Online Appendix for "Targeted Priority Mechanisms in Transplantation: Incentivizing, Not Enforcing, Efficient Matching of Organs"

Ruochen Wang
Queen's University Smith School of Business, Kingston, ON K7L 3N6, Canada, gx25@queensu.ca

Sait Tunç
Virginia Tech Grado Department of Industrial and Systems Engineering, Blacksburg, VA, 24061, sait.tunc@vt.edu

Burhaneddin Sandıkçı
Istanbul Technical University Department of Industrial Engineering, Istanbul, Turkey 34367, sandikcib@itu.edu.tr

Bekir Tanrıöver
The University of Arizona, Division of Nephrology, College of Medicine – Tucson, Tucson, AZ 857244, btanriover@arizona.edu

Matthew J. Ellis
Duke University School of Medicine, Department of Medicine & Department of Surgery, Durham, NC 27710, matthew.ellis@duke.edu

Appendix A: Preliminaries

A.1. Parameter List

Table S-1 Notations used in the analytical study*

Notation	Description
D	Disadvantaged patients
P	Disadvantaged patients who participate in the program
N	Disadvantaged patients who do not participate in the program
O	Non-targeted patients
λ	Overall arrival rate of candidates
λ_i	Arrival rate of type i candidates, $i \in \{O, D\}$
d_i	Death rate of type i candidates while waiting, $i \in \{O, D\}$
q	Organ quality score
$\underline{q}$	Lower bound of organ quality score
$\overline{q}$	Upper bound of organ quality score
μ	Arrival rate of organs at the waiting list (normalized to 1)
$p_i(q)$	Baseline offer probabilities representing the likelihood of an organ with quality q being offered to a type i candidate
μ_i	Arrival rate of organs to type i patients, $i \in \{O, D\}$
Δ_i	Time spent on the waiting list by type i candidates, $i \in \{O, N, P\}$
$T_i(q)$	Average lifespan of a type i candidate after receiving an organ of quality q, $i \in \{O, D\}$
$\overline{T}_i(q)$	Type i's net benefit of receiving a transplant of quality q as opposed to dying on the waitlist
$U_i(q_i)$	Expected QALE for a type i candidate having threshold q_i, $i \in \{O, N, P\}$
$\pi_i(q_i)$	Probability that a type i candidate having threshold q_i receives an organ, $i \in \{O, N, P\}$
α	Quality-of-life score for pre-transplant period
β	Quality-of-life score for post-transplant period
$s(\cdot)$	Social welfare function
$\overset{*}{q}$	Target threshold of the program
$p_i^{\gamma(\overset{*}{q})}(q)$	Allocation probability of non-program organs with quality q for type i candidates, $i \in \{O, N\}$
$\gamma(\overset{*}{q})$	Participating fraction of type D candidates to a targeted priority mechanism ($\overset{*}{q}$
$q_i^{e,\overset{*}{q}}$	Equilibrium offer acceptance threshold of type i candidates under a targeted priority mechanism ($\overset{*}{q}$), $i \in \{O, N, P\}$
$q_i^{e_0}$	Equilibrium offer acceptance thresholds of type i in the status quo, $i \in \{O, D\}$
$\overset{*}{q}_l$	Largest target threshold that ensures no participation of disadvantaged candidates into the program
$\overset{*}{q}_u$	Smallest target threshold that ensures full participation of disadvantaged patients into the program
$q_O^{\overline{e}}$	Equilibrium acceptance threshold of type O in the hypothetical scenario where they have absolute priority in the allocation of any organs

The table summarizes the notation and parameters used in the analytical model. Not all parameters listed here are directly used in the case study.

A.2. Preliminary Derivations

- Let $\mu_i(q_i)$ denote the effective arrival rate of organs to type-i candidates having an offer acceptance threshold q_i. We can express $\mu_i(q_i)$ by incorporating the offer probabilities for type i patients, focusing on the subset of organs that these patients are interested in, as follows:

$$\mu_i(q_i) = \frac{\overline{q} - q_i}{\overline{q} - \underline{q}} \cdot \int_{q_i}^{\overline{q}} p_i(q) \frac{1}{\overline{q} - q_i}\, dq = \frac{\int_{q_i}^{\overline{q}} p_i(q)\, dq}{\overline{q} - \underline{q}}. \tag{S-1}$$

We use the following asymptotic approximations for the expected pre-transplant life of type i and their probability of receiving a transplant:

$$\Delta_i(q_i) \approx \frac{1}{d_i} \cdot [1 - \pi_i(q_i)] \quad \text{and} \quad \pi_i(q_i) \approx \frac{\mu_i(q_i)}{\lambda_i}. \tag{S-2}$$

The formal proofs for the approximations given in (S-2) can be found in Zenios (1999). While the intuition of the approximation for $\pi_i(q_i)$ is clear, the intuition for $\Delta_i(q_i)$ can be better understood as follows. Let $\rho_i(n)$ denote the probability of having n type i candidates on the waiting list under steady-state. Then, we can write the removal rate of type-i candidates due to death as $\sum_{n=0}^{\infty} \rho_i(n) \cdot (n \cdot d_i) = d_i \cdot \sum_{n=0}^{\infty} n \cdot \rho_i(n)$, or equivalently as $\lambda_i \cdot (1 - \pi_i(q_i))$ since $1 - \pi_i(q_i)$ is the probability leaving the waiting list without a transplant. Observing that $\sum_{n=0}^{\infty} n \cdot \rho_i(n)$ corresponds to the average queue length of type i, which equals $\lambda_i \Delta_i(q_i^e)$ using the Little's Law, the approximation $\Delta_i(q_i) \approx \frac{1}{d_i} \cdot [1 - \pi_i(q_i)]$ follows.

- We define q_i' as the organ quality that satisfies the equation $\beta T_i(q) = \frac{\alpha}{d_i}$ as q for type i, $i \in \{O, D\}$, for the proofs of analytical results in Section B. Note that β, α, d_i are constants for type i, and $T_i(q)$ is monotonically increasing in q, therefore for each type i, the solution q_i' is unique.

Appendix B: Proofs of Analytical Results

THEOREM 1. *Under a targeted priority mechanism $(\overset{*}{q})$, candidates set their acceptance thresholds $q_i^{e,\overset{*}{q}}$ in equilibrium by solving the following set of equations:*

$$\beta T_O(q_O^{e,\overset{*}{q}}) = U_O(q_O^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})) := \alpha \Delta_O\left(q_O^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})\right) + \beta \pi_O(q_O^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})) E\left[T_O(q) \mid q \geqslant q_O^{e,\overset{*}{q}}\right], \tag{1a}$$

$$\beta T_D(q_N^{e,\overset{*}{q}}) = U_N(q_N^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})) := \alpha \Delta_D\left(q_N^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})\right) + \beta \pi_N(q_N^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})) E\left[T_D(q) \mid q \geqslant q_N^{e,\overset{*}{q}}\right], \tag{1b}$$

$$\beta T_D(q_P^{e,\overset{*}{q}}) = U_P(q_P^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})) := \alpha \Delta_D\left(q_P^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})\right) + \beta \pi_P(q_P^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})) E\left[T_D(q) \mid \overset{*}{q} \geqslant q \geqslant q_P^{e,\overset{*}{q}}\right]. \tag{1c}$$

Furthermore, the equilibrium participation rate $\gamma^e(\overset{*}{q})$ can be characterized as follows:

(a) *A mixed-participation equilibrium solves* (1a)–(1c) *and the following equation:*

$$U_N(q_N^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})) = U_P(q_P^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})). \tag{1d}$$

(b) *A no-participation equilibrium solves* (1a) *and* (1b) *with $\gamma^e(\overset{*}{q}) = 0$, and satisfies:*

$$U_N(q_N^{e,\overset{*}{q}}, 0) > \lim_{\gamma \to 0^+} U_P(q_P^{e,\overset{*}{q}}, \gamma). \tag{1e}$$

(c) *A full-participation equilibrium solves* (1a) *and* (1c) *with $\gamma^e(\overset{*}{q}) = 1$, and satisfies:*

$$U_P(q_P^{e,\overset{*}{q}}, 1) > \lim_{\gamma \to 1^-} U_N(q_N^{e,\overset{*}{q}}, \gamma). \tag{1f}$$

(d) *Under any participation equilibrium, the equilibrium acceptance thresholds $q_i^{e,\overset{*}{q}}$ are unique when they exist.*

Proof of Theorem 1. Under a symmetric participation equilibrium $\gamma^e(\overset{*}{q})$, the symmetric equilibrium acceptance threshold $q_i^{e,\overset{*}{q}}$ corresponds to the quality of an organ that type i is indifferent between accepting and rejecting. Recall that when an organ is offered, accepting the organ provides the benefit of transplantation, whereas rejecting gives

an opportunity to transplant a higher quality organ at the risk of dying without a transplant. Accordingly, under a participation equilibrium $\gamma^e(\hat{q})$, a type O candidate sets their equilibrium acceptance threshold by solving

$$\beta T_O(q_O^{e,\hat{q}}) = U_O(q_O^{e,\hat{q}}, \gamma^e(\hat{q})) := \alpha\Delta_O\left(q_O^{e,\hat{q}}, \gamma^e(\hat{q})\right) + \beta\pi_O\left(q_O^{e,\hat{q}}, \gamma^e(\hat{q})\right) E\left[T_O(q) \mid q \geqslant q_O^{e,\hat{q}}\right],$$

a non-participating type D candidate, if exists, sets their threshold by solving

$$\beta T_D(q_N^{e,\hat{q}}) = U_N(q_N^{e,\hat{q}}, \gamma^e(\hat{q})) := \alpha\Delta_N\left(q_N^{e,\hat{q}}, \gamma^e(\hat{q})\right) + \beta\pi_{D_N}\left(q_N^{e,\hat{q}}, \gamma^e(\hat{q})\right) E\left[T_D(q) \mid q \geqslant q_N^{e,\hat{q}}\right],$$

and a participating type D candidate, if exists, sets their threshold by solving

$$\beta T_D(q_P^{e,\hat{q}}) = U_P(q_P^{e,\hat{q}}, \gamma^e(\hat{q})) := \alpha\Delta_P\left(q_P^{e,\hat{q}}, \gamma^e(\hat{q})\right) + \beta\pi_{D_P}\left(q_P^{e,\hat{q}}, \gamma^e(\hat{q})\right) E\left[T_D(q) \mid \hat{q} \geqslant q \geqslant q_P^{e,\hat{q}}\right].$$

(a) A mixed-participation equilibrium exists when the eligible candidates are indifferent to participating in the program, which implies that a type D maintains the same utility whether they participate or not. Accordingly, we have

$$\begin{aligned}&\alpha\Delta_N\left(q_N^{e,\hat{q}}, \gamma^e(\hat{q})\right) + \beta\pi_{D_N}\left(q_N^{e,\hat{q}}, \gamma^e(\hat{q})\right) E[T_D(q) \mid q \geqslant q_N^e] \\ &= \alpha\Delta_P\left(q_P^{e,\hat{q}}, \gamma^e(\hat{q})\right) + \beta\pi_{D_P}\left(q_P^{e,\hat{q}}, \gamma^e(\hat{q})\right) E\left[T_D(q) \mid \hat{q} \geqslant q \geqslant q_P^{e,\hat{q}}\right], \end{aligned} \tag{S-3}$$

where the left- and right-hand sides of equation (S-3) correspond to a type D's expected utility from waitlisting by not participating (i.e., $U_N(q_N^{e,\hat{q}}, \gamma^e(\hat{q}))$) and participating (i.e., $U_N(q_N^{e,\hat{q}}, \gamma^e(\hat{q}))$) in the program, respectively.

(b) No disadvantaged patients participate in the program in equilibrium when the expected utility of not participating is greater than that of participating when no one else is participating, i.e.,

$$U_N(q_N^{e,\hat{q}}, 0) > \lim_{\gamma \to 0^+} U_P(q_P^{e,\hat{q}}, \gamma).$$

Under this equilibrium, no type D candidate participates in the program, i.e., $\gamma^e(\hat{q}) = 0$, and accordingly, types O and N set their equilibrium offer acceptance thresholds by solving equations (1a) and (1b) at $\gamma^e(\hat{q}) = 0$, respectively.

(c) All disadvantaged patients participate in the program in equilibrium when the expected utility of not participating is less than that of participating when everyone else is participating, i.e.,

$$U_P(q_P^{e,\hat{q}}, 1) > \lim_{\gamma \to 1^-} U_N(q_N^{e,\hat{q}}, \gamma).$$

Under this equilibrium, all type D candidates participate in the program, i.e., $\gamma^e(\hat{q}) = 1$, and accordingly, type O and type P set their acceptance thresholds by solving equations (1a) and (1c) at $\gamma^e(\hat{q}) = 1$, respectively.

(d) Let $\mathcal{L}(q)$ and $\mathcal{R}(q)$ respectively denote the left- and right-hand sides of equation (1b). For notational simplicity, in the remainder of the proof, we use $\Delta_i\left(q_i^{e,\hat{q}}\right)$ and $\pi_i\left(q_i^{e,\hat{q}}\right)$ to denote $\Delta_i\left(q_i^{e,\hat{q}}, \gamma^e(\hat{q})\right)$ and $\pi_i\left(q_i^{e,\hat{q}}, \gamma^e(\hat{q})\right)$, respectively. Then we have,

$$\begin{aligned}\mathcal{R}\left(q_N^{e,\hat{q}}\right) &= \alpha\Delta_N\left(q_N^{e,\hat{q}}\right) + \beta\pi_N\left(q_N^{e,\hat{q}}\right) E\left(T_D(q) \mid q \geqslant q_N^{e,\hat{q}}\right) \\ &= \alpha\frac{1}{d_D}\left(1 - \pi_N\left(q_N^{e,\hat{q}}\right)\right) + \beta\pi_N\left(q_N^{e,\hat{q}}\right) \int_{q_N^{e,\hat{q}}}^{\bar{q}} T_D(q) \frac{p_N(q)}{\int_{q_N^{e,\hat{q}}}^{\bar{q}} p_N(q)dq} dq \\ &= \alpha\frac{1}{d_D}\left(1 - \frac{\int_{q_N^{e,\hat{q}}}^{\bar{q}} p_N(q)dq}{\lambda_N(\bar{q} - \underline{q})}\right) + \beta\frac{1}{\lambda_N(\bar{q} - \underline{q})} \int_{q_N^{e,\hat{q}}}^{\bar{q}} T_D(q)p_N(q)dq,\end{aligned}$$

and

$$\frac{d\mathcal{R}\left(q_N^{e,\bar{q}^*}\right)}{d\,q_N^{e,\bar{q}^*}} = \frac{\alpha}{d_D}\cdot\frac{p_N(q_N^{e,\bar{q}^*})}{(\bar{q}-\underline{q})\cdot\lambda_N} - \frac{\beta}{\lambda_N(\bar{q}-\underline{q})}T_D(q_N^{e,\bar{q}^*})p_N(q_N^{e,\bar{q}^*})$$
$$= \frac{p_N(q_N^{e,\bar{q}^*})}{\lambda_N(\bar{q}-\underline{q})}\left(\frac{\alpha}{d_D}-\beta T_D(q_N^{e,\bar{q}^*})\right),$$

which implies that $\frac{d\,\mathcal{R}\left(q_N^{e,\bar{q}^*}\right)}{d\,q_N^{e,\bar{q}^*}} = 0$ when $q_N^{e,\bar{q}^*} = q'_D$, where q'_D uniquely solves $\beta T_D(q) = \frac{\alpha}{d_D}$, which has a unique solution because $\beta T_D(q)$ is increasing for $q \in \left(\underline{q},\bar{q}\right)$. Therefore, for $q \in (\underline{q}, q'_D)$, we have $\frac{d\,\mathcal{R}(q)}{d\,q} > 0$ and $\beta T_D(q) < \frac{\alpha}{d_N}$, on the other hand, for $q \in (q'_D, \bar{q})$, we have $\frac{d\,\mathcal{R}(q)}{d\,q} < 0$ and $\beta T_D(q) > \frac{\alpha}{d_D}$.

Then

$$\begin{aligned}
\mathcal{R}(q'_D) &= \frac{\alpha}{d_D}(1-\pi_N(q'_D)) + \frac{1}{\lambda_N(\bar{q}-\underline{q})}\left(\int_{q'_D}^{\bar{q}}\beta T_D(q)p_N(q)dq\right)\\
&\geq \frac{\alpha}{d_D}(1-\pi_N(q'_D)) + \frac{1}{\lambda_N(\bar{q}-\underline{q})}\left(\int_{q'_D}^{\bar{q}}\beta T_D(q'_D)p_N(q)dq\right)\\
&= \frac{\alpha}{d_D}\left(1-\frac{\int_{q'_D}^{\bar{q}}p_N(q)dq}{\lambda_N(\bar{q}-\underline{q})}\right) + \frac{1}{\lambda_N(\bar{q}-\underline{q})}\left(\int_{q'_D}^{\bar{q}}\beta T_D(q'_D)p_N(q)dq\right)\\
&= \frac{\alpha}{d_D} + \frac{1}{\lambda_N(\bar{q}-\underline{q})}\int_{q'_D}^{\bar{q}}\left(\beta T_D(q'_D)-\frac{\alpha}{d_D}\right)p_N(q)dq\\
&= \frac{\alpha}{d_D} = \beta T_N(q'_D) = \mathcal{L}(q'_D)
\end{aligned} \tag{S-4}$$

and

$$\begin{aligned}
\mathcal{R}(\bar{q}) &= \frac{\alpha}{d_D}(1-\pi_N(\bar{q})) + \beta\pi_N(\bar{q})E[T_D(q)\mid q\geqslant\bar{q}]\\
&< \beta T_D(\bar{q})(1-\pi_N(\bar{q})) + \beta\pi_N(\bar{q})T_D(\bar{q})\\
&= \beta T_N(\bar{q}) = \mathcal{L}(\bar{q})
\end{aligned} \tag{S-5}$$

and

$$\begin{aligned}
\mathcal{R}(\underline{q}) &= \frac{\alpha}{d_D}(1-\pi_N(\underline{q})) + \pi_N(\underline{q})\int_{\underline{q}}^{\bar{q}}\beta T_D(q)\frac{p_N(q)}{\int_{\underline{q}}^{\bar{q}}p_N(q)dq}dq\\
&> \beta T_D(\underline{q})(1-\pi_N(\bar{q})) + \pi_N(\underline{q})\int_{\underline{q}}^{\bar{q}}\beta T_D(\underline{q})\frac{p_N(q)}{\int_{\underline{q}}^{\bar{q}}p_N(q)dq}dq\\
&= \beta T_N(\underline{q}) = \mathcal{L}(\underline{q}).
\end{aligned} \tag{S-6}$$

Since $\mathcal{L}(q)$ is increasing for $q \in [\underline{q},\bar{q}]$, and $\mathcal{R}(q)$ is increasing for $q \in [\underline{q}, q'_D)$ and decreasing for $q \in (q'_D,\bar{q}]$, using the intermediate value theorem, (S-4)–(S-6) conclude that if the equilibrium threshold q_N^e exists, it is unique, in $(q'_D,\bar{q}]$.

