the planner's rationing method in our baseline model of Section 3, equilibrium existence follows from that literature (Konishi, 2004). Of course our objective goes beyond this, so we describe the equilibrium and see how its welfare varies with the planner's choice of ration.

Corresponding to our general model (Section 4) a literature on *atomic congestion games* (ACG's) considers (atomic) agents who each control a *mass* of traffic to be routed through a network. Equilibrium existence and uniqueness results can be obtained in such games when they are sufficiently structured, e.g. if all traffic is of a single type and the network is sufficient simple (Bhaskar et al., 2015; Harks and Timmermans, 2018). Unfortunately, without such assumptions these results need not hold. The presence of two types of recipients in our model corresponds to a "two-traffic-type" model, meaning that not only might payoffs violate the common concavity assumption in this literature, but our payoff functions locally violate concavity everywhere. We are not aware of any equilibrium existence results from that literature that apply to our problem. Interestingly however, Wan (2012) considers simple ACG's with two nodes and one traffic type, showing that welfare improves when traffic is split amongst fewer atomic agents. Our Proposition 4 shows the degree to which that idea extends to the two-type case.

2 Model

2.1 Primitives

Though our model is necessarily stylized for tractability, we use terminology related to organ allocation for easier interpretation. There is a continuum of patients of two possible types: a mass $r_\ell > 0$ of low-type (ℓ) patients and a mass $r_h > 0$ of high-type (h) patients. For each patient there are two possible actions labelled N ("Non-treatment") and T ("Treatment"). We consider two scenarios regarding the *decisions* over these actions. In "perfectly competitive" environments (Section 3) each patient decides their own action. In less competitive environments (Section 4) a Transplant Center coordinates decisions across its multiple patients.

There is a positive mass $\phi < r_\ell + r_h$ of scarce objects which we call *organs*, to be assigned to patients in a manner described below.

A patient's welfare depends on their type, action, and whether they receive an organ. Any patient who receives an organ obtains welfare L^*. Otherwise, a type i patient who takes action d obtains welfare L_i^d. To capture the relevant tradeoffs we wish to consider we assume

$$L_h^N < L_h^T < L_\ell^T < L_\ell^N < L^* \tag{1}$$

Interpreting high types as the relatively "sicker" or "high risk" patients, these inequalities embody natural assumptions underlying our motivating examples. Most fundamentally, the Treatment action increases the welfare of high types and reduces the welfare of low types, setting aside any chance of receiving an organ. Therefore we call T a high type's **natural action**, and N a low type's **natural action**. Second is the assumption that high types receive higher marginal benefit from organs. This assumption plays a role only in our welfare statements, but also represents the real world objective of prioritizing high-risk patients (Persad et al., 2009). The assumption that L^* is independent of type is without loss of generality both strategically and in terms of welfare.[5] The assumption that L^* is independent of action represents the idea that treatments in these applications affect short-term hazard rates rather improve long-term quality of life; that is, conditional on

[5]Specifically, our model omits aggregate uncertainty for tractability. Therefore adding a constant to the welfare of each low type has no impact. Furthermore the assumption is a realistic approximation when, for example, L^* represents post-transplant expected life-years in liver (Schaubel et al., 2009) and heart (Meyer et al., 2015) transplantation.

surviving until transplant, any treatment effects are negligible relative to the value of receiving an organ.

2.2 Rationing through Classification

The planner observes patients' actions but not their types. Implicitly assuming that the planner cannot observe patients' identities (or equivalently, must treat patients with identical actions equally), the planner assigns some fraction k of the organs uniformly randomly among patients whose action is T, and similarly uniformly randomly assigns mass $(1-k)\phi$ of organs among those whose action is N. We define the process of *Rationing through Classification* as one where (i) the planner publicly commits to such a fraction $k \in [0,1]$, (ii) actions are decided for each patient (either selfishly in Section 3 or by their Transplant Center in Section 4), and (iii) agents being classified only by their actions, organs are rationed (uniformly randomly) within each classification. Our main question is how *the degree of prioritization of treatment*—modelled through the planner's choice of k—impacts the structure of equilibria and equilibrium welfare.

For example at one extreme, a planner might fully prioritize action T by choosing a maximal value of k. A naive planner (disregarding low types' incentives) might expect this choice to maximally assign organs to high types. Of course if low types are encouraged to choose T, welfare is lower both because of the inefficient choice of action and because fewer high types receive organs. This is analogous to the ICU example discussed in the Introduction, where otherwise identical agents would be strictly prioritized based on differing ICU status. At another extreme imagine a planner entirely ignoring the treatment decision, and randomly allocating organs to all agents with equal probability. The obvious tradeoff is that actions are chosen efficiently, though organs are assigned non-discriminantly. It turns out that, at least in sufficiently competitive environments, this outcome can be replicated in our setup by selecting the proportional value $\hat{k} \equiv r_h/(r_\ell + r_h)$.[6] For intermediate values of k the planner can more finely adjust these tradeoffs. We analyze them under varying degrees of competition.

At a technical level, our focus on RTC appears to be a restriction. More generally the planner could let the ration k depend on the realized profile of action choices. There is a 2-fold response to this point. First it turns out this

[6] With very low degrees of competition this need not hold; see Section 4.

restriction is without loss of generality in our baseline "perfect competition" model since infinitesimal agents are "price takers." That is, any equilibrium profile resulting in some value of k in a general mechanism corresponds to an equilibrium for k under RTC, and thus our restriction is without loss. Second, any general mechanism that is *not* RTC requires the planner to observe a profile of realized actions. While our model is indeed a static one for tractability, a more realistic dynamic setting would not allow a planner to observe an entire profile of actions before making allocation decisions. On the other hand, a mechanism that commits to some (probabilistic) ration k is feasible in any environment.

3 Perfect Competition

3.1 Equilibrium structure

Fixing any choice of k, we imagine each infinitesimal patient selfishly choosing their own action. A **strategy profile** $p = (p_\ell, p_h)$ describes the fractions of low- and high-types that choose action T. A non-wasteful profile p induces two **allocation probabilities** of receiving an organ for patients who have chosen either N or T:[7]

$$\pi^N(p) = \frac{(1-k)\phi}{(1-p_\ell)r_\ell + (1-p_h)r_h} \quad \pi^T(p) = \frac{k\phi}{p_\ell r_\ell + p_h r_h} \tag{2}$$

When p is clear from the context we may simply write π^N and π^T.

A patient's payoff is their expected welfare using the values in (1). A profile p is an **equilibrium** if it satisfies the usual incentive compatibility conditions for both types.

$$\begin{aligned}
p_\ell < 1 \quad &\implies \quad \pi^N L^* + (1-\pi^N)L_\ell^N \geq \pi^T L^* + (1-\pi^T)L_\ell^T \\
p_\ell > 0 \quad &\implies \quad \pi^N L^* + (1-\pi^N)L_\ell^N \leq \pi^T L^* + (1-\pi^T)L_\ell^T \\
p_h < 1 \quad &\implies \quad \pi^N L^* + (1-\pi^N)L_h^N \geq \pi^T L^* + (1-\pi^T)L_h^T \\
p_h > 0 \quad &\implies \quad \pi^N L^* + (1-\pi^N)L_h^N \leq \pi^T L^* + (1-\pi^T)L_h^T
\end{aligned} \tag{3}$$

Either action gives a patient some lottery between the organ payoff (L^*) and the relevant non-organ payoff. Observe that if either type has an incentive not to choose its "natural action" (N for low types, T for high types), it

[7] To avoid division by zero we can define $\pi^N(1,1) = 0$ and $\pi^T(0,0) = 0$, but these particular values are not significant in the analysis.

must be in exchange for a strictly higher probability of receiving an organ. Both types cannot simultaneously have such an incentive, which proves the following. (Formal proofs are in the appendix.)

Lemma 1. *If (p_ℓ, p_h) is an equilibrium then at least one type chooses its natural action with certainty, i.e. $p_\ell = 0$ or $p_h = 1$ (or both).*

It is intuitive that an increase in k should induce a greater number of patients to choose T, and that high types patients should be induced more easily than low types. Indeed in the perfect competition setting this intuition holds. With Lemma 1 this leads to the following description of equilibria.

Proposition 1 (Three regions)**.** *For any $k \in [0,1]$ there exists a unique equilibrium $p^*(k)$. It satisfies*

$$\begin{aligned} k < k' &\Longrightarrow p_\ell^*(k) = 0,\ p_h^*(k) < 1 && \textit{(biased toward N)} \\ k' \leq k \leq k^* &\Longrightarrow p_\ell^*(k) = 0,\ p_h^*(k) = 1 && \textit{(separating)} \\ k > k^* &\Longrightarrow p_\ell^*(k) > 0,\ p_h^*(k) = 1 && \textit{(biased toward T)} \end{aligned} \tag{4}$$

where

$$k' = \max\left\{0, \frac{r_h}{\phi}\frac{\phi(L^* - L_h^N) + r_l(L_h^N - L_h^T)}{r_h(L^* - L_h^N) + r_l(L^* - L_h^T)}\right\} \tag{5}$$

$$k^* = \min\left\{1, \frac{r_h}{\phi}\frac{\phi(L^* - L_l^N) + r_l(L_l^N - L_l^T)}{r_h(L^* - L_l^N) + r_l(L^* - L_l^T)}\right\} \tag{6}$$

Furthermore $p^*()$ is weakly increasing in k, and $k' < \frac{r_h}{r_\ell + r_h} < k^*$.

In fact $p^*()$ is constant only on $[k', k^*]$, i.e. strictly increasing elsewhere. It is possible that $k^* = 1$, i.e. that even if the T action is fully prioritized over N, all low types choose N in equilibrium. Similarly $k' = 0$ is possible. The proof of Proposition 1 implies the following.

$$k' > 0 \quad \Leftrightarrow \quad \frac{L_h^T - L_h^N}{L^* - L_h^N} < \frac{\phi}{r_l} \tag{7}$$

$$k^* < 1 \quad \Leftrightarrow \quad \frac{L_l^N - L_l^T}{L^* - L_l^T} < \frac{\phi}{r_h} \tag{8}$$

Intuition driving (8) is that low types are more easily induced to choose T (via an increase in k) when (i) organ supply is greater, (ii) competing high types are fewer, (iii) the cost to choosing T is lessened, and (iv) the benefit of receiving an organ conditional on choosing T is higher. Analogous intuition drives (7).

3.2 Equilibrium welfare

A change in k directly affects welfare by changing the fraction of organs allocated to low/high types. However it also changes equilibrium treatment decisions, affecting welfare not only through this distortion but through its further impact on organ allocation amongst the two types. The total effect on welfare, or even just on the fraction of organs assigned to high types, can be indeterminate.

An unambiguous case is when k induces separation profiles, i.e. $k \in [k', k^*]$. Since all patients choose natural actions, increasing k within this range merely increases the fraction of organs allocated to high types, increasing total welfare.

When $k > k^*$, an increase in k has two effects: a greater share of organs go to the Treatment group (which contains all high types) but more low types choose Treatment. It turns out that the latter effect always outweighs the former; an analogous result holds for $k < k'$.

Formally, for any $k \in [0, 1]$ denote the **equilibrium fraction of organs allocated to high types** as

$$f(k) = (1-k)\frac{(1-p_h(k))r_h}{(1-p_h(k))r_h + (1-p_\ell(k))r_\ell} + k\frac{p_h(k)r_h}{p_h(k)r_h + p_\ell(k)r_\ell}$$

where $p_\ell(k), p_h(k)$ is the unique equilibrium for k.

