

Northwestern | Kellogg

CMS-EMS

Center for Mathematical Studies in
Economics and Management Sciences

Discussion Paper #1607

“Rationing through Classification”

Edwin Munoz-Rodriguez
El Colegio de Mexico A.C.

James Schummer
Northwestern University

October 15, 2024

Rationing through Classification*

Edwin Muñoz-Rodríguez[†] James Schummer[‡]

October 15, 2024

Abstract

Basing object assignment on recipients’ actions can improve allocations but also distort action choices. The degree to which a planner should utilize such information to make allocation decisions is addressed in previous work on “decentralized” settings, i.e. where individual recipients strategically choose their own actions. We consider less competitive situations (e.g. organ allocation) where strategic agents (transplant centers) choose actions (medical treatments) on behalf of multiple potential recipients (patients). When there are many such agents, we show that equilibrium behavior resembles that of the decentralized setting and that a reduction in competition reduces the threat of distortions, improving welfare. With sufficiently few agents, however, a second form of equilibrium arises where agents inefficiently “invert” action choices for multiple recipient types, diverting resources towards high-value recipients. Computational analysis suggests that such equilibria are welfare-optimal only in some low-stakes allocation problems.

Keywords: strategic classification, mechanism design, imperfect competition.

*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, U.S.C., 13th Conference on Economic Design, and 15th Social Choice and Welfare Conference. We thank Dr. Nitin Katariya and Dr. Daniela Ladner for their valuable insights into the field of transplantation. We have received valuable research assistance from Michelle Avataneo and Isaac Lara. Muñoz-Rodríguez is grateful for funding from the Ford Foundation through a PRAEM program grant at El Colegio de Mexico, A.C. Our views and conclusions are solely ours and may not 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.

1 Introduction

Resources are sometimes allocated to recipients based on their actions. As a running example, patients awaiting organ transplants are prioritized based on the interim medical treatment they (or their care providers) select. Heart transplant patients obtain higher status in wait lists by receiving certain forms of advanced life support. Liver transplant patients historically received higher priority when placed in intensive care units (ICU). Examples beyond organ transplantation include school choice settings where residential location affects priorities (Park and Hahm, 2023), or the allocation of fixed budgets, awards, or grants based on recipients’ costly actions.¹

The degree to which actions should be used as a basis for resource allocation depends on a tradeoff between the possible improvement in allocation decisions versus the welfare cost of distorting incentives in action choices. These distortions can be non-trivial: an October 2018 policy change created a strict priority of heart patients treated with intra-aortic balloon pumps (IABP) over patients treated with high-dose inotropes (HDI); shortly afterwards IABP usage roughly tripled (Ran et al., 2021).² These distortions impact welfare not only in that action choices are suboptimal (e.g. overly aggressive treatments), but because these choices conflate the planner’s intention to favor a different set of agents. While individual recipients can partly internalize the former effect, the latter “congestion effect” spills over to other recipients.

While various factors determine the magnitude of this congestion effect, there is one that has not been addressed in related literature that is a crucial feature in organ allocation: more than one recipient falls under the care of a single strategic decision maker. In particular, a Transplant Center (TC) that chooses actions (treatments) on behalf of multiple patients might internalize part of the congestion effect. In turn this might reduce gaming, or allow a mechanism designer to achieve higher levels of welfare by accounting for this effect. Parker et al. (2018) provide compelling evidence of reduced gaming in the context of heart allocation: a statistical measure of “over-treatment” in heart patients is strongly, positively related to the number of competing Transplant Centers within a transplant region. Our work is motivated by the resulting

¹Certification agencies are a related example, but where the budget constraint is softer than the kind we consider here; see Frankel and Kartik (2021), Ball (forthcoming), Perez-Richet and Skreta (2022), and Perez-Richet and Skreta (2023).

²Similarly in 2002, when ICU status was removed as a prioritization factor for liver transplant patients due to perceived abuse, ICU admission rates dropped by roughly half (Snyder, 2010).

mechanism design question of how to account for this competition effect.

We construct a model to analyze (i) the degree to which recipients should be prioritized based on their actions and, more significantly, (ii) how that degree is affected by “competition,” i.e. the extent to which recipients’ action decisions are (de)centralized as described above. The simplest model that allows us to do this—in the language of organ transplantation—involves two types of patients who obtain Low or High value from an organ and, respectively, negative or positive benefit from taking a “Treatment” action. A planner who *fully* prioritizes the Treatment classification distorts action choices for Low-type patients. At the other extreme, eliminating priorities altogether removes these distortions but fails to utilize information that could improve aggregate welfare. In many organ allocation applications,³ policy discussions typically focus on these two extremes: whether “treated” recipients should receive *full* or *no* prioritization over others. To generalize this binary comparison we allow for *partial* prioritization via what we call rationing through classification (RTC): the planner rations some fraction of organs to the class of patients taking the “Treatment” action, allocating the rest amongst the remaining patients. A maximal ration gives the Treatment class full priority, whereas a sufficiently low ration prioritizes neither class. Our questions above concern the optimal choice of this ration.

The “perfect competition” special case of our model, where individual patients decide their own actions (Section 3), fits within a literature on strategic classification (e.g. Braverman and Garg (2020), Perez-Richet and Skreta (2022)). Consistent with that literature we show that equilibrium welfare is maximized when the planner maximizes the Treatment patients’ ration of organs subject to a no-distortions constraint: Low type patients must not have the incentive to choose Treatment. Any higher ration—inducing Low types to game the system—necessarily lowers welfare.

The novelty of our work is to extend this analysis to the “imperfect competition” setting (Section 4) where n strategic Transplant Centers (TC’s) make Treatment decisions on behalf of their own set of patients. This captures the phenomenon in the examples discussed above that, while allocation is based on recipient actions, multiple recipients’ actions are chosen by a single strategic agent.

The introduction of TC’s has the flavor of reducing competition and might lead to the following intuition. Since TC’s internalize the congestion effect of “wrongly” choosing Treatment for its Low type patients, the planner can further *increase* priori-

³E.g. The IABP heart patients and ICU liver patients discussed above.

tization of Treatment patients without creating distortions. Thus *assuming it remains welfare-optimal to eliminate distortions* (as under perfect competition), reduced competition improves resource allocation without distorting actions, improving welfare.

We show the extent to which this intuition *partially* holds, with the determining factor being whether equilibria under imperfect competition have the same structure as equilibria under perfect competition. When the structure remains the same, it means that the planner again trades off an increased ration for Treatment patients versus distorted action choices for Low types. In this case our main results reestablish the welfare-optimality of no distortions, derive the optimal ration under such equilibria, and verify the above intuition that competition harms welfare. We also compare the extreme cases of *full* and *no* prioritization discussed earlier (Subsection 4.5).

For some parameters, however, the imperfect competition setting yields equilibria with a less-intuitive “Inversion” structure, where action choices are doubly distorted: each TC simultaneously chooses the Treatment action for some *Low* types and *does not* choose Treatment for some *High* types. Convexities in TCs’ payoff functions allow for such equilibria, and complicate the general analysis. Computational analysis suggests two things. First, Inversion equilibria can arise only when there is a low relative cost for Low types to take the Treatment action (or if the planner rations in a way that discriminates *against* Treatment patients). Second, and surprisingly, Inversion equilibria can be welfare-optimal among *all* equilibria, though this phenomenon occurs only in some low-stakes cases where organs are relatively plentiful.

1.1 Related Literature

The work closest to ours is that on strategic classification. This literature focuses on the case where individual recipients strategically choose their own actions, analogous to our baseline “perfect competition” model (Section 3). We distinguish ourselves from this work by allowing “imperfect competition” (Section 4), centralizing multiple decisions under one strategic agent.⁴ Our contribution is to show the extent to which results in the baseline model extend to the general one and how this is impacted by the *level* of competition.

Generally speaking, models in this literature have the following characteristics. A planner wishes to correctly classify an agent’s type as being above or below some

⁴An additional minor difference is that we impose a “classification budget” representing the fixed supply of resources.

threshold (high or low). All agents desire a high classification and can misrepresent their privately known type at some cost. In a continuous-type version of our baseline model (Section 3), [Braverman and Garg \(2020\)](#) maximize equilibrium classification accuracy net of agents’ manipulation costs. Under some assumptions they show that optimal classifiers (i) typically require randomization, and (ii) induce no manipulation. Our setup necessarily induces randomness by the nature of our budgeted rationing problem but its quantification is determined endogenously by equilibrium behavior. Nevertheless our [Theorem 2](#) is analogous to their result.

[Perez-Richet and Skreta \(2022\)](#) allow the planner to commit to a probabilistic testing function that maps (misrepresented) types into randomized signals. The planner uses realized observations to make optimal classification decisions. Under an increasing-returns assumption on misrepresentation costs, accuracy-maximizing mechanisms “raise the bar” by offering the greatest chance of high classification only to observed types above some artificially high threshold. The agents achieving this threshold in equilibrium are precisely those whose true type is above the planner’s desired threshold. Other types engage in no misrepresentation, being compensated with enough probability of high classification to offset the benefit of doing so. [Perez-Richet and Skreta \(2023\)](#) impose this no-misrepresentation condition as a constraint under which they find optimal allocation mechanisms.

In a machine learning context [Hardt et al. \(2016\)](#) provide efficient, near-optimal algorithms for classification accuracy against strategic agents both 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 “high type” agents.

Other work examines variations on mechanism-, scoring-, or ratings-design under costly misrepresentation. [Frankel and Kartik \(2021\)](#) consider agents who vary both in type as above and in misrepresentation costs. This dual heterogeneity leads the planner to under-weight observed information to improve accuracy in equilibrium. When types are multidimensional [Ball \(forthcoming\)](#) shows that the planner benefits by under-weighting some dimensions and over-weighting others. [Lee and Suen \(2023\)](#) consider allocating university seats based on exam scores obtained naturally (high types) or through wasteful tutoring (low types), showing that increasing resources (seats) can increase distortions (wasteful tutoring). Finally, [Akbarpour et al. \(2024\)](#) provide a more distinct setting in which rationing is optimal even when the use of

transfers is available.

Turning to dynamic settings, [Munoz-Rodriguez \(2024\)](#) studies a model that is somewhat like an overlapping generations version of our “perfect competition” model but with partially verifiable types. The optimal dynamic mechanism improves outcomes by granting low types future option value for forgoing early transplants. In a model of dynamic discrete choice, [Sweat \(2024\)](#) studies the impact of costly treatment and organ selectivity in heart transplantation.

While the imperfect competition aspect of our main model ([Section 4](#)) is novel from the perspective of the above literature, it also leads to a generalization of congestion games pioneered by [Wardrop \(1952\)](#). Increasing one’s allocation probability through the “route” of misrepresentation necessarily decreases someone else’s. Fixing the planner’s choice of ration in our baseline model of [Section 3](#), equilibrium existence for example would follow immediately from that literature ([Konishi, 2004](#)), though of course we go beyond this by evaluating welfare as we vary the ration.

Allowing a finite set of Transplant Centers each to control a mass of patients, our general model becomes a type of atomic congestion game (ACG) for any fixed choice of ration. An existing literature derives existence and uniqueness results for such games as long as 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](#)). In our “two-traffic-type” model, however, payoffs violate the typical concavity assumptions that lead to these kind of results; in fact they locally violate concavity everywhere. Despite this technical challenge we provide an existence result under fairly weak additional assumptions on our primitives ([Theorem 4](#)).

In a one-type model, [Wan \(2012\)](#) shows that in ACG’s with two nodes, total equilibrium welfare increases when a fixed amount of traffic is split amongst fewer atomic agents. Here there is no resource to be allocated; agents are merely trying to minimize transportation costs on a fixed network. Nevertheless our [Proposition 3](#) draws an analogous conclusion in our two-type model, the difference being that our planner adjusts the optimal rationing of resources with respect to the number of agents. The intuition discussed earlier applies in both cases: with fewer agents, a greater share of congestion costs are internalized by each agent.

2 Model

2.1 Primitives

Though our model is stylized for tractability, we use the terminology of organ allocation to ease interpretation. There is a continuum of *patients* having two possible types: a mass $r_\ell > 0$ of low type patients and a mass $r_h > 0$ of high type patients. Each patient takes action N (“Non-treatment”) or T (“Treatment”). The strategic choice of a patient’s action is made either by the patient (Section 3) or a *Transplant Center* who chooses actions for a subset of patients (Section 4). A mass $\phi < r_\ell + r_h$ of organs is assigned to patients as described in Subsection 2.2.

A patient’s welfare depends on their type, action, and whether they receive an organ. Any patient receiving an organ obtains welfare L^* , independent of type and action.⁵ Otherwise a patient of type $i \in \{\ell, h\}$ who takes action $a \in \{N, T\}$ obtains welfare L_i^a . To capture the relevant tradeoffs in our problem we assume

$$L_h^N < L_h^T < L_\ell^T < L_\ell^N < L^* \quad (1)$$

Interpreting high types as “high risk” (sicker) patients, these inequalities embody natural assumptions underlying our motivating examples. First, the Treatment action is beneficial (pre-transplant) for high types but not for low types. Second, high types receive higher marginal benefit from organs; this assumption corresponds to the real world objective of prioritizing high-risk patients (Persad et al., 2009), but only plays a role in our welfare statements.