Similarly, letting $\mathcal{L}(q)$ and $\mathcal{R}(q)$ denote the left- and right-hand sides of equation (1c), respectively, we have,

$$\mathcal{R}\left(q_P^{e,\bar{q}^*}\right) = \alpha\Delta_P\left(q_P^{e,\bar{q}^*}\right) + \beta\pi_P\left(q_P^{e,\bar{q}^*}\right)E\left(T_D(q)\mid\bar{q}^*\geqslant q\geqslant q_P^{e,\bar{q}^*}\right)$$
$$= \frac{\alpha}{d_D}\left(1-\frac{\bar{q}^*-q_P^{e,\bar{q}^*}}{\lambda_P(\bar{q}-\underline{q})}\right) + \frac{\beta}{\lambda_P(\bar{q}-\underline{q})}\int_{q_P^{e,\bar{q}^*}}^{\bar{q}^*}T_D(q)dq,$$

and

$$\frac{d\,\mathcal{R}\left(q_P^{e,\,\overset{*}{q}}\right)}{d\,q_P^{e,\,\overset{*}{q}}} = \frac{\alpha}{d_D}\cdot\frac{1}{(\bar{q}-\underline{q})\cdot\lambda_P} - \frac{\beta}{\lambda_P(\bar{q}-\underline{q})}T_D(q_P^{e,\,\overset{*}{q}}) = \frac{1}{\lambda_P(\bar{q}-\underline{q})}\left(\frac{\alpha}{d_D}-\beta T_D(q_P^{e,\,\overset{*}{q}})\right),$$

which implies that $\frac{d\,\mathcal{R}\left(q_P^{e,\,\overset{*}{q}}\right)}{d\,q_P^{e,\,\overset{*}{q}}} = 0$ when $q_P^{e,\,\overset{*}{q}} = q'_D$, where q'_D uniquely solves $\beta T_D(q) = \frac{\alpha}{d_D}$, which has a unique solution as shown above. Therefore, similar to the previous case, for $q \in (\underline{q}, q'_D)$, we have $\frac{\partial \mathcal{R}(q)}{\partial q} > 0$ and $\beta T_D(q) < \frac{\alpha}{d_D}$; on the other hand, for $q \in (q'_D, \bar{q})$, we have $\frac{\partial \mathcal{R}(q)}{\partial q} < 0$ and $\beta T_D(q) > \frac{\alpha}{d_D}$.

Then, following similar steps, we have

$$\begin{aligned}\mathcal{R}(q'_D) &= \frac{\alpha}{d_D}\left(1-\frac{\overset{*}{q}-q'_D}{\lambda_P(\bar{q}-\underline{q})}\right)+\frac{\beta}{\lambda_P(\bar{q}-\underline{q})}\int_{q'_D}^{\overset{*}{q}} T_D(q)dq\\ &\geqslant \frac{\alpha}{d_D}+\frac{1}{\lambda_P(\bar{q}-\underline{q})}\int_{q'_D}^{\overset{*}{q}}\left(\beta T_D(q'_D)-\frac{\alpha}{d_D}\right)dq\\ &= \beta T_P(q'_D) = \mathcal{L}(q'_D),\end{aligned} \tag{S-7}$$

and

$$\begin{aligned}\mathcal{R}(\overset{*}{q}) &= \frac{\alpha}{d_D}(1-\pi_P(\overset{*}{q}))+\beta\pi_P(\overset{*}{q})E[T_D(q)\mid \overset{*}{q}\geqslant q\geqslant \overset{*}{q}]\\ &< \beta T_D(\overset{*}{q})(1-\pi_P(\overset{*}{q}))+\beta\pi_P(\overset{*}{q})T_D(\overset{*}{q})\\ &= \beta T_P(\overset{*}{q}) = \mathcal{L}(\overset{*}{q}),\end{aligned} \tag{S-8}$$

and finally,

$$\begin{aligned}\mathcal{R}(\underline{q}) &= \frac{\alpha}{d_D}\left(1-\frac{\overset{*}{q}-\underline{q}}{\lambda_P(\bar{q}-\underline{q})}\right)+\frac{\beta}{\lambda_P(\bar{q}-\underline{q})}\int_{\underline{q}}^{\overset{*}{q}} T_D(q)dq\\ &> \beta T_D(\underline{q})\left(1-\frac{\overset{*}{q}-\underline{q}}{\lambda_P(\bar{q}-\underline{q})}\right)+\frac{\beta}{\lambda_P(\bar{q}-\underline{q})}\int_{\underline{q}}^{\overset{*}{q}} T_D(\underline{q})dq\\ &= \beta T_P(\underline{q}) = \mathcal{L}\big(\underline{q}\big).\end{aligned} \tag{S-9}$$

Since $\mathcal{L}(q)$ is increasing for $q \in [\underline{q}, \bar{q}]$, and $\mathcal{R}(q)$ is increasing for $q \in [\underline{q}, q'_D)$ and decreasing for $q \in (q'_D, \overset{*}{q}]$, using the intermediate value theorem, (S-7)–(S-9) imply that if the equilibrium threshold $q_P^{e,\,\overset{*}{q}}$ exists, it is unique, in $(q'_D, \overset{*}{q}]$. Finally, the uniqueness of $q_O^{e,\,\overset{*}{q}}$ follows similarly, and its proof is omitted here for brevity.

□

COROLLARY 1. *Under a mixed-participation equilibrium $\gamma^e(\overset{*}{q}) \in (0,1)$, the participating and non-participating type D candidates adopt the same acceptance thresholds in equilibrium, i.e., $q_P^{e,\,\overset{*}{q}} = q_N^{e,\,\overset{*}{q}}$.*

Proof of Corollary 1. Theorem 1 establishes that a mixed-participation equilibrium satisfies equations (1b)–(1d). Using equation (1d) in equations (1b) and (1c), we have $\beta T_D(q_N^{e,\,\overset{*}{q}}) = \beta T_D(q_P^{e,\,\overset{*}{q}})$. And because $T_D(q)$ is increasing in q, we conclude that $q_P^{e,\,\overset{*}{q}} = q_N^{e,\,\overset{*}{q}}$ under any mixed-participation equilibrium. □

PROPOSITION 1. *Under a targeted priority mechanism $(\mathring{q})$, no eligible candidates would participate in the program, i.e., the equilibrium participation rate satisfies $\gamma^e(\mathring{q})=0$, if and only if*

$$\beta T_D(\mathring{q}) < \frac{\alpha}{d_D} + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})} \lim_{\gamma\to 0^+} \int_{\mathring{q}}^{\bar{q}} \left(T_D(q) - \frac{\alpha}{d_D}\right) p_N^{\gamma}(q) dq. \tag{2}$$

Proof of Proposition 1. First, observe that when $\gamma^e(\mathring{q})=0$, i.e., when no one is participating, if an individual type D participates in the program, the expected utility of her participation is $\beta T_D(\mathring{q})$. This is so because when there is no disadvantageous patient in the program, the first person participating will have fully prioritized access to all organs reserved for the program. Therefore, w.p. 1, the patient will receive an organ of quality $\mathring{q}$ with the expected utility $\beta T_D(\mathring{q})$. Note that $\mathring{q}$ should be greater than q'_D to make sure that $\beta T_D(\mathring{q}) > \frac{\alpha}{d_D}$. On the other hand, the expected utility of not participating in the program when $\gamma^e(\mathring{q})=0$, is given by:

$$\begin{aligned}
\beta T_D(q_N^{e,\mathring{q}}) &= \alpha \Delta_N\left(q_N^{e,\mathring{q}}\right) + \beta \pi_{D_N}\left(q_N^{e,\mathring{q}}\right) E\left(T_D(q) \mid q \geqslant q_N^{e,\mathring{q}}\right) \\
&= \frac{\alpha}{d_D}\left(1 - \frac{\int_{q_N^{e,\mathring{q}}}^{\bar{q}} p_D(q)dq}{\lambda_N(\bar{q}-\underline{q})}\right) + \frac{\beta}{\lambda_N(\bar{q}-\underline{q})} \int_{q_N^{e,\mathring{q}}}^{\bar{q}} T_D(q) p_D(q) dq \\
&= \beta T_D(q'_D) + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})} \int_{q_N^{e,\mathring{q}}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_D(q) dq.
\end{aligned} \tag{S-10}$$

Since patients act in their best interest, when $\gamma^e(\mathring{q})=0$, it follows that $q_N^{e,\mathring{q}} > \mathring{q}$, i.e.,

$$\beta T_D(q_N^{e,\mathring{q}}) > \beta T_D(\mathring{q}) \iff \beta T_D(q'_D) + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})} \int_{q_N^{e,\mathring{q}}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_D(q) dq > \beta T_D(\mathring{q}),$$

which implies that

$$\beta T_D(q'_D) + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})} \int_{\mathring{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_D(q) dq > \beta T_D(q'_D) + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})} \int_{q_N^{e,\mathring{q}}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_D(q) dq,$$

and

$$\beta T_D(q'_D) + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})} \int_{\mathring{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_D(q) dq > \beta T_D(\mathring{q}). \tag{S-11}$$

Because

$$\lim_{\gamma\to 0^+} \int_{\mathring{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma}(q) dq = \int_{\mathring{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_D(q) dq,$$

together with (S-11), we have that when $\gamma^e(\mathring{q})=0$,

$$\beta T_D(q'_D) + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})} \lim_{\gamma\to 0^+} \int_{\mathring{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma}(q) dq > \beta T_D(\mathring{q}). \tag{S-12}$$

On the other hand, when (S-12) holds, the expected utility of participation is $\beta T_D(q_P^{e,\mathring{q}})$ satisfying that

$$\beta T_D(q_P^{e,\mathring{q}}) \leq \beta T_D(\mathring{q}) < \beta T_D(q'_D) + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})} \lim_{\gamma\to 0^+} \int_{\mathring{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma}(q) dq,$$

which further implies that

$$\beta T_D(q_P^{e,\mathring{q}}) < \beta T_D(q'_D) + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})} \lim_{\gamma\to 0^+} \int_{\mathring{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma}(q) dq = \beta T_D(q_N^{e,\mathring{q}}). \tag{S-13}$$

The inequality (S-13) implies that not participating in the program leads to a higher expected utility. Thus, no disadvantaged patients would participate in the program in equilibrium, i.e. $\gamma^e(\mathring{q})=0$. $\square$

PROPOSITION 2. *Under a targeted priority mechanism $(\mathring{q})$, all eligible candidates would participate in equilibrium, i.e., the equilibrium participation rate satisfies $\gamma^e(\mathring{q}) = 1$, if and only if*

$$\frac{\alpha}{d_D} + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})}\int_{q_P^{e,\mathring{q}}}^{\mathring{q}}\left(T_D(q) - \frac{\alpha}{d_D}\right)dq > \frac{\alpha}{d_D} + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})}\lim_{\gamma\to 1^-}\frac{\int_{\mathring{q}}^{\bar{q}}\left(T_D(q)-\frac{\alpha}{d_D}\right)p_N^{\gamma}(q)dq}{1-\gamma}. \tag{3}$$

Proof of Proposition 2. The expected utility of an individual type D participating in the program, when all other type D candidates participate in the program, i.e. $\gamma^e(\mathring{q}) \to 1^-$, is given by:

$$\begin{aligned}\beta T_D(q_P^{e,\mathring{q}}) &= \alpha\Delta_P\left(q_P^{e,\mathring{q}}\right) + \beta\pi_{D_P}\left(q_P^{e,\mathring{q}}\right)E\left(T_D(q) \mid \mathring{q} \geqslant q \geqslant q_P^{e,\mathring{q}}\right)\\ &= \beta T_D(q'_D)\left(1 - \frac{\mathring{q} - q_P^{e,\mathring{q}}}{\gamma\lambda_D(\bar{q}-\underline{q})}\right) + \frac{\beta}{\gamma^e(\mathring{q})\lambda_D(\bar{q}-\underline{q})}\int_{q_P^{e,\mathring{q}}}^{\mathring{q}} T_D(q)dq\\ &= \beta T_D(q'_D) + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})}\int_{q_P^{e,\mathring{q}}}^{\mathring{q}}(T_D(q) - T_D(q'_D))dq.\end{aligned} \tag{S-14}$$

On the other hand, the expected utility of not participating in the program is given as follows:

$$\begin{aligned}\beta T_D(q_N^{e,\mathring{q}}) &= \alpha\Delta_N\left(q_N^{e,\mathring{q}}\right) + \beta\pi_{D_N}\left(q_N^{e,\mathring{q}}\right)E\left(T_D(q) \mid q \geqslant q_N^{e,\mathring{q}}\right)\\ &= \beta T_D(q'_D)\left(1 - \frac{\int_{\mathring{q}}^{\bar{q}} p_N^{\gamma^e(\mathring{q})}(q)dq}{(1-\gamma^e(\mathring{q}))\lambda_D(\bar{q}-\underline{q})}\right) + \frac{\beta}{(1-\gamma^e(\mathring{q}))\lambda_D(\bar{q}-\underline{q})}\left(\int_{\mathring{q}}^{\bar{q}} T_D(q)p_N^{\gamma^e(\mathring{q})}(q)dq\right)\\ &= \beta T_D(q'_D) + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})}\lim_{\gamma\to 1^-}\frac{\int_{\mathring{q}}^{\bar{q}}(T_D(q) - T_D(q'_D))p_N^{\gamma}(q)dq}{1-\gamma}.\end{aligned} \tag{S-15}$$

Therefore, if the following inequality holds, then participating in the program is associated with a higher utility than not participating, and all disadvantaged patients will participate in the program in equilibrium.

$$\int_{q_P^{e,\mathring{q}}}^{\mathring{q}}(T_D(q) - T_D(q'_D))dq > \lim_{\gamma\to 1^-}\frac{\int_{\mathring{q}}^{\bar{q}}(T_D(q) - T_D(q'_D))p_N^{\gamma}(q)dq}{1-\gamma}.$$

The converse proof follows similarly, and is omitted here for brevity. □

PROPOSITION 3. *Under a targeted priority mechanism $(\mathring{q})$, a nontrivial fraction of eligible candidates would participate in equilibrium, i.e., there exists a mixed-participation equilibrium $\gamma^e(\mathring{q}) \in (0,1)$, if and only if the following two inequalities hold*

$$\frac{\alpha}{d_D} + \frac{\beta}{\lambda_D(\bar{q}-\underline{q})}\lim_{\gamma\to 0^+}\int_{\mathring{q}}^{\bar{q}}\left(T_D(q) - \frac{\alpha}{d_D}\right)p_N^{\gamma}(q)dq \leq \beta T_D(\mathring{q}), \tag{4a}$$

$$\frac{\alpha}{d_D} + \frac{\beta}{\lambda_D\left(\bar{q}-\underline{q}\right)}\int_{q_P^{e,\mathring{q}}}^{\mathring{q}}\left(T_D(q) - \frac{\alpha}{d_D}\right)dq \leq \frac{\alpha}{d_D} + \frac{\beta}{\lambda_D\left(\bar{q}-\underline{q}\right)}\lim_{\gamma\to 1^-}\frac{\int_{\mathring{q}}^{\bar{q}}\left(T_D(q)-\frac{\alpha}{d_D}\right)p_N^{\gamma}(q)dq}{1-\gamma}. \tag{4b}$$

Proof of Proposition 3. First, observe that (4a) and (4b) are the complements of (2) and (3), respectively. Therefore, if either of (4a) or (4b) does not hold, Propositions 1 and 2 establish that one of the two pure strategy equilibria –no participation or all participation– would be adopted, proving the necessity of the conditions (4a) and (4b). And the sufficiency of these conditions follow by the necessity of the conditions provided in Propositions 1 and 2 along with Theorem 1 and Corollary 1, proving the existence of a mixed-strategy equilibrium in the absence of other types of equilibrium. □

THEOREM 2. *Two thresholds $\mathring{q}_l$ and $\mathring{q}_u$, where $\underline{q} < \mathring{q}_l < \mathring{q}_u < \overline{q}$, determine how the target threshold $\mathring{q}$ regulates the type of the participation equilibrium adopted by eligible patients:*

(a) *Any $\mathring{q} \leq \mathring{q}_l$ is a no-participation target threshold, i.e., $\gamma^e(\mathring{q}) = 0$.*

(b) *Any $\mathring{q} \in (\mathring{q}_l, \mathring{q}_u)$ is a mixed-participation target threshold, i.e., $\gamma^e(\mathring{q}) \in (0, 1)$.*

(c) *Any $\mathring{q} \geq \mathring{q}_u$ is a full-participation target threshold, i.e., $\gamma^e(\mathring{q}) = 1$.*

(d) *The two thresholds $\mathring{q}_l$ and $\mathring{q}_u$ are the respective unique solutions for the inequalities* (2) *and* (3) *to be satisfied as an equality.*

Proof of Theorem 2.

(a) First, observe that the left-hand side of (2) is increasing in $\mathring{q}$ while the right-hand side is decreasing, therefore, there exists a unique solution (see Proposition 1) for the inequality (2) to be satisfied as an equality. Let $\mathring{q}_l$ denote this unique solution. Then the inequality (2) holds if and only if $\mathring{q} \leq \mathring{q}_l$, and thus, the proof follows by Proposition 1.

(c) Similarly, the right-hand side of equation (3) is a constant, while its left-hand side is increasing in $\mathring{q}$, and therefore, there exists a unique solution (see Proposition 2) for the inequality (3) to be satisfied as an equality, which we denote by $\mathring{q}_u$. Therefore, the inequality (3) holds if and only if $\mathring{q} \geq \mathring{q}_u$, and thus, the proof follows by Proposition 2.

