Online Appendix: Not Intended for Print Publication

A An Overloaded Queueing Model of Congestion at a DSA

This appendix derives queueing-theoretic results used in Sections 3 and 4. The first-order (i.e. fluid scale) results below are used to model congestion costs in the selfish routing formulation of Section 3. Similarly, the second order (i.e. diffusion scale) results are used to model the waiting time at a DSA and the resulting patient utility in Section 4. The first two results follow directly from Jennings and Reed (2012).

Depicted in Figure 6 is a multiclass Markovian queue with abandonments, where class k jobs arrive to the system according to a Poisson process with rate x_k $(k = 1, \ldots, K)$. The service times are iid exponential random variables with rate μ. Each job may abandon while waiting for service; the abandonment time is exponentially distributed with rate γ. Arrival, service and abandonment processes are mutually independent. The service discipline is First-Come-First-Served (FCFS). That is, jobs are served in a FCFS fashion without regard to their class designation. The long-run fraction of service effort allocated to class k is denoted by $\alpha_k, k = 1, \ldots, K$.

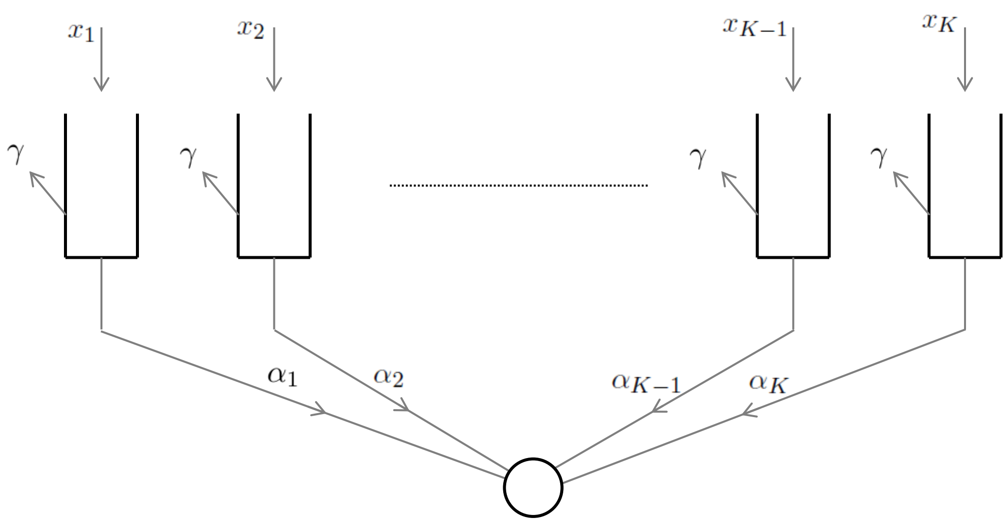


Figure 6: A multiclass overloaded fluid model of an OPO under multiple listing.

The system is overloaded in the sense that its nominal load exceeds its capacity. That is, defining $\rho_k = x_k/\mu$ as the offered load of class k, the total offered load exceeds one: $\rho = \sum_{k=1}^{K} \rho_k > 1$.

An important performance measure is the virtual waiting time process denoted by $V = \{V(s), s \geq 0\}$ which tracks how long a newly arriving job will have to wait to receive service, conditional on not abandoning until then. Characterizing the virtual waiting time analytically is not tractable. Therefore, we consider an approximate yet far more tractable model. The approximation is justified by a limit theorem of Jennings and Reed (2012) in an asymptotic regime where both x_k and μ grow large while keeping the offered load ρ_k constant $(k = 1, \ldots, K)$.

To be more specific, we consider a sequence of systems indexed by n[40], and the asymptotic regime we focus on has $\mu^n = n\mu$ and $x_k^n = n\, x_k$ for $k = 1, \ldots, K$. In words, the arrival and

[40]The superscript n will be attached to the quantities of interest corresponding to the n^{th} system.

service rates grow proportionally large while the time-to-death distribution and the offered loads of various classes remain fixed.

Jennings and Reed (2012) provide both a fluid-scale (i.e. a deterministic, first-order) and a diffusion scale (i.e. a stochastic, second-order) analysis of the system. The following two propositions are Theorems 1 and 2, respectively, of Jennings and Reed (2012), specialized to our setting. The former is a fluid-scale result whereas the latter is a diffusion-scale result.

Proposition 12 *If $V^n(0) \to V(0)$, then $V^n \to V$ as $n \to \infty$, where $V = (V(s), s \geq 0)$ is the unique solution to the following differential equation:*

$$V(s) = V(0) - s + \int_0^s \rho e^{-\gamma V(u)} du, \quad s \geq 0.$$

Moreover, $V(s) \to w^* = \frac{1}{\gamma} \ln \rho$ as $s \to \infty$.

Defining $\tilde{V}^n(s) = \sqrt{n}(V^n(s) - V(s))$ for all n, s and $x = \sum_{k=0}^N x_k$ as the total patient arrival rate, the diffusion scale result of Jennings and Reed (2012) yields the following, where $\Rightarrow$ denotes weak convergence, cf. Billingsley (1999) and Whitt (2002).

Proposition 13 *If $V^n(0) \to V(0) > 0$ and $\tilde{V}^n(0) \Rightarrow \tilde{V}(0)$ as $n \to \infty$, then $\tilde{V}^n \Rightarrow \tilde{V}$ as $n \to \infty$ where $\tilde{V}$ is the unique solution of the following stochastic differential equation:*

$$\tilde{V}(s) = \tilde{V}(0) + \int_0^s \tilde{\sigma}(u) d\tilde{B}(u) - \rho\gamma \int_0^s e^{-\gamma V(u)} \tilde{V}(u) du,$$

where $\tilde{B}$ is a standard Brownian motion and $\tilde{\sigma}^2(s) = e^{-\gamma V(s)}[x/\mu + (x/\mu)e^{-\gamma V(s)} + x(1 - e^{-\gamma V(s)})]/\mu$.

Jennings and Reed (2012) also observe that $\tilde{V}(s) \Rightarrow \tilde{V}(\infty)$ as $s \to \infty$, where $\tilde{V}(\infty)$ is a normal random variable with mean zero and variance $\sigma^2 = 1/\gamma\mu$.

We denote the steady-state virtual waiting time in the n^{th} system by W^n. Motivated by the results of Jennings and Reed (2012), we approximate W^n as follows:

$$W^n \simeq w^* + \frac{1}{\sqrt{n}} \tilde{V}(\infty). \tag{43}$$

That is, W^n is approximated by a normal random variable with mean $w^* = \frac{1}{\gamma} \ln \rho$ and variance $\sigma_n^2 = 1/(n\gamma\mu) = 1/(\gamma\mu^n)$; see page 1293 of Jennings and Reed (2012).

The fluid model (i.e. the deterministic, first-order approximation) associated with the queueing system displayed in Figure 6 is (a single server version of) a special case of that studied in Talreja and Whitt (2008); also see Liu and Whitt (2011a,b), and Ward (2012) for a survey of queueing models with abandonments[41]. Because the service discipline is FCFS, the probability of receiving service and the time spent in the system are identical across various classes, and depend only on the total arrival rate x. Let ϕ and $\bar{W}$ denote the probability of receiving service and the time spent in the system (irrespective of whether the job abandons or receives service), respectively. Also recall that α_k denotes the long-run fraction of service effort allocated to class k patients.

[41] Much of the analysis of Talreja and Whitt (2008) can be imitated for our single-server model with minor modifications.

Proposition 14 *In the fluid model, under the FCFS discipline, we have that*

$$\phi(x) = \frac{\mu}{x}, \quad \bar{W}(x) = \frac{1}{\gamma}\left(1 - \frac{\mu}{x}\right) \quad \textit{and} \quad \alpha_k = \frac{x_k}{x} \quad \textit{for} \quad k = 1, \ldots, K.$$

Proof of Proposition 14. To characterize the allocation rate α_k $(k = 1, \ldots, K)$, first consider the expected waiting time conditional on receiving a transplant w_k, which can be characterized by imitating the proof of Theorem 2.3 of Whitt (2004); see equations (2.24) and (2.26) of Whitt (2004). To be more specific, given α_k one can consider queue k in isolation with the arrival rate x_k, service rate $\mu\alpha_k$ and abandonment rate γ. Imitating Whitt's analysis in our setting gives

$$w_k = \frac{1}{\gamma} \ln\left(\frac{x_k}{\alpha_k \mu}\right) \quad \text{for} \quad k = 1, \ldots, K. \tag{44}$$

Since the FCFS rule is used to allocate the organs, we must have $w_k = w_j$ for all k, j. In other words, it follows from (44) that $x_k/\alpha_k = x_j/\alpha_j$ for all k, j. Since we also have $\sum_{k=1}^{K} \alpha_k = 1$, it follows that

$$\alpha_k = \frac{x_k}{x} \quad \text{for} \quad k = 1, \ldots, K. \tag{45}$$

Then it is straightforward to show that the queue length of class k, denoted by Q^k, is given by $Q^k = \frac{x_k(1-\mu/x)}{\gamma}$. Letting W^k denote the overall average waiting time in class k, since $W^k = Q^k/x_k$ by Little's Law, we have $W^k = (1 - \mu/x)/\gamma$. Similarly,

$$\phi_k = \frac{\alpha_k \mu}{x_k} = \frac{(x_k/x)\mu}{x_k} = \frac{\mu}{x}. \tag{46}$$

■

In applying these results, the queueing system in Figure 6 represents a DSA, say DSA k. Moreover, the terms job, receiving service and abandonment correspond to the terms patient, receiving a transplant and death while waiting, respectively. The arrival rate x_j will correspond to these patients multiple listing to DSA k from DSA j. More specifically, in Sections 3 and 4 a second subscript of k is added (corresponding to DSA k, i.e. the fly-in DSA) to denote the arrival rate of such patients, i.e. x_{jk}. Similarly, the rate x_k above will correspond to local patients in DSA k, who did not multiple list elsewhere. Their arrival rate will be $x_{kk} + (1 - \pi_k)\lambda_k$ where the rate $(1 - \pi_k)\lambda_k$ is the arrival rate of local patients of DSA k who do not have the option to multiple list whereas x_{kk} is the arrival rate of the local DSA k patients who could multiple list elsewhere but choose not to.