Theorem 1. *The fraction $f()$ of organs allocated to high types is*

- *decreasing in k for $k \in [0, k']$,*
- *increasing in k for $k \in [k', k^*]$, and*
- *decreasing in k for $k \in [k^*, 1]$.*

Furthermore $f()$ is maximized at k^*.

For an intuition imagine parameters for which, when $k = 1$, (i) *all* patients choose T in equilibrium, but (ii) each low type patient—facing a lottery between receiving an organ and receiving L_ℓ^T—is indifferent about choosing N instead. Each patient receives an organ with probability $\pi^T = \phi/(r_\ell + r_h)$. Now imagine decreasing k *and* changing the strategy profile so that (i) mass ϵ of low type patients instead choose N and (ii) mass $\pi^T\epsilon$ of organs are instead rationed to patients who choose N.

Observe that after this proportional change, actions N and T both induce the same allocation probability as before ($\phi/(r_\ell + r_h)$). Therefore the new profile cannot be an equilibrium (for the new k): conditional on *not* receiving an organ, N yields higher welfare. That is, in order to maintain equilibrium indifference, low types must "disproportionately follow the organs" that were reallocated to N. This would imply $\pi^N < \pi^T$ at the new equilibrium, and hence increase organ allocation to high types. Conversely, an increase in k would decrease $f()$. The argument extends to $[k^*, 1]$ and symmetrically to $[0, k']$.

Theorem 1 has immediate welfare implications. Within $[k', k^*]$ it is clear that welfare increases in k since actions remain fixed while $f()$ increases. Within $[k^*, 1]$, an increase in k (i) reduces $f()$ and (ii) increases the mass of low types choosing T, necessarily decreasing welfare.

For $k \in [0, k']$ on the other hand, an increase in k (i) decreases $f()$ but (ii) reduces the mass of high types choosing N. It turns out that either effect can dominate, breaking any symmetry with our argument for the case $[k^*, 1]$. Nevertheless, a simple argument relying on the monotonicity of f in this range can be used to prove that welfare for $k \in [0, k']$ is inferior to that at k^*.

Theorem 2. *Utilitarian welfare (the sum of patients' equilibrium payoffs) is*

- *increasing in k for $k \in [k', k^*]$,*
- *decreasing in k for $k \in [k^*, 1]$,*
- *maximized at k^* among all $k \in [0, 1]$.*

We next turn to our "imperfect competition" model, where each Transplant Center decides actions on behalf of its share of patients. Notably, the arguments showing the welfare inferiority of $k \in [0, k']$ in the proof of Theorem 2 do not extend to that model due to the fact that (the analog of) $f()$ need not be monotonic on $[0, k']$ as it was in Theorem 1. More significantly, even the analog of Lemma 1 fails to hold, complicating the description of equilibria. One of our main contributions is to show a form in which Theorem 2 extends to that model nevertheless.

4 Imperfect Competition

4.1 Atomic agents

We capture the idea of imperfect competition in the sense that a finite number of atomic agents choose actions on behalf of their own share of patients. Formally, there are n Transplant Centers (TC's). Each TC decides actions on behalf of its own mass r_ℓ/n of low-type patients and mass r_h/n of high-types. A **strategy** for TC i is a pair $p_i = (p_{i\ell}, p_{ih}) \in [0,1]^2$ describing the percentages of its low- and high-type patients assigned action T. A **strategy profile** $p = (p_i)_{i \in TC} = (p_{i\ell}, p_{ih})_{i \in TC}$. As is standard, p_{-i} denotes strategies for TC's other than i.

A TC's **payoff** (formalized below) is the total expected welfare of its patients as defined earlier (using (1)).[8] We continue to interpret the parameters in (1) as individual patient welfare, suggesting an interpretation of a utilitarian TC that weights only its own patients. However, since individual patients play no strategic role in this section, one can go well beyond that interpretation. For example, the welfare parameters in (1) could instead represent the profitability of treating or transplanting patients of either type. They could represent some combination of welfare and profits, or any (additive) cost or benefit the TC receives from different patient outcomes. Of course such interpretations might lead to different objectives for the planner, which we leave to future analysis.

To generalize concepts from earlier, for a given strategy profile p we denote the resulting **average treatment rates** by classification as

$$\bar{p}_\ell = \sum p_{i\ell}/n \qquad \bar{p}_h = \sum p_{ih}/n$$

Additionally fixing the planner's choice of k, the profile p induces the following two **allocation probabilities** analogous to Equation 2.

$$\pi^N = \min\left\{1, \frac{(1-k)\phi}{(1-\bar{p}_\ell)r_\ell + (1-\bar{p}_h)r_h}\right\} \qquad \pi^T = \min\left\{1, \frac{k\phi}{\bar{p}_\ell r_\ell + \bar{p}_h r_h}\right\}$$

Profile p is **non-wasteful (for k)** when there are no more organs than patients at N or at T:

$$(1-k)\phi \leq (1-\bar{p}_\ell)r_\ell + (1-\bar{p}_h)r_h \quad \text{and} \quad k\phi \leq \bar{p}_\ell r_\ell + \bar{p}_h r_h \tag{9}$$

[8] This sum is deterministic, as our model has no aggregate uncertainty.

Transplant Center i's payoff at profile p is

$$u_i(p) = \frac{1}{n}\Big[(1-p_{i\ell})r_\ell(\pi^N L^* + (1-\pi^N)L_\ell^N) + (1-p_{ih})r_h(\pi^N L^* + (1-\pi^N)L_h^N) \\ +p_{i\ell}r_\ell(\pi^T L^* + (1-\pi^T)L_\ell^T) + p_{ih}r_h(\pi^T L^* + (1-\pi^N)L_h^T)\Big] \tag{10}$$

In standard fashion, p_i is a **best response** to p_{-i} if $p_i \in \arg\max u_i(\cdot, p_{-i})$, and p is a (pure, Nash) **equilibrium** if p_i is a best response to p_{-i} for each i.

4.2 Equilibrium structure and intuition

Though we have numerical analysis suggesting that an equilibrium always exists across a wide range of primitives (Subsection 4.3) we have no analytical proof of equilibrium existence. The primary difficulty is that TC's payoffs need not be quasi-concave; indeed payoffs locally violate concavity everywhere.[9] Notably, much of the literature on atomic congestion games directly makes assumptions on payoff-concavity (rather than on primitives) which guarantee well-behaved problems with existence or uniqueness results. Our interest is in how primitives impact structure and welfare of equilibria without sweepint primitives "under the rug." The downside of this approach is our lack of a formal existence result. Nevertheless we provide partial results arguing that certain conditions are more prone to lead to the existence of certain kinds of equilibria (Subsection 4.3).

We turn to necessary conditions for equilibria. It is immediate that equilibria must be non-wasteful: at any wasteful profile, some TC would be able to change the actions of some patients in order to (i) guarantee those patients receive an otherwise wasted organ, and (ii) increase all other patients' chance of receiving an organ. Formal proof is omitted.

Lemma 2. *Fix k. If p is an equilibrium profile then it is non-wasteful.*

Next, despite the non-concavity of payoff functions, it turns out that equilibrium profiles must be symmetric (or payoff-equivalent to a symmetric equilibrium). Unlike in Section 3, however, there are potentially *two* kinds of equilibria under imperfect competition. In a "Non-inversion" equilibrium either all high types take action T or all low types take action N (analogous

[9] Proof available upon request.

to the conclusion of Lemma 1). In an "Inversion" equilibrium either all *low* types take action T or all *high* types take action N (or both).

Theorem 3. *Fix k and suppose p is an equilibrium. There exists an equilibrium p^* that is symmetric, is Pareto-equivalent to p, and satisfies one of the following.*

- *(Non-inversion) For every TC i, $p^*_{i\ell} = 0$ or $p^*_{ih} = 1$.*
- *(Inversion) For every TC i, $p^*_{i\ell} = 1$ or $p^*_{ih} = 0$.*

Inversion equilibria have the perverse characteristic that each TC is choosing action T for at least some of its low type patients and, simultaneously, choosing action N for at least some of its high type patients. For an intuition behind why this might occur consider TC i's best response when, rationally or not, i's competitors choose action T for a large percentage of their patients. First, excess congestion at T could conceivably cause i to choose action N for (at least some of) its *high* types in order to give those (high organ-value) patients better odds of an organ. Given this, i might also choose to avoid further congesting those high types by placing its *low* types at T, particularly if both the value of an organ and the value of N over T are low for those types. Finally, if doing so leads i to place a large percentage of patients at T we have constructed an equilibrium. The theorem does not specify whether or when either kind of equilibrium exists. Numerical analysis described in Subsection 4.3 demonstrates that, indeed, either or both types may exist, depending on parameters.

The proof of Theorem 3 has two components, one ruling out "interior" equilibria and the other deriving symmetry. The intuition behind the former is that, at any arbitrary strategy profile, all TC's face the same relative incentive to "swap" equal masses of opposite-type agents between actions N and T. Thus at any interior profile, all TC's would strictly prefer executing the same such swaps until reaching a corner solution (which is either a Non-inversion or Inversion strategy).

For an intuition behind symmetry, note that the set of, say, Non-inversion strategies is a monotonic, one-dimensional set: the decision of how many patients to send to Treatment, prioritizing high-type patients over low-types. If TC i sends fewer agents to Treatment than TC j, then i has a greater marginal incentive than j to send *additional* patients to Treatment since doing so crowds out fewer of i's own patients. Since both TC's should face

the same marginal incentive in equilibrium, they must choose symmetrically. The same argument applies to Inversion strategies.

4.3 Numerical analysis

Without an analytical result determining when Inversion (or Non-inversion) equilibria exist, we provide partial, additional insight with the use of some computations. Across a wide range of values for the primitives of our model we approximate equilibria under various rations of k and evaluate the resulting equilibrium welfare. Below we justify the values of primitives we consider as being those which should make Inversion equilibria *most* plausible under our modeling assumptions; namely we consider n and L^* to be relatively small.

Despite this choice of primitives we interpret our initial findings suggest two things. First, while Inversion equilibria can exist—even exclusively—they are atypical across the complete range of parameters. Second, even when they do exist, Inversion equilibria (at any value for k) are typically welfare-dominated by some Non-inversion equilibrium (at the value of k described in Theorem 4). In fact, whenever $2\phi < r_\ell + r_h$ (organs are relatively scarce), *none* of the economies we considered yields an Inversion equilibrium with higher welfare than the highest-welfare Non-inversion equilibrium. These conclusions lead us to focus on Non-inversion equilibria in Subsection 4.4.

We briefly describe the parameters we considered in deriving these conclusions. (Additional details are to be provided in a forthcoming online appendix.) Being restricted to a fine but discrete grid of values, we considered all (normalized) values of $r_\ell + r_h \equiv 1$ and $\phi < 1$. We focused specifically on the case $n = 3$ since (i) we require $n \geq 3$ in Subsection 4.4 while (ii) as n becomes large the model converges to one resembling Section 3 (where Inversion equilibria do not exist).

For patient welfare (Equation 1) we fix (normalize) the values of L_h^N and L_ℓ^N. We (conservatively) fix L^* close to (but slightly larger than) $2L_\ell^N - L_h^N$ because (i) this value is a strict lower bound for L^* in Assumption 1 while (ii) the problem we study becomes less interesting as L^* becomes arbitrarily large (since on a relative scale the distinction between N and T becomes negligible for the patients). With these values fixed, we consider all values (in a discrete grid) for L_h^T and L_ℓ^T satisfying Equation 1.