(b) To prove (b), it suffices to prove that $\mathring{q}_l < \mathring{q}_u$, which is equivalent to showing that there exists no $\mathring{q}$ such that (1e) and (1f) simultaneously holds (see Theorem 1). We prove the result by contradiction. Assume that there exist $\mathring{q}$ satisfying both (1e) and (1f), i.e.,

$$U_P(q_P^{e,\mathring{q}}, 1) > \lim_{\gamma \to 1^-} U_N(q_N^{e,\mathring{q}}, \gamma), \tag{S-16a}$$

$$U_N(q_N^{e,\mathring{q}}, 0) > \lim_{\gamma \to 0^+} U_P(q_P^{e,\mathring{q}}, \gamma). \tag{S-16b}$$

Furthermore, we have

$$\lim_{\gamma \to 0^+} U_P(q_P^{e,\mathring{q}}, \gamma) \geq U_P(q_P^{e,\mathring{q}}, 1), \tag{S-17}$$

which follows by that under a fixed offer acceptance threshold, the utility of participating type P would decrease by increasing participation. Combining (S-16b) and (S-17), we have

$$U_N(q_N^{e,\mathring{q}}, 0) > \lim_{\gamma \to 1^-} U_N(q_N^{e,\mathring{q}}, \gamma),$$

which gives a contradiction, because $U_N(q_N^{e,\mathring{q}}, 0) \leq \lim_{\gamma \to 1^-} U_N(q_N^{e,\mathring{q}}, \gamma)$ as under a fixed offer acceptance threshold, the utility of non-participating type N would increase by increasing participation. Thus, the proof follows.

(d) The proof follows by the definition of $\mathring{q}_l$ and $\mathring{q}_u$, combined with the proofs of parts (a)–(c).

□

PROPOSITION 4. *For any targeted priority mechanism with $\mathring{q} \in (\mathring{q}_l, \mathring{q}_u)$, the equilibrium participation rate of disadvantaged patients $\gamma^e(\mathring{q})$ is unique and increasing in the target threshold $\mathring{q}$.*

Proof of Proposition 4. We begin the proof by showing the uniqueness of $\gamma^e(\mathring{q})$. Theorem 1 and Proposition 3 establish that the equilibrium offer accepting thresholds, $q_i^{e,\mathring{q}}$, are unique under any mixed-participation equilibrium.

For any targeted priority mechanism with $\overset{*}{q} \in (\overset{*}{q}_l, \overset{*}{q}_u)$, the expected utility of a type D participating in the program is given by

$$\beta T_D(q'_D) + \frac{\beta}{\gamma^e(\overset{*}{q})\lambda_D(\bar{q}-\underline{q})} \int_{q_P^{e,\overset{*}{q}}}^{\overset{*}{q}} (T_D(q) - T_D(q'_D))dq. \tag{S-18}$$

Similarly, the expected utility of a disadvantaged patient not participating in the program is given by

$$\beta T_D(q'_D) + \frac{\beta}{(1-\gamma^e(\overset{*}{q}))\lambda_D(\bar{q}-\underline{q})} \int_{\overset{*}{q}}^{\bar{q}} (T_D(q) - T_D(q'_D))p_N^{\gamma^e(\overset{*}{q})}(q)dq. \tag{S-19}$$

Therefore, given that $\frac{p_N^{\gamma}(q)}{1-\gamma}$ is non-decreasing in γ, the expected utility of a participating type D is decreasing in γ while that of a non-participating one is non-decreasing. Thus, the uniqueness of $\gamma^e(\overset{*}{q})$ follows.

Using Theorems 1 and 2, we conclude that for any targeted priority mechanism with $\overset{*}{q} \in (\overset{*}{q}_l, \overset{*}{q}_u)$ the equilibrium offer acceptance threshold of type D, $q_D^{e,\overset{*}{q}}$, satisfies the following equations:

$$\begin{aligned}
F1 &:= \beta T_D(q_D^{e,\overset{*}{q}}) - \alpha\Delta_N\left(q_D^{e,\overset{*}{q}}\right) - \beta\pi_{D_N}\left(q_D^{e,\overset{*}{q}}\right)E(T_D(q) \mid q \geqslant \overset{*}{q}) = 0,\\
F2 &:= \beta T_D(q_D^{e,\overset{*}{q}}) - \alpha\Delta_P\left(q_D^{e,\overset{*}{q}}\right) - \beta\pi_{D_P}\left(q_D^{e,\overset{*}{q}}\right)E\left(T_D(q) \mid \overset{*}{q} \geqslant q \geqslant q_D^{e,\overset{*}{q}}\right) = 0,
\end{aligned}$$

which can be equivalently written as:

$$\begin{aligned}
F1 &= \beta T_D(q_D^{e,\overset{*}{q}}) - \beta T_D(q'_D) - \frac{\beta}{(1-\gamma^e(\overset{*}{q}))\lambda_D(\bar{q}-\underline{q})} \int_{\overset{*}{q}}^{\bar{q}} (T_D(q) - T_D(q'_D))p_N^{\gamma^e(\overset{*}{q})}(q)dq = 0,\\
F2 &= \beta T_D(q_D^{e,\overset{*}{q}}) - \beta T_D(q'_D) - \frac{\beta}{\gamma^e(\overset{*}{q})\lambda_D(\bar{q}-\underline{q})} \int_{q_D^{e,\overset{*}{q}}}^{\overset{*}{q}} (T_D(q) - T_D(q'_D))dq = 0.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\frac{\partial F1}{\partial q_D^{e,\overset{*}{q}}} &= \beta\frac{\partial T_D(q_D^{e,\overset{*}{q}})}{\partial q_D^{e,\overset{*}{q}}},\\
\frac{\partial F2}{\partial q_D^{e,\overset{*}{q}}} &= \beta\frac{\partial T_D(q_D^{e,\overset{*}{q}})}{\partial q_D^{e,\overset{*}{q}}} + \beta\frac{T_D(q_D^{e,\overset{*}{q}}) - T_D(q'_D)}{\gamma^e(\overset{*}{q})\lambda_D(\bar{q}-\underline{q})},\\
\frac{\partial F1}{\partial \gamma^e(\overset{*}{q})} &= \frac{\beta}{\lambda_D(\bar{q}-\underline{q})(1-\gamma^e(\overset{*}{q}))^2}\left(\frac{\partial\left(\int_{\overset{*}{q}}^{\bar{q}}(T_D(q'_D) - T_D(q))p_N^{\gamma^e(\overset{*}{q})}(q)dq\right)}{\partial\gamma^e(\overset{*}{q})}(1-\gamma^e(\overset{*}{q})) + \int_{\overset{*}{q}}^{\bar{q}}(T_D(q'_D) - T_D(q))p_N^{\gamma^e(\overset{*}{q})}(q)dq\right),\\
\frac{\partial F2}{\partial \gamma^e(\overset{*}{q})} &= \frac{\beta}{\lambda_D\left(\bar{q}-\underline{q}\right)(\gamma^e(\overset{*}{q}))^2}\int_{q_D^{e,\overset{*}{q}}}^{\overset{*}{q}}(T_D(q) - T_D(q'_D))dq,\\
\frac{\partial F1}{\partial \overset{*}{q}} &= \frac{\beta}{\lambda_D\left(\bar{q}-\underline{q}\right)(1-\gamma^e(\overset{*}{q}))}(T_D(\overset{*}{q}) - T_D(q'_D))p_N^{\gamma^e(\overset{*}{q})}(\overset{*}{q}),\\
\frac{\partial F2}{\partial \overset{*}{q}} &= \frac{\beta}{\lambda_D\left(\bar{q}-\underline{q}\right)\gamma^e(\overset{*}{q})}(T_D(q'_D) - T_D(\overset{*}{q})).
\end{aligned}$$

Using the implicit function theorem, $\frac{\partial\gamma^e(\overset{*}{q})}{\partial\overset{*}{q}}$ is given by

$$\frac{\partial\gamma^*}{\partial\overset{*}{q}} = -\frac{\frac{\partial(F_1,F_2)}{\partial(\overset{*}{q},q_D^{e,\overset{*}{q}})}}{\frac{\partial(F_1,F_2)}{\partial(\gamma^e(\overset{*}{q}),q_D^{e,\overset{*}{q}})}} = -\frac{\frac{\partial F_1}{\partial\overset{*}{q}}\frac{\partial F_2}{\partial q_D^{e,\overset{*}{q}}} - \frac{\partial F_1}{\partial q_D^{e,\overset{*}{q}}}\frac{\partial F_2}{\partial\overset{*}{q}}}{\frac{\partial F_1}{\partial\gamma^e(\overset{*}{q})}\frac{\partial F_2}{\partial q_D^{e,\overset{*}{q}}} - \frac{\partial F_1}{\partial q_D^{e,\overset{*}{q}}}\frac{\partial F_2}{\partial\gamma^e(\overset{*}{q})}} =: -\frac{(num)}{(deno)}.$$

First, observe that the denominator is as follows:

$$\begin{aligned}
(deno) =& \frac{\beta^2}{\gamma^e(\overset{*}{q})(1-\gamma^e(\overset{*}{q}))\lambda_D^2(\bar{q}-\underline{q})}\left[\left(\frac{\partial\epsilon_1}{\partial\gamma^e(\overset{*}{q})} + \frac{\epsilon_1}{1-\gamma^e(\overset{*}{q})}\right)\frac{T_D(q_D^{e,\overset{*}{q}}) - T_D(q'_D)}{\bar{q}-\underline{q}}\right]\\
&+ \frac{\beta^2}{\gamma^e(\overset{*}{q})(1-\gamma^e(\overset{*}{q}))\lambda_D(\bar{q}-\underline{q})}\cdot\frac{\partial T_D(q_D^{e,\overset{*}{q}})}{\partial(q_D^{e,\overset{*}{q}})}\left[\left(\frac{\partial\epsilon_1}{\partial\gamma^e(\overset{*}{q})} + \frac{\epsilon_1}{1-\gamma^e(\overset{*}{q})}\right)\gamma^e(\overset{*}{q}) - \frac{1-\gamma^e(\overset{*}{q})}{\gamma^e(\overset{*}{q})}\int_{q_D^{e,\overset{*}{q}}}^{\overset{*}{q}}(T_D(q) - T_D(q'_D))dq\right],
\end{aligned}$$

where

$$\epsilon_1 = \int_{\mathring{q}}^{\bar{q}} (T_D(q'_D) - T_D(q)) p_N^{\gamma^e(\mathring{q})}(q) dq.$$

Therefore, $-(deno)$ is positive given that $\left(\frac{\partial \epsilon_1}{\partial \gamma^e(\mathring{q})} + \frac{\epsilon_1}{1-\gamma^e(\mathring{q})}\right)$ is negative, which holds under Assumption 6. Therefore the sign of $\frac{\partial \gamma^e(\mathring{q})}{\partial \mathring{q}}$ is the same as that of the numerator, which satisfies the following

$$\begin{aligned}(num) &= \frac{\partial F_1}{\partial \mathring{q}} \frac{\partial F_2}{\partial q_D^{e,\mathring{q}}} - \frac{\partial F_1}{\partial q_D^{e,\mathring{q}}} \frac{\partial F_2}{\partial \mathring{q}} \\ &= \frac{\beta^2 (T_D(\mathring{q}) - T_D(q'_D))}{\lambda_D(\bar{q} - \underline{q})} \cdot \left[\frac{\partial T_D(q_D^{e,\mathring{q}})}{\partial (q_D^{e,\mathring{q}})} \left(\frac{1}{\gamma^e(\mathring{q})} + \frac{p_N^{\gamma^e(\mathring{q})}(\mathring{q})}{1-\gamma^e(\mathring{q})} \right) + \frac{T_D(q_D^{e,\mathring{q}}) - T_D(q'_D)}{\gamma^e(\mathring{q})(1-\gamma^e(\mathring{q}))\lambda_D(\bar{q} - \underline{q})} \right] \\ &> 0,\end{aligned}$$

and thus, we conclude the proof that $\frac{\partial \gamma^e(\mathring{q})}{\partial \mathring{q}} > 0$. □

PROPOSITION 5. *For any targeted priority mechanism, the impact of the target threshold $(\mathring{q})$ on the equilibrium acceptance behavior of disadvantaged patients $(q_D^{e,\mathring{q}})$ is characterized as follows:*

(a) *If $\mathring{q} \in [\underline{q}, \mathring{q}_l]$, then $q_D^{e,\mathring{q}}$ does not change by $\mathring{q}$.*

(b) *If $\mathring{q} \in (\mathring{q}_l, \mathring{q}_u)$, then $q_D^{e,\mathring{q}} := q_P^{e,\mathring{q}} = q_N^{e,\mathring{q}}$ decreases (increases) in $\mathring{q}$ when the expression given in* (5) *is negative (positive):*

$$(1-\gamma^e(\mathring{q})) \int_{\mathring{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) \cdot \frac{\partial}{\partial \gamma} \left(\frac{p_N^{\gamma}(q)}{1-\gamma} \right)\Big|_{\gamma=\gamma^e(\mathring{q})} dq - \frac{p_N^{\gamma^e(\mathring{q})}(\mathring{q}) \int_{\mathring{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma^e(\mathring{q})}(q) dq}{1-\gamma^e(\mathring{q})}. \quad (5)$$

(c) *If $\mathring{q} \in [\mathring{q}_u, \overline{q}]$, then $q_D^{e,\mathring{q}} := q_P^{e,\mathring{q}}$ increases in $\mathring{q}$.*

Proof of Proposition 5.

(a)&(c) We first prove parts (a) and (c). Theorem 2 establishes that when $\mathring{q} \le \mathring{q}_l$, no type D will participate in the program, and thus the selection of $\mathring{q}$ does not affect the equilibrium acceptance threshold. When, on the other hand, $\mathring{q} \ge \mathring{q}_u$, all type D participates in the program, and any increment of the target threshold of the program $\mathring{q}$ expands the pool of organs that the disadvantaged patients have priority over while keeping their participation rate the same. Therefore, for any given acceptance threshold, the right-hand side of (1f) would increase with increasing $\mathring{q}$, and thus, as per Theorem 1(c) the equilibrium acceptance threshold $q_D^{e,\mathring{q}}$ would also increase in $\mathring{q}$.

(b) To prove part (b), first, recall Theorem 2, which establishes that for any $\mathring{q} \in (\mathring{q}_l, \mathring{q}_u)$, only a nontrivial fraction of type D participates in the program. Then, following similar steps to that of in the proof of Proposition 4, $\frac{\partial q_D^{e,\mathring{q}}}{\partial \mathring{q}}$ can be expressed as:

$$\frac{\partial q_D^{e,\mathring{q}}}{\partial \mathring{q}} = -\frac{\frac{\partial (F_1, F_2)}{\partial (\gamma^e(\mathring{q}), \mathring{q})}}{\frac{\partial (F_1, F_2)}{\partial (\gamma^e(\mathring{q}), q_D^{e,\mathring{q}})}} = -\frac{\frac{\partial F_1}{\partial \gamma^e(\mathring{q})} \frac{\partial F_2}{\partial \mathring{q}} - \frac{\partial F_1}{\partial \mathring{q}} \frac{\partial F_2}{\partial \gamma^e(\mathring{q})}}{\frac{\partial F_1}{\partial \gamma^e(\mathring{q})} \frac{\partial F_2}{\partial q_D^{e,\mathring{q}}} - \frac{\partial F_1}{\partial q_D^{e,\mathring{q}}} \frac{\partial F_2}{\partial \gamma^e(\mathring{q})}} =: -\frac{(num)}{(deno)} \quad (S\text{-}20)$$

And we have shown in the proof of Proposition 4 that $-(deno)$ is positive under Assumption 6. Hence the sign of $\frac{\partial q_D^{e,\mathring{q}}}{\partial \mathring{q}}$ is the same as that of the numerator, which is given by

$$\begin{aligned}(num) &= \frac{\partial F_1}{\partial \gamma^e(\mathring{q})} \frac{\partial F_2}{\partial \mathring{q}} - \frac{\partial F_1}{\partial \mathring{q}} \frac{\partial F_2}{\partial \gamma^e(\mathring{q})}, \\ &= \frac{\beta^2 (T_D(\mathring{q}) - T_D(q'_D))}{\gamma^e(\mathring{q})(1-\gamma^e(\mathring{q}))\lambda_D^2(\bar{q} - \underline{q})^2} \left(-\frac{p_N^{\gamma^e(\mathring{q})}(\mathring{q}) \int_{q_D^{e,\mathring{q}}}^{\mathring{q}} (T_D(q) - T_D(q'_D)) dq}{\gamma^e(\mathring{q})} - \left(\frac{\partial \epsilon_1}{\partial \gamma^e(\mathring{q})} + \frac{\epsilon_1}{1-\gamma^e(\mathring{q})} \right) \right),\end{aligned}$$

where

$$\epsilon_1 := \int_{\overset{*}{q}}^{\bar{q}} (T_D(q'_D) - T_D(q)) p_N^{\gamma^e(\overset{*}{q})}(q) dq.$$

Therefore, the expression controlling the sign of $\frac{\partial q_D^{e,\overset{*}{q}}}{\partial \overset{*}{q}}$ is given by

$$\left(-\frac{p_N^{\gamma^e(\overset{*}{q})}(\overset{*}{q}) \int_{q_D^{e,\overset{*}{q}}}^{\overset{*}{q}} (T_D(q) - T_D(q'_D)) dq}{\gamma^e(\overset{*}{q})} + \left(\frac{\partial \int_{\overset{*}{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma^e(\overset{*}{q})}(q) dq}{\partial \gamma^e(\overset{*}{q})} + \frac{\int_{\overset{*}{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma^e(\overset{*}{q})}(q) dq}{1 - \gamma^e(\overset{*}{q})} \right) \right). \quad \text{(S-21)}$$

Under a mixed equilibrium, the following holds using Theorem 1:

$$\frac{\int_{\overset{*}{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma^e(\overset{*}{q})}(q) dq}{1 - \gamma^e(\overset{*}{q})} = \frac{\int_{q_D^{e,\overset{*}{q}}}^{\overset{*}{q}} (T_D(q) - T_D(q'_D)) dq}{\gamma^e(\overset{*}{q})}.$$

Therefore, the expression in (S-21) can be equivalently written as:

$$\left(-\frac{p_N^{\gamma^e(\overset{*}{q})}(\overset{*}{q}) \int_{\overset{*}{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma^e(\overset{*}{q})}(q) dq}{1 - \gamma^e(\overset{*}{q})} + \left(\frac{\partial \int_{\overset{*}{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma^e(\overset{*}{q})}(q) dq}{\partial \gamma^e(\overset{*}{q})} + \frac{\int_{\overset{*}{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma^e(\overset{*}{q})}(q) dq}{1 - \gamma^e(\overset{*}{q})} \right) \right). \quad \text{(S-22)}$$

Further simplifying the expression in (S-22), we conclude that the impact of the target threshold $\overset{*}{q}$ on the equilibrium acceptance thresholds of disadvantaged patients $q_D^{e,\overset{*}{q}} := q_P^{e,\overset{*}{q}} = q_N^{e,\overset{*}{q}}$ is determined by the following expression

$$(1 - \gamma^e(\overset{*}{q})) \int_{\overset{*}{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) \cdot \frac{\partial}{\partial \gamma} \left(\frac{p_N^{\gamma}(q)}{1 - \gamma} \right) \Big|_{\gamma = \gamma^e(\overset{*}{q})} dq - \frac{p_N^{\gamma^e(\overset{*}{q})}(\overset{*}{q}) \int_{\overset{*}{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma^e(\overset{*}{q})}(q) dq}{1 - \gamma^e(\overset{*}{q})}.$$

As a result, from (S-20), we conclude that the equilibrium acceptance threshold $q_D^{e,\overset{*}{q}}$ decreases (increases) in the target threshold $\overset{*}{q}$ when the expression given in (5) is negative (positive), hence the proof follows.