B An Auxiliary Probability Approximation

In this appendix, we approximate the probability $\mathbb{P}(\min(W_j, W_k) > X)$ for $j, k = 1, \ldots, K$ with $j \neq k$, where X is an exponential random variable with mean $1/\gamma$ and W_i $(i = j, k)$ is a normal random variable with mean

$$w_i^n = \bar{w}_i + \frac{1}{\gamma\sqrt{n}} \frac{\phi_i \theta_i}{\mu_i}$$

and variance $(\sigma_i^n)^2 = 1/(\gamma n \mu_i)$. We study this probability for large n and consider two cases: case i) $\bar{w}_j < \bar{w}_k$; case ii) $\bar{w}_j = \bar{w}_k$.

case i) $\bar{w}_j < \bar{w}_k$.

$$\begin{aligned}\mathbb{P}(\min(W_j, W_k) > X) &= \mathbb{P}(W_j > X, W_k > X) \\ &= \mathbb{P}(W_j > X) - \mathbb{P}(W_j > X, W_k < X) \\ &= \mathbb{P}(W_j > X) - \mathbb{P}(W_j > X > W_k).\end{aligned}$$

Thus, it is easy to see that

$$\begin{aligned}0 \le \mathbb{P}(W_j > X) - \mathbb{P}(\min(W_j, W_k) > X) =& \mathbb{P}(W_j > X > W_k) \\ \le& \mathbb{P}(W_j > W_k) \\ =& \mathbb{P}(W_j - W_k > 0),\end{aligned}$$

where $W_j - W_k$ is a normal random variable with mean and variance:

$$w_j^n - w_k^n = (\bar{w}_j - \bar{w}_k) + \frac{1}{\sqrt{n}\gamma} \left(\phi_j \frac{\theta_j}{\mu_j} - \phi_k \frac{\theta_k}{\mu_k}\right), \text{ and } \sigma^2 = \frac{1}{\gamma n}\left(\frac{1}{\mu_j} + \frac{1}{\mu_k}\right).$$

Therefore, we conclude for large n that

$$\mathbb{P}(W_j - W_k > 0) = \bar{\Phi}\left(-\frac{w_j^n - w_k^n}{\sigma}\right) \le \frac{\phi\left(-\frac{w_j^n - w_k^n}{\sigma}\right)}{-\frac{w_j^n - w_k^n}{\sigma}}, \tag{47}$$

where $\bar{\Phi} = 1 - \Phi$ and ϕ, Φ are the pdf and cdf, respectively, of a standard normal random variable. The inequality in (47) holds because $w_j^n - w_k^n > 0$ for large n.

Note that

$$\begin{aligned}-\frac{w_j^n - w_k^n}{\sigma} &= \sqrt{\frac{\gamma\mu_j\mu_k}{\mu_j + \mu_k}}\left(\sqrt{n}(\bar{w}_j - \bar{w}_k) + \frac{1}{\gamma}\left(\phi_j\frac{\theta_j}{\mu_j} - \phi_k\frac{\theta_k}{\mu_k}\right)\right) \\ &= \hat{c}_1(\sqrt{n}(\bar{w}_j - \bar{w}_k) + \hat{c}_2).\end{aligned}$$

Then we conclude for large n that

$$\begin{aligned}\mathbb{P}(W_j - W_k > 0) &\le \frac{1}{2\pi}\frac{\exp\{-\frac{1}{2}\hat{c}_1^2\ (\sqrt{n}(\bar{w}_j - \bar{w}_k) + \hat{c}_2)^2\}}{-\hat{c}_1\ [\sqrt{n}(\bar{w}_j - \bar{w}_k) + \hat{c}_2]} \\ &\le \frac{\hat{c}_3}{\sqrt{n}\exp\{\hat{c}_4(\bar{w}_k - \bar{w}_j)n\}},\end{aligned}$$

where $\hat{c}_3$, $\hat{c}_4$ are positive constants. Therefore,

$$\mathbb{P}(\min(W_j, W_k) > X) = \mathbb{P}(W_j > X) + o(1/\sqrt{n}).$$

case ii) $\bar{w}_j = \bar{w}_k = w^*$.

$$\mathbb{P}(\min(W_j, W_k) > X) = \int_0^\infty \gamma e^{-\gamma x} \prod_{l=j,k} \bar{\Phi}\left(\frac{x - w_l^n}{\sigma_l^n}\right) dx.$$

We choose Δ sufficiently large so that $\bar{\Phi}\left(\frac{x - w_l^n}{\sigma_l^n}\right) \simeq 1$ for all $x \le \underline{w} = w^* - \Delta/\sqrt{n}$. For example, one can set

$$\Delta \geq \frac{1}{\gamma\Lambda} \ + \max_{k=1,\ldots,K} \frac{3}{\sqrt{\gamma\mu_k}}.$$

We then approximate $\mathbb{P}(\min(W_j, W_k) > X)$ as follows:

$$\begin{aligned}
\mathbb{P}(\min(W_j, W_k) > X) &= \int_0^{\underline{w}} \gamma e^{-\gamma x} dx + \int_{\underline{w}}^{\infty} e^{-\gamma x} \prod_{l=j,k} \bar{\Phi}\left(\frac{x - w_l^n}{\sigma_l^n}\right) dx \\
&= 1 - e^{-\gamma \underline{w}} + e^{-\gamma \underline{w}} \int_0^{\infty} \gamma e^{-\gamma y} \prod_{l=j,k} \bar{\Phi}\left(\frac{y + \underline{w} - w_l^n}{\sigma_l^n}\right) dy \\
&= 1 - e^{-\gamma \underline{w}} + e^{-\gamma \underline{w}} I_1,
\end{aligned}$$

where $\underline{w} = w^* - \Delta/\sqrt{n}$, and

$$\begin{aligned}
I_1 &= \int_0^{\infty} \gamma e^{-\gamma y} \prod_{l=j,k} \bar{\Phi}\left(\frac{y + w^* - \Delta/\sqrt{n} - w^* - \frac{1}{\gamma\sqrt{n}} \frac{\phi_l \theta_l}{\mu_l}}{1/\sqrt{\gamma\mu_l n}}\right) dy \\
&= \int_0^{\infty} \gamma e^{-\gamma y} \prod_{l=j,k} \bar{\Phi}\left(\left(y\sqrt{n} - (\Delta + \frac{1}{\gamma}\phi_l \frac{\theta_l}{\mu_l})\right)\sqrt{\gamma\mu_l}\right) dy,
\end{aligned}$$

The change of variable $z = y\sqrt{n}$ gives

$$\begin{aligned}
I_1 &= \frac{1}{\sqrt{n}} \int_0^{\infty} \gamma e^{-z\gamma/\sqrt{n}} \prod_{l=j,k} \bar{\Phi}\left(\left(z - (\Delta + \frac{1}{\gamma}\phi_l \frac{\theta_l}{\mu_l})\right)\sqrt{\gamma\mu_l}\right) dy, \\
&\simeq \frac{\gamma}{\sqrt{n}} \int_0^{\infty} \prod_{l=j,k} \bar{\Phi}\left(\left(z - (\Delta + \frac{1}{\gamma}\phi_l \frac{\theta_l}{\mu_l})\right)\sqrt{\gamma\mu_l}\right) dy. \qquad (48)
\end{aligned}$$

To simplify this further, we use the following result from Cramer (1946) (see page 374; also see Horowitz (1980)). As $r \to \infty$,

$$\Phi(d_r y + c_r) = 1 - \frac{1}{r} e^{-y} + o(r), \qquad (49)$$

where

$$d_r = \frac{1}{\sqrt{2\ln r}} \quad \text{and} \quad c_r = \sqrt{2\ln r} - \frac{1}{2} \frac{\ln\ln r + \ln 4\pi}{\sqrt{2\ln r}}.$$

Fixing a large r, (49) yields the following approximation:

$$\bar{\Phi}(d_r y + c_r) \simeq \frac{C(r)}{r} e^{-y},$$

Thus,

$$\bar{\Phi}(y) \simeq \frac{C(r)}{r} \exp\left\{-\frac{y - c_r}{d_r}\right\}, \qquad (50)$$

where $C(r)$ is chosen so that $\bar{\Phi}(y) < 1$ for all $y \geq 0$ and that our approximation of the probability of receiving a transplant increases under multiple listing, cf. Equations (52) and (54).

Equation (50) yields

$$\bar{\Phi}\left(\frac{x-w_l}{\sigma_l}\right) \simeq \frac{C(r)}{r} \exp\left\{\frac{-x}{\sigma_l d_r} + \frac{w_l}{\sigma_l d_r} + \frac{c_r}{d_r}\right\}. \tag{51}$$

Substituting (51) into (48) gives

$$\begin{aligned} I_1 &\simeq \frac{\gamma}{\sqrt{n}} \int_0^\infty \prod_{l=j,k} \frac{C(r)}{r} \exp\left\{-\frac{z\sqrt{\gamma\mu_l}}{d_r} + \left(\Delta + \frac{1}{\gamma}\phi_l \frac{\theta_l}{\mu_l}\right)\frac{\sqrt{\gamma\mu_l}}{d_r} + \frac{c_r}{d_r}\right\} dz \\ &= \frac{\gamma}{\sqrt{n}} \frac{C(r)^2}{r^2} \exp\left\{2\frac{c_r}{d_r}\right\} \prod_{l=j,k} \exp\left\{\frac{1}{d_r}\left(\Delta\sqrt{\gamma\mu_l} + \frac{\phi_l\theta_l}{\sqrt{\gamma\mu_l}}\right)\right\} \frac{d_r}{\sqrt{\gamma}\sum_{l=1,2}\sqrt{\mu_l}}. \end{aligned}$$

Thus, we write $I_1 \simeq I_2/\sqrt{n}$, where

$$I_2 = \gamma\frac{C(r)^2}{r^2} \exp\left\{2\frac{c_r}{d_r}\right\} \frac{d_r}{\sqrt{\gamma}(\sqrt{\mu_1}+\sqrt{\mu_2})} \prod_{l=j,k} \exp\left\{\frac{1}{d_r}\left(\Delta\sqrt{\gamma\mu_l} + \frac{\phi_l\theta_l}{\sqrt{\gamma\mu_l}}\right)\right\}. \tag{52}$$