An economy is summarized by primitives $(r_h, r_\ell, \phi, L_h^N, L_h^T, L_\ell^T, L_\ell^N, L^*)$. In Section 4 we add the additional primitive $n > 0$, representing the number of Transplant Centers choosing actions on behalf of their respective patients.

2.2 Rationing through Classification

The planner observes patients’ actions but not their types. Therefore the planner is restricted to assigning some fraction k of the organs to patients who took action T

⁵This normalization is without loss of generality in both our strategic and welfare analysis. However it is a realistic approximation when welfare is measured in post-transplant expected life-years, both in liver (Schaubel et al., 2009) and heart (Meyer et al., 2015) transplantation. “Treatments” in these applications target improvements in short-term hazard rates rather than long-term quality of life.

and assigning the remaining mass $(1 - k)\phi$ of organs to patients whose action is N . We define the process of *Rationing through Classification* as one where (i) the planner publicly commits to a ration $k \in [0, 1]$, (ii) each patient takes an action N or T (chosen individually in [Section 3](#) or by their Transplant Center in [Section 4](#)), and (iii) $k\phi$ organs are (uniformly randomly) assigned to patients taking action T , and the rest assigned to those taking N . The ration k represents the degree to which the planner uses classification information as a basis for allocation. Our main question is how the choice of k impacts the structure of equilibria and equilibrium welfare.

Naively, a planner might expect to maximize utilitarian welfare by choosing k to be maximal, since the first-best organ assignment is to maximize organ allocation to high types. Of course this ignores the possibility that this induces low types also to take action T , achieving neither a first best assignment nor first-best action choices. At another extreme a planner could attempt to induce efficient action choices by choosing a “proportional” value of k . Namely, let

$$\bar{k} \equiv r_h / (r_\ell + r_h)$$

denote the percentage of patients who are of high type. Even when ration \bar{k} induces efficient action choices,⁶ organ assignment is far from optimal since organs are assigned to all patients with equal probability. Our work examines not only *how* adjustments to k can fine-tune this trade-off, but *whether* it does so.

Our restriction to RTC is technically a restriction on feasible mechanisms; e.g. the planner could commit to choosing ration k *ex post*, as a function of all patients’ realized action choices. This restriction is not significant for two reasons. First, RTC is without loss of generality in our baseline “perfect competition” model because infinitesimal agents are “price takers.”⁷ Second, any general mechanism that is *not* RTC requires the planner to observe a *profile* of realized actions before making allocation decisions. While this is feasible within our model (made static for tractability), it is less practical in a dynamic setting.

⁶It may not; see [Section 4](#).

⁷Any equilibrium under a general mechanism that results (ex post) in ration k is an equilibrium under RTC when the planner (ex ante) commits to constant k .

3 Perfect Competition

3.1 Equilibrium structure

Fixing a ration k , each patient selfishly chooses action N or T . A **strategy profile** $p = (p_\ell, p_h)$ denotes the fractions of low- and high-type patients that choose action T . In the equilibrium analysis we can restrict attention to *non-wasteful* profiles, i.e. where $p_\ell r_\ell \geq (1-k)\phi$ and $p_h r_h \geq k\phi$. A (non-wasteful) profile p induces the following **allocation probabilities** for patients who have chosen N or T :⁸

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

When a profile 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 following incentive compatibility conditions.

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

Observe that if some type has an incentive not to choose its “natural action” (N for low types, T for high types), then it receives a strictly higher allocation probability at its non-natural action. Since this cannot hold for both types simultaneously, we have 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).*

Also intuitive is that an increase in k should induce more patients to choose T , and that high type patients are induced more easily than low types. (Interestingly this intuition fails to hold in [Section 4](#).) With [Lemma 1](#) this leads to the following description of equilibria.

⁸Define $\pi^N(1, 1) = 0 = \pi^T(0, 0)$; these particular values are not significant in the analysis.

Proposition 1 (Unique equilibrium). *For any $k \in [0, 1]$ there exists a unique equilibrium $p^*(k)$. Furthermore $p^*(\cdot)$ is weakly increasing in k and satisfies*

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

where

$$k' = \max \left\{ 0, \frac{r_h \phi(L^* - L_h^N) + r_l(L_h^N - L_h^T)}{\phi r_h(L^* - L_h^N) + r_l(L^* - L_h^T)} \right\} < \bar{k} \equiv \frac{r_h}{r_\ell + r_h} \quad (3)$$

$$k^* = \min \left\{ 1, \frac{r_h \phi(L^* - L_l^N) + r_l(L_l^N - L_l^T)}{\phi r_h(L^* - L_l^N) + r_l(L^* - L_l^T)} \right\} > \bar{k} \equiv \frac{r_h}{r_\ell + r_h} \quad (4)$$

In fact $p^*(\cdot)$ is constant only on $[k', k^*]$. For some primitives it is possible that $k' = 0$ or $k^* = 1$. In particular the proof of [Proposition 1](#) implies

$$k' > 0 \iff \frac{L_h^T - L_h^N}{L^* - L_h^N} < \frac{\phi}{r_l} \quad (5)$$

$$k^* < 1 \iff \frac{L_l^N - L_l^T}{L^* - L_l^T} < \frac{\phi}{r_h} \quad (6)$$

Intuition driving (6) is that low types are more easily induced to choose T (via an increase in k) when (i) organ supply is increased, (ii) there are less competing high types, (iii) the cost of choosing T is decreased, or (iv) the benefit of receiving an organ conditional on choosing T is higher. Analogous intuition drives (5).

3.2 Equilibrium welfare

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 . An increase in k affects welfare both by increasing the fraction of organs allocated to patients choosing T , and by (weakly) increasing the percentage of patients choosing T . While the total effect can

be positive or negative, an obvious case is when $k \in [k', k^*]$. Since $p^*(\cdot)$ is constant in this range, an increase in k simply increases $f(\cdot)$, increasing total welfare.

For $k \in [k^*, 1]$, it turns out that an increase in k disproportionately increases the number of low types choosing T to the extent that $f(\cdot)$ decreases. Analogously, for $k \in [0, k']$ a decrease in k disproportionately increases the number of high types choosing N , increasing $f(\cdot)$.

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

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

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

For an intuition, imagine primitives are such that when $k = 1$, (i) *all* patients choose T in equilibrium, but (ii) any low type patient is indifferent between the lottery they face—receiving an organ (L^*) or not (L_ℓ^T)—and deviating to choose N (a payoff of L_ℓ^N). Note that every patient faces the same probability $\pi^T = \phi/(r_\ell + r_h)$ of receiving an organ.

Next consider a small decrease in k and a “proportional” change in the strategy profile such that (i) a mass ϵ of low type patients instead choose N and (ii) a mass $\pi^T \epsilon$ of organs is rationed amongst those low type patients choosing N . Note that once again every patient (at N or T) has probability π^T of receiving an organ. However this means that low types choosing T are strictly worse off than low types choosing N (since their welfare is lower conditional on not receiving an organ). In order to restore equilibrium indifference, a *greater* than proportional number of low types must choose N . In other words, $p_\ell^*(k)$ must be disproportionately sensitive to changes in $k \in [k^*, 1]$. An analogous argument applies to $p_h^*(k)$ for $k \in [0, k']$.

Theorem 1 has immediate welfare implications. For $k \in [k', k^*]$ welfare increases in k since actions remain fixed while $f(\cdot)$ increases. For $k \in [k^*, 1]$, an increase in k reduces $f(\cdot)$ and increases $p_\ell^*(\cdot)$, necessarily decreasing welfare. Finally for $k \in [0, k']$, an increase in k decreases $f(\cdot)$ but also improves welfare by reducing $p_h^*(\cdot)$. Either effect can dominate, breaking symmetry with the case $k \in [k^*, 1]$. Nevertheless once can separately prove that equilibrium welfare at any $k \in [0, k']$ is inferior to that obtained at k^* .

Theorem 2. *Utilitarian welfare (total patient 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 the “imperfect competition” setting where Transplant Centers decide actions on behalf of their own share of patients. Notably [Lemma 1](#) does not extend to that setting, allowing for a second form of equilibria. Even when equilibria do resemble those of [Proposition 1](#), the arguments proving [Theorem 2](#) no longer apply (because $f()$ loses a monotonicity property). We show instead that the conclusions of [Theorem 2](#) extend in the form of [Theorem 5](#).

4 Imperfect Competition

4.1 Atomic agents

We capture the idea of imperfect competition by specifying a number n of atomic agents who choose actions on behalf of their own patients. Formally, there are n Transplant Centers (TC’s), each choosing actions on behalf of a mass r_ℓ/n of low-type patients and a mass r_h/n of high-types. A **strategy** for TC i is a pair $p_i = (p_{i\ell}, p_{ih}) \in [0, 1]^2$ specifying the percentages of its low-type and high-type patients taking action T . A **strategy profile** is denoted $p = (p_i)_{i \in TC} = (p_{i\ell}, p_{ih})_{i \in TC}$. We let p_{-i} denote the list of strategies for all TC’s other than i .

A TC’s **payoff** is the total expected welfare of its patients as in [\(1\)](#). We continue to interpret the parameters in [\(1\)](#) as individual patient welfare, in which case payoffs are that of a utilitarian TC that puts weight only on its own patients. However, since individual patients play no strategic role in this section, one could go well beyond that interpretation. For example, the parameters in [\(1\)](#) could instead represent a TC’s profitability from patients, conditional on their type, treatment decision, and transplant status; or they could represent some combination of welfare, profits, or any cost or benefit the TC receives from different patient outcomes, as long as they are additive across patients. Such interpretations might lead to different planner objectives than those we consider here.

Generalizing concepts from earlier, a profile p is **non-wasteful** (for k) when there

are no more organs than patients at N or at T , i.e.

$$(1 - \sum p_{i\ell}/n) r_\ell + (1 - \sum p_{ih}/n) r_h \geq (1 - k)\phi \text{ and } (\sum p_{i\ell}/n) r_\ell + (\sum p_{ih}/n) r_h \geq k\phi$$

Analogous to (2), a non-wasteful profile p induces **allocation probabilities**

$$\pi^N = \frac{(1 - k)\phi}{(1 - \sum p_{i\ell}/n) r_\ell + (1 - \sum p_{ih}/n) r_h} \quad \pi^T = \frac{k\phi}{(\sum p_{i\ell}/n) r_\ell + (\sum p_{ih}/n) r_h}$$

The **payoff** to Transplant Center i at profile p is

$$u_i(p) = \frac{1}{n} \left[(1 - p_{i\ell}) r_\ell (\pi^N L^* + (1 - \pi^N) L_\ell^N) + (1 - p_{ih}) r_h (\pi^N L^* + (1 - \pi^N) L_h^N) \right. \\ \left. + p_{i\ell} r_\ell (\pi^T L^* + (1 - \pi^T) L_\ell^T) + p_{ih} r_h (\pi^T L^* + (1 - \pi^T) L_h^T) \right] \quad (7)$$

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, for each i , p_i is a best response to p_{-i} .

4.2 Equilibrium structure and intuition

The perfect competition setting yields intuitive equilibria: low types take action N and high types take action T , with the possible exception that patients of *one* type instead take the opposite action. The imperfect competition setting admits the possibility of equilibria with the “inverse” structure: TC’s assign low types to action T and high types to action N , again with a possible exception for only one type.

Theorem 3. *Fix k and suppose p is an equilibrium. There exists an equilibrium p^* that is payoff-equivalent to p , is symmetric, 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 $p_{ih}^* = 0$.

“Interior” equilibria are ruled out since, at any strategy profile, all TC’s face the same linear incentive to “swap” the opposite actions of opposite-type patients. At any interior profile, all TC’s would prefer executing the same such swaps until reaching the same kind of a corner solution (a Non-inversion or Inversion strategy).

For an intuition behind symmetry, note that the set of Non-inversion strategies is a monotonic, one-dimensional set: a decision of how many patients to send to Treatment, prioritizing high-type patients over low types. If TC i sends fewer patients

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 symmetric Non-inversion strategies; the same argument applies to Inversion.

Inversion equilibria appear perverse since each TC simultaneously chooses T for at least some of its low type patients and chooses N for at least some of its high type patients. For an intuition as to how such equilibria arise, consider TC i 's best response when (rationally or not) i 's competitors choose T for a “large” percentage of their patients. First, excess congestion at T could conceivably lead i to choose N for (at least some of) its *high* types in order to give those (high marginal value) patients better odds of an organ. Given this, it is conceivable that i prefers to choose T for its *low* types to avoid congesting its high types at N . This possibility is strongest when low types have low marginal value from organs and low value of N versus T . If this reasoning leads i to choose T for a “large” percentage of patients, we have constructed an equilibrium.