□

PROPOSITION 6. *For any targeted priority mechanism with target threshold $(\overset{*}{q})$:*

(a) *When $\overset{*}{q} \in [\underline{q}, \overset{*}{q}_l]$, the selection of $\overset{*}{q}$ has no impact on $q_O^{e,\overset{*}{q}}$.*

(b) *When $\overset{*}{q} \in (\overset{*}{q}_l, \overset{*}{q}_u)$,*

(b-i) *If $\overset{*}{q}_l < \overset{*}{q} < q_O^{e,\overset{*}{q}}$, then $q_O^{e,\overset{*}{q}}$ is increasing in $\overset{*}{q}$.*

(b-ii) *If $q_O^{e,\overset{*}{q}} \leq \overset{*}{q} < \overset{*}{q}_u$, and $p_N^{\gamma}(q)$ satisfies $\frac{\partial^2 p_N^{\gamma}(q)}{\partial \gamma^2} \geq 0$ for any q, then $q_O^{e,\overset{*}{q}}$ is unimodal in $\overset{*}{q}$, attaining its mode at $q_O^{\max}$ which is the unique solution of $\frac{d\, q_O^{e,\overset{*}{q}}}{d\, \gamma} = 0$.*

(c) *When $\overset{*}{q} \in [\overset{*}{q}_u, \bar{q}]$, $q_O^{e,\overset{*}{q}}$ is non-increasing in $\overset{*}{q}$.*

Proof of Proposition 6. From Theorem 1, we have that the equilibrium offer acceptance threshold of type O, $q_O^{e,\overset{*}{q}}$, solves equation (1a), given by

$$\beta T_O(q_O^{e,\overset{*}{q}}) = U_O(q_O^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q})) := \alpha \Delta_O \left(q_O^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q}) \right) + \beta \pi_O \left(q_O^{e,\overset{*}{q}}, \gamma^e(\overset{*}{q}) \right) E\left[T_O(q) \mid q \geqslant q_O^{e,\overset{*}{q}} \right]. \quad \text{(S-23)}$$

(a) Theorem 2 establishes that any $\overset{*}{q} \leq \overset{*}{q}_l$ is a no-participation target threshold, and thus changing $\overset{*}{q}$ in this interval has no impact on the equilibrium offer acceptance thresholds of type D and O. Hence the proof follows.

(b) Theorem 2 establishes that any $\mathring{q}$ such that $\mathring{q}_l < \mathring{q} < \mathring{q}_u$ is a mixed participation threshold. If, in this case, $\mathring{q} \leq q_O^{e,\mathring{q}}$, then we can write equation (S-23) as follows:

$$F3 := \beta T_O(q_O^{e,\mathring{q}}) - \beta T_O(q'_O) - \frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \int_{q_O^{e,\mathring{q}}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq = 0.$$

We can write its partial derivative in $\mathring{q}$ as follows

$$\begin{aligned}
\frac{\partial F3}{\partial \mathring{q}} &= -\frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \frac{d \int_{q_O^{e,\mathring{q}}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq}{d\,\mathring{q}} \\
&= \frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \int_{q_O^{e,\mathring{q}}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \left(\frac{d\, p_N^{\gamma^e(\mathring{q})}(q)}{d\,\mathring{q}} \right) dq \\
&= \frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \int_{q_O^{e,\mathring{q}}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \left(\frac{d\, p_N^{\gamma^e(\mathring{q})}(q)}{d\,\gamma^e(\mathring{q})} \frac{d\,\gamma^e(\mathring{q})}{d\,\mathring{q}} \right) dq.
\end{aligned}$$

We further have

$$\begin{aligned}
\frac{\partial F3}{\partial q_O^{e,\mathring{q}}} =& \beta \frac{d\, T_O(q_O^{e,\mathring{q}})}{d\, q_O^{e,\mathring{q}}} + \frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \left(\Big(T_O(q_O^{e,\mathring{q}}) - T_O(q'_O)\Big) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q_O^{e,\mathring{q}})\Big) - \int_{q_O^{e,\mathring{q}}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \left(\frac{d\, p_N^{\gamma^e(\mathring{q})}(q)}{d\, q_O^{e,\mathring{q}}} \right) dq \right) \\
=& \beta \frac{d\, T_O(q_O^{e,\mathring{q}})}{d\, q_O^{e,\mathring{q}}} + \frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \left(\Big(T_O(q_O^{e,\mathring{q}}) - T_O(q'_O)\Big) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q_O^{e,\mathring{q}})\Big) - \int_{q_O^{e,\mathring{q}}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \left(\frac{d\, p_N^{\gamma^e(\mathring{q})}(q)}{d\,\gamma^e(\mathring{q})} \frac{d\,\gamma^e(\mathring{q})}{d\, q_O^{e,\mathring{q}}} \right) dq \right).
\end{aligned}$$

Proposition 4 establishes that under Assumption 6, we have $\frac{d\,\gamma^e(\mathring{q})}{d\,\mathring{q}} > 0$, and thus, $\frac{d\,\gamma^e(\mathring{q})}{d\, q_O^{e,\mathring{q}}} = 0$ and $\frac{d\, p_N^{\gamma^e(\mathring{q})}(q)}{d\,\gamma^e(\mathring{q})} < 0$ for any $\mathring{q} \in (\mathring{q}_l, \mathring{q}_u)$. Therefore, we have

$$\frac{\partial F3}{\partial \mathring{q}} < 0, \text{ and, } \frac{\partial F3}{\partial q_O^{e,\mathring{q}}} > 0,$$

which implies that

$$\frac{d\, q_O^{e,\mathring{q}}}{d\,\mathring{q}} = -\frac{\frac{\partial F3}{\partial \mathring{q}}}{\frac{\partial F3}{\partial q_O^{e,\mathring{q}}}} > 0.$$

If, on the other hand, we have $q_O^{e,\mathring{q}} \leq \mathring{q} < \mathring{q}_u$, then we can write equation (1a) as follows:

$$F4 := \beta T_O(q_O^{e,\mathring{q}}) - \beta T_O(q'_O) - \frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \int_{\mathring{q}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq = 0.$$

We can write its partial derivative in γ as follows

$$\begin{aligned}
\frac{\partial F4}{\partial \gamma} &= -\frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \frac{d \int_{\mathring{q}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq}{d\,\gamma} \\
&= \frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \left(\big(T_O(\mathring{q}(\gamma)) - T_O(q'_O)\big) \Big(1 - p_N^{\gamma^e(\mathring{q})}(\mathring{q}(\gamma))\Big) + \int_{\mathring{q}}^{\bar{q}} \big(T_O(q) - T_O(q'_O)\big) \left(\frac{\partial\, p_N^{\gamma^e(\mathring{q})}(q)}{\partial\,\gamma} \right) dq \right).
\end{aligned} \tag{S-24}$$

Let $F4' := (T_O(\mathring{q}(\gamma)) - T_O(q'_O)) \Big(1 - p_N^{\gamma^e(\mathring{q})}(\mathring{q}(\gamma))\Big)$ and $F4'' := \int_{\mathring{q}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \left(\frac{\partial\, p_N^{\gamma^e(\mathring{q})}(q)}{\partial\,\gamma} \right) dq$ respectively denote the two terms in the large parentheses in (S-24). We already showed in the previous parts of the proof that $\frac{\partial F4}{\partial \gamma} < 0$ for $\mathring{q} < q_O^{e,\mathring{q}}$, and $\frac{\partial F4}{\partial \gamma} > 0$ for $\mathring{q} > \mathring{q}_u$. Observe that $F4' > 0$ and is increasing in γ, because $T_O(\mathring{q}(\gamma))$

is increasing in γ and $p_N^{\gamma^e(\mathring{q})}(\mathring{q}(\gamma))$ is decreasing in γ. Moreover, $F4'' \leq 0$ and is increasing in γ when $\frac{\partial^2 p_N^{\gamma}(q)}{\partial \gamma^2} \geq 0$, because $\frac{\partial p_N^{\gamma^e(\mathring{q})}(q)}{\partial \gamma} < 0$ and $\frac{\partial \gamma}{\partial \mathring{q}} > 0$. Therefore, $\frac{\partial F4}{\partial \gamma}$ is increasing in $\mathring{q}$ for any $\mathring{q} \in [\mathring{q}_l, \mathring{q}_u]$. Thus, there exists a unique point $q_O^{max} \in [\mathring{q}_l, \mathring{q}_u]$ such that $\frac{\partial F4}{\partial \gamma} = 0$. Furthermore, we have

$$\frac{\partial F4}{\partial q_O^{e,\mathring{q}}} = \beta \frac{dT_O(q_O^{e,\mathring{q}})}{dq_O^{e,\mathring{q}}} > 0,$$

and

$$\frac{d\, q_O^{e,\mathring{q}}}{d\, \gamma} = -\frac{\frac{\partial F4}{\partial \gamma}}{\frac{\partial F4}{\partial q_O^{e,\mathring{q}}}}.$$

Therefore, we conclude that $q_O^{e,\mathring{q}}$ is increasing in $[\mathring{q}_l, q_O^{max}]$ and decreasing in $[q_O^{max}, \mathring{q}_u]$.

(c) First, recall from Theorem 2 that any $\mathring{q} \geq \mathring{q}_u$ is a full-participation target threshold. There are two possible cases when the system is under full participation equilibrium; namely, $\mathring{q} < q_O^{e,\mathring{q}}$ and $\mathring{q} \geq q_O^{e,\mathring{q}}$. When $\mathring{q} < q_O^{e,\mathring{q}}$, equation (1a) can be further written as

$$\beta T_O(q_O^{e,\mathring{q}}) - \beta T_O(q'_O) - \frac{\beta}{\lambda_O(\bar{q} - \underline{q})} \int_{q_O^{e,\mathring{q}}}^{\bar{q}} (T_O(q) - T_O(q'_O))\, dq = 0, \tag{S-25}$$

and therefore, changing $\mathring{q}$ does not affect $q_O^{e,\mathring{q}}$ when $q_O^{e,\mathring{q}} > \mathring{q} \geq \mathring{q}_u$. When $\mathring{q} \geq q_O^{e,\mathring{q}}$, on the other hand, equation (1a) can be written as

$$\beta T_O(q_O^{e,\mathring{q}}) - \beta T_O(q'_O) - \frac{\beta}{\lambda_O(\bar{q} - \underline{q})} \int_{\mathring{q}}^{\bar{q}} (T_O(q) - T_O(q'_O))\, dq = 0. \tag{S-26}$$

(S-26) implies that $T_O(q_O^{e,\mathring{q}}) = T_O(q'_O) + \frac{1}{\lambda_O(\bar{q} - \underline{q})} \int_{\mathring{q}}^{\bar{q}} (T_O(q) - T_O(q'_O))\, dq$, the right-hand side of which is decreasing in $\mathring{q}$ because $T_O(q)$ monotonically increases in q. Therefore, $q_O^{e,\mathring{q}}$ decreases in $\mathring{q}$ when $\mathring{q} \geq q_O^{e,\mathring{q}}$ and $\mathring{q} \geq \mathring{q}_u$. Combining these two cases, we conclude that $q_O^{e,\mathring{q}}$ is non-increasing in $\mathring{q}$ for $\mathring{q} \geq \mathring{q}_u$.

□

LEMMA 1. *The class-separating allocations satisfy the following:*

(a) *There exists a unique threshold q_O^T such that the utility of class O remains unchanged, gets better or gets worse post the introduction of a class-separating allocation with $q^T = q_O^T$, $q^T < q_O^T$, or $q^T > q_O^T$, respectively.*

(b) *There exists a unique threshold q_D^T such that the utility of class D remains unchanged, gets better or gets worse post the introduction of a class-separating allocation with $q^T = q_D^T$, $q^T > q_D^T$, or $q^T < q_D^T$, respectively.*

Proof of Lemma 1.

(a) Let q_O^T denote the solution q^T to the following equation:

$$\int_{\underline{q}}^{\bar{q}} p_O(q) \overline{T}_O(q) dq = \int_{q^T}^{\bar{q}} \overline{T}_O(q) dq. \tag{S-27}$$

Observe that any solution to (S-27) provides a threshold value that maintains the utility of the class O unchanged post the introduction of class-separating allocation with q^T. The existence and uniqueness of q_O^T simply follow by the monotonicity of $\overline{T}_O(q)$ in q.

For any $q^T < q_O^T$, we have

$$\int_{\underline{q}}^{\bar{q}} p_O(q) \overline{T}_O(q) dq = \int_{q_O^T}^{\bar{q}} \overline{T}_O(q) dq < \int_{q^T}^{\bar{q}} \overline{T}_O(q) dq,$$

which implies that class O is better off under any class-separating allocation with $q^T < q_O^T$. The remainder of the proof follows by symmetry and is omitted here for brevity.

(b) Similar to the previous part, let q_D^T denote the solution q^T to the following equation:

$$\int_{\underline{q}}^{\bar{q}} p_D(q)\overline{T}_D(q)dq = \int_{\underline{q}}^{q^T} \overline{T}_D(q)dq. \tag{S-28}$$

Observe that any solution to (S-28) provides a threshold value that maintains the utility of the class D unchanged post the introduction of class-separating allocation with q^T. The existence and uniqueness of q_D^T follow by the monotonicity of $\overline{T}_D(q)$ in q.

Furthermore, for any $q^T > q_D^T$, we have

$$\int_{\underline{q}}^{\bar{q}} p_D(q)\overline{T}_D(q)dq = \int_{\underline{q}}^{q_D^T} \overline{T}_D(q)dq < \int_{\underline{q}}^{q^T} \overline{T}_D(q)dq,$$

which implies that class D is better off under any class-separating allocation with $q^T > q_D^T$. The remainder of the proof follows by symmetry and is omitted here for brevity.

□

THEOREM 3. *If $\frac{\overline{T}_O(q)}{\overline{T}_D(q)}$ is non-constant non-decreasing in $q \in [\underline{q}, \bar{q}]$, then we have $q_O^T > q_D^T$ and consequently any class-separating allocation with $q^T \in (q_D^T, q_O^T)$ increases the overall utility by improving the utility of both classes.*

Proof of Theorem 3. First, observe that Lemma 1 establishes that when $q_O^T \leq q_D^T$, the overall social welfare can not be improved without hurting the utility of either class. However, when $q_O^T > q_D^T$, any q^T such that $q_O^T > q^T > q_D^T$ improves the overall social welfare by increasing the utilities of both classes because the following holds;

$$\int_{\underline{q}}^{\bar{q}} p_O(q)\overline{T}_O(q)dq = \int_{q_O^T}^{\bar{q}} \overline{T}_O(q)dq < \int_{q^T}^{\bar{q}} \overline{T}_O(q)dq,$$
$$\int_{\underline{q}}^{\bar{q}} p_D(q)\overline{T}_D(q)dq = \int_{\underline{q}}^{q_D^T} \overline{T}_D(q)dq < \int_{\underline{q}}^{q^T} \overline{T}_D(q)dq.$$

Next, we prove that if $\frac{\overline{T}_O(q)}{\overline{T}_D(q)}$ is non-constant non-decreasing in $q \in [\underline{q}, \bar{q}]$, then we have $q_O^T > q_D^T$. To show the desired result, we prove that under a class-separating allocation with $q^T = q_D^T$, the utility of type O increases, which directly implies that $q^T = q_D^T < q_O^T$ using Lemma 1. The following holds by the definition of q_D^T:

$$\int_{\underline{q}}^{\bar{q}} p_D(q)\overline{T}_D(q)dq = \int_{\underline{q}}^{q_D^T} \overline{T}_D(q)dq, \tag{S-29}$$

which, recalling that $p_O(q) + p_D(q) = 1$, can be equivalently written as

$$\int_{q^T}^{\bar{q}} (1 - p_O(q))\overline{T}_D(q)dq - \int_{\underline{q}}^{q_D^T} p_O(q)\overline{T}_D(q)dq = 0. \tag{S-30}$$

Defining $\alpha(q) := \frac{\overline{T}_O(q)}{\overline{T}_D(q)}$, we have

$$\int_{q_D^T}^{\bar{q}} \overline{T}_O(q)dq - \int_{\underline{q}}^{\bar{q}} p_O(q)\overline{T}_O(q)dq = \int_{q_D^T}^{\bar{q}} (1 - p_O(q))\overline{T}_O(q)dq - \int_{\underline{q}}^{q_D^T} p_O(q)\overline{T}_O(q)dq \tag{S-31a}$$
$$= \int_{q_D^T}^{\bar{q}} \alpha(q)(1 - p_O(q))\overline{T}_D(q)dq - \int_{\underline{q}}^{q_D^T} \alpha(q)p_O(q)\overline{T}_D(q)dq \tag{S-31b}$$
$$> \alpha\big(q_D^T\big) \int_{q^T}^{\bar{q}} (1 - p_O(q))\overline{T}_D(q)dq - \alpha\big(q_D^T\big) \int_{\underline{q}}^{q^T} p_O(q)\overline{T}_D(q)dq \tag{S-31c}$$
$$= 0, \tag{S-31d}$$

where (S-31b) follows by the definition of $\alpha(q)$, (S-31c) follows because $\alpha(q)$ is positive, non-constant and non-decreasing, and finally, (S-31d) follows by (S-30). From (S-31d), we have $\int_{q_D^T}^{\bar{q}} \overline{T}_O(q)dq > \int_{\underline{q}}^{\bar{q}} p_O(q)\overline{T}_O(q)dq$, which implies that the utility of type O increases under the class-separating allocation with $q^T = q_D^T$. Therefore, using Lemma 1 we conclude that $q^T = q_D^T < q_O^T$, hence the proof follows.