Focusing on the higher order terms, we further approximate $\log I_2$ as follows:

$$\log I_2 \simeq 2\frac{c_r}{d_r} + \log\left(\frac{\gamma C(r)^2}{r^2}\right) + \log\left(\frac{d_r}{\sqrt{\gamma}\, 2\sqrt{\bar{\mu}}}\right) + \frac{1}{d_r}\sum_{l=j,k}\left(\Delta\sqrt{\gamma\mu_l} + \frac{\phi_l\theta_l}{\sqrt{\gamma\mu_l}}\right). \tag{53}$$

Also note that[42]

$$\begin{aligned} u_k(j;q^n) &= \mathbb{P}(X > \min(W_j, W_k)) = 1 - \mathbb{P}(\min(W_j,W_k) > X) = e^{-\gamma \underline{w}}(1-I_1) \\ &= e^{-\gamma w^*} e^{\gamma\Delta/\sqrt{n}}(1-I_1) \\ &\simeq e^{-\gamma w^*}\left(1+\frac{\gamma\Delta}{\sqrt{n}}\right)\left(1-\frac{I_2}{\sqrt{n}}\right) \\ &\simeq e^{-\gamma w^*}\left(1+\frac{\gamma\Delta}{\sqrt{n}} - \frac{I_2}{\sqrt{n}}\right) \\ &= \phi\left(1+\frac{\gamma\Delta}{\sqrt{n}} - \frac{I_2}{\sqrt{n}}\right) \end{aligned} \tag{54}$$

where $\phi = e^{-\gamma w^*}$. That is,

$$1 - \mathbb{P}(\min(W_j,W_k) > X) = \phi\left(1+\frac{\gamma\Delta}{\sqrt{n}} - \frac{I_2}{\sqrt{n}}\right). \tag{55}$$

Rearranging terms gives

$$\frac{1}{\phi}[1 - \mathbb{P}(\min(W_j,W_k) > X)] = 1 + \frac{1}{\sqrt{n}}(\gamma\Delta - I_2), \tag{56}$$

[42] Recall that $C(r)$ is chosen so that our approximation of the patient's utility increases under multiple listing, cf. Equation (52).

from which it follows that

$$\left[\frac{1}{\phi}\mathbb{P}(X > \min(W_j, W_k)) - 1\right] \sqrt{n} = \gamma\Delta - I_2.$$

Equivalently, we write

$$\begin{aligned} I_2 &= \gamma\Delta + \sqrt{n}\left[1 - \frac{1}{\phi}\mathbb{P}(X > \min(W_j, W_k))\right], \\ \log I_2 &= \log\left\{\gamma\Delta + \sqrt{n}\left[1 - \frac{1}{\phi}\mathbb{P}(X > \min(W_j, W_k))\right]\right\}. \end{aligned} \tag{57}$$

Defining $C_1 = 2\frac{c_r}{d_r} + \log\left(\gamma\frac{C(r)^2}{r^2}\right) + \log\left(\frac{d(r)}{2\sqrt{\gamma\mu}}\right)$ and $C_2 = \frac{1}{d_r}$, equations (51) and (57) gives

$$\log\left(\sqrt{n}\left[1 - \frac{1}{\phi}\mathbb{P}(X > W_{kj})\right] + \gamma\Delta\right) \simeq C_1 + C_2\left(\sum_{l=j,k}\Delta\sqrt{\gamma\mu_l} + \frac{\phi_l\theta_l}{\sqrt{\gamma\mu_l}}\right).$$

From this it is straightforward to deduce that the improvement due to multiple listing is $O(1/\sqrt{n})$.

C Proofs

Proof of Theorem 1. Without loss of generality, we assume $G(x)$ is connected. Otherwise, the proof below can be repeated for each subgraph. By Corollary 1, since $G(x)$ has no directed cycles, it is topologically sortable. This observation is crucial for the induction argument (on the levels of the graph) used in the proof.

First, we prove that $\bar{z}_{kj} = 0$ for all k, $j \in A(k)$ such that $j \neq k$. Suppose not. Then let L be the lowest level of the graph which contains a node, say node k, such that $\bar{z}_{kj} > 0$, with j in level $L+1$ or higher. Then consider the following two cases. Case i) $L = 0$, Case ii) $L > 0$.

Case i) $L = 0$. Note that since $x_{kj} > 0$, we must have $w_j \leq w_k$ and $\phi_j \geq \phi_k$, where (by (31))

$$\phi_k = \frac{\mu_k}{\lambda_k - \sum_{l\in A(k), l\neq k} x_{kl}} \quad \text{and} \quad \phi_j = \frac{\mu_j}{\lambda_j + \sum_{l\in A(j), l\neq j} x_{lj} - \sum_{l\in A(j), l\neq j} x_{jl}}. \tag{58}$$

At the same time, since $\bar{z}_{kj} > 0$, it follows from (22) that

$$\bar{w}_k \leq \bar{w}_j \quad \text{and} \quad \bar{\phi}_j \leq \bar{\phi}_k, \tag{59}$$

(otherwise $\bar{z}_{kj} = 0$), where by (28) and (58)

$$\bar{\phi}_k = \frac{\mu_k - \sum_{l\in A(k), l\neq k} \bar{z}_{kl}}{\lambda_k - \sum_{l\in A(k), l\neq k} x_{kl}} < \phi_k \leq \phi_j. \tag{60}$$

The first inequality in Equation (60) uses $L = 0$ crucially (because $x_{lk} = 0$ for $l \in A(k)$ by definition of $L = 0$ and that node k is in level 0). Also note that

$$\bar{\phi}_j = \frac{\mu_j + \sum_{l\in A(j), l\neq j} \bar{z}_{lj}\mathbb{1}_{\{\bar{\phi}_l \leq \bar{\phi}_j\}} - \sum_{l\in A(j), l\neq j} \bar{z}_{jl}}{\lambda_j + \sum_{l\in A(j), l\neq j} x_{lj}\mathbb{1}_{\{\bar{\phi}_l \leq \bar{\phi}_j\}} - \sum_{l\in A(j), l\neq j} x_{jl}}. \tag{61}$$

Note that (59)-(60) imply that

$$\bar{\phi}_j < \phi_j, \tag{62}$$

which combined with (61) and (58) implies that

$$\sum_{l \in A(j), l \neq j} \bar{z}_{lj} \mathbb{1}_{\{\bar{\phi}_l \leq \bar{\phi}_j\}} - \sum_{l \in A(j), l \neq j} \bar{z}_{jl} < 0,$$

or that $\sum_{l \in A(j), l \neq j} \bar{z}_{jl} > 0$. In particular, $\bar{z}_{jl} > 0$ for some l in a level greater than or equal to $L + 2 = 2$.

Since $\bar{z}_{jl} > 0$, we must have by (22) that

$$\bar{w}_j \leq \bar{w}_l \quad \text{and} \quad \bar{\phi}_l \leq \bar{\phi}_j. \tag{63}$$

But since $x_{jl} > 0$, we must also have $w_j \geq w_l$ and

$$\phi_j \leq \phi_l. \tag{64}$$

We conclude from (62), (63) and (64) that

$$\bar{\phi}_l \leq \bar{\phi}_j < \phi_j \leq \phi_l. \tag{65}$$

In particular,

$$\bar{\phi}_l < \phi_l. \tag{66}$$

We also have from (58) and (28), respectively, that

$$\phi_l = \frac{\mu_l}{\lambda_l + \sum_{i \in A(l), i \neq l} x_{il} - \sum_{i \in A(l), i \neq l} x_{li}} \tag{67}$$

and

$$\bar{\phi}_l = \frac{\mu_l + \sum_{i \in A(l), i \neq l} \bar{z}_{il} \mathbb{1}_{\{\bar{\phi}_i \leq \bar{\phi}_l\}} - \sum_{i \in A(l), i \neq l} \bar{z}_{li}}{\lambda_l + \sum_{i \in A(l), i \neq l} x_{il} \mathbb{1}_{\{\bar{\phi}_i \leq \bar{\phi}_l\}} - \sum_{i \in A(l), i \neq l} x_{li}}. \tag{68}$$

Then (66), (67) and (68) imply that

$$\sum_{i \in A(l), i \neq l} \bar{z}_{il} \mathbb{1}_{\{\bar{\phi}_i \leq \bar{\phi}_l\}} - \sum_{i \in A(l), i \neq l} \bar{z}_{li} < 0.$$

In particular, $\sum_{i \in A(l), i \neq l} \bar{z}_{li} > 0$, which implies that there exists a node i such that $\bar{z}_{li} > 0$ for some i in a level greater than or equal to $L + 3 = 3$.

By induction, we conclude that there exists a terminal node t and its predecessor u such that

$$\bar{\phi}_u < \phi_u \quad \text{and} \quad \bar{z}_{ut} > 0, \tag{69}$$

which implies by (22) that

$$\bar{w}_u \leq \bar{w}_t \quad \text{and} \quad \bar{\phi}_t \leq \bar{\phi}_u. \tag{70}$$

But, since $x_{ut} > 0$,

$$w_u \geq w_t \quad \text{and} \quad \phi_u \leq \phi_t. \tag{71}$$

From the induction step, we also have $\bar{\phi}_u < \phi_u$ which implies

$$\bar{\phi}_t \leq \bar{\phi}_u < \phi_u \leq \phi_t,$$

where the first and third inequalities follow from (70), and the second inequality follows from (69). Therefore,

$$\bar{\phi}_t < \phi_t. \tag{72}$$

We also have from (58) and (28), respectively, that

$$\phi_t = \frac{\mu_t}{\lambda_t + \sum_{m \in A(t), m \neq t} x_{mt}}, \tag{73}$$

and

$$\bar{\phi}_t = \frac{\mu_t + \sum_{m \in A(t), m \neq t} \bar{z}_{mt} \mathbb{1}_{\{\bar{\phi}_m \leq \bar{\phi}_l\}}}{\lambda_t + \sum_{m \in A(t), m \neq t} x_{mt} \mathbb{1}_{\{\bar{\phi}_m \leq \bar{\phi}_t\}}}. \tag{74}$$

Then, (73) and (74) imply that

$$\bar{\phi}_t \geq \phi_t, \tag{75}$$

which contradicts (72). Therefore, $\bar{z}_{kj} = 0$ for all $k, j \in A(k), j \neq k$ in case i.