Computational analysis demonstrates that one or both forms of equilibria might exist (Subsection 4.3). One factor contributing to this—and complicating equilibrium analysis in general—is that payoff functions (7) exhibit convexities⁹ contrasting typical assumptions made in the literature on atomic congestion games. While our computational results suggest a general existence result, our analytical result below specifically shows that Non-inversion equilibria exist under mild additional assumptions. The first is that any patient’s benefit of receiving an organ exceeds the welfare difference between any two non-receiving patients.

Assumption 1 (Organs are sufficiently valuable). $L^* - L_\ell^N > L_\ell^N - L_h^N$.

Under this assumption, and for “reasonable” ratios k (i.e. no less than the proportional value $\bar{k} = r_h/(r_\ell + r_h)$), exactly one Non-inversion equilibrium exists as long as, relative to organ scarcity, high types benefit significantly from Treatment (8), and Treatment makes high and low type patients similar (9).¹⁰

Theorem 4. Fix $n \geq 3$ and suppose that Assumption 1 holds. If $k \geq \bar{k}$ and

$$\frac{L_h^T - L_h^N}{L_\ell^N - L_h^N} \geq \frac{\phi}{(r_\ell + r_h)n - \phi(n - 1)} \quad (8)$$

⁹In fact they locally violate concavity everywhere; see the online appendix.

¹⁰Even weaker assumptions are used in the proof, but require concepts from Subsection 4.4.

and

$$\frac{L_\ell^T - L_h^T}{(L_h^T - L_h^N) + (L_\ell^N - L_\ell^T)} < 1 - \frac{\phi}{(r_\ell + r_h)^{\frac{n-1}{n}}} \quad (9)$$

then there exists a unique Non-inversion equilibrium.

The proof cannot be applied to the full range of the model's primitives, and does not apply to Inversion equilibria due to complications arising from convexities in payoff functions. We therefore turn to computational analysis to investigate the prevalence of either type of equilibrium.

4.3 Computational analysis: equilibria and welfare

Across a wide range of economies and rations k we search for all (approximate) equilibria, classify each one (as Non-inverting or Inverting), and evaluate its welfare. We find that while Inversion equilibria can exist—even exclusively—they are atypical, arising when organs are relatively plentiful and when $L_\ell^T - L_h^T$ is relatively large. Not surprisingly these conditions oppose conditions (8) and (9).

To strengthen the argument that Inversion equilibria are atypical, we focus on primitives that should make Inversion equilibria *most* plausible. In these cases, our computational results suggest not only that Inversion equilibria have low prevalence, but that the equilibrium-welfare maximizing ration k typically leads to a Non-inversion equilibrium, with exceptions only in what we view as extreme cases. We next describe these primitives, providing technical details and additional computations in the online appendix.

We consider a full (discretized) range of values for $0 < \phi < r_\ell + r_h \equiv 1$ and rations $0 \leq k \leq 1$. We consider $n = 3$ TC's since (i) [Theorem 3](#) and [4](#) require $n \geq 3$ and (ii) additional computations confirm that Inversion equilibria become less prevalent as n increases (see [Observation 5](#) below).

Normalizing patient welfare values $L_h^N \equiv 0$ and $L_\ell^N \equiv 0.5$, we set L^* equal to $2L_\ell^N - L_h^N = 1$ since (i) this is a lower bound for L^* in [Assumption 1](#) and (ii) additional computations confirm Inversion equilibria become less prevalent as L^* increases (see [Observation 4](#) below). With these three values fixed, we consider the full (discretized) range of values for L_h^T and L_ℓ^T satisfying [Equation 1](#).

For all such instances we compute approximate equilibria, classifying each as an

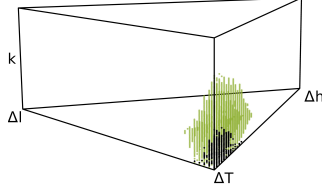


Figure 1: Equilibrium structure when $r_\ell = 0.3, \phi = 0.5$. Black: unique Inversion equilibrium. Unshaded: unique Non-inversion equilibrium. Green: one of equilibrium type.

Inversion or Non-inversion profile.¹¹ We present these results using the following notation.

$$\begin{aligned} \Delta_h &\equiv L_h^T - L_h^N \geq 0 & \Delta_T &\equiv L_\ell^T - L_h^T \geq 0 \\ \Delta_\ell &\equiv L_\ell^N - L_\ell^T \geq 0 & \Delta_* &\equiv L^* - L_\ell^N \geq 0 \end{aligned} \quad (10)$$

Our normalizations of L_h^N , L_ℓ^N , and L^* imply $\Delta_h + \Delta_T + \Delta_\ell = 0.5$ (and $\Delta_* = 0.5$). Thus the set of feasible choices of L_h^T and L_ℓ^T can be visualized as a 2-dimensional simplex representing feasible choices of the triplet $(\Delta_\ell, \Delta_T, \Delta_h)$, as in Figure 1. The vertical dimension of the prism represents the ration $k \in [0, 1]$.

As one example of our computations, the prism in Figure 1 shows which (approximate) equilibrium types exist when $r_\ell = 0.3$ and $\phi = 0.5$. More generally Figure 2 displays prisms for various r_ℓ and ϕ . Consistent with Theorem 4, a unique Non-inversion equilibrium exists for sufficiently large k and sufficiently small Δ_T . However they exist beyond the set of primitives assumed in that theorem (with uniqueness implied by Proposition 2). Even when one does not exist, a unique Inversion equilibrium does; we find (approximate) equilibrium existence across the full range of parameters despite the poorly behaved payoff functions that complicate more general analytical results. We now summarize our computational findings.

Observations. For the parameters discussed above we find:

1. For each instance considered, there exists a unique Non-inversion equilibrium or a unique Inversion equilibrium (or both).
2. Inversion equilibria are more prevalent as Δ_T becomes large. Since a large Δ_T ($\Delta_\ell, \Delta_h \approx 0$) implies a low cost to choosing the “wrong” action for either type. A TC may be induced to do so if this increases the rate at which its high types receive organs. If in addition $k = 0.5$, a Non-inversion strategy profile is

¹¹The use of a high-performance numerical computing Python library (Bradbury et al., 2018) was crucial in using GPU’s to overcome the computational challenges of searching for all equilibria in each of a large number of economies.

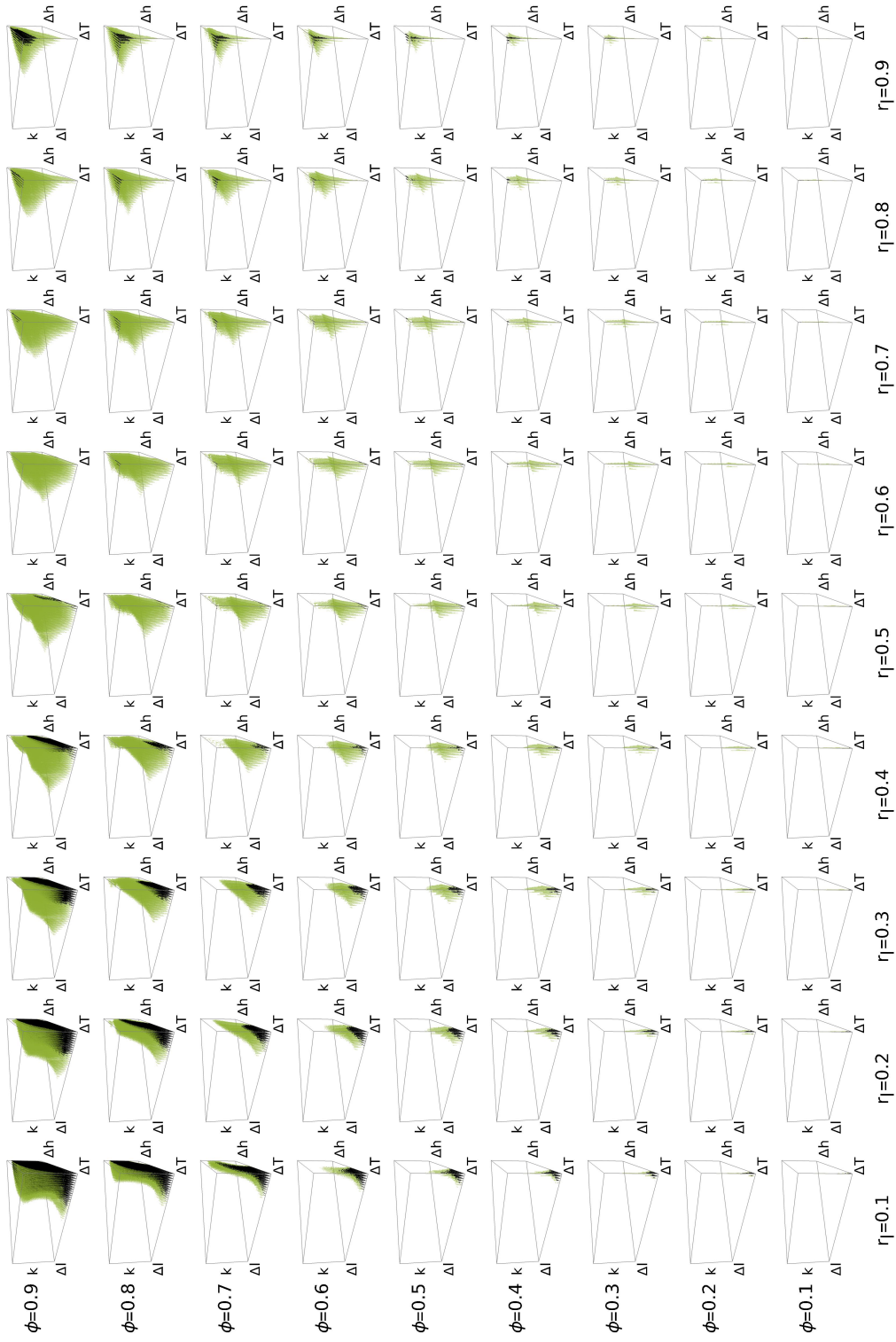


Figure 2: Equilibrium forms in various economies. Black: one Inversion equilibrium only. Unshaded: one Non-inversion equilibrium only. Green: one of each.

- almost payoff equivalent to its mirror (Inversion) profile obtained by reversing all patients’ actions; [Figure 2](#) exhibits equilibrium multiplicity in these cases.
3. Inversion equilibria are more prevalent when organs are more plentiful. When ϕ is large, the intuition for Inversion equilibria ([Subsection 4.2](#)) becomes stronger.
 4. Inversion equilibria are less prevalent as L^* increases (see online appendix). Intuitively, as the distinction between types and actions disappears we converge to a one-type congestion model with known existence and uniqueness results.
 5. Inversion equilibria are less prevalent as n increases (see online appendix), consistent with their non-existence under Perfect Competition ([Proposition 1](#)).

Observations 3 and 4 suggest that Inversion equilibria arise in cases where the planner’s allocation problem has lower stakes, i.e. organs are more plentiful or provide lower value. In remaining cases (where Δ_T is large) it makes little sense to Ration through Classification in the first place, since action choices have little screening power: N and T are almost cheap-talk messages.

The remaining question is on welfare: What ration k (and equilibrium form) maximizes welfare? Though intuition suggests that welfare maximizing equilibria should be Non-inverting, it is logically possible for Inversion equilibria to be optimal. To see why, suppose $\Delta_\ell \approx 0$ (low types have a low cost for Treatment), and start from some arbitrary Non-inversion profile at which the allocation probabilities satisfy $\pi^N < \pi^T$. If TC i increases $p_{i\ell}$, its payoff changes primarily in two ways: (i) it gains since its low-type patients receive organs more frequently and (ii) it loses because it crowds out its own high-type patients, who receive organs less frequently. If we had started from an Inversion profile where $\pi^N > \pi^T$, the same conclusion would follow from a decrease in $p_{i\ell}$, *but with a stronger effect (ii)*. The asymmetry in effect (ii) is because crowding out high types from receiving organs is *more costly* when those high types are taking the “wrong” action, N . Interestingly, this introduces the possibility for Inversion profiles to provide a *stronger* incentive for a TC *not* to crowd out its own (and others’) high types from receiving organs, because doing so is more costly than if actions were reversed.

Can this improvement in the high types’s share of organs outweigh the “wrong” choices being made over actions? Our computations illustrate that it can in some cases where Δ_ℓ is small ([Figure 3](#)). However the fact that it happens *only* when Δ_ℓ is small—even under primitives that make Inversion equilibria most plausible—justify our focus on Non-inversion equilibria for the remainder of the paper.