Across this entire range of parameter values we consider a values of $k \in (0, 1)$ roughly in increments of 0.1. For all such values we search for

symmetric profiles that are (approximately) equilibria and classify each as either Inversion or Non-inversion. Remarkably, every economy we considered yields (i) at least one (approximate) equilibrium, (ii) at most one of which is Inversion, and (iii) at most one which is Non-inversion.[10] In any economy where $\phi < (r_\ell + r_h)/2$, the highest equilibrium welfare is obtained under a Non-inversion equilibrium (for a ration k^* described in Theorem 4).

4.3.1 Plausibility of Inversion Equilibria

Intuition suggests that various conditions on primitives might make Inversion equilibria less plausible. Observe that for a given set of primitives an Inversion equilibrium cannot exist if, at all symmetric strategy profiles, a TC's payoff is decreasing in $p_{i\ell}$ (i.e. the corresponding partial derivative is negative); otherwise a TC would benefit from reducing any positive mass of low types assigned to N. While this condition is overly strong, it turns out that two kinds of parameter changes that *decrease* this partial derivative are consistent with a lower prevalence of Inversion equilibria in our numerical examples.

To formalize them, define the following differences.

$$\Delta_* = L^* - L_\ell^N \qquad \Delta_T = L_\ell^T - L_h^T$$
$$\Delta_\ell = L_\ell^N - L_\ell^T \qquad \Delta_h = L_h^T - L_h^N$$

The proof of the following result is available upon request.

Proposition 2. *At any symmetric, non-wasteful profile $\tilde{p}$ and holding all other Δ's constant, the partial $\partial u_i(\tilde{p})/\partial p_{i\ell}$ is*

- *decreasing in $\Delta_\ell = L_\ell^N - L_\ell^T$;*
- *increasing in $\Delta_h = L_h^T - L_h^N$.*

One might therefore suspect Inversion equilibria to be less prevalent when Δ_ℓ is relatively large or when Δ_h is relatively small. Even though Inversion equilibria are generally uncommon in our numerical analysis to begin with, we find that this pattern holds indeed. (Details will be added in a future version of the paper.)

In fact in the few cases where Inversion equilibria exist exclusively it is typical that Δ_T is large. Intuitively, such cases are those where TC's become

[10] The latter observation is consistent with Proposition 3; however we have no result corresponding to (ii) due to non-concavities in payoffs.

relatively indifferent about treatment decisions (N vs. T) for both of their patient types, and care only about which types get organs. Thus N and T become mere labels for two approximately equivalent actions. When $k = 0.5$ for example, any Non-inverting equilibrium is mirrored by a payoff-equivalent Inverting equilibrium by simply switching the actions of all patients.

4.4 Non-inversion equilibrium

As demonstrated by the numerical results of Subsection 4.3, the existence and uniqueness of a "non-inversion" kind of equilibrium obtained in the perfect competition setting (Section 3) does not generally extend to the imperfect competition setting. To begin with, however, we can show that at most Non-inversion equilibrium exists in the general case, and that the results of Theorem 1 and Theorem 2 extend to such equilibria when they exist in the imperfect competition case.

To do this we focus on symmetric, Non-inversion profiles that satisfy *local* IC constraints, namely, only checking deviations to nearby Non-inversion strategies.[11]

Definition 1 (NI-candidate)**.** *Fixing k, a symmetric profile p^* is an NI-candidate for k when any one of the following holds.*

- *(1-NI) $p^*_{i\ell} \equiv 0$ and $\frac{\partial u_i}{\partial p_{ih}}(p^*) \equiv 0$.*
- *(1-NI corner) $p^*_{i\ell} \equiv 0$, $p^*_{ih} r_h \equiv k\phi$, and $\frac{\partial u_i}{\partial p_{ih}}(p^*) \leq 0$.*
- *(2-NI) $p^*_{i\ell} \equiv 0$, $p^*_{ih} \equiv 1$, $\frac{\partial u_i}{\partial p_{ih}}(p^*) \geq 0$, and $\frac{\partial u_i}{\partial p_{i\ell}}(p^*) \leq 0$.*
- *(3-NI) $p^*_{ih} \equiv 1$ and $\frac{\partial u_i}{\partial p_{i\ell}}(p^*) \equiv 0$.*
- *(3-NI corner) $p^*_{ih} \equiv 1$, $(1 - p^*_{i\ell}) r_\ell \equiv (1-k)\phi$, and $\frac{\partial u_i}{\partial p_{i\ell}}(p^*) \geq 0$.*

The next result proving the uniqueness of NI-candidates implies there is at most one NI-equilibrium. Furthermore we show that NI-candidates satisfy an analog of Proposition 1: they are monotonic in k and yield separation whenever k lies within some intermediate range of values. Finally we prove an analog of Subsection 3.2 for NI candidates, deriving the value of k that yields the NI-candidate with higher welfare than any candidate for some other value of k.

[11] Definition 1 only checks FOC's, making it even weaker than this verbal definition; however proofs in the Appendix show the definitions are equivalent.

Some of these proofs require two mild assumptions: that $n \geq 3$ and that the benefit of receiving an organ exceeds any welfare differences between non-recipients.[12]

Assumption 1 (Valuable organs)**.** $L^* - L_\ell^N > L_\ell^N - L_h^N$.

Proposition 3 (Three NI regions)**.** *Fix $n \geq 3$ and suppose Assumption 1 holds. For any $k \in [0,1]$ there exists a unique NI-candidate $p^*(k)$. Furthermore $p^*()$ is weakly increasing in k, and*

$$\begin{aligned} k < k' &\implies \forall i,\ p_{i\ell}^*(k) = 0 \text{ and } p_{ih}^*(k) < 1 && \textit{(Region NI-1)} \\ k' \leq k \leq k^* &\implies \forall i,\ p_{i\ell}^*(k) = 0 \text{ and } p_{ih}^*(k) = 1 && \textit{(Region NI-2)} \\ k > k^* &\implies \forall i,\ p_{i\ell}^*(k) > 0 \text{ and } p_{ih}^*(k) = 1 && \textit{(Region NI-3)} \end{aligned}$$

where

$$k' = \max\left\{0, \frac{-(L_h^T - L_h^N) + \frac{\phi}{r_\ell}\left(\frac{n-1}{n}L^* + \frac{1}{n}L_\ell^N - L_h^N\right)}{\frac{\phi}{r_\ell}\left(\frac{n-1}{n}L^* + \frac{1}{n}L_\ell^N - L_h^N\right) + \frac{\phi}{r_h}\frac{n-1}{n}(L^* - L_h^T)}\right\} \tag{11}$$

$$k^* = \min\left\{1, \frac{(L_\ell^N - L_\ell^T) + \frac{\phi}{r_\ell}\frac{n-1}{n}(L^* - L_\ell^N)}{\alpha}\right\} > k' \tag{12}$$

$$\alpha = \frac{\phi}{r_\ell}\frac{n-1}{n}(L^* - L_\ell^N) + \frac{\phi}{r_h}\left[\frac{n-1}{n}L^* + \frac{1}{n}L_h^T - L_\ell^T\right] > 0 \tag{13}$$

It is easily verified that as $n \to \infty$, k' and k^* converge to their corresponding values in the perfect competition case of Proposition 1.

The derivation of k' and k^* in Proposition 3 is possible since, even though payoffs need not be generally concave, we prove a "limited concavity" result (Lemma 6). As a special case of it, when TC's use the separation strategy $p_{i\ell} \equiv 0$ and $p_{ih} \equiv 1$, each TC's payoff is concave in p_{ih} and either decreasing or concave in $p_{i\ell}$. Therefore k' and k^* can be derived only from the two partial derivatives of payoffs at that profile.

Our main result on NI-candidates under imperfect competition is only partially analogous to our results on perfect competition equilibria (Subsection 3.2). Analogously, both the fraction of organs allocated to high types and utilitarian welfare are single-peaked (at k^*) for values $k \in [k', 1]$. However neither is necessarily monotonic on $[0, k']$. Nevertheless, a more involved proof recovers our main conclusion: among all NI-candidates, both the fraction of organs to high types and total welfare are maximized at k^*.

[12] Some results require only the weaker assumption that $L^* - L_\ell^T > L_\ell^T - L_h^T$.

Theorem 4. *Fix $n \geq 3$ and suppose that Assumption 1 holds. Among all NI-candidates as a function of $k \in [0, 1]$, both the fraction of organs allocated to high types and the sum of TCs' payoffs are*

- *increasing in k for $k \in [k', k^*]$,*
- *decreasing in k for $k \in [k^*, 1]$, and*
- *maximized at k^*,*

where k^* is defined in Equation 12.

Under the planner's welfare-optimal choice of k^* (and the assumption that NI-candidate profiles are realized), optimal welfare is decreasing in the degree of competition, n. This is a corollary of Theorem 4 along with the following observation.

Proposition 4. *The fraction $k^*(n)$ (defined in Equation 12) is decreasing in n.*

5 Conclusion

By using agents' observable actions as a basis for resource allocation, a planner induces a tradeoff between (i) making better allocation decisions and (ii) distorting choice over recipients' actions. Though a planner might want to bias allocation in favor of recipients prone to take one particular action, doing so encourages other recipients to "wrongly" take that action. Furthermore the intensity of this tradeoff is impacted by the level of "competition"—the extent to which recipients' decisions are coordinated or centralized.

Evaluations of this tradeoff have led to organ allocation policy changes that have given or removed *full* prioritization of recipients who take a particular "treatment" action over those who do not. In order to offer a more nuanced approach that compromises between the binary choice between full- and no-prioritization, we introduce and analyze the concept of *rationing through classification*; a planner partially prioritizes recipients taking one action by setting aside only a fraction of resources for them, reserving the rest for the remaining recipients.

When we consider the special case of our model that yields the most-competitive environments (where action choices are fully decentralized) the

planner maximizes aggregate welfare with a rationing of resources that eliminates distorted actions. This special case of our model and the flavor of the result are reminiscent of related work in the literature.

Our broader focus, and the novelty of our approach, is to consider less-competitive environments where an agent (e.g. a Transplant Center) chooses actions on behalf of many patients. The general model yields a so-called atomic congestion game that fails to satisfy standard payoff assumptions (e.g. quasi-concavity) that would yield equilibrium existence or uniqueness results. Indeed, in addition to equilibria analogous to those in more competitive environments, this model allows for the theoretical possibility of an "inverted" form of equilibria in which action choices are severely distorted. Numerical analysis confirms that such equilibria can exist.

Despite this, our preliminary numerical investigation also suggests that these latter type of equilibria are (i) atypical, being ruled out by reasonable assumptions on primitives, and (ii) typically welfare-dominated by the more intuitive form of equilibria analogous to those necessarily describing behavior in more competitive environments. Focusing on these intuitive, "Non-inversion" equilibria we generalize the above baseline result by characterizing the welfare-maximizing rationing level under such equilibria, showing that it eliminates distortions, and show that welfare decreases in the degree of competition.

This work underscores at least two considerations for policy makers who design institutions for resource allocation based on action choices. First the model demonstrates how a rationing approach can serve as a welfare-improving compromise between the all-or-nothing decision of whether to fully prioritize one class of recipients over another. Second, the novelty of considering both high and low competition levels in our model highlights the importance of considering the degree decentralized decision-making when designing such institutions. In allocation systems where agents—such as Transplant Centers—take allocation-relevant actions on behalf of many potential recipients, policy needs to be tailored to the competitive environment. Indeed, despite the political difficulty of doing so, this work even demonstrates how optimal allocation policy should differ across regions of varying competitiveness.

References

Ian Ball. Scoring strategic agents, 2023.

Umang Bhaskar, Lisa Fleischer, Darrell Hoy, and Chien-Chung Huang. On the uniqueness of equilibrium in atomic splittable routing games. *Mathematics of Operations Research*, 40(3):634–654, 2015. URL http://www.jstor.org/stable/24540968.