Case ii) $L \geq 1$. Let k be the node in level L such that $\bar{z}_{kj} > 0$ for some $j \in A(k), j \neq k$. Since $\bar{z}_{lk} = 0$ for all predecessors of k, we must have $\bar{\phi}_k \geq \bar{\phi}_l$ for all such l. Then by (28) (and that $\bar{z}_{lk} = 0$ for all predecessors of k) we conclude that

$$\bar{\phi}_k = \frac{\mu_k - \sum_{j \in A(k), j \neq k} \bar{z}_{kj}}{\lambda_k + \sum_{j \in A(k), j \neq k} x_{jk} - \sum_{j \in A(k), j \neq k} x_{kj}}.$$

Next, comparing this with ϕ_k and noting that $\bar{z}_{kj} > 0$ for some j, where

$$\phi_k = \frac{\mu_k}{\lambda_k + \sum_{j \in A(k), j \neq k} x_{jk} - \sum_{j \in A(k), j \neq k} x_{kj}},$$

we conclude that $\bar{\phi}_k < \phi_k$. Starting with this, we can repeat the induction argument as in the case of $L = 0$.

To conclude the proof, it remains to show that $\bar{\phi}_k = \phi_k$ and $\bar{z}_{kk} = \phi_k x_{kk}$ for all k and that for all k, $j \in A(k), j \neq k$,

$$\bar{y}_{kj} = \begin{cases} \phi_j \ x_{kj} & \text{if } x_{kj} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

For all k, consider

$$\bar{\phi}_k = \frac{\mu_k}{\lambda_k + \sum_{j \in A(k), j \neq k} x_{jk} \mathbb{1}_{\{\bar{\phi}_j \leq \bar{\phi}_k\}} - \sum_{j \in A(k), j \neq k} x_{kj}}, \tag{76}$$

$$\phi_k = \frac{\mu_k}{\lambda_k + \sum_{j \in A(k), j \neq k} x_{jk} - \sum_{j \in A(k), j \neq k} x_{kj}}. \tag{77}$$

Note that $\bar{z}_{kj} = 0$ implies (by (23) and (24)) that $\bar{\phi}_j \leq \bar{\phi}_k$ for all k, whenever $x_{jk} > 0$. Therefore, it follows from (76) and (77) that

$$\bar{\phi}_k = \frac{\mu_k}{\lambda_k + \sum_{j\in A(k), j\neq k} x_{jk} - \sum_{j\in A(k), j\neq k} x_{kj}} = \phi_k.$$

Then, since $\bar{\phi}_k = \phi_k$ for all k, the result follows from (25)-(27). ■

Proof of Proposition 1. To simplify the formulation (6)-(8) further, note from (4)-(5) and Proposition 14 that

$$g_{s_k,\tau}(x) = -\frac{1}{\gamma}x - \big(L - \frac{1}{\gamma}\big)H_k(x).$$

Substituting this in (6) reduces the objective to the following:

$$-\frac{1}{\gamma}\sum_{k=1}^{K}\sum_{j\in A(k)} x_{jk} - \big(L - \frac{1}{\gamma}\big)\sum_{k=1}^{K} H_k\big(\sum_{j\in A(k)} x_{jk}\big).$$

To simplify the objective further, note that

$$\sum_{k=1}^{K}\sum_{j\in A(k)} x_{jk} = \sum_{k=1}^{K}\sum_{j\in A(k)} x_{kj} = \sum_{k=1}^{K}\pi_k\lambda_k. \tag{78}$$

Thus, the objective function becomes

$$-\frac{1}{\gamma}\sum_{k=1}^{K}\pi_k\lambda_k - 1\big(L - \frac{1}{\gamma}\big)\sum_{k=1}^{K} H_k\big(\sum_{j\in A(k)} x_{jk}\big).$$

Clearly, maximizing this is equivalent to minimizing $\sum_{k=1}^{K} H_k(\sum_{j\in A(k)} x_{jk})$. ■

Proof of Proposition 2. We prove that

$$\max_{i,j}(\tilde{w}_i - \tilde{w}_j) \leq \max_{i,j}(w_i - w_j). \tag{79}$$

The proof of the other assertion follows similarly. Suppose (79) fails, then we must have one of the following two cases: Case i) $\max \tilde{w}_i > \max w_i$; Case ii) $\min \tilde{w}_i < \min w_i$. We assume without loss of generality that $G(x)$ is topologically sorted, cf. Corollary 1. First, consider case i), and let j be such that $\tilde{w}_j = \max \tilde{w}_i$. Consider all predecessors of j in $G(x)$, denoted by $P(j)$. Note that for all $l \in P(j)$, $\tilde{w}_l \geq \tilde{w}_j$. Otherwise, there would not be a directed path from l to j. Moreover, since $\tilde{w}_j = \max \tilde{w}_i$, we conclude that

$$\tilde{w}_l = \max_i \tilde{w}_i \ \text{ for all } \ l \in P(j).$$

Let $m_h \in P(j)$ be a predecessor of j in the highest level (level 0) of the topologically sorted graph $G(x)$. Since m_h is in the highest level, no patient multiple lists to DSA m_h. Hence, $\tilde{\lambda}_{m_h} \leq \lambda_{m_h}$ and therefore, $\tilde{w}_{m_h} \leq w_{m_h}$ from which we conclude that

$$\max \tilde{w}_i = \tilde{w}_{m_h} \leq w_{m_h} \leq \max w_i,$$

which contradicts the defining assumption of Case i).

Next, consider case ii). Let k be such that $\tilde{w}_k = \min \tilde{w}_i$, and consider all successors of k denoted by $S(k)$, and note that $\tilde{w}_l \leq \tilde{w}_k$ for $l \in S(k)$. Otherwise, there would not be a directed path from k to l. Moreover, since $\tilde{w}_k = \min \tilde{w}_i$, we conclude that

$$\tilde{w}_l = \min \tilde{w}_i \quad \text{for all } l \in S(k).$$

Let $m_l \in S(k)$ be a successor of k at the lowest level of the topologically sorted graph $G(x)$. Because m_l is at the lowest level, no patient of DSA m_l multiple lists elsewhere, and hence $\tilde{\lambda}_{m_l} \geq \lambda_{m_l}$ and therefore, $\tilde{w}_{m_l} \geq w_{m_l}$ from which we conclude that

$$\min \tilde{w}_i = \tilde{w}_{m_l} \geq w_{m_l} \geq \min w_i,$$

which contradicts Case ii). Thus, (79) follows. ■

Proof of Proposition 3. Under the assumption that $A(k) = \{1, \ldots, K\}$ for all k, the formulation (10)-(11) reduces to choosing x_{kj} for $j, k = 1, \ldots, K$ so as to

$$\text{Maximize} \quad \sum_{k=1}^{K} \mu_k \ \ln\Big(\sum_{j=1}^{K} x_{jk} + (1 - \pi_k)\lambda_k\Big) \tag{80}$$

subject to

$$\sum_{j=1}^{K} x_{kj} = \pi_k \lambda_k, \quad k = 1, \ldots, K, \tag{81}$$

$$x_{kj} \geq 0, \qquad k, j = 1, \ldots, K. \tag{82}$$

Because our main interest is characterizing the arrival rate $\tilde{\lambda}_k = (1 - \pi_k)\lambda_k + \sum_{j=1}^{K} x_{jk}$ to DSA k under multiple listing and that there are no travel restrictions, we can equivalently study the following formulation: Choose y_k for $k = 1, \ldots, K$ so as to

$$\text{Maximize} \quad \sum_{k=1}^{K} \mu_k \ \ln\big((1 - \pi_k)\lambda_k + y_k\big) \tag{83}$$

subject to

$$\sum_{k=1}^{K} y_k = \sum_{k=1}^{K} \pi_k \lambda_k, \tag{84}$$

$$y_k \geq 0, \tag{85}$$

where y_k denotes the incremental arrival rate to DSA k in addition to its local patients who do not have the option to multiple list. That is, $y_k = \sum_{j=1}^{K} x_{jk}$; and $\tilde{\lambda}_k = (1 - \pi_k)\lambda_k + y_k$.

Formulation (83)-(85) is a straightforward convex optimization problem. It is easy to verify that

$$y_k = (\frac{1}{\eta}\mu_k - (1 - \pi_k)\lambda_k)^+, \qquad k = 1, \ldots, K, \tag{86}$$

where $\eta > 0$ is a Lagrange multiplier such that (84) holds. The solution structure (86) arises in other contexts and is sometimes referred to as the "water-filling" solution; and the parameter η can be interpreted as the desired "target utilization" for the DSAs. (This target is achieved only for some DSAs).

Under the assumptions of the proposition (by temporarily relabeling the DSAs for the proof) we write

$$\lambda_1 \geq \ldots \geq \lambda_K, \tag{87}$$

$$\frac{\lambda_1}{\mu_1} \geq \ldots \geq \frac{\lambda_K}{\mu_K}. \tag{88}$$

Then let $\underline{\mathrm{n}} = \min\{k : y_k > 0\}$ and note that

$$\tilde{\lambda}_k = (1 - \pi_k)\lambda_k > \eta\mu_k \ \text{ for } k < \underline{\mathrm{n}},$$

which relies on (88). Also let

$$\bar{n} = \min\{n \geq \underline{\mathrm{n}} : \frac{\lambda_k}{\mu_k} \leq \eta\},$$

and observe that

$$\tilde{\lambda}_1 + \ldots + \tilde{\lambda}_k \leq \lambda_1 + \ldots + \lambda_k \ \text{ for } \ k < \bar{n}, \tag{89}$$

which follows because $\tilde{\lambda}_k = (1 - \pi_k)\lambda_k$ for $k < \bar{n}$. Clearly,

$$\sum_{k=1}^{K} \tilde{\lambda}_k = \sum_{k=1}^{K} \lambda_k.$$

Rewriting this as

$$\sum_{j=1}^{\bar{n}-1} \tilde{\lambda}_j + \sum_{j=\bar{n}}^{K} \tilde{\lambda}_j = \sum_{j=1}^{\bar{n}-1} \lambda_j + \sum_{j=\bar{n}}^{K} \lambda_j$$

leads to the following:

$$\begin{aligned}\sum_{j=1}^{\bar{n}-1} \tilde{\lambda}_j - \sum_{j=1}^{\bar{n}-1} \lambda_j &= \sum_{j=\bar{n}}^{K} \lambda_j - \sum_{j=\bar{n}}^{K} \tilde{\lambda}_j \\ &\leq \sum_{j=\bar{n}}^{k} \lambda_j - \sum_{j=\bar{n}}^{k} \tilde{\lambda}_j \ \text{ for } k \geq \bar{n},\end{aligned} \tag{90}$$

where the inequality follows because $\eta \geq \frac{\lambda_k}{\mu_k}$ and $\eta = \frac{\tilde{\lambda}_k}{\mu_k}$ for $k \geq n$, i.e. $\lambda_k \leq \eta\mu_k$ and $\tilde{\lambda}_k = \eta\mu_k$ for $k \geq \bar{n}$, from which we conclude that $\lambda_k - \tilde{\lambda}_k \leq 0$ for $k \geq \bar{n}$, and hence, that the inequality in (90) follows.