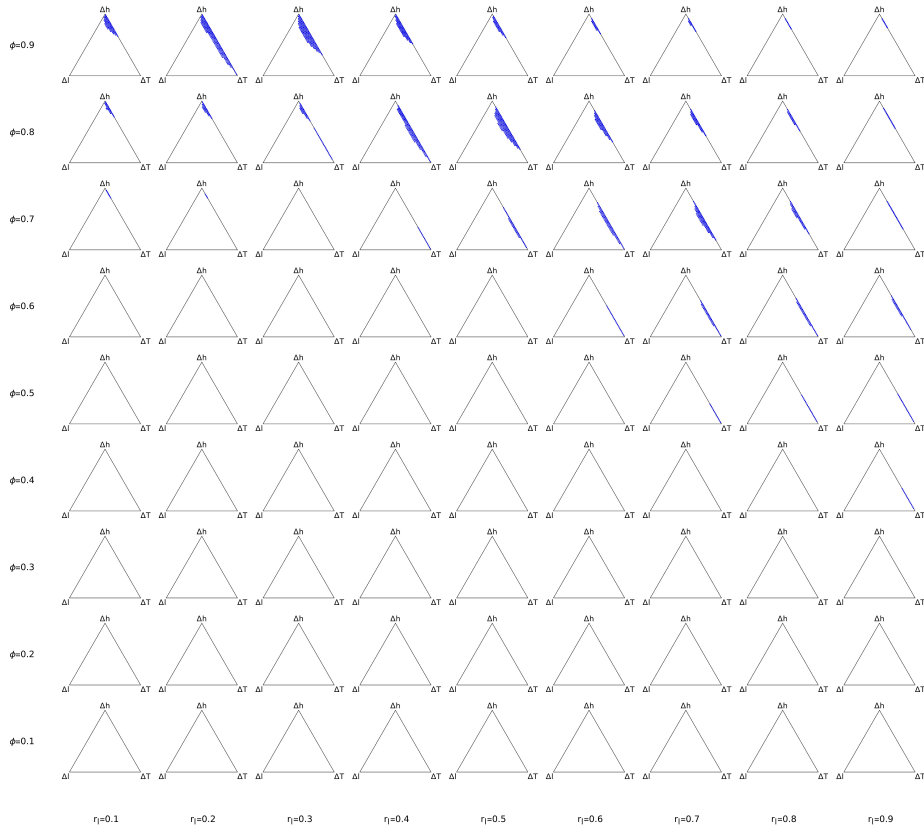


Figure 3: Across all values of k , the welfare-maximizing equilibrium is an Inversion equilibrium for economies in the blue regions.

4.4 Non-inversion equilibrium

We next show how the ideas of [Section 3](#) extend to Imperfect Competition via Non-inversion equilibria, the challenge being that in exceptional cases such equilibria may neither exist ([Figure 2](#)) nor be optimal (when Δ_ℓ is small; [Figure 3](#)). To overcome this, we first show general existence and uniqueness of “NI-candidates:” symmetric, Non-inversion strategy profiles satisfying only certain *local* IC constraints. Since any NI equilibrium is an NI-candidate, this proves at most one NI equilibrium can exist (as in our computations). We then prove our main results: the conclusions of [Theorem 1](#) and [Theorem 2](#) hold for any NI-candidate, and thus for any NI equilibrium.

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 following result extends [Proposition 1](#) to NI-candidates.

Proposition 2 (Unique NI-candidate). *Fix $n \geq 3$ and suppose [Assumption 1](#) holds. For any $k \in [0, 1]$ there exists a unique NI-candidate $p^*(k)$. Furthermore $p^*(\cdot)$ is weakly increasing in k , and*

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

where

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

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

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

It can be checked that k'_n and k_n^* converge to (3) and (4) as $n \rightarrow \infty$. The result is obtained by proving a limited form of concavity among Non-Inversion profiles (Lemma 5). Notably, a version of that lemma cannot hold for Inversion profiles, so we do not have corresponding results for such equilibria.

Extending the main conclusions of Theorem 1 and Theorem 2 to NI-candidates, both the fraction of organs allocated to high types and utilitarian welfare are single-peaked on $k \in [k'_n, 1]$ and are maximized at $k = k_n^*$.¹²

Theorem 5. *Fix $n \geq 3$ and suppose that Assumption 1 holds. Among all NI-candidates for $k \in [0, 1]$, both the equilibrium fraction of organs allocated to high types and TCs' total equilibrium payoffs are*

- *increasing in k for $k \in [k'_n, k_n^*]$,*
- *decreasing in k for $k \in [k_n^*, 1]$, and*
- *maximized at k_n^* (defined in Equation 12).*

Note that the intuition provided after Theorem 1 does not apply here. Namely, as $k \in [k_n^*, 1]$ increases, low types shift to T at a slower rate since a TC internalizes the congestion effect on its high types. Plausibly, an increase in k might then increase the equilibrium fraction of organs allocated to high types. Nevertheless Theorem 5 rules this out. It also leads to the following result.

Proposition 3 (Competition lowers welfare). *The ration k_n^* is decreasing in n . Therefore maximal welfare across all NI-candidates is decreasing in n .*

The simple proof of this (which we omit) is apparent in the following comparison between the perfect and imperfect competition scenarios. In the perfect competition case we can rewrite k^* as defined in (4) using the welfare differences defined in (10).

$$k^* = \frac{\Delta_\ell + \frac{\phi}{r_\ell} \Delta_*}{\frac{\phi}{r_\ell} \Delta_* + \frac{\phi}{r_h} (\Delta_* + \Delta_\ell)}$$

We also can rewrite k_n^* (see proof of Proposition 4).

$$k_n^* = \frac{\Delta_\ell + \frac{\phi}{r_\ell} \frac{n-1}{n} \Delta_*}{\frac{\phi}{r_h} \frac{n-1}{n} \Delta_\ell + \left(\frac{\phi}{r_\ell} + \frac{\phi}{r_h} \right) \frac{n-1}{n} \Delta_* - \frac{\phi}{r_h} \frac{1}{n} \Delta_T}$$

¹²One inconsequential contrast to Theorem 1 is a possible non-monotonicity on $k \in [0, k'_n]$.

Thus imperfect competition shrinks the numerator of k^* by less than a factor of $(1/n)$, and the denominator by more than that factor; hence $k_n^* > k^*$. However this expression also highlights the role of Δ_T : the TC internalizes the congestion its low types impose on its own high types at Treatment via Δ_T , the relative welfare difference between the types.

4.5 All-or-nothing prioritization

Though fractional rationing ($k < 1$) is typically optimal, in practice this approach may be infeasible for a variety of reasons (institutional or political constraints, complexity, etc.). Regardless of the reason, we compare the following two extremes of our approach more commonly seen in practice.

(FP) *Full prioritization* of one classification of recipients over the other.

(NP) *No Prioritization* of either classification over the other.

In our model FP corresponds to setting $k = 1$ (as long as $\phi \leq r_h$). NP uniformly rations organs across all patients regardless of action, i.e. each patient receives an organ with probability $\phi/(r_\ell + r_h)$. Under perfect competition, NP is the equilibrium result of setting $k = \bar{k} = r_h/(r_\ell + r_h)$ (see [Proposition 1](#)). Under imperfect competition, the same is typically true (e.g. when [Equation 8](#) holds).

It follows intuitively (assuming Non-inversion equilibria) that FP achieves close to the optimal level of welfare when k_n^* is larger (closer to 1) and NP is closer to optimal when k_n^* is lower (closer to \bar{k}). Therefore we reframe the comparison between FP and NP as a question of whether k_n^* is large or small. The following comparative statics use the welfare differences defined in [\(10\)](#).¹³

Proposition 4 (Conditions justifying FP over NP). *Consider varying primitives $L_h^N, L_h^T, L_\ell^T, L_\ell^N, L^*$ in a way that varies only one of the differences $\Delta_\ell, \Delta_T, \Delta_h, \Delta_*$, defined in [Equation 10](#), keeping the rest constant.*

- k_n^* is increasing in Δ_ℓ .
- k_n^* is increasing in Δ_T .
- k_n^* is constant in Δ_h .

¹³An alternative approach is to directly compute welfare at (i) the NI-candidate profile when $k = 1$ and (ii) the separation profile $(p_{i\ell}, p_{ih}) \equiv (0, 1)$ when $k = \bar{k}$. It tediously involves two roots when solving for $p_{i\ell}$ in case (i), and offers little additional insight to [Proposition 4](#).

- k_n^* is decreasing in Δ_* .

Furthermore k_n^* decreases in ϕ .

Natural intuition suggests why FP should be better than NP under these conditions. Large Δ_ℓ makes it costly for a low type receiving Treatment to fail to receive an organ. Large Δ_T increases the value of organs to high types, increasing the TC’s internalized cost or congesting them with low types. At k_n^* , a TC’s marginal decision (at Non-inversion profiles) does not involve high types, so Δ_h is irrelevant. Increasing Δ_* is analogous to reducing differences between types and actions; if the TC’s objective becomes organ share maximization it is harder to maintain type separation. This latter effect implies that an improvement in transplant technology benefiting both patient types could have the negative impact of increasing distortions. Finally when ϕ decreases there is less to be gained by choosing Treatment for low types.

5 Conclusion

Basing resource allocations on recipients’ observable actions allows a planner allocate resources more effectively but distorts action choices. When actions are chosen by the individual recipients (Section 3) this tradeoff yields a resource allocation problem that fits within a literature on strategic classification. Our work is the first to consider this problem in scenarios where action choices are partially centralized. In the language of organ allocation our strategic agents (Transplant Centers) choose actions (Treatment) on behalf of many patients.

Intuitively, fewer competing Transplant Centers internalize more of the congestion effect they impose on other patients by over-utilizing Treatment. This intuition holds in cases where equilibria of our general model are “Non-inverted,” having the same structure as in the setting where individual patients choose their own actions. Under such equilibria, action choices are distorted in at most *one* direction depending on the degree to which the planner biases allocation decisions; optimal welfare is achieved by eliminating these distortions (Theorem 5); and confirming the above intuition, optimal welfare increases as competition is lessened (Proposition 3).

In some cases, however, equilibria can be “inverted:” a Transplant Center’s action choices can be simultaneously distorted in two directions, allowing it to disproportionately improve allocation rates for its high-value patients. In exceptional cases such

counter-intuitive equilibria can even maximize equilibrium welfare. Computational exercises suggest that this occurs only in “low stakes” cases of our model.

Our work underscores two considerations for policy makers who allocate resources on the basis of actions. First, a rationing approach allows for welfare-improving compromises in the all-or-nothing choice between *full* or *no* prioritization of one classification of recipients over another. Second, the novelty of incorporating competition into our model highlights how optimal allocation policies can be sensitive to market concentration when agents make allocation-relevant decisions on behalf of many recipients. In applications such as organ allocation, optimal policies are likely to vary by region when market concentration does (Parker et al., 2018). Relatedly, Agarwal and Budish (2021) highlight market power as an understudied by necessary component of market design analysis. The literature where this interaction has been studied (e.g. auctions) shows how market power negatively impacts the designer’s objectives. Interestingly our setting yields the opposite conclusion.

References

- Nikhil Agarwal and Eric Budish. Market design. In *Handbook of Industrial Organization*, pages 1–79. Elsevier, 2021.
- Mohammad Akbarpour, Piotr Dworzak, and Scott Duke Kominers. Redistributive allocation mechanisms. *Journal of Political Economy*, 132(6):1831–1875, 2024.
- Ian Ball. Scoring strategic agents. *American Economic Journal: Microeconomics*, forthcoming.
- 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.
- James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. Version: 0.3.13.
- Mark Braverman and Sumegha Garg. The Role of Randomness and Noise in Strategic Classification. In *1st Symposium on Foundations of Responsible Computing (FORC*

- 2020), volume 156 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 9:1–9:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020.
- Alex Frankel and Navin Kartik. Improving Information from Manipulable Data. *Journal of the European Economic Association*, 20(1):79–115, 2021.
- 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.
- Frances Xu Lee and Wing Suen. Gaming a selective admissions system. *International Economic Review*, 64(1):413–443, 2023.
- 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.
- Minseon Park and Dong Woo Hahm. Location choice, commuting, and school choice. *Commuting, and School Choice (January 18, 2023)*, 2023.
- 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 Cardiology*, 71(16):1715–1725, 2018.

- Eduardo Perez-Richet and Vasiliki Skreta. Test design under falsification. *Econometrica*, 90(3):1109–1142, 2022.
- Eduardo Perez-Richet and Vasiliki Skreta. Fraud-proof non-market allocation mechanisms. 2023.
- 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*, 21(11):3684–3693, 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.
- Kurt R. Sweat. Endogenous priority in centralized matching markets: The design of the heart transplant waitlist. 2024.
- Cheng Wan. Coalitions in nonatomic network congestion games. *Mathematics of Operations Research*, 37(4):654–669, 2012.
- John Glen Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3):325–362, 1952.

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 $p_h = 1$. \square

Lemma 2. *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 2. 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 $(p_\ell, p_h) \leq (\tilde{p}_\ell, \tilde{p}_h)$ which also implies uniqueness ($k = \tilde{k}$).