Mark Braverman and Sumegha Garg. The role of randomness and noise in strategic classification, 2020.

Alex Frankel and Navin Kartik. Improving Information from Manipulable Data. *Journal of the European Economic Association*, 20(1):79–115, 2021. doi: 10.1093/jeea/jvab017. URL https://doi.org/10.1093/jeea/jvab017.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, pages 111–122, 2016.

Tobias Harks and Veerle Timmermans. Uniqueness of equilibria in atomic splittable polymatroid congestion games. *Journal of Combinatorial Optimization*, 36:812–830, 2018.

Hideo Konishi. Uniqueness of user equilibrium in transportation networks with heterogeneous commuters. *Transportation Science*, 38(3):315–330, 2004. URL http://www.jstor.org/stable/25769203.

Frances Xu Lee and Wing Suen. Gaming a selective admissions system. *International Economic Review*, 64(1):413–443, 2023. doi: https://doi.org/10.1111/iere.12591. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/iere.12591.

D. M. Meyer, J. G. Rogers, L. B. Edwards, E. R. Callahan, S. A. Webber, M. R. Johnson, J. D. Vega, M. J. Zucker, and J. C. Cleveland, Jr. The future direction of the adult heart allocation system in the united states. *American Journal of Transplantation*, 15(1):44–54, 2015.

Smitha Milli, John Miller, Anca D Dragan, and Moritz Hardt. The social cost of strategic classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 230–239, 2019.

Edwin Munoz-Rodriguez. Ethical organ allocation under dynamic asymmetric information, 2024.

William F. Parker, Allen S. Anderson, Donald Hedeker, Elbert S. Huang, Edward R. Garrity, Jr, Mark Siegler, and Matthew M. Churpek. Geographic variation in the treatment of u.s. adult heart transplant candidates. *Journal of the American College of Cardioloy*, 71(16):1715–1725, 2018.

Eduardo Perez-Richet and Vasiliki Skreta. Test design under falsification. *Econometrica*, 90(3):1109–1142, 2022. doi: https://doi.org/10.3982/ECTA16346. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA16346.

Eduardo Perez-Richet and Vasiliki Skreta. Fraud-proof non-market allocation mechanisms. 2023. URL http://eduardo.perez.free.fr/papers/FalsproofAllocationV4.3.pdf.

Govind Persad, Alan Wertheimer, and Ezekiel J Emanuel. Principles for allocation of scarce medical interventions. *The Lancet*, 373(9661):423–431, 2009.

Gege Ran, Kevin Chung, Allen S. Anderson, Robert D. Gibbons, Nikhil Narang, Matthew M. Churpek, and William F. Parker. Between-center variation in high-priority listing status under the new heart allocation policy. *American Journal of Transplantation*, 2021.

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

Jason Snyder. Gaming the liver transplant market. *Journal of Law, Economics, & Organization*, 26(3):546–568, 2010. URL http://www.jstor.org/stable/40925623.

Cheng Wan. Coalitions in nonatomic network congestion games. *Mathematics of Operations Research*, 37(4):654–669, 2012. URL http://www.jstor.org/stable/23358641.

John Glen Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3):325–362, 1952.

doi: 10.1680/ipeds.1952.11259. URL https://doi.org/10.1680/ipeds.1952.11259.

6 Proofs Appendix

6.1 Perfect Competition

Proof of Lemma 1. If $p_\ell > 0$ then low types weakly prefer choosing Treatment:

$$\pi^N L^* + (1-\pi^N)L_\ell^N \leq \pi^T L^* + (1-\pi^T)L_\ell^T$$

Since $L^* > L_\ell^N > L_\ell^T$ (and $\min\{\pi^N, \pi^T\} < 1$) this would imply $\pi^N < \pi^T$. Similarly $p_h < 1$ would imply $\pi^N > \pi^T$. Hence $p_\ell = 0$ or $p_h = 1$. □

Lemma 3. *For any k there is a unique equilibrium $p_\ell(k), p_h(k)$. Furthermore $p_\ell()$ and $p_h()$ are weakly increasing in k.*

Proof of Lemma 3. For any k, equilibrium existence follows from standard arguments and is omitted. To prove uniqueness and monotonicity, fix $k, \tilde{k}$ with $k \leq \tilde{k}$ and let (p_ℓ, p_h) and $(\tilde{p}_\ell, \tilde{p}_h)$ be arbitrary equilibria for k and $\tilde{k}$ respectively, with allocation probabilities $\pi^N, \pi^T, \tilde{\pi}^N, \tilde{\pi}^T$. We show monotonicity $(p_\ell, p_h) \leqq (\tilde{p}_\ell, \tilde{p}_h)$ which also implies uniqueness ($k = \tilde{k}$).

Claim: either $(p_\ell, p_h) \leqq (\tilde{p}_\ell, \tilde{p}_h)$ or $(p_\ell, p_h) \geqq (\tilde{p}_\ell, \tilde{p}_h)$. If $p_\ell = \tilde{p}_\ell$ (or $p_h = \tilde{p}_h$) the claim follows immediately. If $p_\ell < \tilde{p}_\ell$ then Lemma 1 implies $\tilde{p}_h = 1 \geq p_h$. Similarly $p_\ell > \tilde{p}_\ell$ implies $p_h = 1 \geq \tilde{p}_h$, proving the claim.

Claim: $(p_\ell, p_h) \leqq (\tilde{p}_\ell, \tilde{p}_h)$. First suppose instead that $p_\ell > \tilde{p}_\ell$ and hence $p_h \geq \tilde{p}_h$. Since $k \leq \tilde{k}$ this implies $\pi^N > \tilde{\pi}^N$ and $\pi^T < \tilde{\pi}^T$. Since p is an equilibrium for k, low types weakly prefer Treatment in that equilibrium:

$$\pi^T L^* + (1-\pi^T)L_\ell^T \geq \pi^N L^* + (1-\pi^N)L_\ell^N$$

This implies a strict such preference at $\tilde{p}$ under $\tilde{k}$:

$$\tilde{\pi}^T L^* + (1-\tilde{\pi}^T)L_\ell^T > \tilde{\pi}^N L^* + (1-\tilde{\pi}^N)L_\ell^N$$

This strict preference requires $\tilde{p}_\ell = 1$ in equilibrium, contradicting $p_\ell > \tilde{p}_\ell$. Supposing $p_h > \tilde{p}_h$ leads to a similar contradiction. □

Proof of Proposition 1. By Lemmas 1 and 3 there exist $0 \leq k' \leq k^* \leq 1$ that define the three cases of Equation 4. When $k = \bar{k}r_h/(r_\ell + r_h)$, a separating profile ($p_\ell = 0$, $p_h = 1$) yields $\pi^N = \pi^T$, so ($p_\ell = 0$, $p_h = 1$) is an equilibrium where each agent has *strict* incentive to choose their natural action. By continuity this would hold for small perturbations of k, thus

$$k' < \frac{r_h}{r_\ell + r_h} < k^* \tag{14}$$

Next, by continuity, k' is the lowest value of k at which the separation profile ($p_\ell = 0$, $p_h = 1$) induces a high type to choose T, i.e. at which

$$\pi^T L^* + (1 - \pi^T)L_h^T \geq \pi^N L^* + (1 - \pi^N)L_h^N$$

Substituting $\pi^N = (1 - k)\phi/r_\ell$ and $\pi^T = k\phi/r_h$ this becomes

$$k \geq \frac{(L_h^N - L_h^T) + \frac{\phi}{r_\ell}\phi(L^* - L_h^N)}{\frac{\phi}{r_h}(L^* - L_h^T) + \frac{\phi}{r_\ell}(L^* - L_h^N)}$$

which yields k' as in (5). Thus $k' > 0$ whenever $\phi(L^* - L_h^N) + r_l(L_h^N - L_h^T) > 0$ yielding (7).

Low types are induced to choose N at the separation profile when

$$\pi^T L^* + (1 - \pi^T)L_\ell^T \leq \pi^N L^* + (1 - \pi^N)L_\ell^N$$

i.e. simply reversing the above inequality and changing h to ℓ. Similarly this leads to (6) and (8). □

Proof of Theorem 1. The result is obvious in the range $k \in [k', k^*]$ where $p_l(k) \equiv 0$, $p_h(k) \equiv 1$, and hence $f(k) \equiv k$.

For any $k \in (k^*, 1)$, Proposition 1 implies $p_\ell(k) > 0$ and $p_h(k) = 1$; furthermore $p_\ell(k) < 1$ (otherwise a low type guarantees an organ deviating to N). This implies an equilibrium indifference condition for low types. Writing equilibrium allocation probabilities π^N, π^T as functions of k, it is

$$\pi^T(k)L^* + (1 - \pi^T(k))L_\ell^T = \pi^N(k)L^* + (1 - \pi^N(k))L_\ell^N, \text{ or}$$
$$\frac{1 - \pi^N(k)}{1 - \pi^T(k)} = \frac{L^* - L_\ell^T}{L^* - L_\ell^N} > 1$$

where $L_\ell^T < L_\ell^N$ implies the inequality. Therefore $\pi^N(k) < \pi^T(k)$, and $\pi^N(k), \pi^T(k)$ vary in the same direction with a change in $k \in (k^*, 1)$. We show $\pi^T(k)$ (hence f) is decreasing on this range.

Fix $k^* < k < k + \epsilon < 1$ and let $\delta = p_\ell(k+\epsilon) - p_\ell(k) \geq 0$. If instead we have $\frac{k\phi+\epsilon\phi}{p_\ell(k)r_\ell+\delta r_\ell+r_h} = \pi^T(k+\epsilon) \geq \pi^T(k) = \frac{k\phi}{p_\ell(k)r_\ell+r_h}$ then $(\epsilon\phi)/(\delta r_\ell) \geq \pi^T(k) > \pi^N(k)$. This also means $\frac{(1-k)\phi}{(1-p_\ell(k))r_\ell} = \pi^N(k) > \pi^N(k+\epsilon) = \frac{(1-k)\phi-\epsilon\phi}{(1-p_\ell(k))r_\ell-\delta r_\ell}$. (In words, if an increase in k moves "disproportionately" few low types to T to increase π^T, this must decrease $\pi^N < \pi^T$.) This contradicts the fact that π^N, π^T covary; the indifference condition cannot hold at $k+\epsilon$. Therefore (with continuity arguments) π^T decreases in $k \in [k^*, 1]$.

A symmetric argument applies to $k \in [0, k']$ (where $\pi^N > \pi^T$). An increase in k disproportionately increases p_h, increasing π^N, the rate at which *low* types receive organs, hence decreasing $f()$. □

Proof of Theorem 2. On the interval $[k', 1]$, welfare is clearly single-peaked (with peak at k^*) following the arguments made in the text. The rest of the proof covers $[0, k']$.

At $k = 0$ we know that (i) all organs go to the agents choosing N, (ii) all low types choose N ($p_\ell = 0$), and (iii) at most all high types choose N ($p_h \leq 1$). Thus the fraction of organs going to high types at $k = 0$ is

$$f(0) = \frac{(1 - p_h(0))r_h}{(1 - p_h(0))r_h + r_\ell} \leq \frac{r_h}{r_h + r_\ell} \equiv \bar{k}$$

i.e. high types receive less than their "proportional share" $\bar{k}$.

At $k = k^*$, agents use a separating profile and thus $f(k^*) = k^* > \bar{k}$ (where the inequality follows (14)). Thus when comparing k^* to $k = 0$, (i) high types receive more organs and (ii) treatment decisions are more efficient. Welfare is thus higher at k^*.