It follows from (90) that

$$\tilde{\lambda}_1 + \ldots + \tilde{\lambda}_k \leq \lambda_1 + \ldots + \lambda_k \ \text{ for } \ k \geq \bar{n}.$$

Combining this with (89) gives $\tilde{\lambda} < \lambda$ because $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K$ by our labeling convention (87) and that $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_K$, which follows from (86) and (88).

Recall from Propositions 12 and 14 (see Appendix A) that

$$w_k = \frac{1}{\gamma} \ln(\frac{\lambda_k}{\mu_k}) \ \text{ and } \ \tilde{w}_k = \frac{1}{\gamma} \ln(\frac{\tilde{\lambda}_k}{\mu_k}), \ k = 1, \ldots, K,$$

and that

$$\phi_k = \frac{\mu_k}{\lambda_k} \quad \text{and} \quad \tilde{\phi}_k = \frac{\mu_k}{\tilde{\lambda}_k}.$$

Because w_k is a concave function of λ_k and ϕ_k is a convex function of λ_k, it follows from Theorem 4.A.1 of Marshall et al. (2009) (p. 165) that $\tilde{w} \prec^w w$ and $\tilde{\phi} \prec_w \phi$. ■

Proof of Proposition 4. To establish the result, we consider a relaxed version of (10)-(11), whose solution provides an upper bound on the objective of (10)-(11). We verify that the solution characterized in Proposition 4 achieves this upper bound and hence, is optimal. Then (15) is immediate from (14), and Propositions 12 and 14.

The relaxed version of (10)-(11) is given by

$$\text{maximize} \quad \sum_{k=1}^{K} H_k\left(\sum_{j=1}^{K} x_{jk}\right) \tag{91}$$

$$\text{subject to} \quad \sum_{k=1}^{K}\sum_{j=1}^{K} x_{kj} = \sum_{k=1}^{K} \pi_k\lambda_k, \quad k = 1, \ldots, K, \tag{92}$$

Then we introduce a change of variable $y_k = \sum_{j=1}^{K} x_{jk}$ and rewrite (91)-(92) as follows:

$$\text{maximize} \quad \sum_{k=1}^{K} H_k\left(y_k\right) \tag{93}$$

$$\text{subject to} \quad \sum_{k=1}^{K} y_k = \sum_{k=1}^{K} \pi_k\lambda_k. \tag{94}$$

Since (93)-(94) is a concave optimization problem, the following KKT conditions (cf. Boyd and Vandenberghe (2004)) provide the necessary and sufficient conditions for optimality: There exists a lagrange multiplier ν such that ν and any optimal solution $\{y_k : k = 1, \ldots, K\}$ jointly satisfy the following conditions (cf. (2) and (9)):

$$\frac{\mu_k(y_k + (1-\pi_k)\lambda_k)}{y_k + (1-\pi_k)\lambda_k} = \nu, \quad \text{for} \quad k = 1, \ldots, K, \tag{95}$$

$$\sum_{k=1}^{K} y_k = \sum_{k=1}^{K} \pi_k\lambda_k. \tag{96}$$

By definition of $\mu_k(\cdot)$ (see equation (2)), (95) is equivalent to

$$\frac{f\left(\frac{y_k + (1-\pi_k)\lambda_k}{m_k}\right)}{\frac{y_k + (1-\pi_k)\lambda_k}{m_k}} = \nu. \tag{97}$$

Since $h(x) = f(x)/x$ is a decreasing function, it is invertible. Denoting that inverse by $h^{-1}(\cdot)$, (97) implies

$$\frac{y_k + (1-\pi_k)\lambda_k}{m_k} = h^{-1}(\nu) \quad \text{for} \quad k = 1, \ldots, K. \tag{98}$$

Then, equations (96) and (98) imply that $\nu = h(\Lambda/m)$, and substituting this into (98) gives $y_k =$

$\Lambda m_k/m - \lambda_k(1-\pi_k)$. Thus, the objective of the relaxed problem is given by $\sum_{k=1}^K H_k\left(\Lambda\, m_k/m - \lambda_k(1-\pi_k)\right)$. It is straightforward to verify that the rates $\{x_{kj}\}$ characterize this objective when they are substituted in (10).

Then (15) follows from (14) and Propositions 12 and 14, and since $w_k = w_j$ and $\phi_k = \phi_j$ for all k, j, we conclude $GCV_\phi = GCV_w = 0$.

To conclude the proof, we next consider the national organ supply. Let Λ_k denote the total patient arrival rate to the DSA after multiple listing. That is $\Lambda_k = (1-\pi_k)\lambda_k + \sum_{j=1}^k x_{jk}$. It follows from (14) that $\Lambda_k/m_k = \Lambda/m$ for all k. Note that the supply of organs before multiple listing is given by $\sum_k m_k f(\lambda_k/m_k)$, whereas that after multiple listing is

$$\sum_k m_k f\left(\frac{\Lambda_k}{m_k}\right) = \sum_k m_k f\left(\frac{\Lambda}{m}\right) = mf\left(\frac{\Lambda}{m}\right).$$

Then, note that

$$\sum_k m_k f\left(\frac{\lambda_k}{m_k}\right) = m\sum_k \frac{m_k}{m} f\left(\frac{\lambda_k}{m_k}\right) < mf\left(\sum_k \frac{m_k}{m}\frac{\lambda_k}{m_k}\right) = mf\left(\frac{\Lambda}{m}\right),$$

where the inequality follows from Jensen's inequality. Thus, the supply of organs increases after multiple listing. ■

Proof of Proposition 5. Since $W_{kj} = \min(W_k, W_j)$ and W_k, W_j and X are modeled as mutually independent random variables, we conclude that

$$\mathbb{P}(\min(X, W_{kj}) > x) = e^{-\gamma x}\ \bar{\Phi}\left(\frac{x - w_k^n}{\sigma_k^n}\right)\bar{\Phi}\left(\frac{x - w_j^n}{\sigma_j^n}\right).$$

Then assuming $W_{kj} > 0$ with probability one (which is a reasonable approximation since $\mathbb{P}(W < 0)$ is negligible in our setting) it follows that

$$\mathbb{E}[\min\{X, W_{kj}\}] = \int_0^\infty \mathbb{P}(\min(X, W_{kj}) > x)dx = \int_0^\infty e^{-\gamma x}\ \bar{\Phi}\left(\frac{x - w_k^n}{\sigma_k^n}\right)\bar{\Phi}\left(\frac{x - w_j^n}{\sigma_j^n}\right)dx. \quad (99)$$

Also note that $\mathbb{P}(W_{kj} < X) = 1 - \mathbb{P}(W_{kj} \geq X)$, where

$$\mathbb{P}(W_{kj} > X) = \int_0^\infty \gamma\, e^{-\gamma x}\bar{\Phi}\left(\frac{x - w_k^n}{\sigma_k^n}\right)\bar{\Phi}\left(\frac{x - w_j^n}{\sigma_j^n}\right)dx. \quad (100)$$

Recall that the patient's life expectancy is given by $\mathbb{E}[\min\{X, W_{kj}\}] + L\,\mathbb{P}(W_{kj} < X)$. Substituting (99)-(100) into this gives

$$\int_0^\infty e^{-\gamma x}\bar{\Phi}\left(\frac{x - w_k^n}{\sigma_k^n}\right)\bar{\Phi}\left(\frac{x - w_j^n}{\sigma_j^n}\right)dx + L - L\int_0^\infty \gamma\, e^{-\gamma x}\bar{\Phi}\left(\frac{x - w_k^n}{\sigma_k^n}\right)\bar{\Phi}\left(\frac{x - w_j^n}{\sigma_j^n}\right)dx.$$

Then rearranging the terms and using (100) gives the result. ■

Proof of Proposition 6. The result follows from Theorem 1 of Mas-Colell (1984). In Mas-Colell's notation, let $A = \{1, \ldots, K\}$ and $u_A = \{\widetilde{u}_1, \ldots, \widetilde{u}_K\}$, where

$$\widetilde{u}_k(j; q^n) = \begin{cases} u_k(j; q^n) & \text{if } j \in A(k)\backslash\{k\}, \\ -1 & \text{otherwise.} \end{cases}$$

Then consider the game given by $\widetilde{\mu}$ such that $\widetilde{\mu}(\widetilde{u}_k) = \lambda_k$. Note that the continuity of $u(j; q^n) \in \mathcal{U}_A$ in q^n is immediate from that of w_l^n in q for all l. The set of strategies is compact since we assume $|q^n| \leq C^n < \infty$. Then Theorem 1 of Mas-Colell (1984) yields the existence of a pure strategy Nash equilibria. From this (by integrating frequencies of actions for each DSA as probabilities) we conclude the existence of a symmetric mixed-strategy equilibrium. ∎

Proof of Proposition 7. In what follows, we characterize the net inflow θ_k to DSA k for all k. Evidently, once θ_k's are given, one can characterize $\tilde{x}_{kj}$ for $j \in A(k)$ consistent with them, which, in turn, pin down the patients' strategy p^n. To characterize θ_k, we will observe in essence that waiting times at certain subsets[43] of DSAs must be identical. Otherwise patients find it attractive to deviate. Building on this observation, we derive a set of equations that characterize $\theta_k, k = 1, \ldots, K$. In doing so, we use (36)-(37) to approximate patients' utility.