Claim: either $(p_\ell, p_h) \leq (\tilde{p}_\ell, \tilde{p}_h)$ or $(p_\ell, p_h) \geq (\tilde{p}_\ell, \tilde{p}_h)$. 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: $(p_\ell, p_h) \leq (\tilde{p}_\ell, \tilde{p}_h)$. 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. \square

Proof of Proposition 1. By Lemmas 1 and 2 there exist $0 \leq k' \leq k^* \leq 1$ that define the three cases of Proposition 1. 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 $k' < \bar{k} < k^*$.

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

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

yielding k' as in (3). This is positive whenever the numerator is, yielding (5).

Similarly, low types are induced to choose N at the separation profile when

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

Analogous arguments lead to (4) and (6). \square

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

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

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 $f()$. \square

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

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 holds from Proposition 1). 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 $k \in [0, k']$. \square

6.2 Imperfect Competition

It is immediate that any equilibrium profile must be non-wasteful, so we restrict attention to non-wasteful profiles henceforth.

6.2.1 Equilibrium structure: Non-inversion/Inversion

We first observe that a TC i 's best response must be either a Non-inversion or Inversion *strategy*: a point on the boundary of $[0, 1]^2$. Fix a profile p with interior $p_i \in (0, 1)^2$, resulting in allocation probabilities π^N, π^T . Consider deviation p'_i obtained from p_i by “swapping” $\epsilon > 0$ mass of low types at T to N with ϵ mass of high types at N to T , i.e.

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

Since this deviation does not change the total masses of patients at N and T it does not change π^N and π^T . Therefore this deviation affects neither the other TCs'

payoffs nor i 's total consumption of organs. The deviation changes i 's payoff only in that, among i 's patients who fail to receive an organ, some who were assigned to N transform from high types into low types and some who were assigned to T turn from low types into high types. The magnitude of this change in payoff is

$$\epsilon[(1 - \pi^N)(L_\ell^N - L_h^N) + (1 - \pi^T)(L_h^T - L_\ell^T)] \quad (13)$$

While (13) can have any sign, its linearity in ϵ means that payoff functions are ruled surfaces, so a best response is a corner solution (or payoff-equivalent to one).

Lemma 3 (No double-mixing). *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 also 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 3. Fix i and p as in the Lemma. If (13) is positive, i would have the incentive to swap equal masses of low types at T with 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 (13) is negative then $p_{i\ell} = 1$ or $p_{ih} = 0$. In either case the result follows immediately by letting $p'_i = p_i$.

Suppose (13) equals zero. If $p_i \notin (0, 1)^2$, setting $p'_i = p_i$ proves the result. Otherwise let $(p'_{i\ell}, p'_{ih}) = (p_{i\ell} - \epsilon n/r_\ell, p_{ih} + \epsilon n/r_h)$ where, choosing ϵ maximally, $p'_i \notin (0, 1)^2$. Since (13) is zero p'_i is also a best response to p_{-i} , proving (i). Furthermore this deviation preserves i 's total masses of patients assigned N and T , implying (ii). \square

Fact 1. The following facts about (13) are used below.

- (i) Since $|(L_h^T - L_\ell^T)| < (L_\ell^N - L_h^N)$, $\pi^N = \pi^T$ implies that (13) is positive.
- (ii) $\pi^N = 1$ ($\pi^T = 1$) implies that (13) is negative (positive).
- (iii) When a change in strategy profile increases the total mass of patients at T , π^N increases, π^T decreases, and thus (13) decreases.

By Lemma 3, a best response is either Non-inverting or Inverting, or it can be replaced with a payoff equivalent such strategy without affecting other agents' payoffs. Figure 4 illustrates these best responses, though we henceforth ignore interior ones. In addition we can prove the following.

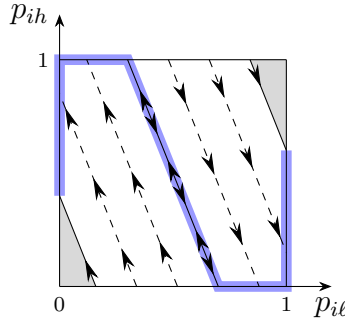


Figure 4: Changing TC i 's strategy along a dashed line (northwestly) changes i 's payoff according to (13); arrows represent a payoff increase. Best responses lie within the thick blue line. Depending on parameters, there may exist (shaded) regions of wasteful strategies.

Lemma 4. *Fix k . If profile p is an equilibrium, there exists a payoff-equivalent equilibrium p^* where either*

- (Non-inversion) for every TC i , $p_{i\ell}^* = 0$ or $p_{ih}^* = 1$; or
- (Inversion) for every TC i , $p_{i\ell}^* = 1$ or $p_{ih}^* = 0$.

Proof. Fixing such p , if (13) is positive (or negative) all TC's are using Non-inversion (or Inversion) strategies (Figure 4). If (13) is zero we can construct a payoff equivalent, Non-inverting profile p' (as in the proof of Lemma 3) at which each individual TC sends the same mass of patients to T at both profiles. Therefore for all i , all best responses to p_{-i} remain best responses to p'_{-i} , and thus p' also is an equilibrium. \square

6.2.2 Equilibrium structure: symmetry

To make the rest of our proofs more concise, we express strategies and the planner's rationing decision in terms of masses rather than percentages. To represent a planner's choice of k we denote the masses of organs rationed to N and T as

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

Similarly for a given strategy profile p and TC i 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, high types, and competitors that take action N ; D_i , E_i , and F_i correspond to T . We can write (7) (i 's payoff u_i) as

$$\begin{aligned} & A_i L_\ell^N + \frac{A_i}{A_i + B_i + C_i} \phi_N(L^* - L_\ell^N) + B_i L_h^N + \frac{B_i}{A_i + B_i + C_i} \phi_N(L^* - L_h^N) \\ & + D_i L_\ell^T + \frac{D_i}{D_i + E_i + F_i} \phi_T(L^* - L_\ell^T) + E_i L_h^T + \frac{E_i}{D_i + E_i + F_i} \phi_T(L^* - L_h^T) \end{aligned} \quad (14)$$

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} and, in some special cases, with respect to $p_{i\ell}$. The proof of the lemma also contains partial derivatives of payoffs utilized in later proofs.

Lemma 5 (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 $u_i(p)$ is either decreasing or 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 (14) is concave in E_i . Omitting subscript i , its derivative with respect to E (noting $B = r_h/n - E$) is

$$\begin{aligned} \frac{\partial u_i}{\partial E} &= \frac{A}{(A + B + C)^2} \phi_N(L^* - L_\ell^N) - L_h^N - \frac{A + C}{(A + B + C)^2} \phi_N(L^* - L_h^N) \\ &\quad - \frac{D}{(D + E + F)^2} \phi_T(L^* - L_\ell^T) + L_h^T + \frac{D + F}{(D + E + F)^2} \phi_T(L^* - L_h^T) \\ &= (L_h^T - L_h^N) + \frac{\phi_N}{A + B + C} \left(\frac{A}{A + B + C} (L_h^N - L_\ell^N) - \frac{C}{A + B + C} (L^* - L_h^N) \right) \\ &\quad + \frac{\phi_T}{D + E + F} \left(\frac{D}{D + E + F} (L_\ell^T - L_h^T) + \frac{F}{D + E + F} (L^* - L_h^T) \right) \\ &= (L_h^T - L_h^N) - \phi_N \frac{A(L_\ell^N - L_h^N) + C(L^* - L_h^N)}{(A + r_h/n - E + C)^2} + \phi_T \frac{D(L_\ell^T - L_h^T) + F(L^* - L_h^T)}{(D + E + F)^2} \end{aligned} \quad (15)$$

Since all bracketed terms are positive, (15) is decreasing in E . Therefore u_i is concave in E (i.e. in p_{ih}).

¹⁴The assumption that both $C_i \geq r_h/n$ and $p_{ih} = 0$ also is sufficient.

To show (ii) and (iii), the derivative of (14) with respect to D is

$$\begin{aligned}
\frac{\partial u_i}{\partial D} &= (L_\ell^T - L_\ell^N) + \frac{\phi_N}{A+B+C} \left(\frac{B}{A+B+C} (L_\ell^N - L_h^N) - \frac{C}{A+B+C} (L^* - L_\ell^N) \right) \\
&\quad + \frac{\phi_T}{D+E+F} \left(\frac{E}{D+E+F} (L_h^T - L_\ell^T) + \frac{F}{D+E+F} (L^* - L_\ell^T) \right) \\
&= \underbrace{(L_\ell^T - L_\ell^N)}_{\text{treatment effect}} + \underbrace{\phi_N \frac{B(L_\ell^N - L_h^N) - C(L^* - L_\ell^N)}{(r_\ell/n - D + B + C)^2}}_{N\text{-reallocation effect}} + \underbrace{\phi_T \frac{-E(L_\ell^T - L_h^T) + F(L^* - L_\ell^T)}{(D + E + F)^2}}_{T\text{-reallocation effect}}
\end{aligned} \tag{16}$$

While the treatment effect is negative, the overall sign of (16) depends on the signs of two “reallocation effects.” Denote

$$X = B(L_\ell^N - L_h^N) - C(L^* - L_\ell^N) \quad X' = -E(L_\ell^T - L_h^T) + F(L^* - L_\ell^T)$$

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).

Assumption 1 implies $X < X'$ (see online appendix); thus there are three cases.

- $X < X' \leq 0$: it is immediate that (16) 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 (16) 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 (16)’s sign and its direction of change w.r.t. D are indeterminate.

To prove statement (ii) of the lemma observe that if $p_{ih} = 1$ (i.e. $B = 0$) 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. \square

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

Lemma 6. For any k , any TCs i and j , and any non-wasteful profile p ,

$$\begin{aligned} \frac{\partial u_i}{\partial D_i} - \frac{\partial u_j}{\partial D_j} &= \frac{\partial u_i}{\partial E_i} - \frac{\partial u_j}{\partial E_j} \\ &= (D_j - D_i) \left[\frac{\phi_N(L^* - L_\ell^N)}{[r_\ell + r_h - (D_i + E_i + F_i)]^2} + \frac{\phi_T(L^* - L_\ell^T)}{(D_i + E_i + F_i)^2} \right] \\ &\quad + (E_j - E_i) \left[\frac{\phi_N(L^* - L_h^N)}{[r_\ell + r_h - (D_i + E_i + F_i)]^2} + \frac{\phi_T(L^* - L_h^T)}{(D_i + E_i + F_i)^2} \right] \end{aligned} \quad (17)$$

Proof. Rewriting (15) with $A_i = r_\ell/n - D_i$ and $C_i = (n-1)(r_\ell/n + r_h/n) - F_i$,

$$\begin{aligned} \frac{\partial u_i}{\partial E_i} &= (L_h^T - L_h^N) - \phi_N \frac{(r_\ell/n - D_i)(L_\ell^N - L_h^N) + ((n-1)(r_\ell/n + r_h/n) - F_i)(L^* - L_h^N)}{(r_\ell + r_h - D_i - E_i - F_i)^2} \\ &\quad + \phi_T \frac{D_i(L_\ell^T - L_h^T) + F_i(L^* - L_h^T)}{(D_i + E_i + F_i)^2} \end{aligned} \quad (18)$$

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 (18) are the same as those in the analogous expression for j . Hence

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

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

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

which equals (17). The same argument (in the online appendix) yields the same expression for $\partial u_i/\partial D_i - \partial u_j/\partial D_j$. \square

Proof of Theorem 3. By Lemma 4 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 $(p_{i\ell}, p_{ih}) \leq (p_{j\ell}, p_{jh})$, 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 (17) 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. \square

6.2.3 NI-candidate existence and uniqueness

To prove Proposition 2 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 (15) and (16) at such profiles.

$$\begin{aligned} \delta^W(E, k) &\equiv \frac{\partial u}{\partial E_i} \Big|_{\forall j D_j=0, E_j=E} & (19) \\ &= (L_h^T - L_h^N) - \phi_N \frac{r_\ell(L_\ell^N - L_h^N)}{n(r_\ell + r_h - nE)^2} - \phi_N \frac{(n-1)(L^* - L_h^N)}{n(r_\ell + r_h - nE)} + \phi_T \frac{(n-1)(L^* - L_h^T)}{n^2 E} \end{aligned}$$

$$\begin{aligned} \delta^N(D, k) &\equiv \frac{\partial u}{\partial D_i} \Big|_{\forall j D_j=D, E_j=\frac{r_h}{n}} & (20) \\ &= (L_\ell^T - L_\ell^N) - \phi_N \frac{(n-1)(L^* - L_\ell^N)}{n(r_\ell - nD)} - \phi_T \frac{r_h(L_\ell^T - L_h^T)}{n(nD + r_h)^2} + \phi_T \frac{(n-1)(L^* - L_\ell^T)}{n(nD + r_h)} \end{aligned}$$

With this notation we rewrite 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) $p_{ih}^* \equiv 1$, $r_\ell - nD_i = (1-k)\phi$, 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 high type-patients but requires mild assumptions for low-type patients since they congest their TC's high-type patients.