Finally the same conclusion can be drawn for any $k \in (0, k']$: By Theorem 1 high types receive even fewer organs at such k than at $k = 0$, and thus fewer than at k^*. Furthermore treatment decisions remain less efficient than at k^*. Therefore welfare is higher at k^* than at any $k \in [0, k']$. □

6.2 Imperfect Competition

6.2.1 Symmetry, I/NI equilibria

The proof of symmetry and the non-inversion/inversion structure of equilibria involves supplemental results—helpful in later proofs—that focus on a particular kind of deviation by a TC i.

Fix a non-wasteful profile p at which i is "double mixing," i.e. $p_i \in (0,1)^2$, resulting in allocation probabilities π^N, π^T. Consider any (non-wasteful) deviation from p, where i moves $\epsilon > 0$ mass of low types from T to N and ϵ mass of high types from N to T. This results in the strategy

$$(p'_{i\ell}, p'_{ih}) = (p_{i\ell} - \epsilon/r_\ell, p'_{ih} + \epsilon/r_h)$$

Note that this deviation does not change the total masses of patients assigned actions N and T; therefore it also does not change π^N and π^T. Therefore this deviation has no effect on other TCs' payoffs *regardless* of what strategies p'_{-i} they are using. In addition, the deviation impacts TC i's payoff only in how it affects the (expected) welfare of the 2ϵ mass of patients that were swapped:[13]

$$\begin{aligned}
&\epsilon[\text{Payoff effect moving a low type } T \to N \text{ and a high type } N \to T] \\
&= \epsilon[L_\ell^N + \pi^N(L^* - L_\ell^N) - L_\ell^T - \pi^T(L^* - L_\ell^T) \\
&\qquad + L_h^T + \pi^T(L^* - L_h^T) - L_h^N - \pi^N(L^* - L_h^N)] \\
&= \epsilon[(1-\pi^N)(L_\ell^N - L_h^N) + (1-\pi^T)(L_h^T - L_\ell^T)] \qquad (15)
\end{aligned}$$

The effect can be read as the combination of facts that, among i's patients who fail to receive an organ, some who were assigned to N turn from high types into low types and some who were assigned to T turn from low types into high types. The convex combination of these positive and negative effects could have any sign. Regardless, it means that a best response is either a corner solution or is Pareto-equivalent to one.

Lemma 4 (No double-mixing is w.l.o.g.)**.** *Fix k, a TC i, and a profile p at which p_i is a best response to p_{-i}. There exists $p'_i \in [0,1]^2 \setminus (0,1)^2$ such that*

(i) p'_i is a best response to p_{-i}, and

(ii) for any TC j and any p'_{-i}, $u_j(p_i, p'_{-i}) = u_j(p'_i, p'_{-i})$.

Proof of Lemma 4. Fix i and p as in the Lemma. Suppose (15) is positive. Then i would have the incentive to swap equal masses of low types at T and high types at N if feasible. Since p_i is a best response this must be infeasible: either $p_{i\ell} = 0$ or $p_{ih} = 1$. Similarly if (15) is negative then $p_{i\ell} = 1$ or $p_{ih} = 0$. In either case letting $p'_i = p_i$ proves the result.

[13] The resulting linearity in ϵ means that payoff functions are ruled surfaces.

Suppose (15) is zero. If $p_i \notin (0,1)^2$, again setting $p'_i = p_i$ proves the result. Otherwise we can let $(p'_{i\ell}, p'_{ih}) = (p_{i\ell} - \epsilon/r_\ell, p'_{ih} + \epsilon/r_h)$ where, by choosing ϵ maximally, $p'_i \notin (0,1)^2$. Since (15) is zero, p'_i is also a best response to p_{-i}. Furthermore this deviation preserves i's masses of patients assigned N and T, so for any TC j and any p'_{-i}, $u_j(p_i, p'_{-i}) = u_j(p'_i, p'_{-i})$. □

Observation 1. *The following facts about Equation 15 are helpful.*

(i) *The negative coefficient $(L_h^T - L_\ell^T)$ is of lower absolute magnitude than $(L_\ell^N - L_h^N)$. Hence $\pi^N = \pi^T$ implies that (15) is positive.*

(ii) *$\pi^N = 1$ ($\pi^T = 1$) implies that (15) is negative (positive).*

(iii) *When π^n increases and π^T decreases (e.g. by increasing the total mass of agents receiving Treatment) (15) decreases.*

This narrows the search for i's best responses to an arbitrary p_{-i} as follows, and illustrated in Figure 1.

- By Lemma 2 and Lemma 4 we restrict attention to non-wasteful strategies at the edges of $[0,1]^2$.
- The downward sloping (dashed) lines represent "swaps" of the kind described above; equivalently these are isoquants along which π^N and π^T are constant.
- Moving to the upper-left along any such line changes i's payoff at a rate described by (15).
- That rate (15) decreases as $(p_{i\ell}, p_{ih})$ increases.
- In any region where (15) is positive a best response must be "non-inverting."
- In any region where (15) is negative a best response must be "inverting."
- The set of candidate best responses is therefore reduced to the set highlighted in yellow.

These arguments prove the following part of Theorem 3.

Lemma 5. *Fix k. For any equilibrium profile p there exists a Pareto-equivalent equilibrium p^* for which one of the following is true.*

- *(Non-inversion) For every TC i, $p^*_{i\ell} = 0$ or $p^*_{ih} = 1$.*
- *(Inversion) For every TC i, $p^*_{i\ell} = 1$ or $p^*_{ih} = 0$.*

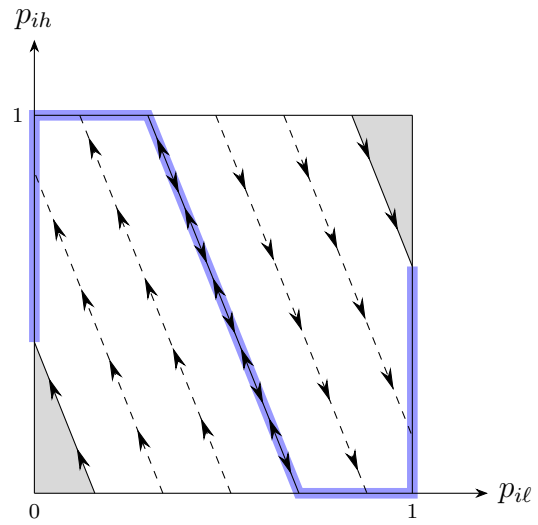

Figure 1: Candidate best responses are in blue; arrows indicate directions along which expression (15) increases. The triangular "wasteful" regions and the double-arrow segment may not exist depending on parameters.

Proof. As in the proof of Lemma 4, if (15) is non-zero, either all TC's are using Non-inversion strategies or all are using Inversion strategies (depending on its sign). In the special case that (15) is zero for some interior equilibrium, we can construct a Pareto-equivalent equilibrium while shifting each TC's strategy to, say, a Non-inversion strategy as in Lemma 4. □

The remainder of our proofs are more concise when expressing strategies and payoffs in terms of masses (rather than percentages) of organs and patients. Fixing some k we denote the masses of organs rationed to N and T as

$$\phi_N = (1-k)\phi \qquad \phi_T = k\phi$$

Similarly when a strategy profile p is clear from context we write

$$\begin{aligned} A_i &= (1-p_{i\ell})r_\ell/n & D_i &= p_{i\ell}r_\ell/n \\ B_i &= (1-p_{ih})r_h/n & E_i &= p_{ih}r_h/n \\ C_i &= \sum_{j\neq i}[(1-p_{j\ell})r_\ell/n + (1-p_{jh})r_h/n] & F_i &= \sum_{j\neq i}[p_{j\ell}r_\ell/n + p_{jh}r_h/n] \end{aligned}$$

Here A_i, B_i, and C_i are i's low types, i's high types, and i's competitors' patients that take action N; D_i, E_i, and F_i correspond to T.

TC i's payoff as given in (10) becomes

$$A_i L_\ell^N + \frac{A_i}{A_i+B_i+C_i}\phi_N(L^* - L_\ell^N) + B_i L_h^N + \frac{B_i}{A_i+B_i+C_i}\phi_N(L^* - L_h^N)$$
$$+ D_i L_\ell^T + \frac{D_i}{D_i+E_i+F_i}\phi_T(L^* - L_\ell^T) + E_i L_h^T + \frac{E_i}{D_i+E_i+F_i}\phi_T(L^* - L_h^T) \tag{16}$$

keeping in mind that $A_i = r_\ell/n - D_i$ and $B_i = r_h/n - E_i$.

While a TC's payoff is not generally concave in p_i, it is concave with respect to p_{ih}; under certain conditions it can be concave with respect to $p_{i\ell}$ along edges of the strategy space. The proof of the following lemma also provides the partial derivatives of payoffs, (17) and (18), utilized in later proofs.

Lemma 6 (Limited concavity.)**.** *Fix k, a non-wasteful profile p, and a TC i.*

(i) $u_i(p)$ is concave in p_{ih}.

(ii) If $p_{ih} = 1$ then $u_i(p)$ either is decreasing or is concave in (non-wasteful) $p_{i\ell} \in [0,1]$.

(iii) If Assumption 1 holds, $n \geq 2$, and $p_{jh} = p_{kh}$ for all j,k,[14] then $u_i(p)$ is concave in $p_{i\ell}$.

Proof. To prove (i) we show (16) is concave in E_i. Omitting subscript i, its derivative with respect to E (noting $B = r_h/n - E$) is

$$\begin{aligned}\frac{\partial u_i}{\partial E} &= \frac{A}{(A+B+C)^2}\phi_N(L^* - L_\ell^N) - L_h^N - \frac{A+C}{(A+B+C)^2}\phi_N(L^* - L_h^N) \\ &\quad - \frac{D}{(D+E+F)^2}\phi_T(L^* - L_\ell^T) + L_h^T + \frac{D+F}{(D+E+F)^2}\phi_T(L^* - L_h^T) \\ &= (L_h^T - L_h^N) + \frac{\phi_N}{A+B+C}\left(\frac{A}{A+B+C}(L_h^N - L_\ell^N) - \frac{C}{A+B+C}(L^* - L_h^N)\right) \\ &\quad + \frac{\phi_T}{D+E+F}\left(\frac{D}{D+E+F}(L_\ell^T - L_h^T) + \frac{F}{D+E+F}(L^* - L_h^T)\right) \\ &= (L_h^T - L_h^N) - \phi_N\frac{A(L_\ell^N - L_h^N) + C(L^* - L_h^N)}{(A + r_h/n - E + C)^2} + \phi_T\frac{D(L_\ell^T - L_h^T) + F(L^* - L_h^T)}{(D+E+F)^2}\end{aligned} \tag{17}$$

[14] In fact the combination $C_i \geq r_h/n$ and $p_{ih} = 0$ is sufficient.

Since all bracketed terms are positive, (17) is decreasing in E. Therefore u_i is concave in E (i.e. in p_{ih}).