□

THEOREM 4. *The impact of the targeted priority mechanisms on social welfare is characterized as follows:*

(a) *For any system such that $\mathring{q}_u \le q_O^{\overline{e}}$, the targeted priority mechanism with $\mathring{q} = q_O^{\overline{e}}$ leads to a class-separating equilibrium. Furthermore, this mechanism maximizes social welfare if $\frac{\overline{T}_O(q)}{\overline{T}_D(q)} \ge 1$ is non-decreasing in $q \in [q_O^{e_0}, \overline{q}]$, and the following inequality holds for all $\mathring{q}$:*

$$\overline{q} - q_O^{\overline{e}} \ge \int_{q_O^{e,\mathring{q}}}^{\overline{q}} \left(1 - p_N^{\gamma^e(\mathring{q})}(q)\right) dq. \tag{9}$$

(b) *For any system such that $\mathring{q}_u > q_O^{\overline{e}}$, no targeted priority mechanism can guarantee a class-separating equilibrium. However, if $\frac{\overline{T}_O(q)}{\overline{T}_D(q)} \ge 1$ for $q \in [q_O^{e_0}, \overline{q}]$, then there exists a targeted priority mechanism that improves social welfare compared to the status quo.*

Proof of Theorem 4. Define q'_D and q'_O as the unique solutions to $\beta T_D(q'_D) = \frac{\alpha}{d_D}$ and $\beta T_O(q'_O) = \frac{\alpha}{d_O}$, respectively. Define $q_D^{e,q_O^{\overline{e}}}$ to be the acceptance threshold of disadvantaged patients when $\mathring{q}$ is set to be $q_O^{\overline{e}}$.

(a) We analyze the systems having $\mathring{q}_u \le q_O^{\overline{e}}$ under two scenarios; $\mathring{q}_u = q_O^{\overline{e}}$ and $\mathring{q}_u < q_O^{\overline{e}}$.

- Under the first scenario, i.e. $\mathring{q}_u = q_O^{\overline{e}}$, setting $\mathring{q} = q_O^{\overline{e}}$ would result in all type D participating in the program ($\gamma^e(\mathring{q}) = 1$, see Theorem 2), while type O getting fully prioritized access to all organs that they are interested in, and therefore it would lead to a class-separating equilibrium.

  Any target threshold $\mathring{q} > q_O^{\overline{e}} = \mathring{q}_u$ would result in all type D participating in the program, i.e., $\gamma^e(\mathring{q}) = 1$ (see Theorem 2). Further, we have $q_O^{\overline{e}} > q_O^{e,\mathring{q}}$ for any $\mathring{q} > q_O^{\overline{e}}$ by Theorem 1 because

$$\beta T_O(q_O^{\overline{e}}) = \beta T_O(q'_O) + \frac{\beta}{\lambda_O(\overline{q} - \underline{q})} \int_{q_O^{\overline{e}}}^{\overline{q}} (T_O(q) - T_O(q'_O))dq,$$
$$\beta T_O(q_O^{e,\mathring{q}}) = \beta T_O(q'_O) + \frac{\beta}{\lambda_O(\overline{q} - \underline{q})} \int_{q_O^{e,\mathring{q}}}^{\overline{q}} (T_O(q) - T_O(q'_O))dq.$$

  If for all $\mathring{q} > q_O^{\overline{e}}$, we have $q_D^{e,q_O^{\overline{e}}} > q_D^{e,\mathring{q}}$, then both patient groups are better off under the mechanism with $\mathring{q} = q_O^{\overline{e}}$, and therefore, $s\left(q_D^{e,\mathring{q}}, q_O^{e,\mathring{q}}\right) < s\left(q_D^{e,q_O^{\overline{e}}}, q_O^{\overline{e}}\right)$ by the definition of social welfare (see equation (6)). If, on the other hand, there exists $\mathring{q} > q_O^{\overline{e}}$ such that $q_D^{e,q_O^{\overline{e}}} \le q_D^{e,\mathring{q}}$, then using Theorem 1, we have $q_O^{e,\mathring{q}}$ and $q_D^{e,\mathring{q}}$ satisfy:

$$\beta T_O(q_O^{e,\mathring{q}}) = \beta T_O(q'_O) + \frac{\beta}{\lambda_O(\overline{q} - \underline{q})} \int_{\mathring{q}}^{\overline{q}} (T_O(q) - T_O(q'_O))dq, \tag{S-32a}$$
$$\beta T_D(q_D^{e,\mathring{q}}) = \beta T_D(q'_D) + \frac{\beta}{\lambda_D(\overline{q} - \underline{q})} \int_{q_D^{e,\mathring{q}}}^{\mathring{q}} (T_D(q) - T_D(q'_D))dq, \tag{S-32b}$$

  and $q_O^{\overline{e}}$ and $q_D^{e,q_O^{\overline{e}}}$ satisfy:

$$\beta T_O(q_O^{\overline{e}}) = \beta T_O(q'_O) + \frac{\beta}{\lambda_O(\overline{q} - \underline{q})} \int_{q_O^{\overline{e}}}^{\overline{q}} (T_O(q) - T_O(q'_O))dq, \tag{S-33a}$$
$$\beta T_D(q_D^{e,q_O^{\overline{e}}}) = \beta T_D(q'_D) + \frac{\beta}{\lambda_D(\overline{q} - \underline{q})} \int_{q_D^{e,q_O^{\overline{e}}}}^{q_O^{\overline{e}}} (T_D(q) - T_D(q'_D))dq. \tag{S-33b}$$

Combining (S-32b) and (S-33b) with Theorem 1, we obtain

$$\begin{aligned}
&s\Big(q_D^{e,q_O^{\overline{e}}}, q_O^{\overline{e}}\Big) - s\Big(q_D^{e,\mathring{q}}, q_O^{e,\mathring{q}}\Big)\\
=&\Big(\lambda_O\beta T_O(q_O^{e,\mathring{q}}) + \lambda_D\beta T_D(q_D^{e,\mathring{q}})\Big) - \Big(\lambda_O\beta T_O(q_O^{e,\mathring{q}}) + \lambda_D\beta T_D(q_D^{e,\mathring{q}})\Big)\\
=&\frac{\beta}{\overline{q}-\underline{q}}\left(\int_{q_D^{e,q_O^{\overline{e}}}}^{q_D^{e,\mathring{q}}}(T_D(q)-T_D(q'_D))dq + \int_{q_O^{\overline{e}}}^{\mathring{q}}((T_O(q)-T_O(q'_O))-(T_D(q)-T_D(q'_D)))dq\right)\\
=&\frac{\beta}{\overline{q}-\underline{q}}\left(\int_{q_D^{e,q_O^{\overline{e}}}}^{q_D^{e,\mathring{q}}}\overline{T}_D(q)dq + \int_{q_O^{\overline{e}}}^{\mathring{q}}(\overline{T}_O(q)-\overline{T}_D(q))dq\right)\\
>&0,
\end{aligned} \tag{S-34a}$$

where (S-34a) follows from that $q_D^{e,\mathring{q}} > q_D^{e,q_O^{\overline{e}}}$, $\overline{T}_D(q) > 0$ for $q \in [\underline{q}, \overline{q}]$, and $\frac{\overline{T}_O(q)}{\overline{T}_D(q)} \geq 1$ for $q \in [q_O^{e_0}, \overline{q}]$. Therefore, for any targeted priority mechanism with $\mathring{q} > q_O^{\overline{e}}$, the overall social welfare is less than that of $\mathring{q} = q_O^{\overline{e}}$.

On the other hand, any target threshold $\mathring{q} < q_O^{\overline{e}} = \mathring{q}_u$ would result in $q_O^{e,\mathring{q}} < q_O^{\overline{e}}$, because under this scenario not all type D participate in the program (i.e., $\gamma^e(\mathring{q}) < 1$), and therefore, type O has no longer full priority over organs $[q_O^{\overline{e}}, \overline{q}]$. If for this scenario, we have $q_D^{e,\mathring{q}} < q_D^{e,q_O^{\overline{e}}}$, then the mechanism with $\mathring{q} = q_O^{\overline{e}}$ increases the utility of both type D (given by $q_D^{e,\mathring{q}} < q_D^{e,q_O^{\overline{e}}}$ and Theorem 1) and type O (given by $q_O^{e,\mathring{q}} < q_O^{\overline{e}}$ and Theorem 1), it maximizes the social welfare. We next analyze the case when there exists a target threshold $\mathring{q} < q_O^{\overline{e}}$ such that $q_D^{e,\mathring{q}} > q_D^{e,q_O^{\overline{e}}}$. Under this case ($q_D^{e,\mathring{q}} > q_D^{e,q_O^{\overline{e}}}$), we have $\mathring{q} \leq q_O^{e,\mathring{q}} < q_O^{\overline{e}}$ or $q_O^{e,\mathring{q}} < \mathring{q} < q_O^{\overline{e}}$. First, if $\mathring{q} \leq q_O^{e,\mathring{q}} < q_O^{\overline{e}}$, then, from Theorem 1, we have $q_O^{e,\mathring{q}}$, $\mathring{q}$, and $q_D^{e,\mathring{q}}$ satisfy the following equations:

$$\begin{aligned}
\beta T_O(q_O^{e,\mathring{q}}) &= \beta T_O(q'_O) + \frac{\beta}{\lambda_O(\overline{q}-\underline{q})}\int_{q_O^{e,\mathring{q}}}^{\overline{q}}(T_O(q)-T_O(q'_O))\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq,\\
\beta T_D(q_D^{e,\mathring{q}}) &= \beta T_D(q'_D) + \frac{\beta}{\gamma^e(\mathring{q})\lambda_D(\overline{q}-\underline{q})}\int_{q_D^{e,\mathring{q}}}^{\mathring{q}}(T_D(q)-T_D(q'_D))dq,\\
\beta T_D(q_D^{e,\mathring{q}}) &= \beta T_D(q'_D) + \frac{\beta}{(1-\gamma^e(\mathring{q}))\lambda_D(\overline{q}-\underline{q})}\int_{\mathring{q}}^{\overline{q}}(T_D(q)-T_D(q'_D))p_N^{\gamma^e(\mathring{q})}(q)dq.
\end{aligned}$$

In this case, the overall social welfare can be expressed as follows:

$$\begin{aligned}
s\Big(q_D^{e,\mathring{q}}, q_O^{e,\mathring{q}}\Big) =&\lambda_O\beta T_O(q_O^{e,\mathring{q}}) + \lambda_D\beta T_D(q_D^{e,\mathring{q}}),\\
=&\beta\Big(\lambda_O T_O(q_O^{e,\mathring{q}}) + \gamma\lambda_D T_D(q_D^{e,\mathring{q}}) + (1-\gamma)\lambda_D T_D(q_D^{e,\mathring{q}})\Big)\\
=&\beta\lambda_O T_O(q'_O) + \frac{\beta}{\overline{q}-\underline{q}}\int_{q_O^{e,\mathring{q}}}^{\overline{q}}(T_O(q)-T_O(q'_O))\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq + \beta\lambda_D T_D(q'_D)\\
&+\frac{\beta}{\overline{q}-\underline{q}}\int_{q_D^{e,\mathring{q}}}^{\mathring{q}}(T_D(q)-T_D(q'_D))dq + \frac{\beta}{\overline{q}-\underline{q}}\int_{\mathring{q}}^{\overline{q}}(T_D(q)-T_D(q'_D))p_N^{\gamma^e(\mathring{q})}(q)dq.
\end{aligned} \tag{S-36a}$$

Similar to (S-36a), we can write

$$\begin{aligned}
s\Big(q_D^{e,q_O^{\overline{e}}}, q_O^{\overline{e}}\Big) =&\beta\lambda_O T_O(q'_O) + \frac{\beta}{\overline{q}-\underline{q}}\int_{q_O^{\overline{e}}}^{\overline{q}}(T_O(q)-T_O(q'_O))dq + \beta\lambda_D T_D(q'_D)\\
&+\frac{\beta}{\overline{q}-\underline{q}}\int_{q_D^{e,q_O^{\overline{e}}}}^{q_O^{\overline{e}}}(T_D(q)-T_D(q'_D))dq.
\end{aligned} \tag{S-37a}$$

Combining (S-36a) and (S-37a) with $\mathring{q} < q_O^{e,\mathring{q}} < q_O^{\overline{e}}$, $q_D^{e,\mathring{q}} > q_D^{e,q_O^{\overline{e}}}$, we obtain the following

$$s\Big(q_D^{e,q_O^{\overline{e}}}, q_O^{\overline{e}}\Big) - s\Big(q_D^{e,\mathring{q}}, q_O^{e,\mathring{q}}\Big) = \frac{\beta}{\overline{q}-\underline{q}}\left(\int_{q_D^{e,q_O^{\overline{e}}}}^{q_D^{e,\mathring{q}}}\overline{T}_D(q)dq + \int_{\mathring{q}}^{q_O^{e,\mathring{q}}}\overline{T}_D(q)\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq\right)$$

$$+\frac{\beta}{\bar{q}-\underline{q}}\int_{q_O^{e,\mathring{q}}}^{q_O^{\overline{e}}}\big(\overline{T}_D(q)-\overline{T}_O(q)\big)\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq$$
$$+\frac{\beta}{\bar{q}-\underline{q}}\int_{q_O^{\overline{e}}}^{\bar{q}}\big(\overline{T}_O(q)-\overline{T}_D(q)\big)p_N^{\gamma^e(\mathring{q})}(q)dq. \tag{S-38a}$$

Since $\frac{\overline{T}_O(q)}{\overline{T}_D(q)}\geq 1$ is non-decreasing for $q\in\Big[q_O^{e,\mathring{q}},\overline{q}\Big]$, the following holds

$$\frac{\partial\Big(\frac{\overline{T}_O(q)}{\overline{T}_D(q)}\Big)}{\partial q}=\frac{\frac{\partial\overline{T}_O(q)}{\partial q}\overline{T}_D(q)-\frac{\partial\overline{T}_D(q)}{\partial q}\overline{T}_O(q)}{\overline{T}_D(q)^2}\geq 0, \tag{S-39}$$

which implies that

$$\frac{\partial\overline{T}_O(q)}{\partial q}\geq\frac{\partial\overline{T}_D(q)}{\partial q}. \tag{S-40}$$

Thus, $\overline{T}_O(q)-\overline{T}_D(q)\geq 0$ is non-decreasing for $q\in\Big[q_O^{e,\mathring{q}},\overline{q}\Big]$. Further, condition (10) implies that

$$\begin{aligned}\int_{q_O^{\overline{e}}}^{\overline{q}}p_N^{\gamma^e(\mathring{q})}(q)dq&=\int_{q_O^{\overline{e}}}^{\overline{q}}1\,dq-\int_{q_O^{\overline{e}}}^{\overline{q}}(1-p_N^{\gamma^e(\mathring{q})}(q))\,dq\\&=\overline{q}-q_O^{\overline{e}}-\int_{q_O^{\overline{e}}}^{\overline{q}}(1-p_N^{\gamma^e(\mathring{q})}(q))\,dq\\&\geq\int_{q_O^{e,\mathring{q}}}^{\overline{q}}\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)\,dq-\int_{q_O^{\overline{e}}}^{\overline{q}}(1-p_N^{\gamma^e(\mathring{q})}(q))\,dq &\text{(S-41a)}\\&=\int_{q_O^{e,\mathring{q}}}^{q_O^{\overline{e}}}\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq. &\text{(S-41b)}\end{aligned}$$

Therefore, the following set of inequalities hold

$$\begin{aligned}&\int_{q_O^{\overline{e}}}^{\bar{q}}\big(\overline{T}_O(q)-\overline{T}_D(q)\big)p_N^{\gamma^e(\mathring{q})}(q)dq+\int_{q_O^{e,\mathring{q}}}^{q_O^{\overline{e}}}\big(\overline{T}_D(q)-\overline{T}_O(q)\big)\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq\\&\geq\big(\overline{T}_O(q_O^{\overline{e}})-\overline{T}_D(q_O^{\overline{e}})\big)\int_{q_O^{\overline{e}}}^{\bar{q}}p_N^{\gamma^e(\mathring{q})}(q)dq+\int_{q_O^{e,\mathring{q}}}^{q_O^{\overline{e}}}\big(\overline{T}_D(q)-\overline{T}_O(q)\big)\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq &\text{(S-42a)}\\&\geq\big(\overline{T}_O(q_O^{\overline{e}})-\overline{T}_D(q_O^{\overline{e}})\big)\int_{q_O^{e,\mathring{q}}}^{q_O^{\overline{e}}}\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq+\int_{q_O^{e,\mathring{q}}}^{q_O^{\overline{e}}}\big(\overline{T}_D(q)-\overline{T}_O(q)\big)\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq &\text{(S-42b)}\\&\geq\int_{q_O^{e,\mathring{q}}}^{q_O^{\overline{e}}}\big(\overline{T}_O(q)-\overline{T}_D(q)\big)\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq+\int_{q_O^{e,\mathring{q}}}^{q_O^{\overline{e}}}\big(\overline{T}_D(q)-\overline{T}_O(q)\big)\Big(1-p_N^{\gamma^e(\mathring{q})}(q)\Big)dq &\text{(S-42c)}\\&=0, &\text{(S-42d)}\end{aligned}$$

where (S-42a) and (S-42c) follow from that $\overline{T}_O(q)-\overline{T}_D(q)\geq 0$ is non-decreasing for $q\in\Big[q_O^{e,\mathring{q}},\overline{q}\Big]$, and (S-42b) follows from (S-41b). Observe that, the first line of the summation in (S-38a) is always positive, and therefore (S-42d) implies that $s\Big(q_D^{e,q_O^{\overline{e}}},q_O^{\overline{e}}\Big)-s\Big(q_D^{e,\mathring{q}},q_O^{e,\mathring{q}}\Big)>0$ when $\mathring{q}\leq q_O^{e,\mathring{q}}<q_O^{\overline{e}}$. Following similar steps, we can also show that $s\Big(q_D^{e,q_O^{\overline{e}}},q_O^{\overline{e}}\Big)-s\Big(q_D^{e,\mathring{q}},q_O^{e,\mathring{q}}\Big)>0$ when $q_O^{e,\mathring{q}}<\mathring{q}<q_O^{\overline{e}}$; we omitted those steps here for brevity.