Lemma 7 (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

$$\frac{\partial \delta^W}{\partial k} = \phi \frac{r_\ell(L_\ell^N - L_h^N)}{n(r_\ell + r_h - nE)^2} + \phi \frac{(n-1)(L^* - L_h^N)}{n(r_\ell + r_h - nE)} + \phi \frac{(n-1)(L^* - L_h^T)}{n^2 E} > 0 \quad (21)$$

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,

$$\frac{\partial \delta^N}{\partial k} = \phi \frac{(n-1)(L^* - L_\ell^N)}{n(r_\ell - nD)} + \phi \frac{-r_h(L_\ell^T - L_h^T) + (n-1)(r_h + nD)(L^* - L_\ell^T)}{n(r_h + nD)^2} \quad (22)$$

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 . \square

Lemma 8 (Region 2-NI). *Fix k and let p^s denote the “NI-separating” profile, $p_{i\ell}^s \equiv 0$ and $p_{ih}^s \equiv 1$. Then p^s is an NI-candidate for k if and only if $k'_n \leq k \leq k_n^*$, where $k'_n < k_n^*$ are defined by [\(11\)](#) and [\(12\)](#).*

Proof. Fixing k , p^s is an 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 requirement is $\delta^W(r_h/n, k) \geq 0$ which, by Equation 19, 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

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

establishing (11).

The latter requirement is $\delta^N(0, k) \leq 0$ which, by Equation 20, is

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

which holds when

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

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

Assumption 1 and the assumption that $n \geq 3$ imply $\alpha > 0$. Dividing both sides of the inequality by α yields $k \leq k_n^*$ as defined in (12). \square

The following implies that separation (2-NI) occurs for non-degenerate values of k .

Lemma 9. For k'_n, k_n^* defined in (11)–(12), $k'_n < k_n^*$ and $\frac{r_h}{r_\ell + r_h} \equiv \bar{k} < k_n^*$.

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

Suppose $k'_n > 0$ and $k_n^* < 1$, hence $\delta^W(r_h/n, k'_n) = 0$ and $\delta^N(0, k_n^*) = 0$. Since δ^W is increasing in k , we prove the result by showing $\delta^W(r_h/n, k_n^*) > 0 = \delta^N(0, k_n^*)$. 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 $\bar{k} \equiv \frac{r_h}{r_\ell + r_h} < k_n^*$.

To show (i) we evaluate (21)–(22) at $(r_h/n, k)$ and $(0, k)$ (reordering the first two

terms of the first expression).

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

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} .

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

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

Secondly,

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

Note that

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

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

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

implying $k_n^* > k_n'$. □

Lemma 10 (Region 3-NI). *If $k > k_n^*$ then there exists a unique NI-candidate. It satisfies $p_{ih} \equiv 1$.*

Proof. Let p^s be defined as in Lemma 8 and (with a slight abuse of notation) recall $\delta^N(p^s, k_n^*) = 0$ by definition of k_n^* . By Lemma 8, $k > k_n^* > k_n'$ 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 8 similarly implies $\delta^N(p^s, k) > 0$. By Lemma 7, $\delta^N(\cdot, k)$ 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 5 (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 11 (Region 1-NI). *If $k < k_n'$ then there exists a unique NI-candidate. It satisfies $p_{i\ell} \equiv 0$.*

The omitted proof mirrors that of Lemma 10 with the simplification that, in reference to Lemma 5, payoffs are always concave in p_{ih} .

Proof of Proposition 2. NI-candidate existence, uniqueness, and their description follow from the above lemmas. Monotonicity of $p^*(\cdot)$ follows from Lemma 7. □

6.2.4 Optimal NI-candidate

The proof of [Theorem 5](#) relies on the following lemma, stating that in region 3-NI we have $\pi^T > \pi^N$.

Lemma 12 ($\pi^T > \pi^N$ 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 5](#), $p_{ih}^* = 1$ 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 (20) is zero, or the equilibrium is at a corner (where the N-nonwastefulness constraint binds and $\pi^N = 1$). However [Fact 1\(ii\)](#) rules out the latter, hence (20) 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 20](#) is zero we have

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

Thus

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

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

Proof of Theorem 5. 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](#) under Perfect Competition.

It is immediate that $f(\cdot)$ is increasing on $[k'_n, k_n^*]$ since the strategy profile is the same for all NI-candidates on this range. The remainder of the proof consists of showing (i) f is decreasing on $[k_n^*, 1]$, and (ii) $f(k) < f(k_n^*)$ for $k \in [0, k'_n]$.

Step (i). For any $k \in (k_n^*, 1]$, there is at most one symmetric profile (namely the NI-candidate $p(k)$) satisfying $\delta^N(D, k) = 0$ by [Lemma 10](#). 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 7](#) $D(k)$ is increasing in k ; hence the values of $k > k_n^*$ for which such $\delta^N(D(k), k) = 0$ exist are an interval (of the form $(k_n^*, x]$ 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 $\pi^T(\cdot)$ necessarily decreases $f(\cdot)$, proving (i).

We implicitly differentiate $\delta^N(D(k), k) = 0$ ([Equation 20](#)) 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 $nD = r_\ell$.) This yields

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

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

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

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

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

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 12](#). Hence $\pi^T()$ and $f()$ are decreasing on $[k_n^*, 1]$.

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

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

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

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: $f(\tilde{k}) \leq \frac{r_h}{r_\ell+r_h} < k_n^*$, where the second inequality follows from [Lemma 9](#) ($\bar{k} < k_n^*$).

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

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

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

we need to show

$$E'(\tilde{k}) \leq \frac{r_\ell + r_h - nE(\tilde{k})}{(1 - \tilde{k})n} \quad (23)$$

We implicitly differentiate $\delta^W(E(k), k) = 0$ (Equation 19) w.r.t. k and evaluate at \tilde{k} . Writing $E = E(k)$ and $E' = E'(k)$ we obtain

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

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 (23) 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 (23) holds. \square

Proof of Proposition 4. When $k_n^* < 1$ and $\phi \leq r_h$, Equation 12 takes the form

$$\begin{aligned} k_n^* &= \frac{(L_\ell^N - L_\ell^T) + \frac{\phi}{r_\ell} \frac{n-1}{n} (L^* - L_\ell^N)}{\frac{\phi}{r_\ell} \frac{n-1}{n} (L^* - L_\ell^N) + \frac{\phi}{r_h} \left[\frac{n-1}{n} L^* + \frac{1}{n} L_h^T - L_\ell^T \right]} \\ &= \frac{\Delta_\ell + \frac{\phi}{r_\ell} \frac{n-1}{n} \Delta_*}{\frac{\phi}{r_\ell} \frac{n-1}{n} \Delta_* + \frac{\phi}{r_h} \left[\frac{n-1}{n} (\Delta_* + \Delta_\ell) - \frac{1}{n} \Delta_T \right]} \\ &= \frac{\Delta_\ell + \frac{\phi}{r_\ell} \frac{n-1}{n} \Delta_*}{\frac{\phi}{r_h} \frac{n-1}{n} \Delta_\ell + \left(\frac{\phi}{r_\ell} + \frac{\phi}{r_h} \right) \frac{n-1}{n} \Delta_* - \frac{\phi}{r_h} \frac{1}{n} \Delta_T} = \frac{a\Delta_\ell + b\Delta_*}{a'\Delta_\ell + b'\Delta_* + c\Delta_T} \end{aligned}$$

where $a > a'$, $b < b'$, and $c < 0$. It is clearly decreasing in ϕ and increasing in Δ_T . Differentiating the last expression yields the remaining results since $k_n^* < 1$. \square

6.2.5 Non-inversion equilibrium existence

The proof of [Theorem 4](#) is presented last as it makes use of [Proposition 2](#). We prove a (technically) stronger result since (8) implies $\bar{k} \geq k'$ (proven in the Online Appendix).

Theorem. *Suppose $n \geq 3$ and that [Assumption 1](#) holds. If $k \geq \max\{k'_n, \bar{k}\}$ and (9) holds then there exists a unique Non-inversion equilibrium.*

Proof. Make the assumptions of the theorem and let p^* be the unique NI-candidate for k . We first show that TC i 's best response to p_{-i}^* must be a Non-inversion strategy, then show p_i^* is optimal among all such strategies.

Claim 1: any best response to p_{-i}^ satisfies $p_{i\ell} = 0$ or $p_{ih} = 1$.*

To prove Claim 1 it is sufficient to show that (13) is positive for any profile (p_i, p_{-i}^*) . Since p_{-i}^* is fixed throughout let $\pi^N(p_i)$ and $\pi^T(p_i)$ denote the allocation probabilities when i uses strategy p_i . We want to show that for any p_i ,

$$(1 - \pi^N(p_i))(\Delta_h + \Delta_T + \Delta_\ell) - (1 - \pi^T(p_i))\Delta_T > 0$$

If $\pi^N(p_i) \leq \pi^T(p_i)$ the inequality is immediate; if $\pi^N(p_i) > \pi^T(p_i)$ we must show

$$\frac{\Delta_T}{\Delta_h + \Delta_\ell} < \frac{1 - \pi^N(p_i)}{\pi^N(p_i) - \pi^T(p_i)} \quad (24)$$

Since $k \geq k'_n$ implies $p_{jh}^* = 1$ for all $j \neq i$, for all p_i we have

$$\pi^N(p_i) = \frac{(1-k)\phi}{\frac{n-1}{n}(1-p_{j\ell}^*)r_\ell + (1-p_{i\ell})\frac{r_\ell}{n} + (1-p_{ih})\frac{r_h}{n}} \leq \frac{(1-k)\phi}{\frac{n-1}{n}(1-p_{j\ell}^*)r_\ell} \quad (25)$$

Separately, $k \geq \max\{k'_n, \bar{k}\}$ implies

$$\frac{\phi(1-k)}{(1-p_{i\ell}^*)r_\ell} = \pi^N(p_i^*) < \pi^T(p_i^*) = \frac{\phi k}{r_h + p_{i\ell}^* r_\ell}$$

since either $k \geq k_n^*$ (in which case [Lemma 12](#) applies) or $k \in [k'_n, k_n^*]$ (in which case $p_{i\ell}^*(k) = 0$, $p_{ih}^*(k) = 1$, and $k \geq \bar{k}$ imply the inequality). The inequality can be rewritten as

$$p_{i\ell}^* < \frac{kr_\ell - (1-k)r_h}{r_\ell}$$

With (25) this means that for any p_i ,

$$\pi^N(p_i) \leq \frac{(1-k)\phi}{\frac{n-1}{n}(1-p_{i\ell}^*)r_\ell} < \frac{(1-k)\phi}{\frac{n-1}{n}\left(1 - \frac{kr_\ell - (1-k)r_h}{r_\ell}\right)r_\ell} = \frac{\phi}{\frac{n-1}{n}(r_\ell + r_h)}$$

Hence (9) implies $1 - \pi^N(p_i) > \Delta_T/(\Delta_H + \Delta_\ell)$, implying (24) and the claim.

Claim 2: p_i^ is a best response to p_{-i}^* .*

By Claim 1 it suffices to compare p_i^* only to other Non-inversion strategies. We show that $u_i(\cdot, p_{-i}^*)$ is concave across the entire range of such (non-wasteful) strategies, proving the result (since p_i^* is a local maximizer).

Lemma 5(i) implies $u_i(p_i, p_{-i}^*)$ is concave over the range where $p_{i\ell} = 0$ and $p_{ih} \in [0, 1]$. Lemma 5(iii) implies $u_i(p_i, p_{-i}^*)$ is concave over the range where $p_{i\ell} \in [0, 1]$ and $p_{ih} = 1$. Consider their intersection, $p'_i = (0, 1)$. At profile (p'_i, p_{-i}^*) ,

$$\begin{aligned} \frac{\partial u_i}{\partial E_i} - \frac{\partial u_i}{\partial D_i} &= (L_\ell^N - L_h^N) \left(1 - \frac{\phi_N}{\frac{r_\ell}{n} + C_i}\right) - (L_\ell^T - L_h^T) \left(1 - \frac{\phi_T}{\frac{r_h}{n} + F_i}\right) \\ &= (L_\ell^N - L_h^N)(1 - \pi^N(p'_i)) - (L_\ell^T - L_h^T)(1 - \pi^T(p'_i)) \end{aligned} \quad (26)$$

Note also that

$$\begin{aligned} \pi^T(p'_i) &= \frac{\phi_T}{r_h + \frac{n-1}{n}p_{j\ell}^*r_\ell} > \frac{\phi_T}{r_h + p_{j\ell}^*r_\ell} = \pi_T(p_i^*) \\ \pi^N(p'_i) &= \frac{\phi_N}{\frac{1}{n}r_\ell + \frac{n-1}{n}(1-p_{j\ell}^*)r_\ell} < \frac{\phi_N}{(1-p_{j\ell}^*)r_\ell} = \pi^N(p_i^*) \end{aligned}$$

Lemma 12 implies $\pi^T(p_i^*) > \pi^N(p_i^*)$, $\pi^T(p'_i) > \pi^N(p'_i)$. Combining with (26) we have $\frac{\partial u_i}{\partial E} > \frac{\partial u_i}{\partial D}$, i.e. $u_i(\cdot, p_{-i}^*)$ is concave at p'_i . \square

A Online Appendix

(This online appendix will ultimately be separated from the main paper and given its own title.)