To show (ii) and (iii), the derivative of (16) with respect to D is

$$\begin{aligned}\frac{\partial u_i}{\partial D} &= (L_\ell^T - L_\ell^N) + \frac{\phi_N}{A+B+C}\left(\frac{B}{A+B+C}(L_\ell^N - L_h^N) - \frac{C}{A+B+C}(L^* - L_\ell^N)\right) \\ &\quad + \frac{\phi_T}{D+E+F}\left(\frac{E}{D+E+F}(L_h^T - L_\ell^T) + \frac{F}{D+E+F}(L^* - L_\ell^T)\right) \\ &= \underbrace{(L_\ell^T - L_\ell^N)}_{\text{treatment effect}} + \underbrace{\phi_N \frac{B(L_\ell^N - L_h^N) - C(L^* - L_\ell^N)}{(r_\ell/n - D + B + C)^2}}_{N\text{-reallocation effect}} + \underbrace{\phi_T \frac{-E(L_\ell^T - L_h^T) + F(L^* - L_\ell^T)}{(D+E+F)^2}}_{T\text{-reallocation effect}}\end{aligned} \tag{18}$$

While the treatment effect is negative, the overall sign of (18) depends on the signs of two "reallocation effects." Denote

$$X = B(L_\ell^N - L_h^N) - C(L^* - L_\ell^N) \qquad X' = -E(L_\ell^T - L_h^T) + F(L^* - L_\ell^T) \tag{19}$$

When $X > 0$ ($X < 0$) the "N-reallocation effect" is convex and increasing in D (concave, decreasing in D); when $X' > 0$ ($X' < 0$) the "T-reallocation effect" is convex and decreasing in D (concave, increasing in D).

Under Assumption 1 one can show $X < X'$; this yields three cases.

- $X < X' \leq 0$: it is immediate that (18) is negative, so u_i is decreasing in D (i.e. in $p_{i\ell}$).
- $X \leq 0 < X'$: both reallocation effects are decreasing in D so (18) is decreasing in D; hence u_i is concave in D (in $p_{i\ell}$).
- $0 < X < X'$: both treatment effects are positive and convex in D, but change in opposite directions with respect to D. Therefore (18)'s sign and its direction of change w.r.t. D are indeterminate.

To prove statement (ii) of the lemma observe that if $p_{ih} = 1$ ($B = 0$) then $X < 0$ yielding the first and second cases above.

To prove (iii) observe that if $p_{ih} = p_{jh}$ for all j then $B \leq (n-1)C$ and $E \leq (n-1)F$. If $n \geq 2$ then Assumption 1 implies $X < 0 < X'$ yielding the second case above. □

The next lemma implies the intuitive idea that a TC who has already sent more patients to Treatment than another derives lower marginal benefit from sending additional patients of either type to Treatment due to crowding out more of its own patients.

Lemma 7. *For any k, any TCs i and j, and any non-wasteful profile p,*

$$\begin{aligned}\frac{\partial u_i}{\partial D_i} - \frac{\partial u_j}{\partial D_j} &= \frac{\partial u_i}{\partial E_i} - \frac{\partial u_j}{\partial E_j} \\ &= (D_j - D_i)\left[\frac{\phi_N(L^* - L_\ell^N)}{[r_\ell + r_h - (D_i + E_i + F_i)]^2} + \frac{\phi_T(L^* - L_\ell^T)}{(D_i + E_i + F_i)^2}\right] \\ &\quad + (E_j - E_i)\left[\frac{\phi_N(L^* - L_h^N)}{[r_\ell + r_h - (D_i + E_i + F_i)]^2} + \frac{\phi_T(L^* - L_h^T)}{(D_i + E_i + F_i)^2}\right] \qquad (20)\end{aligned}$$

Proof. Rewriting the partial Equation 17 with $A_i = r_\ell/n - D_i$ and $C_i = (n-1)(r_\ell/n + r_h/n) - F_i$,

$$\begin{aligned}\frac{\partial u_i}{\partial E_i} = (L_h^T - L_h^N) - \phi_N \frac{(r_\ell/n - D_i)(L_\ell^N - L_h^N) + ((n-1)(r_\ell/n + r_h/n) - F_i)(L^* - L_h^N)}{(r_\ell + r_h - D_i - E_i - F_i)^2} \\ + \phi_T \frac{D_i(L_\ell^T - L_h^T) + F_i(L^* - L_h^T)}{(D_i + E_i + F_i)^2} \qquad (21)\end{aligned}$$

An analogous expression holds for j. Since $D_i + E_i + F_i = D_j + E_j + F_j$ (the total mass of patients receiving Treatment is fixed), the two denominators in (21) are the same as those in the analogous expression for j. Hence

$$\begin{aligned}\frac{\partial u_i}{\partial E_i} - \frac{\partial u_j}{\partial E_j} = \phi_N \frac{(D_i - D_j)(L_\ell^N - L_h^N) + (F_i - F_j)(L^* - L_h^N)}{(r_\ell + r_h - D_i - E_i - F_i)^2} \\ + \phi_T \frac{(D_i - D_j)(L_\ell^T - L_h^T) + (F_i - F_j)(L^* - L_h^T)}{(D_i + E_i + F_i)^2}\end{aligned}$$

Since $F_i - F_j = -(D_i - D_j) + (E_j - E_i)$,

$$\begin{aligned}\frac{\partial u_i}{\partial E_i} - \frac{\partial u_j}{\partial E_j} = \phi_N \frac{(D_i - D_j)(L_\ell^N - L_h^N - L^* + L_h^N) + (E_j - E_i)(L^* - L_h^N)}{(r_\ell + r_h - D_i - E_i - F_i)^2} \\ + \phi_T \frac{(D_i - D_j)(L_\ell^T - L_h^T - L^* + L_h^T) + (E_j - E_i)(L^* - L_h^T)}{(D_i + E_i + F_i)^2}\end{aligned}$$

which equals (20).

A parallel argument shows that $\partial u_i/\partial D_i - \partial u_j/\partial D_j$ equals this same expression. □

Proof of Theorem 3. By Lemma 5 it is without loss to restrict attention to Inverting and Non-inverting equilibria. Consider any Non-inverting equilibrium profile p. Observe that for any i, j, either $(p_{i\ell}, p_{ih}) \geq (p_{j\ell}, p_{jh})$ or $(p_{i\ell}, p_{ih}) \leq (p_{j\ell}, p_{jh})$.

Suppose $(p_{i\ell}, p_{ih}) \lneq (p_{j\ell}, p_{jh})$, i.e. using the above notation suppose $D_j \geq D_i$ and $E_j \geq E_i$ with at least one inequality being strict. By (20) i has a greater marginal incentive to send patients (of either type) to T than j does. This implies either that i has the strict incentive to (feasibly) increase p_i or that j has the strict incentive to (feasibly) strictly decrease p_j, contradicting the equilibrium assumption. A parallel argument applies to Inversion equilibria. □

6.2.2 Unique NI-candidates

To prove Proposition 3 we write the partial derivatives of TC payoffs u_i also as a function of k. For any (symmetric, non-wasteful) non-inversion strategy profile and k, define δ^W and δ^N by evaluating (17) and (18) at such profiles.

$$\delta^W(E,k) \equiv \left.\frac{\partial u}{\partial E_i}\right|_{\forall j\ D_j=0,\, E_j=E} \tag{22}$$

$$= (L_h^T - L_h^N) - \phi_N \frac{r_\ell (L_\ell^N - L_h^N)}{n(r_\ell + r_h - nE)^2} - \phi_N \frac{(n-1)(L^* - L_h^N)}{n(r_\ell + r_h - nE)} + \phi_T \frac{(n-1)(L^* - L_h^T)}{n^2 E}$$

$$\delta^N(D,k) \equiv \left.\frac{\partial u}{\partial D_i}\right|_{\forall j\ D_j=D,\, E_j=\frac{r_h}{n}} \tag{23}$$

$$= (L_\ell^T - L_\ell^N) - \phi_N \frac{(n-1)(L^* - L_\ell^N)}{n(r_\ell - nD)} - \phi_T \frac{r_h (L_\ell^T - L_h^T)}{n(nD + r_h)^2} + \phi_T \frac{(n-1)(L^* - L_\ell^T)}{n(nD + r_h)}$$

With this notation we write the definition of NI-candidate (Definition 1 as follows.

Definition (NI-candidate)**.** *A symmetric profile p^* (inducing strategies $D_i = p_{i\ell}^* r_\ell / n$, $E_i = p_{ih}^* r_h / n$) is an NI-candidate if*

- *(1-NI) $p_{i\ell}^* \equiv 0$ and $\delta^W(E,k) = 0$; or*
- *(1-NI corner solution) $p_{i\ell}^* \equiv 0$, $nE_i = k\phi$, and $\delta^W(E,k) \leq 0$; or*
- *(3-NI) $p_{ih}^* \equiv 1$ and $\delta^N(D,k) = 0$; or*
- *(3-NI corner solution) $p_{ih}^* \equiv 1$, $r_\ell/n - D_i = (1-k)\phi$, and $\delta^N(D,k) \geq 0$; or*
- *(2-NI) $p_{i\ell}^* \equiv 0$, $p_{ih}^* \equiv 1$, $\delta^W(E,k) \geq 0$, and $\delta^N(D,k) \leq 0$.*

The following lemma conveys the intuition that the benefit of assigning more patients to T increases in k and decreases in the total mass of patients assigned to T. This intuition is always true for (i) assigning more high type patients to T, but requires mild assumptions for (ii) low-type patients (who congest their own TC's high-type patients).

Lemma 8 (Properties of δ^N, δ^W)**.**

(i) $\delta^W(E,k)$ *is linearly increasing in* $k \in [0,1]$ *and decreasing in* $E \in [0, r_h/n]$.

(ii) If Assumption 1 holds, $n \geq 2$ *implies* $\delta^N(D,k)$ *is linearly increasing in* $k \in [0,1]$*, and* $n \geq 3$ *implies* $\delta^N(D,k)$ *is decreasing in* $D \in [0, r_\ell/n]$.

Proof. To prove the first claim, note that δ^W is continuous and differentiable. Differentiating $\delta^W(E,k)$ with respect to k yields

$$\frac{\partial \delta^W}{\partial k} = \phi \frac{r_\ell (L_\ell^N - L_h^N)}{n(r_\ell + r_h - nE)^2} + \phi \frac{(n-1)(L^* - L_h^N)}{n(r_\ell + r_h - nE)} + \phi \frac{(n-1)(L^* - L_h^T)}{n^2 E} > 0 \tag{24}$$

which is a sum of positive terms independent of k; so δ^W is linearly increasing in k. Likewise,

$$\frac{\partial \delta^W}{\partial E} = -\phi_N \frac{2r_\ell (L_\ell^N - L_h^N)}{(r_\ell + r_h - nE)^3} - \phi_N \frac{(n-1)(L^* - L_h^N)}{(r_\ell + r_h - nE)^2} - \phi_T \frac{(n-1)(L^* - L_h^T)}{(nE)^2}$$

which for any $E \in (0, r_h/n]$ is a sum of three strictly negative terms; so δ^W is decreasing in E.

Analogously for the second claim,

$$\frac{\partial \delta^N}{\partial k} = \phi \frac{(n-1)(L^* - L_\ell^N)}{n(r_\ell - nD)} + \phi \frac{-r_h(L_\ell^T - L_h^T) + (n-1)(r_h + nD)(L^* - L_\ell^T)}{n(r_h + nD)^2} \tag{25}$$

If $n \geq 2$ and Assumption 1 holds, then the second term is strictly positive. Since the first term is positive, δ^N is linearly increasing in k. Likewise

$$\frac{\partial \delta^N}{\partial D} = \phi_N \frac{-(n-1)(L^* - L_\ell^N)}{(r_\ell - nD)^2} + \phi_T \frac{2r_h(L_\ell^T - L_h^T)}{(nD + r_h)^3} - \phi_T \frac{(n-1)(L^* - L_\ell^T)}{(nD + r_h)^2}$$

Since $r_h/(nD + r_h) < 1$,

$$\frac{\partial \delta^N}{\partial D} < \phi_N \frac{-(n-1)(L^* - L_\ell^N)}{(r_\ell - nD)^2} + \phi_T \frac{2(L_\ell^T - L_h^T)}{(nD + r_h)^2} - \phi_T \frac{(n-1)(L^* - L_\ell^T)}{(nD + r_h)^2}$$

If $n \geq 3$ and Assumption 1 holds, then the magnitude of the third term exceeds that of the second term; so δ^N is decreasing in D. □

Lemma 9 (Candidates in Region 2-NI)**.** *Fix k and denote the "NI-separating" profile p^s by $p^s_{i\ell} \equiv 0$ and $p^s_{ih} \equiv 1$. Then (p^s, k) is an NI-candidate if and only if $k' \leq k \leq k^*$ where $k' < k^*$ are defined by (11) and (12).*

Proof. Fixing k, (p^s, k) is a NI-candidate if and only if a TC has no incentive to decrease E from its value r_h/n and has no incentive to increase D above 0.