As a preliminary to our analysis, first note that

$$\sum_{j \in A(k), j \neq k} \tilde{x}_{kj} \leq \sqrt{n}(\pi_k \lambda_k - \sum_{j \in A(k)} \bar{x}_{kj}).$$

This follows because of the constraints on the fraction of patients multiple listing, i.e. π_k for DSA k. In particular, if

$$\sum_{j \in A(k), j \neq k} \bar{x}_{kj} = \pi_k \lambda_k, \tag{101}$$

i.e. everyone in DSA k, who can multiple list, has multiple listed in the equilibrium of the selfish routing problem, we must have

$$\sum_{j \in A(k), j \neq k} \tilde{x}_{kj} = 0$$

because the constraint on the number of patients who can multiple list is reached. This implies that

$$\theta_k = \sum_{j \in A(k), j \neq k} \tilde{x}_{jk} \geq 0. \tag{102}$$

Secondly, we deduce from the approximation (36)-(37) of $\mathbb{P}(X > W_{kj})$ that a (positive) second-order distortion of the first-order flow $\bar{x}_{kj}$ may happen only if $\bar{w}_i = \bar{w}_j$. Otherwise, consider two possibilities: On the one hand, if $\bar{w}_j < \bar{w}_k$, then it must be that the number of DSA k patients multiple listing is already maxed out in the selfish routing game (i.e. in the first order) and (101) holds. Thus, $\tilde{x}_{kj} = 0$. On the other hand, if $\bar{w}_k < \bar{w}_j$, then by (36), DSA k patients do not gain anything (up to first two orders) by multiple listing at DSA j. Hence $\tilde{x}_{kj} = 0$.

Given these observations, we focus on the groups of DSAs with identical fluid-scale waiting times. Let $\tilde{w}_1, \ldots, \tilde{w}_m$ be the distinct waiting times in equilibrium of the selfish routing problem. Then partition the set of DSAs $\{1, \ldots, K\}$ with respect to these, say $\mathcal{A}_1, \ldots, \mathcal{A}_m$. It suffices to

[43] These subsets of DSAs depend on the sets $A(k)$ $(k = 1, \ldots, K)$ and the equilibrium flows for the selfish routing problem.

study each of $\mathcal{A}_1, \ldots, \mathcal{A}_m$ in isolation. Define

$$\mathcal{A}_j^+ = \{j \in \mathcal{A}_j : \theta_j \geq 0\}, \quad j = 1, \ldots, m.$$

All patients in these DSAs, who can multiple list, multiple listed in the (first-order) selfish routing equilibrium. Consider each $\mathcal{A}_j \backslash \mathcal{A}_j^+$ (for $j = 1, \ldots, m$) and the associated graph such that for $i, k \in \mathcal{A}_j \backslash \mathcal{A}_j^+$ there is an (undirected) edge connecting DSAs i and k if $i \in A(k)$. Then let $\mathcal{G}_1^{(j)}, \ldots, \mathcal{G}_{m_j}^{(j)}$ denote its connected (distinct) subgraphs such that $\mathcal{G}_k^{(j)}$ and $\mathcal{G}_l^{(j)}$ have no nodes in common[44] if $k \neq l$ and partition $\mathcal{A}_j \backslash \mathcal{A}_j^+$. Also define $\mathcal{B}_i^{(j)} \subset \mathcal{A}_j^+$ as follows[45]:

$$\mathcal{B}_i^{(j)} = \{l \in \mathcal{A}_j^+ : l \in A(k) \text{ for some } k \in \mathcal{G}_i^{(j)}\}, \quad i = 1, \ldots, m_j.$$

Lastly, define

$$\beta_l = \Delta\sqrt{\gamma\mu_l} + \phi_l \frac{\theta_l}{\sqrt{\gamma\mu_l}} \quad \text{for all } l. \tag{103}$$

Note that β_l will be used as a proxy for congestion at DSA l by the approximation (37). In particular, the higher β_l, the higher the waiting time at DSA l.

To characterize the equilibrium net inflow rates we look for a solution $\psi_i^{(j)}$ $(i = 1, \ldots, m_j)$ and θ_l $(l \in \mathcal{A}_j \backslash \mathcal{A}_j^+)$ to the following set of equations for each $j = 1, \ldots, m$:

$$\beta_l = \psi_i^{(j)} \text{ for every node } l \text{ of } \mathcal{G}_i^{(j)}, \ i = 1, \ldots, m_j, \tag{104}$$

$$\psi_i^{(j)} \leq \beta_l \text{ for every node } l \text{ of } \mathcal{B}_i^{(j)}, \ i = 1, \ldots, m_j, \tag{105}$$

$$\sum_{i \in \mathcal{A}_j} \theta_l = 0, \tag{106}$$

$$\theta_l \geq 0, \ l \in \mathcal{A}_j^+. \tag{107}$$

The constraint (104) states that the congestion in the nodes of $\mathcal{G}_i^{(j)}$ should be identical. Otherwise, if there was a node with lower congestion some additional patients would deviate and multiple list there. Note by costruction that each DSA in $\mathcal{G}_i^{(j)}$ have some patients who can but have not multiple listed. So such distortions to the rates is feasible. If the second constraint is violated patients in DSAs that constitute the nodes of $\mathcal{G}_i^{(j)}$ can deviate and multiple list to DSA l. Note that the reverse inequality is not needed because all patients in DSA $l \in \mathcal{B}_i^{(j)}$, who can multiple list, have already done so. The third equation is simply a flow conversation equation. The last equation simply restates equation (102). Equations (104)-(107) ensure that the resulting θ's correspond to an equilibrium.

Next, we argue that (104)-(107) admit a solution. Combining (103) and (104) gives

$$\theta_l = \frac{\psi_i^{(j)}\sqrt{\gamma}\mu_l - \Delta\gamma\mu_l}{\phi_l} \quad \text{for } l \in \mathcal{G}_i^{(j)}, \ i = 1, \ldots, m_j. \tag{108}$$

[44] Clearly, $\mathcal{G}_k^{(j)}$ and $\mathcal{G}_l^{(j)}$ are not connected to each other.

[45] A graphical illustration of these definitions is given in Figure 7.

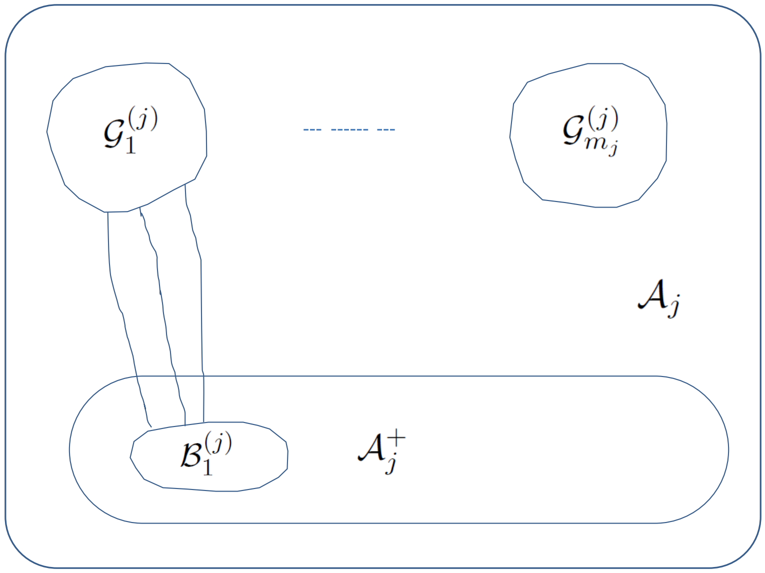


Figure 7: A Graphical illustration of $\mathcal{G}_i^{(j)}$, $\mathcal{B}_i^{(j)}$, $\mathcal{A}_j^+$.

Similarly, combining (103) and (105) gives

$$\psi_i^{(j)} \le \beta_l = \Delta\sqrt{\gamma\mu_l} + \phi\frac{\theta_l}{\sqrt{\gamma\mu_l}} \qquad \forall l \in \mathcal{B}_i^{(j)}. \tag{109}$$

Substituting (108) into (106) gives

$$\sum_{i=1}^{m_j}\sum_{l\in\mathcal{G}_i^{(j)}} \frac{\psi_i^{(j)}\sqrt{\gamma}\mu_l - \Delta\gamma\mu_l}{\phi_l} + \sum_{l\in\mathcal{A}_j^+}\theta_l = 0. \tag{110}$$

Note that if $\mathcal{A}_j^+ = \phi$, then we can choose $\psi_i^{(j)}$ so that (110) holds, and all other constraints are satisfied. If $\mathcal{A}_j^+ \neq \phi$, then the second term on the left hand side of (110) is nonnegative, i.e. $\sum_{l\in\mathcal{A}_j^+}\theta_l \ge 0$ by (107), but we can find $\psi_i^{(j)}$ to solve (110) by lowering $\psi_i^{(j)}$ if necessary (see (105)). Therefore, (104)-(107) admit a solution. Clearly, there exists a solution such that $\theta_l = O(1)$ for all l.

To conclude the proof, note that

$$w_k^d = \bar{w}_k + \frac{1}{\gamma\sqrt{n}}\frac{\theta_k}{\bar{\Lambda}_k}$$

so that $w_k^d = w_k^f + O(1/\sqrt{n})$, because $w_k^f = \bar{w}_k$. Similarly,

$$\phi_k^d = \frac{\mu_k}{\bar{\Lambda}_k + \frac{1}{\sqrt{n}}\theta_k} \quad \text{and} \quad \phi_k^f = \frac{\mu_k}{\bar{\Lambda}_k}.$$

Thus, $\phi_k^d - \phi_k^f = O(1/\sqrt{n})$.