In sum, the social welfare under the targeted priority mechanism with $\mathring{q}=q_O^{\overline{e}}$ is greater than that of any mechanism with $\mathring{q}>q_O^{\overline{e}}$, and those with $\mathring{q}<q_O^{\overline{e}}$. Therefore, the mechanism with $\mathring{q}=q_O^{\overline{e}}$ maximizes the social welfare.

- If, on the other hand, $\mathring{q}_u < q_O^{\overline{e}}$, then setting $\mathring{q} = q_O^{\overline{e}} > \mathring{q}_u$ would still result in all type D participating in the program ($\gamma^e(\mathring{q}) = 1$), and leads to a class-separating equilibrium. Therefore, the proof follows by the proof of the first scenario, i.e. $\mathring{q}_u = q_O^{\overline{e}}$, since any target threshold $\mathring{q} > q_O^{\overline{e}}$ would still result in a donor-recipient mismatch, whereas any target threshold $\mathring{q} < q_O^{\overline{e}}$ would result in the nonuse of organs of quality $q \in (\mathring{q}, q_O^{\overline{e}})$.

(b) For any system such that $\mathring{q}_u > q_O^{\overline{e}}$, any targeted priority mechanism with $\mathring{q} < \mathring{q}_u$ results in a nonempty set of non-participating type D ($\gamma^e(\mathring{q}) < 1$, see Theorem 2). On the other hand, by the definition of $q_O^{\overline{e}}$, any mechanism with $\mathring{q} \geq \mathring{q}_u > q_O^{\overline{e}}$ results in type O lowering their equilibrium offer acceptance threshold below $q_O^{\overline{e}} < \mathring{q}$. Therefore, in this case, no targeted priority mechanism can guarantee a class-separating equilibrium.

Next, we prove that if $\frac{\overline{T}_O(q)}{\overline{T}_D(q)} \geq 1$ for $q \in [q_O^{e_0}, \overline{q}]$, which is equivalent to $T_O(q) - \frac{\alpha}{\beta d_O} \geq T_D(q) - \frac{\alpha}{\beta d_D}$, then setting $\mathring{q} = q_O^{e_0}$ improves the overall social welfare. Observe that for the targeted priority mechanism with $\mathring{q} = q_O^{e_0}$, if we have $q_D^{e_0} < q_D^{e,\mathring{q}}$, then $s\Big(q_D^{e,\mathring{q}}, q_O^{e,\mathring{q}}\Big) > s(q_D^{e_0}, q_O^{e_0})$ because $q_O^{e_0} < q_O^{e,\mathring{q}}$ (see Theorem 1), and thus we conclude that the mechanism improves the social welfare.

If, on the other hand, we have $q_D^{e_0} \geq q_D^{e,\mathring{q}}$, then $\mathring{q} = q_O^{e_0}$. Further, $q_O^{e,\mathring{q}}$, $\mathring{q}$, and $q_D^{e,\mathring{q}}$ satisfy the following equations

$$\beta T_O(q_O^{e,\mathring{q}}) = \beta T_O(q'_O) + \frac{\beta}{\lambda_O(\overline{q} - \underline{q})} \int_{q_O^{e,\mathring{q}}}^{\overline{q}} (T_O(q) - T_O(q'_O)) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq, \tag{S-43a}$$

$$\beta T_D(q_D^{e,\mathring{q}}) = \beta T_D(q'_D) + \frac{\beta}{\gamma^e(\mathring{q})\lambda_D(\overline{q} - \underline{q})} \int_{q_D^{e,\mathring{q}}}^{\mathring{q}} (T_D(q) - T_D(q'_D)) dq, \tag{S-43b}$$

$$\beta T_D(q_D^{e,\mathring{q}}) = \beta T_D(q'_D) + \frac{\beta}{(1 - \gamma^e(\mathring{q}))\lambda_D(\overline{q} - \underline{q})} \int_{\mathring{q}}^{\overline{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma^e(\mathring{q})}(q) dq. \tag{S-43c}$$

On the other hand, $q_O^{e_0}$ and $q_D^{e_0}$ satisfy

$$\beta T_O(q_O^{e_0}) = \beta T_O(q'_O) + \frac{\beta}{\lambda_O(\overline{q} - \underline{q})} \int_{q_O^{e_0}}^{\overline{q}} (T_O(q) - T_O(q'_O))(1 - p_D(q)) dq, \tag{S-44a}$$

$$\beta T_D(q_D^{e_0}) = \beta T_D(q'_D) + \frac{\beta}{\lambda_D(\overline{q} - \underline{q})} \int_{q_D^{e_0}}^{\overline{q}} (T_D(q) - T_D(q'_D)) p_D(q) dq. \tag{S-44b}$$

Using (6), the change in overall social welfare after the introduction of the targeted priority program with a target threshold $\mathring{q} = q_O^{e_0}$ can be written as

$$\begin{aligned} s\Big(q_D^{e,\mathring{q}}, q_O^{e,\mathring{q}}\Big) - s(q_D^{e_0}, q_O^{e_0}) =& \lambda_O T_O(q_O^{e,\mathring{q}}) + \gamma \lambda_D T_D(q_D^{e,\mathring{q}}) + (1 - \gamma)\lambda_D T_D(q_D^{e,\mathring{q}}) \\ & - \lambda_O T_O(q_O^{e_0}) - \lambda_D T_D(q_D^{e_0}). \end{aligned} \tag{S-45}$$

Using (S-43c), we can write the following set of equations

$$\begin{aligned} & \lambda_O T_O(q_O^{e,\mathring{q}}) + \gamma \lambda_D T_D(q_D^{e,\mathring{q}}) + (1 - \gamma)\lambda_D T_D(q_D^{e,\mathring{q}}) \\ =& \lambda_O T_O(q'_O) + \frac{1}{\overline{q} - \underline{q}} \int_{q_O^{e,\mathring{q}}}^{\overline{q}} \overline{T}_O(q) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq + \lambda_D T_D(q'_D) + \frac{1}{\overline{q} - \underline{q}} \int_{q_D^{e,\mathring{q}}}^{\mathring{q}} \overline{T}_D() dq \\ & + \frac{1}{\overline{q} - \underline{q}} \int_{\mathring{q}}^{\overline{q}} \overline{T}_D(q) p_N^{\gamma^e(\mathring{q})}(q) dq. \\ =& \lambda_O T_O(q'_O) + \frac{1}{\overline{q} - \underline{q}} \int_{q_O^{e,\mathring{q}}}^{\overline{q}} \overline{T}_O(q) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq + \lambda_D T_D(q'_D) + \frac{1}{\overline{q} - \underline{q}} \int_{q_D^{e,\mathring{q}}}^{q_O^{e_0}} \overline{T}_D(q) dq \\ & + \frac{1}{\overline{q} - \underline{q}} \int_{q_O^{e_0}}^{\overline{q}} \overline{T}_D(q) p_N^{\gamma^e(\mathring{q})}(q) dq, \end{aligned} \tag{S-46}$$

where (S-46) follows because we have $\mathring{q} = q_O^{e_0}$.

Similarly, using (S-44b) we can write

$$\begin{aligned}\lambda_O T_O(q_O^{e_0}) + \lambda_D T_D(q_D^{e_0}) =& \lambda_O T_O(q'_O) + \frac{1}{\bar{q}-\underline{q}} \int_{q_O^{e_0}}^{\bar{q}} \overline{T}_O(q)(1-p_D(q))dq \\ &+ \lambda_D T_D(q'_D) + \frac{1}{(\bar{q}-\underline{q})} \int_{q_D^{e_0}}^{\bar{q}} \overline{T}_D(q) p_D(q) dq. \end{aligned} \tag{S-47}$$

Combining (S-45)–(S-47), we have

$$\begin{aligned} s\Big(q_D^{e,\mathring{q}}, q_O^{e,\mathring{q}}\Big) - s(q_D^{e_0}, q_O^{e_0}) =& \frac{\beta}{\bar{q}-\underline{q}} \Big(\int_{q_O^{e_0}}^{\bar{q}} \big(\overline{T}_O(q) - \overline{T}_D(q)\big)\Big(p_D(q) - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq \\ &+ \int_{q_D^{e_0}}^{q_O^{e_0}} \overline{T}_D(q)(1-p_D(q))dq + \int_{q_D^{e,\mathring{q}}}^{q_D^{e_0}} \overline{T}_D(q) dq \Big) \\ =& \frac{\beta}{\bar{q}-\underline{q}} \Big(\int_{q_O^{e_0}}^{\bar{q}} \big(\overline{T}_O(q) - \overline{T}_D(q)\big)\Big(p_D(q) - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq \\ &+ \int_{q_D^{e_0}}^{q_O^{e_0}} \overline{T}_D(q)(1-p_D(q))dq + \int_{q_D^{e,\mathring{q}}}^{q_D^{e_0}} \overline{T}_D(q) dq \Big) \\ >& 0, \end{aligned} \tag{S-48}$$

where (S-48) follows because $\overline{T}_O(q) - \overline{T}_D(q) > 0$ for $q \in [q_O^{e_0}, \bar{q}]$. Hence, we have $s\Big(q_D^{e,\mathring{q}}, q_O^{e,\mathring{q}}\Big) > s(q_D^{e_0}, q_O^{e_0})$, and the proof follows.

□

COROLLARY 2. *There exists a targeted priority mechanism such that $\mathring{q} = q_O^{e,\mathring{q}}$, which satisfies the following:*

(a) *If $q_D^{e,\mathring{q}} > q_D^{e_0}$, then the mechanism improves the social welfare by increasing the utility of both patient groups.*

(b) *If $q_D^{e,\mathring{q}} \le q_D^{e_0}$, then the mechanism increases the utilization of available organs. Further, if $\frac{\overline{T}_O(q)}{\overline{T}_D(q)} \ge 1$ is non-decreasing in $q \in [q_O^{e_0}, \bar{q}]$, and the offer probabilities satisfy*

$$\int_{q_O^{e,\mathring{q}}}^{\bar{q}} \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big)\, dq \ge \int_{q_O^{e_0}}^{\bar{q}} p_O(q)\, dq, \tag{10}$$

then it increases the overall social welfare by providing non-targeted patients improved access to higher quality organs and increasing their utility.

Proof of Corollary 2. The existence of a targeted priority mechanism such that $\mathring{q} = q_O^{e,\mathring{q}}$ follows from Proposition 6, because we have $\mathring{q} > q_O^{e,\mathring{q}}$ for $\mathring{q} = \bar{q}$, while $q_O^{e,\mathring{q}} = q_O^{e_0} > \mathring{q}$ for $\mathring{q} = \underline{q} < \mathring{q}_l$.[2]

(a) For this case, using Theorem 1 and the definition of social welfare given in (6), we have the following inequalities

$$s(q_D^{e,\mathring{q}}, q_O^{e,\mathring{q}}) = \lambda_D \beta T_D(q_D^{e,\mathring{q}}) + \lambda_O \beta T_O(q_O^{e,\mathring{q}}) \tag{S-49a}$$

$$> \lambda_D \beta T_D(q_D^{e_0}) + \lambda_O \beta T_O(q_O^{e_0}) \tag{S-49b}$$

$$= s(q_D^{e_0}, q_O^{e_0}), \tag{S-49c}$$

where (S-49a) and (S-49c) follow by combining the results of Theorem 1 with (6), and (S-49b) follows because $q_D^{e,\mathring{q}} > q_D^{e_0}$ and $q_O^{e,\mathring{q}} = \mathring{q} > q_O^{e_0}$ (which follows by a similar set of steps to those provided in (S-32b) and (S-33b) in the proof of Theorem 4), and the overall arrival rates of types D and O are fixed. Hence, the proof follows.

[2] Under the assumption that $p_N^{\gamma}(q)$ satisfies $\frac{\partial^2 p_N^{\gamma}(q)}{\partial \gamma^2} \ge 0$ for any q, Proposition 6 implies the uniqueness of such a mechanism.

(b) First, we show that $\gamma^e(\mathring{q}) > 0$ for $\mathring{q} = q_O^{e,\mathring{q}}$ under this case. Assume, on the contrary that $\gamma^e(\mathring{q}) = 0$, which implies that $\mathring{q} \leq \mathring{q}_l$ (see Theorem 2). If $q_D^{e_0} < \mathring{q} \leq \mathring{q}_l$, then any type D participating would enjoy an expected utility of $\beta T_D(\mathring{q}) > \beta T_D(q_D^{e_0})$, which contradicts with that $\gamma^e(\mathring{q}) = 0$ (see Proposition 1). And thus, we have $q_O^{e,\mathring{q}} = \mathring{q} \leq q_D^{e_0} < q_O^{e_0} = q_O^{e,\mathring{q}}$ (see Proposition 6(a)), which leads to a contradiction.

Therefore, we have that $\mathring{q} > \mathring{q}_l$ and $\gamma^e(\mathring{q}) > 0$, which together with the assumption that $p_N^{\gamma(\mathring{q})}(q)$ is decreasing in $\gamma(\mathring{q})$ imply that $p_N^{\gamma^e(\mathring{q})}(q) < p_N^{\gamma=0}(q) = p_D(q)$. Furthermore, following similar steps to those in the proof of Theorem 4, we can prove that $q_O^{e,\mathring{q}} > q_O^{e_0}$, using Theorem 1, since $p_N^{\gamma^e(\mathring{q})}(q) < p_D(q)$ and

$$\beta T_O(q_O^{e,\mathring{q}}) = \beta T_O(q'_O) + \frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \int_{q_O^{e,\mathring{q}}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq,$$
$$\beta T_O(q_O^{e_0}) = \beta T_O(q'_O) + \frac{\beta}{\lambda_O(\bar{q}-\underline{q})} \int_{q_O^{e_0}}^{\bar{q}} (T_O(q) - T_O(q'_O))(1 - p_D(q)) dq.$$

Observe that having $q_O^{e,\mathring{q}} > q_O^{e_0}$ implies that the utility of non-targeted patients group improves after the introduction of the program (see Theorem 1).

When $q_D^{e,\mathring{q}} \leq q_D^{e_0}$, we can write the flows of organs (see (S-1)) to non-targeted, participating disadvantaged, and non-participating disadvantaged patients, respectively, as follows:

$$\mu_O\Big(q_O^{e,\mathring{q}}\Big) = \frac{\int_{\mathring{q}}^{\bar{q}} \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq}{\bar{q}-\underline{q}}, \quad \mu_P\Big(q_D^{e,\mathring{q}}\Big) = \frac{\mathring{q} - q_D^{e,\mathring{q}}}{\bar{q}-\underline{q}}, \quad \text{and} \quad \mu_N\Big(q_D^{e,\mathring{q}}\Big) = \frac{\int_{\mathring{q}}^{\bar{q}} p_N^{\gamma^e(\mathring{q})}(q) dq}{\bar{q}-\underline{q}}.$$

Similarly, the flow of organs to non-targeted and disadvantaged patients before the introduction of the program, respectively, are given as follows:

$$\mu_O(q_O^{e_0}) = \frac{\int_{q_O^{e_0}}^{\bar{q}} (1 - p_D(q)) dq}{\bar{q}-\underline{q}}, \quad \text{and} \quad \mu_D(q_D^{e_0}) = \frac{\int_{q_D^{e_0}}^{\bar{q}} p_D(q) dq}{\bar{q}-\underline{q}}.$$

Therefore, the following holds:

$$\begin{aligned} \mu_O\Big(q_O^{e,\mathring{q}}\Big) &= \frac{\int_{\mathring{q}}^{\bar{q}} \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq}{\bar{q}-\underline{q}} \\ &> \frac{\int_{q_O^{e_0}}^{\bar{q}} (1 - p_D(q)) dq}{\bar{q}-\underline{q}} \\ &= \mu_O(q_O^{e_0}), \end{aligned} \tag{S-50a}$$

where (S-50a) follows from the condition (10). Thus, the flow of organs allocated to non-targeted patients increases after the initiation of the program, and together with $q_O^{e,\mathring{q}} > q_O^{e_0}$, we have that non-targeted patients' utility improves along with their better access to higher quality organs.