The former is the requirement $\delta^W(r_h/n, k) \geq 0$ which, by Equation 22, is

$$(L_h^T - L_h^N) - \frac{\phi_N}{r_\ell}\frac{1}{n}(L_\ell^N - nL_h^N + (n-1)L^*) + \frac{\phi_T}{r_h}\frac{n-1}{n}(L^* - L_h^T) \geq 0$$

Substituting for $\phi_N = (1-k)\phi$ and $\phi_T = k\phi$ this inequality holds when

$$k \geq \frac{-(L_h^T - L_h^N) + \frac{\phi}{r_\ell}\frac{1}{n}\left((n-1)L^* + L_\ell^N - nL_h^N\right)}{\frac{\phi}{r_\ell}\frac{1}{n}\left((n-1)L^* + L_\ell^N - nL_h^N\right) + \frac{\phi}{r_h}\frac{n-1}{n}(L^* - L_h^T)} \equiv k' \tag{26}$$

establishing (11).

The latter requirement is $\delta^N(0, k) \leq 0$, i.e. by Equation 23

$$(L_\ell^T - L_\ell^N) - \phi_N\frac{\frac{n-1}{n}(L^* - L_\ell^N)}{r_\ell} + \phi_T\frac{-\frac{1}{n}(L_\ell^T - L_h^T) + \frac{n-1}{n}(L^* - L_\ell^T)}{r_h} \leq 0$$

which holds when

$$k\alpha \leq (L_\ell^N - L_\ell^T) + \frac{\phi}{r_\ell}\frac{n-1}{n}(L^* - L_\ell^N) \tag{27}$$

$$\text{where } \alpha = \left[\frac{\phi}{r_\ell}\frac{n-1}{n}(L^* - L_\ell^N) + \frac{\phi}{r_h}\left[-\frac{1}{n}(L_\ell^T - L_h^T) + \frac{n-1}{n}(L^* - L_\ell^T)\right]\right]$$

The inequality obviously holds if $\alpha \leq 0$. However Assumption 1 and the assumption that $n \geq 3$ imply $\alpha > 0$. Dividing both sides of (27) by α yields $k \leq k^*$ as defined in (12). □

The following implies that region 2-NI is non-degenerate: equilibrium separation occurs for a range of values of k.

Lemma 10. *For k', k^* defined in (11)–(12), $k' < k^*$ and $\frac{r_h}{r_\ell + r_h} \equiv \bar{k} < k^*$.*

Proof. It is clear from (11) and (12) that $k' < 1$ and $k^* > 0$. Hence if $k' = 0$ or $k^* = 1$ the conclusion is immediate.

Suppose $k' > 0$ and $k^* < 1$, hence $\delta^W(r_h/n, k') = 0$ and $\delta^N(0, k^*) = 0$. Since δ^W is increasing in k, we prove the result by showing $\delta^W(r_h/n, k^*) > 0 = \delta^N(0, k^*)$. We do this by showing that (i) $\delta^W(r_h/n, k) - \delta^N(0, k)$ increases in k, and (ii) $\delta^W(r_h/n, \bar{k}) > \delta^N(0, \bar{k})$ at the "proportional" value $\bar{k} \equiv \frac{r_h}{r_\ell + r_h} < k^*$.

To show (i) we evaluate (24)–(25) at $(r_h/n, k)$ and $(0, k)$ (reordering the first two terms of the first expression).

$$\frac{\partial \delta^W}{\partial k}(r_h/n, k) = \phi \frac{(n-1)(L^* - L_h^N)}{n r_\ell} + \phi \frac{(L_\ell^N - L_h^N)}{n r_\ell} + \phi \frac{(n-1)(L^* - L_h^T)}{n r_h}$$
$$\frac{\partial \delta^N}{\partial k}(0, k) = \phi \frac{(n-1)(L^* - L_\ell^N)}{n r_\ell} + \phi \frac{-(L_\ell^T - L_h^T)}{n r_h} + \phi \frac{(n-1)(L^* - L_\ell^T)}{n r_h}$$

It is easy to see that the three terms in the first expression are greater than the respective terms in the second expression, proving (i).

To prove (ii), evaluate the two derivatives at $\bar{k}$.

$$\begin{aligned}
\delta^N(0, \bar{k}) &= (L_\ell^T - L_\ell^N) - \phi \frac{r_\ell}{r_h + r_\ell} \frac{(n-1)(L^* - L_\ell^N)}{n r_\ell} + \phi \frac{r_h}{r_h + r_\ell} \frac{-(L_\ell^T - L_h^T) + (n-1)(L^* - L_\ell^T)}{n r_h} \\
&= (L_\ell^T - L_\ell^N) - \phi \frac{(n-1)(L^* - L_\ell^N)}{n(r_h + r_\ell)} + \phi \frac{-(L_\ell^T - L_h^T) + (n-1)(L^* - L_\ell^T)}{n(r_h + r_\ell)} \\
&= (L_\ell^T - L_\ell^N) - \phi \frac{(L_\ell^T - L_h^T)}{n(r_h + r_\ell)} + \phi \frac{(n-1)(L_\ell^N - L_\ell^T)}{n(r_h + r_\ell)} < 0
\end{aligned}$$

which is negative since the magnitude of the first (negative) term exceeds that of the third (positive) term. Additionally, since $\delta^N(0, k^*) = 0$ and is increasing in k this implies $\bar{k} < k^*$.

Secondly,

$$\begin{aligned}
\delta^W(r_h/n, \bar{k}) &= (L_h^T - L_h^N) - \phi \frac{r_\ell}{r_h + r_\ell} \frac{(L_\ell^N - L_h^N)}{n r_\ell} \\
&\quad - \phi \frac{r_\ell}{r_h + r_\ell} \frac{(n-1)(L^* - L_h^N)}{n r_\ell} + \phi \frac{r_h}{r_h + r_\ell} \frac{(n-1)(L^* - L_h^T)}{n r_h} \\
&= (L_h^T - L_h^N) - \phi \frac{(L_\ell^N - L_h^N)}{n(r_h + r_\ell)} - \phi \frac{(n-1)(L^* - L_h^N)}{n(r_h + r_\ell)} + \phi \frac{(n-1)(L^* - L_h^T)}{n(r_h + r_\ell)} \\
&= (L_h^T - L_h^N) - \phi \frac{(L_\ell^N - L_h^N)}{n(r_h + r_\ell)} - \phi \frac{(n-1)(L_h^T - L_h^N)}{n(r_h + r_\ell)}
\end{aligned}$$

Note that

$$\begin{aligned}\delta^W(r_h/n,\bar{k})-\delta^N(0,\bar{k}) &= (L_h^T-L_h^N)-(L_\ell^T-L_\ell^N)-\phi\frac{(L_\ell^N-L_h^N)-(L_\ell^T-L_h^T)}{n(r_h+r_\ell)}\\ &\quad-\phi(n-1)\frac{L_h^T-L_h^N+L_\ell^N-L_\ell^T}{n(r_h+r_\ell)}\\ &= \left[L_h^T-L_h^N-L_\ell^T+L_\ell^N\right]\left[1-\frac{\phi}{r_h+r_\ell}\right]>0\end{aligned}$$

since $L_h^T > L_h^N$, $L_\ell^N > L_\ell^T$, and $\phi < r_h + r_\ell$. Therefore at $k^* > \bar{k}$, (i) implies

$$\delta^W(r_h/n,k^*) > \delta^N(0,k^*) = 0 = \delta^W(r_h/n,k')$$

implying $k^* > k'$. □

Lemma 11 (Candidates in Region 3-NI)**.** *If $k > k^*$ then there exists a unique NI-candidate. It satisfies $p_{ih} \equiv 1$.*

Proof. Let p^s be defined as in Lemma 9 and recall $\delta^N(p^s,k^*)=0$ by definition of k^*. By Lemma 9, $k > k^* > k'$ implies $\delta^W(p^s,k) > 0$. The lemma furthermore implies $\delta^W(p,k) > 0$ for any symmetric profile satisfying $p_{i\ell} \equiv 0$, i.e. there can be no NI-candidate in region 1-NI.

Lemma 9 similarly implies $\delta^N(p^s,k) > 0$. By Lemma 9, $\delta^N(\cdot,k)$ continuously decreases as we increase D ($p_{i\ell}$) from zero. Either $\delta^N(D,k)=0$ at some unique D or we have (corner solution) $\delta^N(r_\ell/n,k) > 0$. In the latter case we clearly have a unique NI-candidate. In the former (interior) case, recall by Lemma 6 (statement (ii)) that at such a profile, a TC's payoffs are either decreasing or concave in $p_{i\ell}$. Since $\delta^N(D,k)=0$ we must have concavity with respect to $p_{i\ell}$, hence this point uniquely satisfies the local first- and second-order conditions. □

Lemma 12 (Candidates in Region 1-NI)**.** *If $k < k'$ then there exists a unique NI-candidate. It satisfies $p_{i\ell} \equiv 0$.*

We omit the proof. It mirrors that of Lemma 11 with the simplification that, in reference to Lemma 6, payoffs are always concave in p_{ih}.

Proof of Proposition 3. Existence, uniqueness, and the description of the NI-candidates are proven by the above three lemmas. Monotonicity of $p^*()$ w.r.t. k follows from Lemma 8. □

6.2.3 Optimal NI-candidate

The proof of Theorem 4 relies on the following lemma, stating that in region 3-NI we have $\pi^T > \pi^N$.

Lemma 13 ($\pi^T > \pi^N$ in Region NI-3)**.** *Fix k, and suppose p^* is a NI-3 equilibrium: for all i, $p^*_{i\ell} = p^*_\ell > 0$ (and hence $p^*_{ih} = 1$). Then $k > (p^*_\ell r_\ell + r_h)/(r_\ell + r_h)$, that is, the equilibrium allocation probability is higher in T than in N: $\pi^T > \pi^N$.*

Proof. By Lemma 6, $p^*_{ih} = 1$ implies $u_i(p^*)$ is either decreasing or concave in $p_{i\ell}$. Since $p^*_\ell > 0$ it must be concave. Therefore either the partial derivative (23) is zero, or the equilibrium is at a corner (where the N-nonwastefulness constraint binds and $\pi^N = 1$). However Observation 1 (Fact 2) rules out the latter, hence (23) is zero.