A.1 Payoffs are nowhere-concave

A necessary condition for (weak) concavity of U_i is for the determinant of the Hessian matrix to be negative. Firstly, one can confirm that

$$\begin{aligned}\frac{\partial^2 U_i}{\partial D \partial E} &= \phi_N \frac{-(r_\ell + r_h - D - E - F)(L_\ell^N - L_h^N) + 2[(r_h/n - E)(L_\ell^N - L_h^N) - ((n-1)(r_\ell/n + r_h/n) - F)(L^* - L_\ell^N)]}{(r_\ell + r_h - D - E - F)^3} \\ &\quad - \phi_T \frac{(D + E + F)(L_\ell^T - L_h^T) + 2(-E(L_\ell^T - L_h^T) + F(L^* - L_\ell^T))}{(D + E + F)^3} \\ \frac{\partial^2 U_i}{\partial E^2} &= -2\phi_N \frac{(r_\ell/n - D)(L_\ell^N - L_h^N) + [(n-1)(r_\ell/n + r_h/n) - F](L^* - L_h^N)}{(r_\ell + r_h - D - E - F)^3} - 2\phi_T \frac{D(L_\ell^T - L_h^T) + F(L^* - L_h^T)}{(D + E + F)^3} \\ \frac{\partial^2 U_i}{\partial D^2} &= 2\phi_N \frac{(r_h/n - E)(L_\ell^N - L_h^N) - [(n-1)(r_\ell/n + r_h/n) - F](L^* - L_\ell^N)}{(r_\ell + r_h - D - E - F)^3} - 2\phi_T \frac{-E(L_\ell^T - L_h^T) + F(L^* - L_\ell^T)}{(D + E + F)^3}\end{aligned}$$

The determinant is $\frac{\partial^2 U_i}{\partial E^2} \cdot \frac{\partial^2 U_i}{\partial D^2} - \left(\frac{\partial^2 U_i}{\partial D \partial E}\right)^2$. Substituting with the above expressions and rearranging terms, the determinant can be written as

$$\frac{-S^2}{(D + E + F)^4 (r_h + r_\ell - (D + E + F))^4}$$

where (letting $r = r_\ell + r_h$)

$$S = -\phi_T (L_\ell^T - L_h^T) (r - (D + E + F))^2 - \phi_N (L_\ell^N - L_h^N) (D + E + F)^2$$

which is always negative. Therefore the determinant is negative for any parameter values (setting aside the two degenerate combinations where $D + E + F = 0 = k$ and where $(D + E + F)/r = 1 = k$), so U_i is not concave (nor convex) at any point.

A.2 Omitted argument in proof of Lemma 6

The claim is made that “a parallel argument shows that $\partial u_i / \partial D_i - \partial u_j / \partial D_j$ equals” the expression given in the proof. To formalize this argument here, recall $A_i = r_\ell/n - D_i$, $B_i = r_h/n - E_i$, and $C_i = (n-1)(r_\ell/n + r_h/n) - F_i$. The derivative of

(14) with respect to D_i is

$$\begin{aligned}
& (L_\ell^T - L_\ell^N) + \frac{\phi_N}{A+B+C} \left(\frac{B}{A+B+C} (L_\ell^N - L_h^N) - \frac{C}{A+B+C} (L^* - L_\ell^N) \right) \\
& + \frac{\phi_T}{D+E+F} \left(\frac{E}{D+E+F} (L_h^T - L_\ell^T) + \frac{F}{D+E+F} (L^* - L_\ell^T) \right) \\
& = (L_\ell^T - L_\ell^N) + \phi_N \frac{(r_h/n - E_i)(L_\ell^N - L_h^N) - ((n-1)(r_\ell/n + r_h/n) - F_i)(L^* - L_\ell^N)}{(r_\ell + r_h - D_i - E_i - F_i)^2} \\
& \quad + \phi_T \frac{-E_i(L_\ell^T - L_h^T) + F_i(L^* - L_\ell^T)}{(D_i + E_i + F_i)^2} \tag{27}
\end{aligned}$$

An analogous expression holds for j ,

$$\begin{aligned}
\frac{\partial u_j}{\partial D_j} &= (L_\ell^T - L_\ell^N) + \phi_N \frac{(r_h/n - E_j)(L_\ell^N - L_h^N) - ((n-1)(r_\ell/n + r_h/n) - F_j)(L^* - L_\ell^N)}{(r_\ell + r_h - D_j - E_j - F_j)^2} \\
& \quad + \phi_T \frac{-E_j(L_\ell^T - L_h^T) + F_j(L^* - L_\ell^T)}{(D_j + E_j + F_j)^2}
\end{aligned}$$

Again since $D_i + E_i + F_i = D_j + E_j + F_j$ this yields

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

as desired.

A.3 Assumption 1 implies $X < X'$ in proof of Lemma 5.

$$\begin{aligned}
X - X' &= B(L_\ell^N - L_h^N) - C(L^* - L_\ell^N) + E(L_\ell^T - L_h^T) - F(L^* - L_\ell^T) \\
&= \left(\frac{r_h}{n} - E\right)(L_\ell^N - L_h^N) - \left(\frac{n-1}{n}(r_\ell + r_h) - F\right)(L^* - L_\ell^N) + E(L_\ell^T - L_h^T) - F(L^* - L_\ell^T) \\
&= \frac{r_h}{n}(L_\ell^N - L_h^N) - \frac{n-1}{n}(r_\ell + r_h)(L^* - L_\ell^N) + F(L^* - L_\ell^N - L^* + L_\ell^T) - E(L_\ell^N - L_h^N - L_\ell^T + L_h^T) \\
&< \left[\frac{r_h}{n}(L_\ell^N - L_h^N) - \frac{n-1}{n}(r_\ell + r_h)(L_\ell^N - L_h^N)\right] + F(L_\ell^T - L_\ell^N) + E((L_\ell^T - L_h^T) - (L_\ell^N - L_h^N))
\end{aligned}$$

The inequality follows Assumption 1, and the final expression is the sum of three negative terms.

A.4 Inequality $\bar{k} \geq k'_n$

Rewriting the expression for k'_n (Equation 11) using the definitions in Equation 10, the inequality $\bar{k} > k'_n$ becomes

$$\frac{r_h}{r_\ell + r_h} \geq \frac{-\Delta_h + \frac{\phi}{r_\ell} \left(\Delta - \frac{1}{n}\Delta_*\right)}{\frac{\phi}{r_\ell} \left(\Delta - \frac{1}{n}\Delta_*\right) + \frac{\phi}{r_h} \frac{n-1}{n}(\Delta - \Delta_h)}$$

which is equivalent to

$$\phi \frac{n-1}{n} \Delta - \phi \left(\Delta - \frac{1}{n}\Delta_*\right) \geq \phi \frac{n-1}{n} \Delta_h - \Delta_h(r_\ell + r_h)$$

Simplifying and rearranging this expression leads to Equation 8.

A.5 Computational analysis: details

Here we summarize methodology and details of the computational analysis described in Subsection 4.3. The code is available at <http://www.kellogg.northwestern.edu/faculty/schummer/ftp/research/RTC/RTC-code.zip>.

The computations were performed on a workstation equipped with dual Intel Xeon Gold 5220R processors, each operating at 2.20 GHz, 256 GB of RAM, and two NVIDIA GV100 GPUs, each with 32 GB memory.

Fixing primitives that satisfy our assumptions, the goal of the computational exercises is to identify the equilibria that exist for various values of k , identify their structure, and compare welfare among all of them. The primitives are the pa-

tient masses (normalized to $r_\ell + r_h = 1$), organ mass $\phi < 1$, patient welfare levels $L^* > L_\ell^N > L_\ell^T > L_h^T > L_h^N$, and number of TC's n . First, to find equilibria for a given k we find (approximate) best response functions in a discretized strategy space and their (approximate) fixed points, leveraging some analytical results to simplify the search. Second, to find the planner's optimal ration k , we compare welfare under any Inversion equilibrium found in the previous step to welfare obtained at the Non-inversion equilibrium obtained when $k = k_n^*$, ignoring all other NI equilibria by [Theorem 5](#).

The following summary of details gives an overview of the code structure and additional technical details including how primitives and k are discretized. The result is one of the “prisms” found in [Figure 2](#).

Step 1: Compute TC i 's best responses (Python/Jax)

- Configure GPUs; Set precision to float64.
- Initialize set of economies.
 - Fix $n, r_\ell, r_h = 1 - r_\ell, \phi < 1$.
 - Normalize $L_h^N = 0, L_\ell^N = 0.5, L^* \geq 1$ ([Assumption 1](#)).
 - Consider values $0 < L_h^T < L_\ell^T < 0.5$ in increments of 0.02.
 - Consider values $0 \leq k \leq 1$ in increments of 0.01.
- Create grids for best response calculation.
 - Determine ‘admissible’ range of profiles p_{-i} (specifically, F_i in the notation of the paper’s appendix) that gives i a non-empty set of non-wasteful strategies.
 - Discretize domain of F_i (brmesh=100).
- [Parallelized] For each p_{-i} (F_i) and each of the four edges of $[0, 1]^2$ ([Lemma 3](#)) determine an edge-constrained best response. For each edge:
 - Find feasibility (non-wastefulness) constraints.
 - Find edge-constrained best responses (if edge is feasible).
 - * Optimization method: jaxopt.LBFGSB
 - * Tolerance: 10^{-10} ; Max Iterations: 1000
- Choose edge-constrained best response(s) yielding the largest payoff.

Step 2: Construct a best response function (Mathematica)

- Round numerically indistinguishable zeros and ones using a tolerance of Machine epsilon ($\approx 2.22e^{-16}$). Filter out repeated edge-constrained best responses.
- For a relatively small number of instances with multiple best response edges:

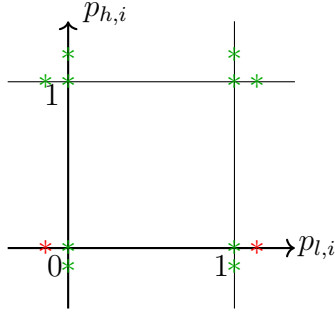


Figure 5: The figure depicts all the possible types of violations of probability constraints allowed by the optimization algorithm. Obviously, since we find one optimum per edge, at most one violation per edge is possible.

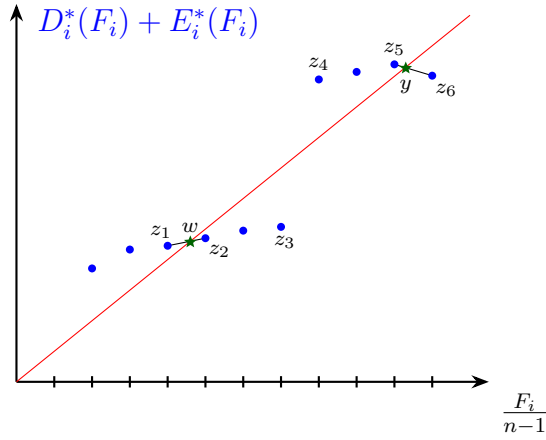


Figure 6: Masses

- Check for violations of probability constraints (expected in numerical optimization) and correct using analytical properties of the problem. Violation types are depicted in [Figure 5](#).
 - * Green violations can be removed by limited concavity ([Lemma 5](#)).
 - * Red violations: if origin belongs to best responses, ignore the violation and delete it.
- For all the remaining non-singletons, both best responses in the $(p_{l,i}, p_{h,i})$ space yield the same mass $D + E$, so preserve both.

Step 3: Find and classify equilibria (Mathematica)

- To find BR fixed points, find all critical pairs: pairs of consecutive points which are at opposite sides of the 45° line. For example, pairs (z_1, z_2) , (z_3, z_4) , (z_5, z_6) in [Figure 6](#)) which illustrates i 's best response to the per-capita mass of patients

i 's opponents send to $T (F_i/(n - 1))$.

- If both members of a critical pair yield the same (Inversion/Non-inversion) type of BR, interpolate a fixed point. E.g. if i 's best response is Non-inverting at both z_1, z_2 , label w as a fixed point, and hence a Non-inversion equilibrium. Otherwise classify as a jump discard the pair (e.g. (z_3, z_4)).
- If no equilibrium is found, refine the best response within the critical pair found.

Step 4: Welfare comparisons (in Mathematica)

- Compute welfare for all Inversion equilibria under all considered values of k .
- Explicitly calculate k_n^* (Equation 12) and compute welfare under corresponding Non-inversion equilibrium.
- Find the welfare-optimal choice of k (and equilibrium) among these candidates (Theorem 5).