Similar to the proof of Theorem 3, the overall social welfare before and after the introduction can be respectively written as follows:

$$s(q_O^{e_0}, q_D^{e_0}) = \lambda_O \beta T_O(q_O^{e_0}) + \lambda_D \beta T_D(q_D^{e_0}), \tag{S-51a}$$
$$s\Big(q_O^{e,\mathring{q}}, q_D^{e,\mathring{q}}\Big) = \lambda_O \beta T_O(q_O^{e,\mathring{q}}) + \lambda_D \beta T_D(q_D^{e,\mathring{q}}). \tag{S-51b}$$

Therefore, using (S-51b) as well as (S-43c) and (S-44b) from the proof of Theorem 4, after some algebra, the program with $\mathring{q} = q_O^{e,\mathring{q}}$ satisfies the following:

$$\begin{aligned}
s\Big(q_O^{e,\mathring{q}}, q_D^{e,\mathring{q}}\Big) =& \lambda_O \beta T_O(q_O^{e,\mathring{q}}) + \lambda_D \beta T_D(q_D^{e,\mathring{q}}) \\
=& \lambda_O \beta T_O(q'_O) + \frac{\beta}{\bar{q}-\underline{q}} \int_{q_O^{e,\mathring{q}}}^{\bar{q}} (T_O(q) - T_O(q'_O)) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq + \lambda_D \beta T_D(q'_D) \\
&+ \frac{\beta}{\bar{q}-\underline{q}} \int_{q_D^{e,\mathring{q}}}^{\mathring{q}} (T_D(q) - T_D(q'_D)) dq + \frac{\beta}{\bar{q}-\underline{q}} \int_{\mathring{q}}^{\bar{q}} (T_D(q) - T_D(q'_D)) p_N^{\gamma^e(\mathring{q})}(q) dq \\
=& \lambda_O \beta T_O(q'_O) + \frac{\beta}{\bar{q}-\underline{q}} \int_{q_O^{e,\mathring{q}}}^{\bar{q}} \overline{T}_O(q) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq + \lambda_D \beta T_D(q'_D) + \frac{\beta}{\bar{q}-\underline{q}} \int_{q_D^{e,\mathring{q}}}^{q_O^{e,\mathring{q}}} \overline{T}_D(q) dq \\
&+ \frac{\beta}{\bar{q}-\underline{q}} \int_{q_O^{e,\mathring{q}}}^{\bar{q}} \overline{T}_D(q) p_N^{\gamma^e(\mathring{q})}(q) dq \\
>& \lambda_O \beta T_O(q'_O) + \lambda_D \beta T_D(q'_D) + \frac{\beta}{\bar{q}-\underline{q}} \int_{q_D^{e_0}}^{q_O^{e_0}} \overline{T}_D(q) p_D(q) dq \\
&+ \frac{\beta}{\bar{q}-\underline{q}} \left[\int_{q_O^{e,\mathring{q}}}^{\bar{q}} \overline{T}_O(q) \Big(1 - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq + \int_{q_O^{e_0}}^{q_O^{e,\mathring{q}}} \overline{T}_D(q) dq + \int_{q_O^{e,\mathring{q}}}^{\bar{q}} \overline{T}_D(q) p_N^{\gamma^e(\mathring{q})}(q) dq \right] \qquad \text{(S-52a)} \\
=& \lambda_O \beta T_O(q'_O) + \lambda_D \beta T_D(q'_D) + \frac{\beta}{\bar{q}-\underline{q}} \int_{q_D^{e_0}}^{q_O^{e_0}} \overline{T}_D(q) p_D(q) dq + \frac{\beta}{\bar{q}-\underline{q}} \left[\int_{q_O^{e_0}}^{\bar{q}} \overline{T}_O(q)(1 - p_D(q)) dq \right. \\
&+ \int_{q_O^{e_0}}^{\bar{q}} \overline{T}_D(q) p_D(q) dq + \int_{q_O^{e,\mathring{q}}}^{\bar{q}} \big(\overline{T}_O(q) - \overline{T}_D(q)\big) \Big(p_D(q) - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq \\
&\left. + \int_{q_O^{e_0}}^{q_O^{e,\mathring{q}}} \big(\overline{T}_D(q) - \overline{T}_O(q)\big)(1 - p_D(q)) dq \right] \\
=& \lambda_O \beta T_O(q_O^{e_0}) + \lambda_D \beta T_D(q_D^{e_0}) + \frac{\beta}{\bar{q}-\underline{q}} \int_{q_D^{e_0}}^{q_O^{e_0}} \overline{T}_D(q) p_D(q) dq \\
&+ \frac{\beta}{\bar{q}-\underline{q}} \left[\int_{q_O^{e,\mathring{q}}}^{\bar{q}} \big(\overline{T}_O(q) - \overline{T}_D(q)\big) \Big(p_D(q) - p_N^{\gamma^e(\mathring{q})}(q)\Big) dq \right. \\
&\left. + \int_{q_O^{e_0}}^{q_O^{e,\mathring{q}}} \big(\overline{T}_D(q) - \overline{T}_O(q)\big)(1 - p_D(q)) dq \right] \\
>& \lambda_O \beta T_O(q_O^{e_0}) + \lambda_D \beta T_D(q_D^{e_0}) + \frac{\beta}{\bar{q}-\underline{q}} \int_{q_D^{e_0}}^{q_O^{e_0}} \overline{T}_D(q) p_D(q) dq \qquad \text{(S-52b)} \\
>& \lambda_O \beta T_O(q_O^{e_0}) + \lambda_D \beta T_D(q_D^{e_0}) \\
=& s(q_O^{e_0}, q_D^{e_0}) \qquad \text{(S-52c)}
\end{aligned}$$

where (S-52a) follows from that $q_D^{e,\mathring{q}} < q_D^{e_0}$ and $q_O^{e,\mathring{q}} > q_O^{e_0}$, and (S-52b) follows from condition (10) and that $\frac{\overline{T}_O(q)}{\overline{T}_D(q)} \geq 1$ is non-decreasing in $q \in [q_O^{e_0}, \bar{q}]$. Hence, the mechanism with $\mathring{q} = q_O^{e,\mathring{q}}$ increases the overall social welfare by providing non-targeted patients improved access to higher quality organs and increasing their utility.

□

Appendix C: Details of the Numerical Illustrations of the Impact of $\mathring{q}^*$ on Equilibrium Behavior

In §5.3, we present a numerical example to illustrate the analytical results concerning the impact of $\mathring{q}$ on patient equilibrium behavior. This section details the specifics of that numerical example. In particular, we set the organ allocation probability for non-participating type D individuals as $p_N^{\gamma}(q) = (0.5 - 0.04q)(1-\gamma)^{0.8}$ for any $q \in [0,1]$,

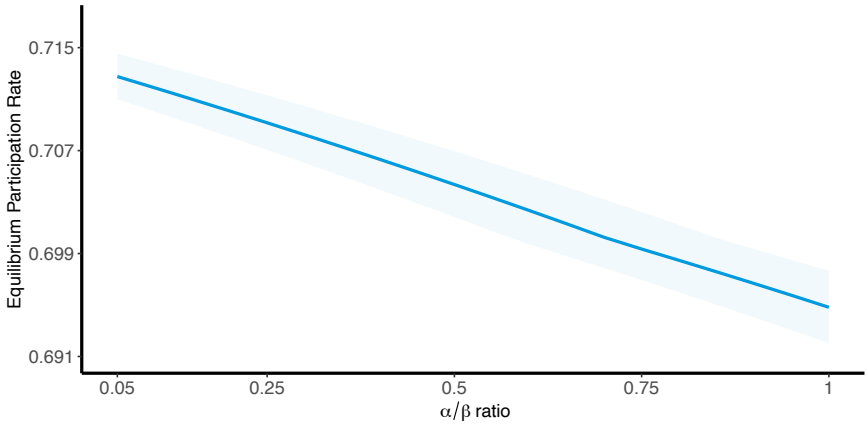


Figure S-1 Equilibrium participation rates of elderly candidates in the targeted priority program with a KDPI threshold of 84%, under varying ratios of pre- to post-transplant quality-of-life scores (α/β).

adjusting based on the level of participation γ and the organ quality q. For both patient types D and O, the post-transplant life expectancy is modeled as a linear function of the organ quality, defined as $T_D(q) = T_O(q) = q$, the simplicity of which allows for straightforward comparisons. The parameters α and β represent the pre-transplant and post-transplant life factors, respectively, set with a ratio of $\frac{\alpha}{\beta} = \frac{3}{4}$. Moreover, the pre-transplant mortality rates are set to differ significantly between disadvantaged (type D) and non-targeted (type O) patients, with rates denoted as $d_D = \frac{20}{3}$ and $d_O = \frac{20}{9}$, respectively, to highlight the disparities in such health outcomes. With these selcetions, the equations $\beta T_D(q_D) = \frac{\alpha}{d_D}$ and $\beta T_O(q_O) = \frac{\alpha}{d_O}$ yield unique solutions for the quality $q_D = 0.2$ and $q_O = 0.6$, respectively. These solutions indicate the threshold levels of organ quality that equate the adjusted life expectancies to the respective pre-transplant life for each patient type.

Appendix D: Robustness Analyses

To ensure the reliability and practical relevance of our findings, we examine the robustness of the proposed targeted priority mechanisms under variations in key modeling assumptions. In this section, we present two extensions: (i) a sensitivity analysis on the relative valuation of pre- versus post-transplant quality of life (§D.1), and (ii) an analysis of voluntary disenrollment behavior from the targeted program (§D.3). Together, these analyses test the stability of equilibrium participation outcomes and the broader performance of the mechanisms under more flexible behavioral and system-level assumptions.

D.1. Sensitivity Analysis

In this section, we perform a sensitivity analysis on key system parameters used in our case study (§7). Recall that the case study evaluates the impact of the proposed targeted priority mechanisms using the comprehensive and validated simulation framework developed by Sandıkçı et al. (2019). As discussed in §7, key simulation events and outcome metrics are either directly linked to national transplant databases or calibrated using SRTR's predictive models. A central modeling assumption in our study is the differentiation between the quality of life before and after transplantation. Consistent with estimates from the medical literature, we adopt a base-case ratio of $\alpha/\beta = 3/4$ to capture this trade-off (Wang et al. 2021, Tonelli et al. 2011). However, recognizing variability in these estimates across studies, we conduct a sensitivity analysis by varying this ratio from 0.05 to 1. Specifically, we examine how changes in the relative

valuation of pre- and post-transplant health states affect equilibrium participation decisions under the targeted priority mechanism.

We focus on equilibrium participation as it is a key driver of system-level outcomes under the proposed policy, as demonstrated in both our theoretical and numerical analyses. Figure S-1 presents the equilibrium participation rates of disadvantaged patients (specifically, elderly candidates) in a targeted priority program with a target KDPI threshold of 84%, assuming that patients optimize for quality-adjusted life expectancy. The results demonstrate that equilibrium participation remains highly stable across the range of tested values. Even when the ratio of pre- to post-transplant quality-of-life scores varies substantially (from 0.05 to 1), the equilibrium participation rate remains within a narrow band of 0.695 to 0.713, indicating the robustness of our findings to this modeling choice.

D.2. Numerical Evaluation of the Equilibrium Program Participation Under Alternative Objectives

Patients in the real allocation systems may potentially utilize metrics other than the expected utility when making their participation decisions. For example, elderly patients may particularly prioritize the reduction of mortality. Figure S-2 illustrates the identification of the equilibrium participation rate when mortality is the primary concern of patients. Similar to the case of expected utility, there is a significant discrepancy in mortality when the participation rate is near 0. As the participation rate increases, the normalized mortality of participating patients grows with an increasingly slower pace while remaining the lower mortality option until the participation rate reaches 0.92. It is noteworthy that the equilibrium participation rate of 0.92 identified by normalized mortality is higher than that of utilizing expected utility, which highlights the predominant benefit of the program in reducing pre-transplant mortality.

Patients may also prioritize their likelihood of receiving transplantation when making participation decisions. Figure S-3 illustrates the equilibrium participation decisions when patients prioritize normalized transplantation rate as their primary objective. It demonstrates that the program consistently yields a higher normalized transplantation rate for participating elderly patients when the participation rate is below 0.79. Note that the equilibrium participation rate identified by normalized transplantation is even more aggressive than that of expected utility or normalized mortality, which highlights the improved transplant access as the most prominent benefit of targeted priority mechanisms for elderly patients.

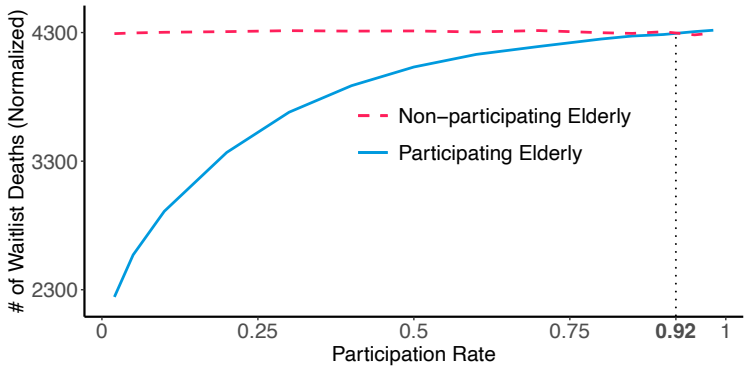


Figure S-2 Waitlist deaths among participating and non-participating elderly under the targeted priority program with a KDPI threshold of 84%. The counts are normalized to the arrival rate of each patient group.

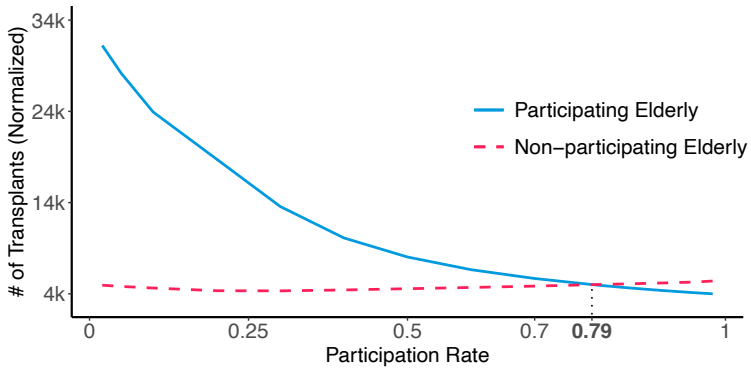
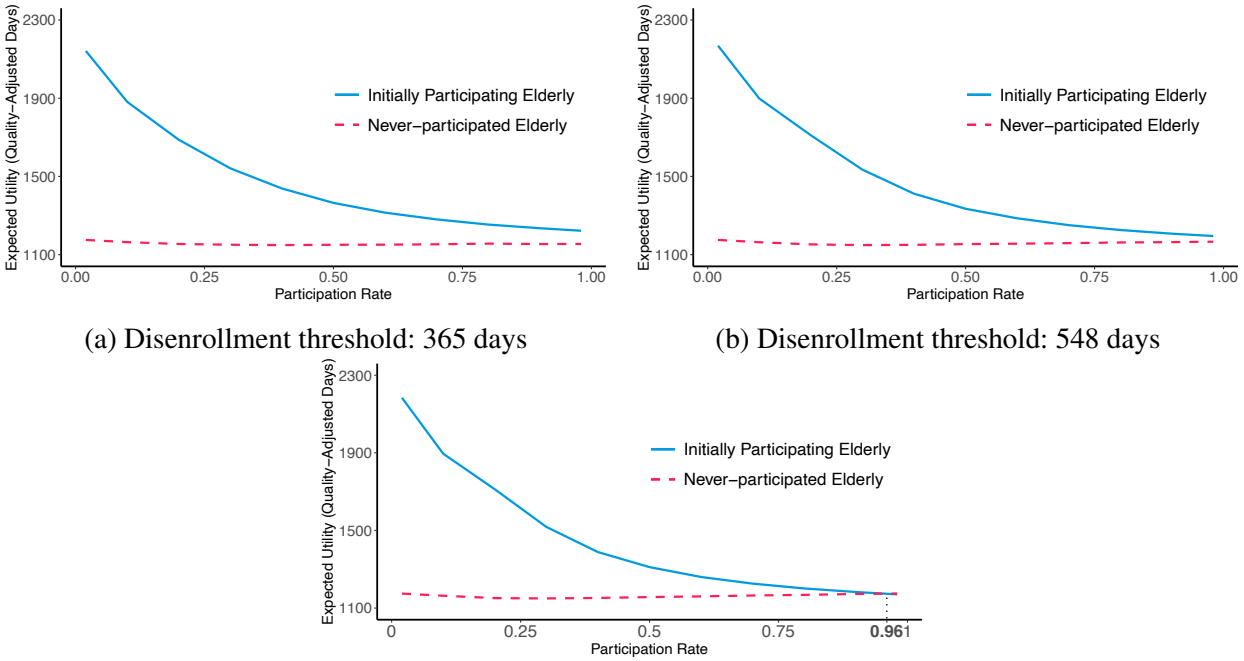


Figure S-3 Transplant counts of participating and non-participating elderly under the targeted priority program with a KDPI threshold of 84%. The counts are normalized to the arrival rate of each patient group.



(a) Disenrollment threshold: 365 days

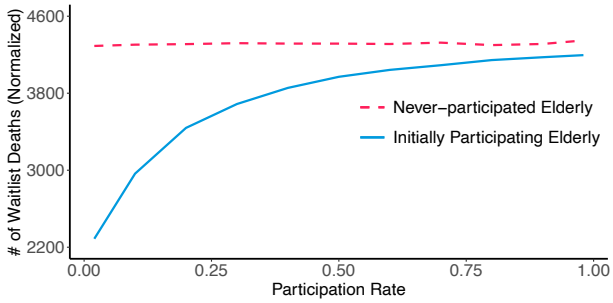
(b) Disenrollment threshold: 548 days

(c) Disenrollment threshold: 731 days

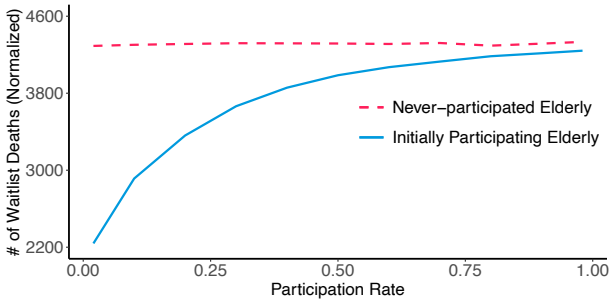
Figure S-4 Expected utility of participating and non-participating elderly patients under the targeted priority program with a KDPI threshold of 84%, assuming voluntary disenrollment from the program after a fixed waiting time.

D.3. Voluntary Disenrollment from the Targeted Program

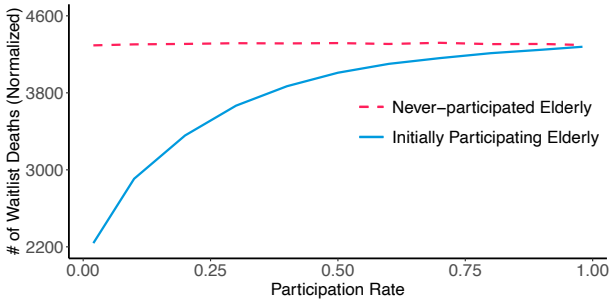
As discussed in §7.1, identifying equilibrium behavior within the organ allocation system depends critically on how patient preferences are modeled. In our main analysis, we explored equilibrium participation under three decision-making criteria: expected utility (quality-adjusted life expectancy, QALE), transplant probability, and pre-transplant mortality. This section examines the robustness of those findings under a more flexible policy that allows for voluntary disenrollment from the targeted priority program.