Recall for NI-3 equilibria that $A = r_\ell/n - D$, $B = 0$, $C = (n-1)A$, $E = r_h/n$, $F = (n-1)(D+E)$. So $\pi^N = \phi_N/(A+B+C) = \phi_N/(r_\ell - nD)$ and $\pi^T = \phi_T/(D+E+F) = \phi_T/(r_h + nD)$. Let $\lambda = r_h/(r_h + nD)$. Since Equation 23 is zero we have

$$
\begin{aligned}
&(L^N_\ell - L^T_\ell) + \phi_N \frac{(n-1)(L^* - L^N_\ell)}{n(r_\ell - nD)} \\
&\qquad = \phi_T \frac{-r_h(L^T_\ell - L^T_h) + (n-1)(nD + r_h)(L^* - L^T_\ell)}{n(nD + r_h)^2} \\
&L^N_\ell + \pi^N \frac{(n-1)}{n}(L^* - L^N_\ell) \\
&\qquad = L^T_\ell + \pi^T \frac{-r_h(L^T_\ell - L^T_h) + (n-1)(nD + r_h)(L^* - L^T_\ell)}{n(nD + r_h)}
\end{aligned}
$$

Thus

$$\begin{aligned}
&(1-\pi^N)L_\ell^N + \pi^N\left(\frac{(n-1)}{n}L^* + \frac{1}{n}L_\ell^N\right)\\
&\qquad = L_\ell^T + \pi^T\left(\frac{(n-1)(L^*-L_\ell^T)}{n} + \frac{-r_h(L_\ell^T-L_h^T)}{n(nD+r_h)}\right)\\
&\qquad = L_\ell^T + \pi^T\left(\frac{n-1}{n}(L^*-L_\ell^T) + \frac{-\lambda(L_\ell^T-L_h^T)}{n}\right)\\
&\qquad = (1-\pi^T)L_\ell^T + \pi^T\left(\frac{n-1}{n}L^* + \frac{(1-\lambda)L_\ell^T+\lambda L_h^T}{n}\right)\\
&\qquad < (1-\pi^T)L_\ell^N + \pi^T\left(\frac{n-1}{n}L^* + \frac{1}{n}L_\ell^N\right)
\end{aligned}$$

Since $L^* > L_\ell^N$ we have $\pi^T > \pi^N$; equivalently $k > (p_\ell^* r_\ell + r_h)/(r_\ell + r_h)$. □

Proof of Theorem 4. For any k let $f(k)$ and $\pi^T(k)$ respectively denote the fraction of organs allocated to high types and the probability that a patient assigned to T receives an organ, under k's NI-candidate. We prove the results regarding f. The results regarding TCs' total payoffs follow directly using the same arguments made in Subsection 3.2.

It is immediate that $f()$ is increasing on $[k', k^*]$ since (as under perfect competition) the strategy profile is constant across all such NI-candidates. The remainder of the proof consists of showing (i) f is decreasing on $[k^*, 1]$, and (ii) $f(k) < f(k^*)$ whenever $k \in [0, k']$.

Step (i). For any $k \in (k^*, 1]$, there is at most one symmetric profile (namely the NI-candidate $p(k)$) satisfying $\delta^N(D, k) = 0$ by Lemma 11. Whenever such $p(k)$ exists (i.e. the NI-candidate is not a corner solution), let $D(k) = p(k)r_\ell/n$ denote the corresponding mass of low types each TC sends to T.

By Lemma 8 $D(k)$ is increasing in k; hence the values of $k > k^*$ for which such $\delta^N(D(k), k) = 0$ exist are an interval (of the form $(k^*, x]$ by continuity). We show that $\pi^T(k)$ is decreasing in k on this interval. Since $p_{ih}(k) \equiv 1$ on this range, a decrease in $\pi^T()$ necessarily decreases $f()$, proving (i).

We implicitly differentiate $\delta^N(D(k), k) = 0$ (Equation 23) w.r.t. k after substituting $\phi_N = (1-k)\phi$ and $\phi_T = k\phi$. (Write $D = D(k)$ and $D' =$

$\partial D(k)/\partial k$, and ignore the corner case $p_{i\ell} \equiv 1$, where $nD = r_\ell$.)

$$\phi\frac{(n-1)(L^*-L_\ell^N)}{n(r_\ell-nD)} - \phi(1-k)\frac{(n-1)(L^*-L_\ell^N)}{n(r_\ell-nD)^2}nD' - \phi\frac{r_h(L_\ell^T-L_h^T)}{n(nD+r_h)^2}$$
$$+ 2\phi k\frac{r_h(L_\ell^T-L_h^T)}{n(nD+r_h)^3}nD' + \phi\frac{(n-1)(L^*-L_\ell^T)}{n(nD+r_h)} - \phi k\frac{(n-1)(L^*-L_\ell^T)}{n(nD+r_h)^2}nD' = 0$$

or

$$\frac{(n-1)(L^*-L_\ell^N)}{n(r_\ell-nD)} - \frac{r_h(L_\ell^T-L_h^T)}{n(nD+r_h)^2} + \frac{(n-1)(L^*-L_\ell^T)}{n(nD+r_h)}$$
$$= (1-k)\frac{(n-1)(L^*-L_\ell^N)}{n(r_\ell-nD)^2}nD' + k\frac{(n-1)(L^*-L_\ell^T)}{n(nD+r_h)^2}nD' - 2k\frac{r_h(L_\ell^T-L_h^T)}{n(nD+r_h)^3}nD'$$

Denoting $r = r_\ell + r_h$ and $S = nD + r_h < r$,

$$D' = \frac{\frac{(n-1)(L^*-L_\ell^N)}{n(r_\ell-nD)} + \frac{(n-1)(L^*-L_\ell^T)}{n(nD+r_h)} - \frac{r_h(L_\ell^T-L_h^T)}{n(nD+r_h)^2}}{(1-k)\frac{(n-1)(L^*-L_\ell^N)}{(r_\ell-nD)^2} + k\frac{(n-1)(L^*-L_\ell^T)}{(nD+r_h)^2} - 2k\frac{r_h(L_\ell^T-L_h^T)}{(nD+r_h)^3}}$$
$$= \frac{\frac{(n-1)(L^*-L_\ell^N)}{n(r-S)} + \frac{(n-1)(L^*-L_\ell^T)}{nS} - \frac{r_h(L_\ell^T-L_h^T)}{nS^2}}{(1-k)\frac{(n-1)(L^*-L_\ell^N)}{(r-S)^2} + k\frac{(n-1)(L^*-L_\ell^T)}{S^2} - 2k\frac{r_h(L_\ell^T-L_h^T)}{S^3}}$$
$$= \frac{(n-1)(L^*-L_\ell^N)S^2 + (n-1)(L^*-L_\ell^T)(r-S)S - r_h(L_\ell^T-L_h^T)(r-S)}{(1-k)(n-1)(L^*-L_\ell^N)S^3 + k(n-1)(L^*-L_\ell^T)S(r-S)^2 - 2kr_h(L_\ell^T-L_h^T)(r-S)^2}\frac{(r-S)S}{n}$$

To show that the derivative of $\pi^T(k) \equiv \frac{k\phi}{nD+r_h}$ is negative, i.e. that

$$\frac{\phi}{nD+r_h} - \frac{nk\phi}{(nD+r_h)^2}D' = \frac{\phi}{S} - \frac{nk\phi}{S^2}D' < 0$$

we need to show $D' > S/(nk)$. Using the derivation of D' above, this inequality becomes (see tex comments)

$$(L^*-L_\ell^N)S^2(n-1)(kr-S) > -kr_h(L_\ell^T-L_h^T)(r-S)^2$$

Since $r > S$ this is true whenever $k \geq S/r$, i.e. whenever $\pi^T(k) \geq \pi^N(k)$, which is true by Lemma 13. Hence $\pi^T()$ and $f()$ are decreasing on $[k^*, 1]$.

Step (ii): consider the case $k \in [0, k']$.[15] By previous arguments, NI-candidate profiles vary continuously in k; therefore $f()$ is continuous. Hence

[15] This case is mostly symmetric to the previous one, except that the possibility that $\bar{k} < k'$ necessitates additional arguments.

we can choose

$$\tilde{k} = \arg\max_{[0,k']} f(k)$$

We show $f(\tilde{k}) < f(k^*) \equiv k^*$.

Case 1: $\pi^N(\tilde{k}) \geq \pi^T(\tilde{k})$. A low type receives an organ with probability $\pi^N(\tilde{k})$, whereas a high type receives an organ with a weakly lower probability of

$$(1 - p_{ih})\pi^N(\tilde{k}) + p_{ih}\pi^T(\tilde{k})$$

where $(0, p_{ih})$ is the NI-candidate for $\tilde{k}$. Since high types receive organs with lower probability than low types, they collectively receive no more than the (unconditional) organ allocation rate: $f(\tilde{k}) \leq \frac{\phi}{r_\ell + r_h} < k^*$, where the second inequality follows from Lemma 10 ($\bar{k} < k^*$).

Case 2: $\pi^N(\tilde{k}) < \pi^T(\tilde{k})$. We show that f is increasing at $\tilde{k}$. This means $\tilde{k} = k'$, implying the desired conclusion.

Since the mass of organs allocated to *low* types is $\pi^N(\tilde{k})r_\ell$, $f(\tilde{k}) = 1 - \frac{\pi^N}{\phi}r_\ell$. To show f is increasing we show $\pi^N()$ is decreasing at $\tilde{k}$.

To show the derivative of $\pi^N(k) \equiv \frac{(1-k)\phi}{r_\ell + r_h - nE}$ is negative at $\tilde{k}$, i.e. that

$$\frac{-\phi}{r_\ell + r_h - nE} + \frac{(1-\tilde{k})\phi n E'}{(r_\ell + r_h - nE)^2} = \left(\frac{-\phi}{r_\ell + r_h - nE}\right)\left(1 - \frac{(1-\tilde{k})nE'}{r_\ell + r_h - nE}\right) \leq 0$$

we need to show

$$E'(\tilde{k}) \leq \frac{r_\ell + r_h - nE(\tilde{k})}{(1-\tilde{k})n} \tag{28}$$

We implicitly differentiate $\delta^W(E(k), k) = 0$ (Equation 22) w.r.t. k and evaluate at $\tilde{k}$. Writing $E = E(k)$ and $E' = E'(k)$ we obtain

$$\phi\frac{(n-1)(L^* - L_h^T)}{nE} - \phi(1-\tilde{k})\frac{(n-1)(L^* - L_h^N)}{(r_\ell + r_h - nE)^2}nE' + \phi\frac{r_\ell(L_\ell^N - L_h^N)}{(r_\ell + r_h - nE)^2}$$
$$- 2\phi(1-\tilde{k})\frac{r_\ell(L_\ell^N - L_h^N)}{(r_\ell + r_h - nE)^3}nE' + \phi\frac{(n-1)(L^* - L_h^N)}{(r_\ell + r_h - nE)} - \phi\tilde{k}\frac{(n-1)(L^* - L_h^T)}{(nE)^2}E'n = 0$$

Denoting $S = r_\ell + r_h - nE(\tilde{k})$ and $r = r_h + r_\ell$ this yields

$$E'(\tilde{k}) = \frac{r_\ell(L_\ell^N - L_h^N)S^{-2} + (n-1)(L^* - L_h^N)S^{-1} + (n-1)(L^* - L_h^T)(r-S)^{-1}}{2(1-\tilde{k})r_\ell(L_\ell^N - L_h^N)S^{-3}n + (1-\tilde{k})(n-1)(L^* - L_h^N)S^{-2}n + \tilde{k}(n-1)(L^* - L_h^T)(r-S)^{-2}n}$$

Therefore one can show that (28) is equivalent to

$$(n-1)(L^* - L_h^T)[(1-\tilde{k})(r-S) - \tilde{k}S] \leq r_\ell (L_\ell^N - L_h^N)(1-\tilde{k}) \left(\frac{r-S}{S}\right)^2$$

Note that $\frac{\phi(1-\tilde{k})}{S} = \pi^N(\tilde{k}) < \pi^T(\tilde{k}) = \frac{\phi\tilde{k}}{r-S}$ implies that the LHS is non-positive. Since the RHS is non-negative, (28) holds. □

6.3 Numerical exercises: details

Specifics and summaries of our numerical exercises will be provided in a future version.