A.6 Additional computational results

As L^* increases (fixing other parameters as in the main text), Inversion equilibria are replaced by Non-inversion equilibria.

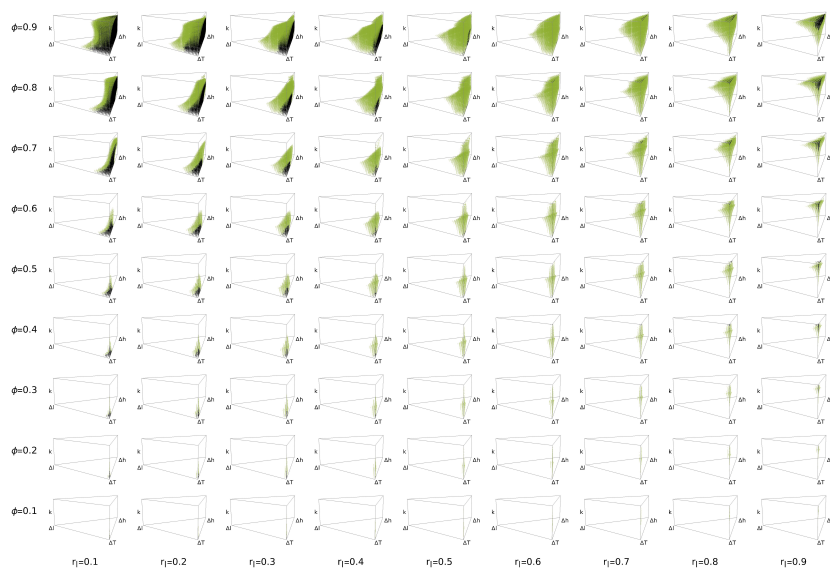


Figure 7: $L^* = 1.0$

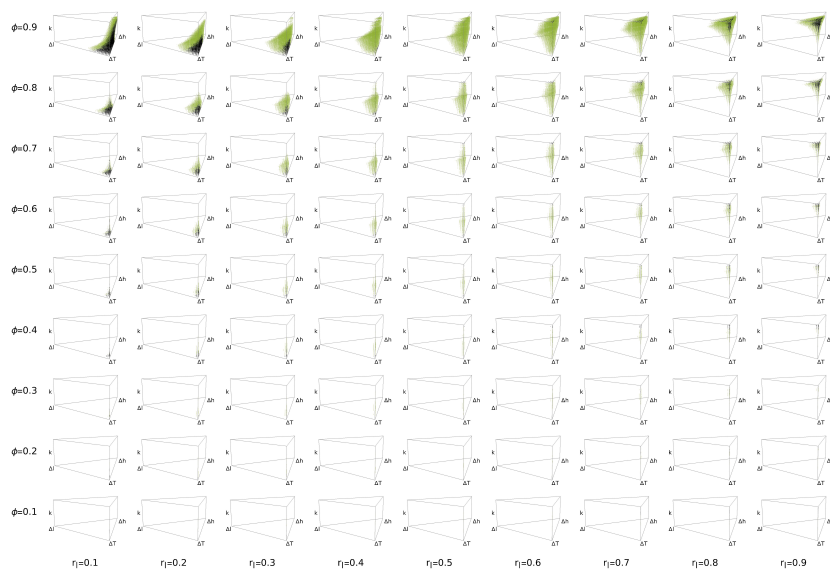


Figure 8: $L^* = 2.0$

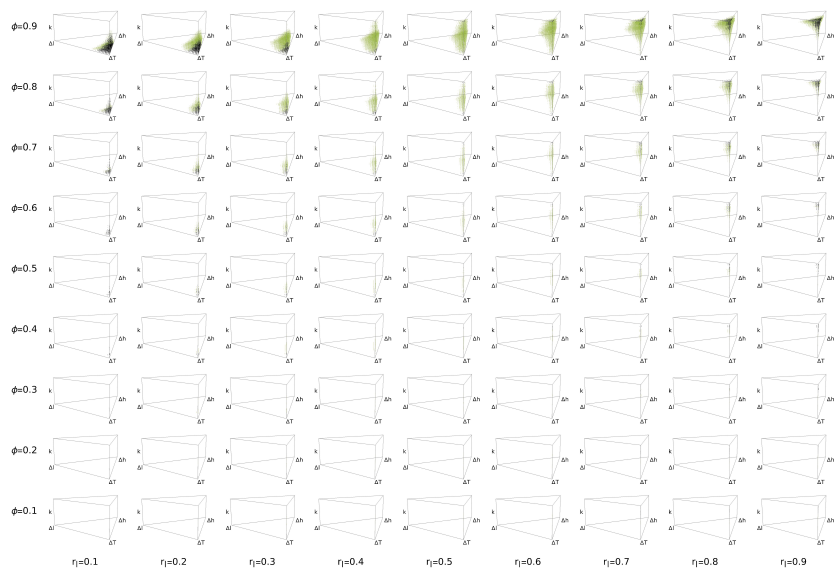


Figure 9: $L^* = 1.0$

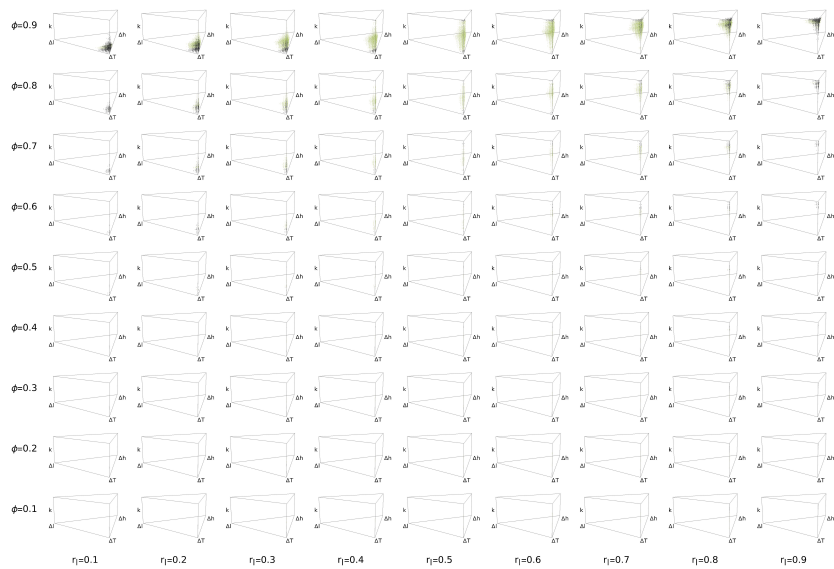


Figure 10: $L^* = 2.0$

As n increases (fixing other parameters as in the main text), Inversion equilibria are replaced by Non-inversion equilibria.

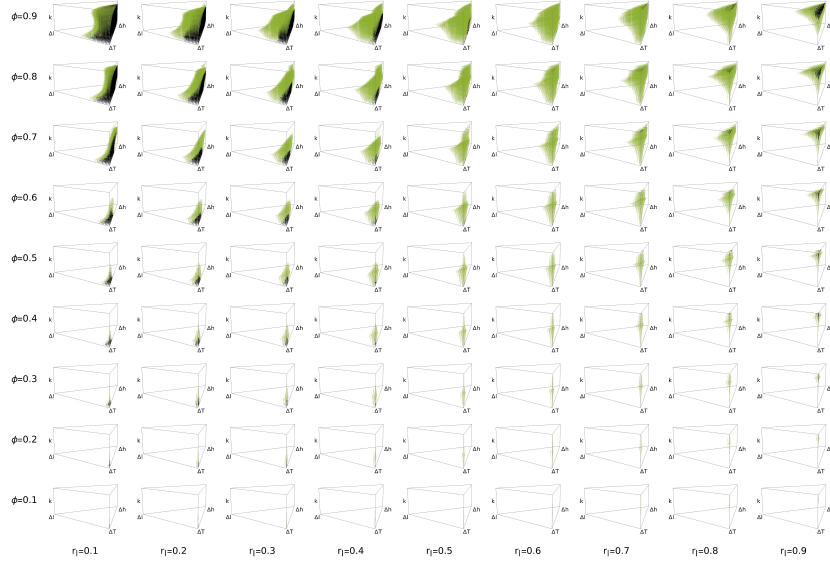


Figure 11: $n = 3$

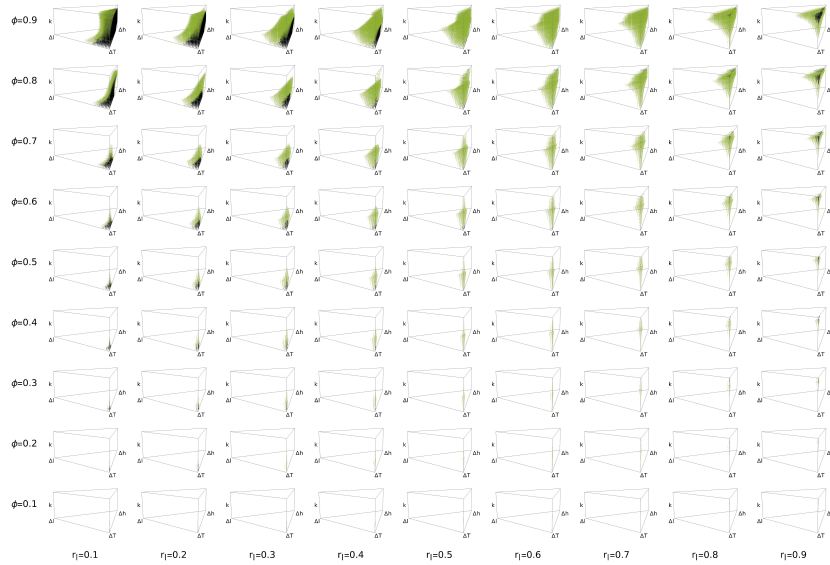


Figure 12: $n = 4$

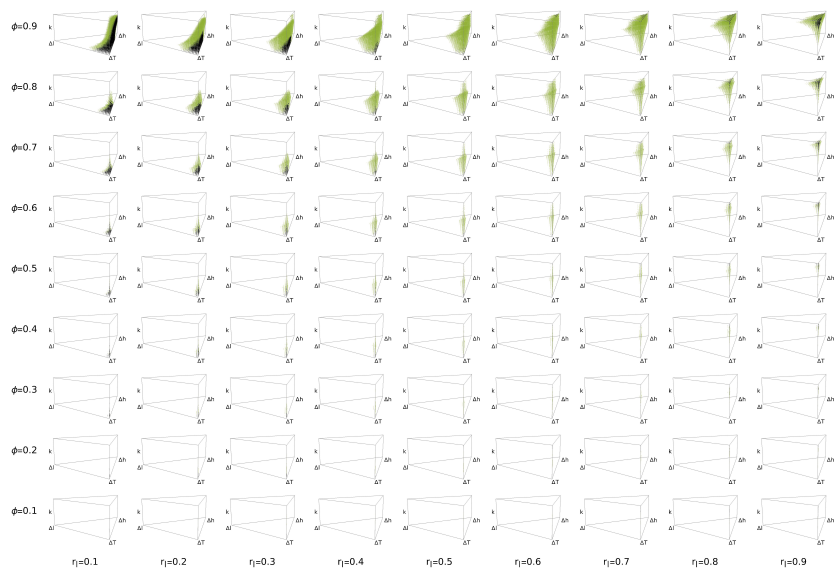


Figure 13: $n = 6$

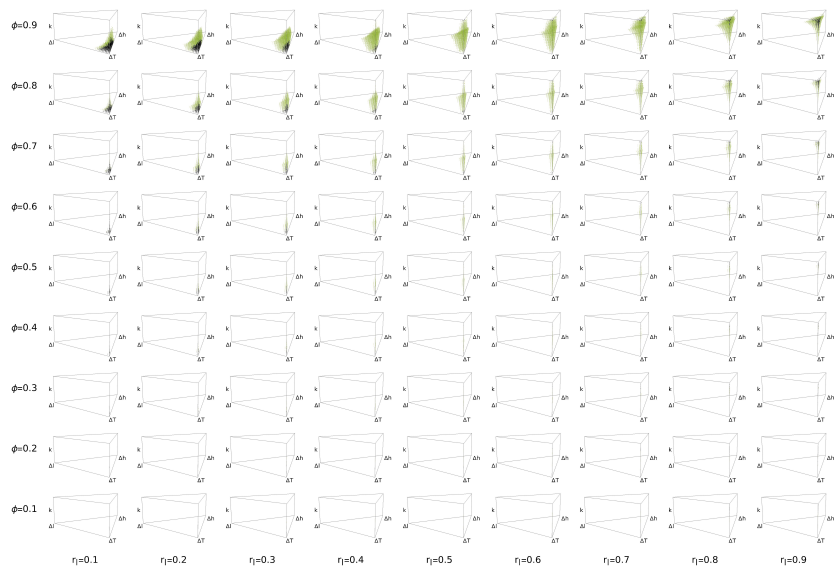


Figure 14: $n = 10$