Lastly, we consider GCV_i^d for $i = \phi, w$. To that end, recall that the overall average time to transplant in the selfish routing formulation of Section 3, denoted by $\bar{w}$, is given by

$$\bar{w} = \sum_{k=1}^{K} \frac{\mu_k}{M} \bar{w}_k,$$

whereas the corresponding quantity in the diffusion model, denoted by $\widetilde{w}^n$, is given below:

$$\begin{aligned}
\widetilde{w}^n &= \sum_{k=1}^{K} \frac{\mu_k}{M} w_k^n = \sum_{k=1}^{K} \frac{\mu_k}{M} \big(\bar{w}_k + \frac{1}{\gamma\sqrt{n}} \frac{\theta_k}{\bar{\Lambda}_k}\big) \\
&= \sum_{k=1}^{K} \frac{\mu_k}{M} \bar{w}_k + \sum_{k=1}^{K} \frac{\mu_k}{M} \frac{\theta_k}{\bar{\Lambda}_k} \frac{1}{\gamma\sqrt{n}} \\
&= \bar{w} + \sum_{k=1}^{K} \frac{\mu_k}{\bar{\Lambda}_k} \frac{\theta_k}{M} \frac{1}{\gamma\sqrt{n}} \\
&= \bar{w} + \sum_{k=1}^{K} \phi_k \frac{\theta_k}{M} \frac{1}{\gamma\sqrt{n}} \\
&= \bar{w} + \frac{1}{\sqrt{n}\gamma M} \sum_{k=1}^{K} \phi_k \, \theta_k \\
&= \bar{w} + \frac{1}{\sqrt{n}\gamma M} \sum_{i=1}^{M} \sum_{j \in \mathcal{A}_i} \phi_i \, \theta_j \\
&= \bar{w} + \frac{1}{\sqrt{n}\gamma M} \sum_{i=1}^{M} \phi_i \sum_{j \in \mathcal{A}_i} \theta_j \\
&= \bar{w},
\end{aligned}$$

where the last equality follows because $\sum_{j \in \mathcal{A}_i} \theta_j = 0$ by (106). Then note that

$$GCV_w^d = \frac{\sqrt{\sum_k \mu_k^2 (w_k^n - \widetilde{w})^2 / \sum_k \mu_k^2}}{\widetilde{w}} = \frac{\sqrt{\sum_k \mu_k^2 (w_k^n - \bar{w})^2 / \sum_k \mu_k^2}}{\bar{w}}.$$

$$GCV_w^f = \frac{\sqrt{\sum_k \mu_k^2 (\bar{w}_k - \bar{w})^2}}{\bar{w}}.$$

Thus, it follows that

$$\begin{aligned}
GCV_w^d - GCV_w^f &= \frac{\sqrt{\sum_k \mu_k^2 (w_k^n - \bar{w})^2} - \sqrt{\sum_k \mu_k^2 (\bar{w}_k - \bar{w})^2}}{\bar{w}\sqrt{\sum_k \mu_k^2}}. \\
&= \frac{\sum_k \mu_k^2 (w_k^n - \bar{w})^2 - \sum_k \mu_k^2 (\bar{w}_k - \bar{w})^2}{\bar{w}\sqrt{\sum_k \mu_k^2}\left(\sqrt{\sum_k \mu_k^2 (w_k^n - \bar{w})^2} + \sqrt{\sum_k \mu_k^2 (\bar{w}_k - \bar{w})^2}\right)}.
\end{aligned}$$

Therefore, it suffices to show that the numerator is $O(1/\sqrt{n})$:

$$\sum_k \mu_k^2\big[(w_k^n - \bar{w})^2 - (\bar{w}_k - \bar{w})^2\big] = \sum_k \mu_k^2\big[(w_k^n + \bar{w}_k - 2\bar{w})(w_k^n - \bar{w}_k)\big] \leq O(1/\sqrt{n}),$$

because $(w_k^n + \bar{w}_k - 2\bar{w}) \leq O(1)$ and $(w_k^n - \bar{w}_k) = O(1/\sqrt{n})$.

Finally, we prove that $GCV_\phi^d - GCV_\phi^f \leq O(1/\sqrt{n})$. Note that

$$GCV_\phi^f = \frac{\sqrt{\sum_k \bar{\Lambda}_k^2(\bar{\phi}_k - \bar{\phi})^2/\sum_k \bar{\Lambda}_k^2}}{\bar{\phi}},$$

$$GCV_\phi^d = \frac{\sqrt{\sum_k \bar{\Lambda}_k^2(\tilde{\phi}_k - \bar{\phi})^2/\sum_k \bar{\Lambda}_k^2}}{\bar{\phi}},$$

where $\bar{\phi}_k = \mu_k/\bar{\Lambda}_k$ and $\tilde{\phi}_k = \mu_k/(\bar{\Lambda}_k + \theta_k\sqrt{n})$, and

$$\bar{\phi} = \sum_{k=1}^K \frac{\lambda_k}{\Lambda}\frac{\mu_k}{\lambda_k} = \frac{\mu}{\Lambda}.$$

Also note that

$$GCV_\phi^f - GCV_\phi^d = \frac{\sqrt{\sum_k \bar{\Lambda}_k^2(\bar{\phi}_k - \bar{\phi})^2} - \sqrt{\sum_k \bar{\Lambda}_k^2(\tilde{\phi}_k - \bar{\phi})^2}}{\bar{\phi}\sqrt{\sum_k \bar{\Lambda}_k^2}}.$$

Therefore, it suffices to show that the numerator is of $O(1/\sqrt{n})$:

$$\sqrt{\sum_k \bar{\Lambda}_k^2(\tilde{\phi}_k - \bar{\phi})^2} - \sqrt{\sum_k \bar{\Lambda}_k^2(\bar{\phi}_k - \bar{\phi})^2} = \frac{\sum_k \bar{\Lambda}_k^2\big[(\tilde{\phi}_k - \bar{\phi})^2 - (\bar{\phi}_k - \bar{\phi})^2\big]}{\sqrt{\sum_k \bar{\Lambda}_k^2(\tilde{\phi}_k - \bar{\phi})^2} + \sqrt{\sum_k \bar{\Lambda}_k^2(\bar{\phi}_k - \bar{\phi})^2}}.$$

Once again, it suffices to show that the numerator is of $O(1/\sqrt{n})$:

$$\sum_k \bar{\Lambda}_k^2\big[(\tilde{\phi}_k - \bar{\phi})^2 - (\bar{\phi}_k - \bar{\phi})^2\big] = \sum_k \bar{\Lambda}_k^2\big[(\tilde{\phi}_k + \bar{\phi}_k - 2\bar{\phi})(\tilde{\phi}_k - \bar{\phi}_k)\big].$$

Note that $\tilde{\phi}_k + \bar{\phi}_k - 2\bar{\phi} = O(1)$. Thus, it suffices to show that $\tilde{\phi}_k - \bar{\phi}_k = O(1/\sqrt{n})$, which was done above. ■

Proof of Proposition 8. In equilibrium, patients in DSA k must be indifferent among their choices in $A(k)$. By virtue of (37), this implies that for all $j \in A(k), j \neq k$,

$$\sum_{l=k,j}\left(\Delta\sqrt{\gamma\mu_l} + \phi^*\ \frac{\theta_l^*}{\sqrt{\gamma\mu_l}}\right)$$

Should be the same, which, in turn, is equivalent to having

$$\Delta\sqrt{\gamma\mu_l} + \phi^*\ \frac{\theta_l^*}{\sqrt{\gamma\mu_l}} = \tilde{\beta}\ \text{ for all } l \neq k$$

for some constant $\tilde{\beta}$. Since this can be repeated for all DSAs and patients in different DSAs have overlapping options, we conclude that

$$\Delta\sqrt{\gamma\mu_l} + \phi^* \, \frac{\theta_l^*}{\sqrt{\gamma\mu_l}} = \beta^* \text{ for all } l.$$

Combining this with $\sum_{l=1}^K \theta_l^* = 0$ yields $\tilde{\beta} = \gamma M\Delta / \sum_{l=1}^K \sqrt{\gamma\mu_l}$, from which (41) follows. Since all equilibria leads to the same θ_l^* for all l, they also lead to the same set of virtual waiting times at the various DSAs. Then by virtue of (37), we also conclude that every equilibrium leads to the same utility for the (multiple-listing) patients. ■

Proof of Proposition 9. First, we show that $\theta_1 < 0$. Suppose not, i.e. $\theta_1^* \geq 0$. Let $\mu_2 = \alpha\,\mu_1$ where $\alpha < 1$. Then $\theta_2^* = \sqrt{\mu_2}/\hat{M} - \mu_2/M = \sqrt{\alpha}\sqrt{\mu_1}/\hat{M} - \alpha\mu_1/M$, where $\hat{M} = \sum_{j=1}^K \sqrt{\mu_j}$, $M = \sum_{j=1}^K \mu_j$. Then

$$\theta_2^* = \frac{\sqrt{\alpha}\sqrt{\mu_1}}{\hat{M}} - \frac{\sqrt{\alpha}\mu_1}{M} + \frac{\sqrt{\alpha}\mu_1}{M} - \frac{\alpha\mu_1}{M} = \sqrt{\alpha}\theta_1^* + (\sqrt{\alpha} - \alpha)\frac{\mu_1}{M} > 0.$$

Then we conclude by induction that $\theta_l^* > 0$ for all $l \geq 2$, which contradicts that $\sum_{l=1}^K \theta_l^* = 0$. Therefore, we must have $\theta_l^* < 0$. The preceding argument also shows that if $\theta_j^* \geq 0$, then $\theta_k^* > 0$ for all $k \geq j$. Then defining l^* to be the smallest j such that $\theta_j^* \geq 0$ yields $\theta_j^* < 0$ for all $j < l^*$, and (by the preceding argument) $\theta_j^* > 0$ for all $j > l^*$.