(a) Disenrollment threshold: 365 days



(b) Disenrollment threshold: 548 days

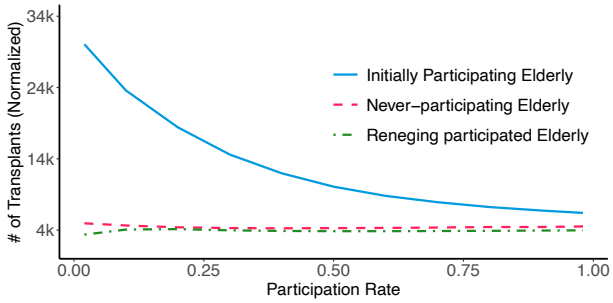


(c) Disenrollment threshold: 731 days

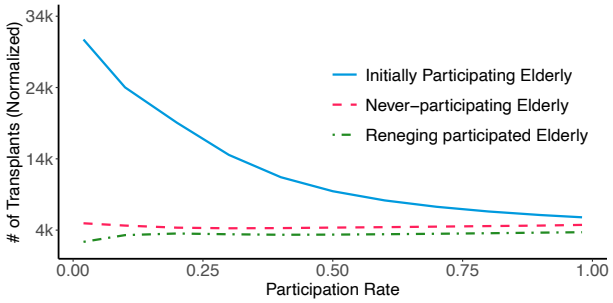
Figure S-5 Normalized pre-transplant deaths among participating and non-participating elderly patients under the targeted priority program with a KDPI threshold of 84%, assuming voluntary disenrollment after a fixed waiting time. Death counts are normalized by the arrival rate of each patient group.

Specifically, we relax the assumption of irrevocable participation by introducing a policy feature that permits patients to opt out of the targeted program after waiting for a fixed duration, termed the "disenrollment threshold", and rejoin the general waiting list with their accrued waiting time preserved. This mirrors the structure of the Eurotransplant Senior Program (ESP), which allows elderly patients to return to standard allocation after participation. We evaluate three disenrollment thresholds: 365 days (1 year), 548 days (1.5 years), and 731 days (2 years). Figures S-4a–S-4c present the expected utility of participating and non-participating elderly patients as a function of the overall participation rate, under each disenrollment policy. Across all three thresholds, we assume a KDPI threshold of 84% and use QALE as the primary decision criterion.

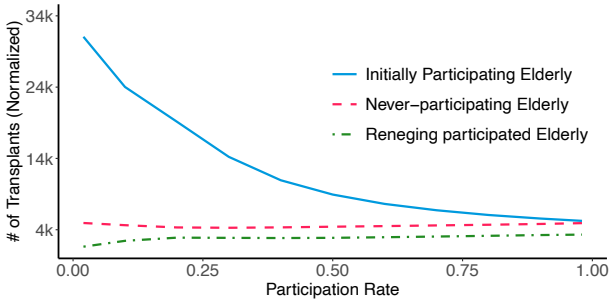
Strategic Implications of Voluntary Disenrollment. We find that when patients are allowed to opt out after 365 days, participating elderly achieve higher expected utility than non-participants across the entire range of nontrivial participation rates (Figure S-4a). Consequently, full participation becomes the unique equilibrium outcome: joining the program strictly dominates non-participation for all disadvantaged patients. A similar conclusion holds under a 548-day disenrollment threshold (Figure S-4b). When the disenrollment threshold is extended to 731 days (Figure S-4c), the expected utility of initial participants continues to dominate that of non-participants, although the equilibrium participation rate slightly drops to 0.96. This near-complete participation suggests that permitting voluntary disenrollment, even after a longer waiting period, does not discourage program engagement. Interestingly, relative to the no-disenrollment baseline (Figure 3), the introduction of a disenrollment option appears to increase equilibrium participation rates. In our implementation, patients who disenroll return to the non-participating queue without any advantage



(a) Disenrollment threshold: 365 days



(b) Disenrollment threshold: 548 days



(c) Disenrollment threshold: 731 days

Figure S-6 Normalized transplant counts for initially participating, non-participating, and initially participating but subsequently disenrolled elderly patients under the targeted priority program with a KDPI threshold of 84%, assuming voluntary disenrollment after a fixed waiting time. Counts are normalized by the arrival rate of each patient group.

in waitlist position or priority. As a result, disenrollment does not improve access to higher-quality kidneys and can reduce access to program kidneys that would otherwise be prioritized for them. This helps explain why disenrollment is not beneficial *on average* to those who exit the program and why allowing an opt-out option does not erode participation incentives. As noted in the main body, individual patients may still reconsider participation as their clinical status

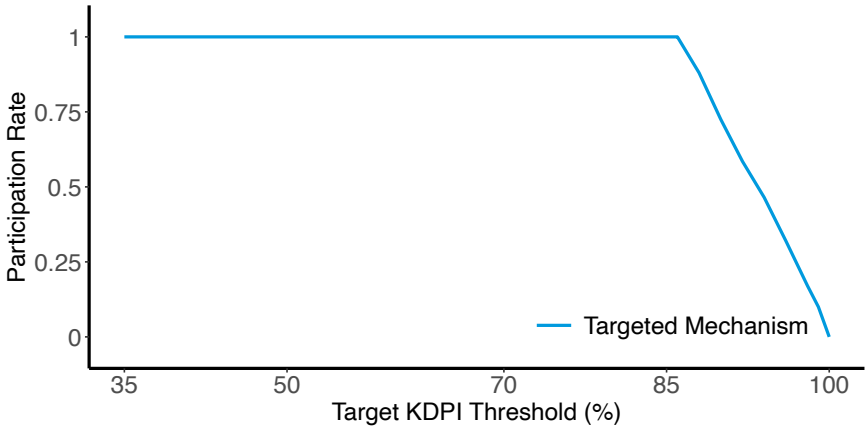


Figure S-7 Equilibrium participation rates of elderly patients in targeted priority mechanisms under varying KDPI thresholds, assuming a disenrollment threshold of 548 days.

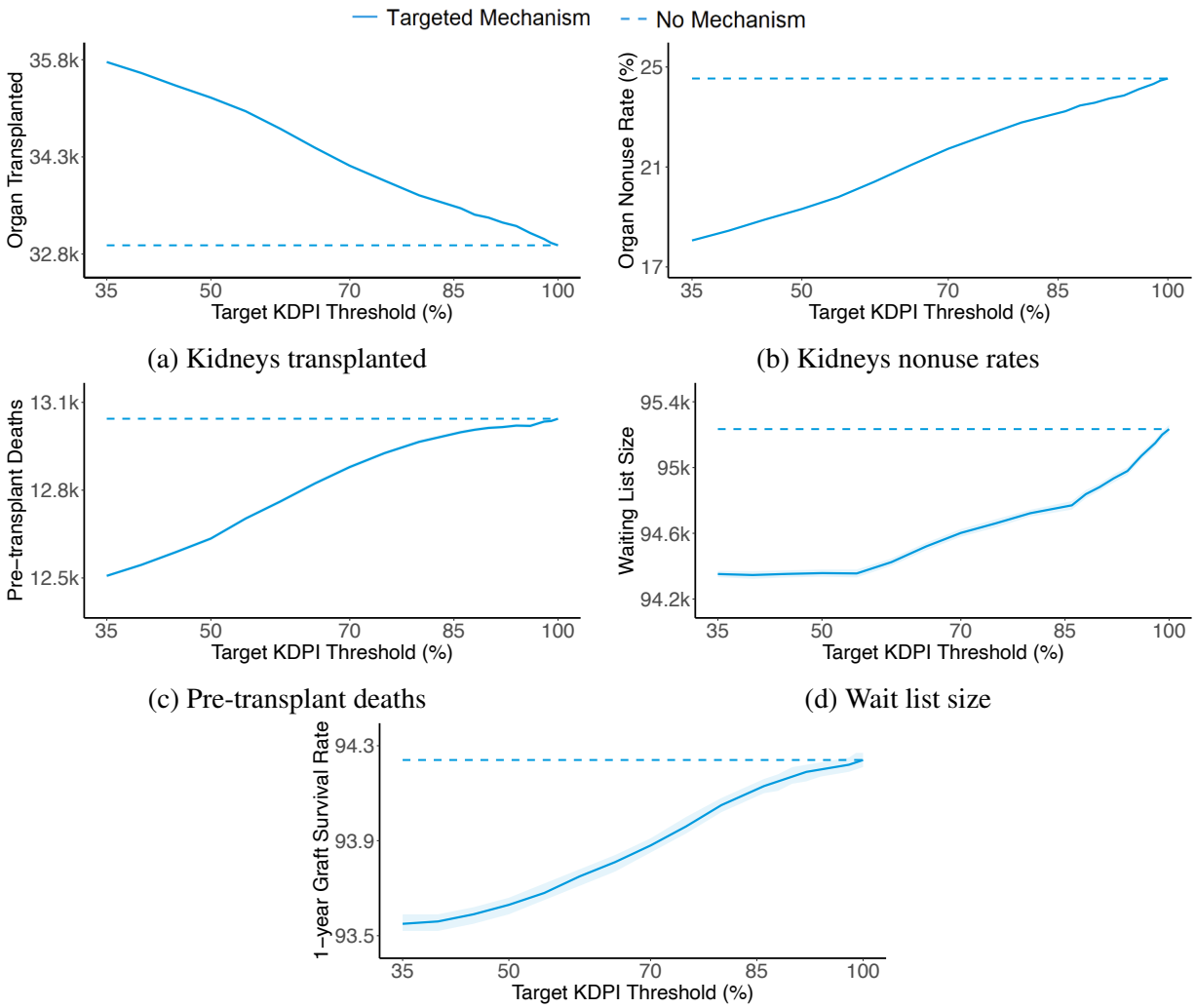
evolves, but modeling fully individualized, endogenous disenrollment decisions is beyond the scope of our simulation analysis.

We also examine how group-level outcomes vary under different decision criteria. Across all three disenrollment thresholds, elderly patients who initially participate in the program consistently experience significantly lower pre-transplant mortality rates compared to non-participants Figures S-5a–S-5c). Likewise, participants consistently achieve increased access to transplantation than non-participants under any nontrivial participation rate (Figures S-6a–S-6c). These patterns indicate that the benefits of voluntary disenrollment for participating elderly patients persist across different objective functions. While this suggests that the added flexibility of disenrollment may improve the perceived value of the program, it is not entirely clear whether the increase in participation is driven directly by the availability of the disenrollment provision or by the resulting dynamics it induces; further investigation is needed to understand this effect.

To better understand how voluntary disenrollments influence the equilibrium participation rate in the program, we vary the KDPI threshold under the 548-day disenrollment policy and examine the resulting changes in participation (Figure S-7). In line with our theoretical results and baseline numerical analysis, participation remains non-increasing in the KDPI threshold. Notably, full participation is sustained up to a target threshold of 86%, after which it gradually declines, reaching zero at a threshold of 100%. These findings reinforce the conclusion that allowing voluntary disenrollments does not undermine the attractiveness of targeted priority mechanisms; on the contrary, it broadens participation in equilibrium.

Impact of Targeted Priority Mechanisms with Voluntary Disenrollments on Key Transplant Outcomes. Although the results indicate that voluntary disenrollment does not confer a strategic advantage to individuals, real-life patients may still exhibit impatience, violating our assumption of full rationality. Accordingly, we evaluate the system-level impact of targeted priority mechanisms when patients withdraw from the program after waiting 1.5 years without receiving a transplant. Figure S-8 summarizes the system-wide effects of such a mechanism across key outcomes: number of transplants, organ nonuse rate, pre-transplant mortality, waiting list size, and one-year post-transplant graft survival. At a KDPI threshold of 84%, the program increases annual kidney transplants by 207 (Figure S-8a) and reduces the organ nonuse rate to 23.12%, down from the 24.54% baseline (Figure S-8b). These gains yield significant life-saving benefits, with up to 530 additional patients surviving the waitlist annually, including 60 direct lives saved (Figure S-8c). The higher transplant activity also decreases the waiting list size by approximately 470 patients (Figure S-8d). Importantly, these improvements in access and efficiency do not come at the expense of transplant quality: one-year graft survival remains high at 94.11%, nearly matching the 94.24% baseline (Figure S-8e).

The Effect of Targeted Priority Mechanisms with Voluntary Disenrollments on Ineligible Patients. We also assess the impact on non-elderly patients, who are not eligible for the program but may be indirectly affected by changes in allocation dynamics. As shown in Figure S-9a, their pre-transplant mortality falls below baseline levels at KDPI thresholds of 82% and above. Additionally, average one-year graft survival among non-elderly recipients improves as the threshold decreases, reaching 96.01% at a KDPI of 76% (Figure S-9b). These outcomes suggest a favorable shift in donor-recipient match quality for non-elderly patients, as targeted programs redirect disadvantaged patients away from higher-quality organs. To further evaluate the overall benefit, we compute the expected utility for non-elderly patients across different KDPI thresholds (Figure S-9c). Utility increases monotonically up to a KDPI of 88%, after



(a) Kidneys transplanted

(b) Kidneys nonuse rates

(c) Pre-transplant deaths

(d) Wait list size

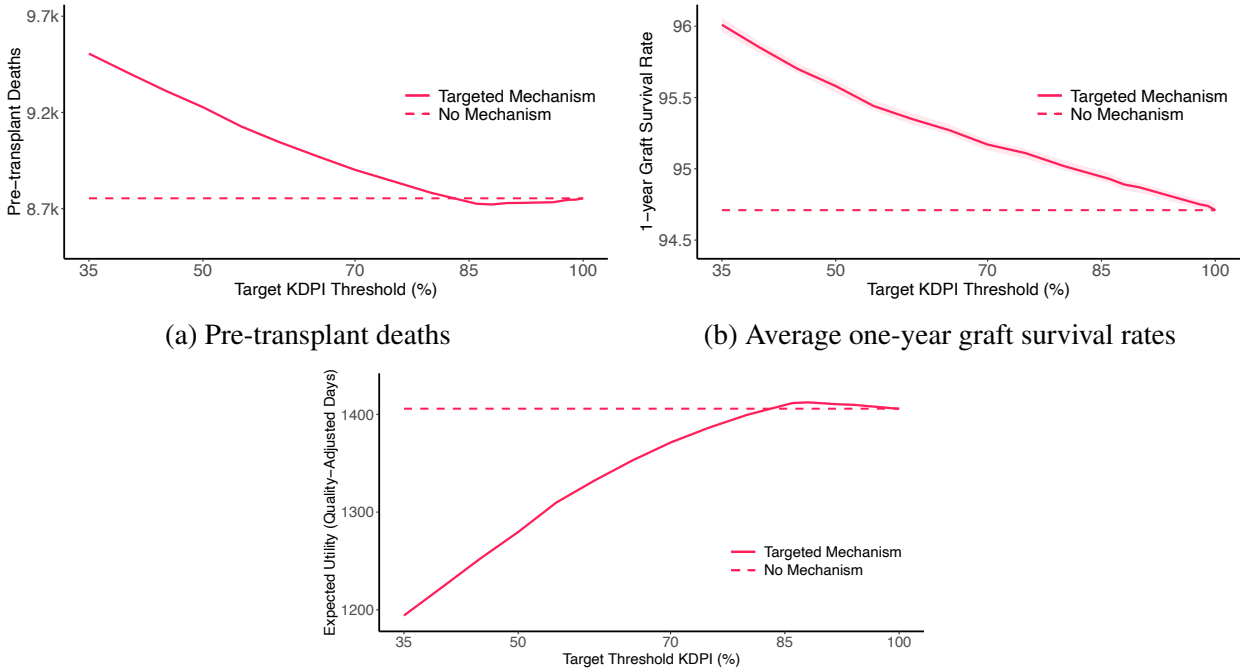
(e) Overall one-year graft survival rates

Figure S-8 Impact of targeted priority mechanisms on key transplant outcomes, assuming a disenrollment threshold of 548 days. Shaded regions represent 99% confidence intervals around the point estimates.

which it begins to decline due to reduced participation among elderly patients. Nevertheless, for any threshold above 84%, non-elderly patients attain higher expected utility compared to the baseline system. These findings reinforce the theoretical insights from §5 and the baseline numerical analysis, demonstrating that well-designed targeted priority programs can significantly enhance access, efficiency, and fairness—without imposing costs on ineligible patients.

Appendix E: Value of the Forfeiture Requirement

As discussed in the main body of the paper, the requirement that participating patients forfeit access to higher-quality organs in exchange for priority over lower-quality ones is central to the effectiveness and fairness of targeted priority mechanisms. This design feature serves two purposes: it ensures efficient organ–recipient matching by directing lower-quality organs to patients more likely to accept them, and it protects ineligible patients from being crowded out. Without forfeiture, participation in the program would be a strictly dominant strategy for all disadvantaged patients, leading to near-universal enrollment. In such a scenario, ineligible patients would lose access to program-designated organs without receiving compensating access to higher-quality organs, thereby worsening their outcomes.



(a) Pre-transplant deaths

(b) Average one-year graft survival rates

(c) Expected utility

Figure S-9 Impact of targeted priority mechanisms on key transplant outcomes among non-elderly transplant recipients, assuming a disenrollment threshold of 548 days. Shaded areas indicate 99% confidence intervals around the point estimates.

To quantify the value of forfeiture, we use our simulation model to compare targeted priority programs with a KDPI threshold of 84%, under two scenarios: one with forfeiture (the baseline design) and one without. With forfeiture, ineligible patients (under age 65) experience 8,757 pre-transplant deaths in equilibrium and achieve an average of 580 expected post-transplant survival days. In contrast, without forfeiture, the number of pre-transplant deaths increases to 8,846, and expected post-transplant survival drops to 542 days. Interestingly, the graft survival rate for ineligible patients is slightly higher in the no-forfeiting case (94.99% vs. 94.95%). This indicates that, without forfeiture, ineligible patients gain access only to a limited share of even higher quality organs, but at the cost of overall reductions in both access and survival.

Taken together, these results highlight the critical importance of the forfeiting requirement. By maintaining access to a sufficient pool of organs for ineligible patients while improving utilization of lower-quality organs, forfeiture safeguards the implementability and system-wide benefits of targeted priority mechanisms.

Appendix F: Data Availability

The empirical analysis in this paper relies on data from the Organ Procurement and Transplantation Network (OPTN)/United Network for Organ Sharing (UNOS), provided under a data use agreement that does not permit the authors to share or redistribute the underlying data files. Therefore, the data used in this study cannot be publicly posted by the authors.

Public information describing the relevant data files and the corresponding data dictionary is available through the Scientific Registry of Transplant Recipients (SRTR) Standard Analysis Files documentation at https:

//www.srtr.org/requesting-srtr-data/about-srtr-standard-analysis-files/ and https://www.srtr.org/requesting-srtr-data/saf-data-dictionary/. Researchers interested in obtaining access to the underlying data may apply directly to OPTN/UNOS, subject to applicable review procedures and data use agreement terms, at https://hrsa.unos.org/data/view-data-reports/request-data/.