To conclude the proof, it suffices to show that $\theta_1^* < \theta_2^* < \ldots < \theta_{l^*-1}^* < 0$. Note that $\theta_{l^*-1}^* < 0$ by definition of l^*. Then we let $\mu_{l-2} = \tilde{\alpha}\,\mu_{l-1}$, where $\tilde{\alpha} > 1$. Then

$$\theta_{l^*-2}^* = \sqrt{\tilde{\alpha}}\;\theta_{l^*-1}^* + (\sqrt{\alpha} - \alpha)\;\frac{\mu_{l^*-1}}{M} < 0,$$

since $\theta_{l^*-1}^* < 0$ and $\sqrt{\tilde{\alpha}} - \tilde{\alpha} < 0$. Then the result follows by induction. ■

Proof of Proposition 10. Let $j < k$. To show that $w_j < w_k$, it suffices to show that $\theta_j/\mu_j < \theta_k/\mu_k$ because of (40). Also, it follows from Proposition 8 that

$$\frac{\theta_l}{\mu_l} = \Delta\gamma \wedge \left(\frac{1}{\sqrt{\mu l}}\;\frac{1}{\sum_{i=1}^K \sqrt{\mu_i}} - \frac{1}{\sum_{i=1}^K \mu_i} \right), \quad l = j, k. \tag{111}$$

Since $j < k$, $\mu_j > \mu_k$ (by (42)), and thus it follows from (111) that $\theta_j/\mu_j < \theta_k/\mu_k$. ■

Proof of Proposition 11. Note that

$$\bar{w} = \sum_{k=1}^K \frac{\mu_k}{M}\bar{w}_k = \sum_{k=1}^K \frac{\mu_k}{M}\left(w^* + \frac{1}{\gamma\sqrt{n}}\;\frac{M}{\Lambda}\frac{\theta_k^*}{\mu_k} \right) = w^*, \tag{112}$$

Then,

$$GCV_w^d = \frac{\sqrt{\sum_{k=1}^K \mu_k^2 \left(\frac{1}{\gamma^2 n}\;\frac{M^2}{\Lambda^2}\frac{(\theta_k^*)^2}{\mu_k^2} \right) / \sum_k \mu_k^2}}{w^*} = \frac{1}{\gamma\sqrt{n}}\;\frac{1}{w*}\;\frac{M}{\Lambda}\sqrt{\frac{\sum_k (\theta_k^*)^2}{\sum_k \mu_k^2}}.$$

Substituting θ_k^* into this and rearranging the terms gives the result. Derivation of GCV_ϕ^d follows similarly. ■

D Auxiliary Graph Theoretic Results

Proposition 15 *There exists an equilibrium x of the selfish routing problem (10)-(11) such that $G(x)$ contains no directed cycles.*

Proof of Proposition 15. Suppose not. Then fix any equilibrium of the selfish routing problem, say x. Suppose x^0 contains m directed cycles. Consider one such cycle, e.g. the longest one, say $\mathcal{C} = \{i_1, i_2, \ldots, i_l\}$ (with $i_1 = i_l$). Since $\mathcal{C}$ is a cycle we conclude that

$$w_{i1} \geq w_{i2} \geq \ldots \geq w_{il-1} \geq w_{il} = w_{i1}.$$

That is, $w_{i1} = w_{i2} = \ldots = w_{il}$. Then we define a new flow x^1 by updating the flows on the cycle while keeping all other flows the same:

$$x^1_{ik,ik+1} = x^0_{ik,ik+1} - \min_j \{x^0_{ij}, i_{j+1}\}, \quad k = 0, 1, \ldots, l-1.$$

The new flow x^1 is an equilibrium for the selfish routing problem too, because the deleted flows on the cycle do not change the time-to-transplant w at the nodes on the cycle (or at any other node in the graph), and the patients corresponding to the deleted flows are indifferent between staying put or multiple listing. Clearly, $G(x^1)$ does not have any new cycles. In particular, it has at most $m-1$ directed cycles. Iterating on x^1 in the same manner until all cycles are eliminated yields an equilibrium x to the selfish routing problem (in at most $m-1$ steps) such that $G(x)$ contains no directed cycles. ■

The following corollary is immediate from Proposition 15 and has important consequences; see Section 22.4 of Cormen et al. (2009) for details of the relevant linear-time topological sorting algorithms.

Corollary 1 *If x is a feasible flow for the selfish routing problem such that $G(x)$ has no directed cycles, then $G(x)$ is topologically sortable.*

E An Alternative Approximation for Proving Proposition 10

We derive an alternative approximation using a result of Clark (1961) who provides formulas for the mean and variance of the minimum of two normally distributed variables[46]. The formula for the mean allows us to study the patients' multiple listing game.

Consider a patient living in DSA k and choosing DSA l to multiple list. Recall that his utility is given by $\mathbb{P}(X > W)$ where $W = \min(W_k, W_l)$ and W_k and W_l are the virtual waiting times at DSAs k and l, respectively. Also recall that

$$w_i^n = w^* + \frac{1}{\gamma\sqrt{n}}\, \phi^* \,\frac{\theta_i}{\mu_i} \quad \text{and} \quad \sigma_i = \frac{1}{\sqrt{n\gamma\mu_i}} \quad \text{for} \;\; i = k, l.$$

Hence, we write $W = w^* + Z/\sqrt{n}$ where $Z = \min(Z_k, Z_l)$ and Z_i is normally distributed with mean z_i and standard deviation $\hat{\sigma}_i$:

$$z_i = \frac{1}{\gamma}\, \phi^* \,\frac{\theta_i}{\mu_i} \quad \text{and} \quad \hat{\sigma}_i = \frac{1}{\sqrt{\gamma\mu_i}} \quad \text{for} \;\; i = k, l.$$

[46] Also see Kella (1986) and Cain (1994) for the moment generating function of the maximum and minimum, respectively, of two normal random variables.

Thus, we conclude that

$$\mathbb{P}(X > W) = e^{-\gamma w^*}[e^{-\gamma Z/\sqrt{n}}] \simeq e^{-\gamma w^*}\mathbb{E}\left[1 - \frac{\gamma Z}{\sqrt{n}}\right]$$
$$= \frac{1}{\phi} - \frac{1}{\phi}\,\frac{\gamma}{\sqrt{n}}\,\mathbb{E}[Z].$$

The patient wishes to maximize $\mathbb{P}(X > W)$, which corresponds to minimizing $\mathbb{E}[Z] = \mathbb{E}[\min(Z_k, Z_l)]$. Clark (1961) proves that

$$\mathbb{E}[Z] = z_k\Phi\left(\frac{z_l - z_k}{\sigma}\right) + z_l\Phi\left(\frac{z_k - z_l}{\sigma}\right) - \sigma\phi\left(\frac{z_l - z_k}{\sigma}\right),$$

where $\sigma = \sqrt{\hat{\sigma}_l^2 + \hat{\sigma}_k^2}$. Rearranging the terms gives

$$\mathbb{E}[Z] = z_k - (z_k - z_l)\Phi\left(\frac{z_k - z_l}{\sigma}\right) - \sigma\phi\left(\frac{z_k - z_l}{\sigma}\right).$$

Defining $\psi(x) = x\Phi(x) + \phi(x)$ for $x \in \mathbb{R}$, the patient's objective is to choose DSA l to maximize $\sigma\psi\left(\frac{z_k - z_l}{\sigma}\right)$. The following lemma is helpful in understanding the patient's objective.

Lemma 1 *The function ψ is convex increasing, and $\sigma\psi(y/\sigma)$ is increasing in σ.*

Proof. This lemma shows that when the patient chooses between two DSAs with equal mean delays, then he will prefer the one with larger variance. This intuition leads to an "alternative proof" of Proposition 10. ■

An alternative proof of Proposition 10. Note that in equilibrium the patient living in DSA k must be indifferent between all other DSAs. Thus, if $\mu_j > \mu_l$, then we must have $z_j < z_l$ for $j, l \neq k$, cf. Lemma 1. Since we can repeat this argument for all DSAs k, we conclude that $z_1 < z_2 < \ldots < z_K$. The result follows from this because $w_l^n = w^* + z_l/\sqrt{n}$ for $l = 1, \ldots, K$.

F An equilibrium of the Game in Section 3

We construct an equilibrium of the selfish routing (or multiple listing) game in Section 3. In this equilibrium, the multiple listing rates are minimal. To that end, for each $k = 1, \ldots, K$, let $x_{kk} = \min\{\Lambda\, m_k/m - (1-\pi_k)\lambda_k, \pi_k\lambda_k\}$, and set $\alpha_k := \pi_k\lambda_k - x_{kk}$, and $\beta_k := \Lambda\, m_k/m - (1-\pi_k)\lambda_k - x_{kk}$. The following algorithm determines the remaining flows x_{kj} with $k \neq j$.

for $k = 1 \to K$ **do**
  for $j = 1 \to K \;\; (j \neq k)$ **do**
    if $\beta_k > 0$ **then**
      if $\alpha_j > 0$ **then**
        $x_{jk} = \min\{\beta_k, \alpha_j\},\; \beta_k = \beta_k - x_{jk},\; \alpha_j = \alpha_j - x_{jk};$
      end if
    end if
  end for
end for

Because $\sum_{k=1}^{K}\sum_{j=1}^{K} x_{kj} = \sum_{k=1}^{K}\sum_{j=1}^{K} x_{jk}$, this algorithm will terminate with $\alpha_k = \beta_k = 0$. Moreover, for $k = 1, \ldots, K$, $\beta_k = 0$ means $\sum_{j=1}^{K} x_{jk} = \Lambda\frac{m_k}{m} - (1-\Pi_k)\lambda_k$, and $\alpha_k = 0$ means $\sum_{j=1}^{K} x_{kj} = \pi_k\lambda_k$. Also note that the algorithm assigns flow rates such that $x_{jk} \geq 0$ for all j, k. Thus, these flow rates constitute an equilibrium.

G Supplementary Numerical Results

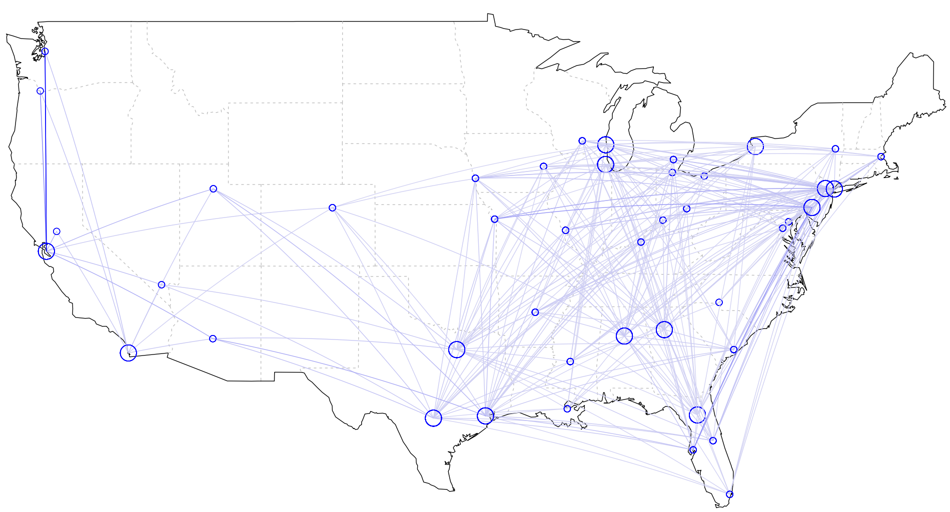


Figure 8: Equilibrium multiple listing rates for blood type B on a US map for the case of $\pi = 0.25$ and $d = 1200$. Each line corresponds to a particular flow; and for each flow, the large circle indicates the fly-out DSA whereas the small circle corresponds to the associated y-in DSA.

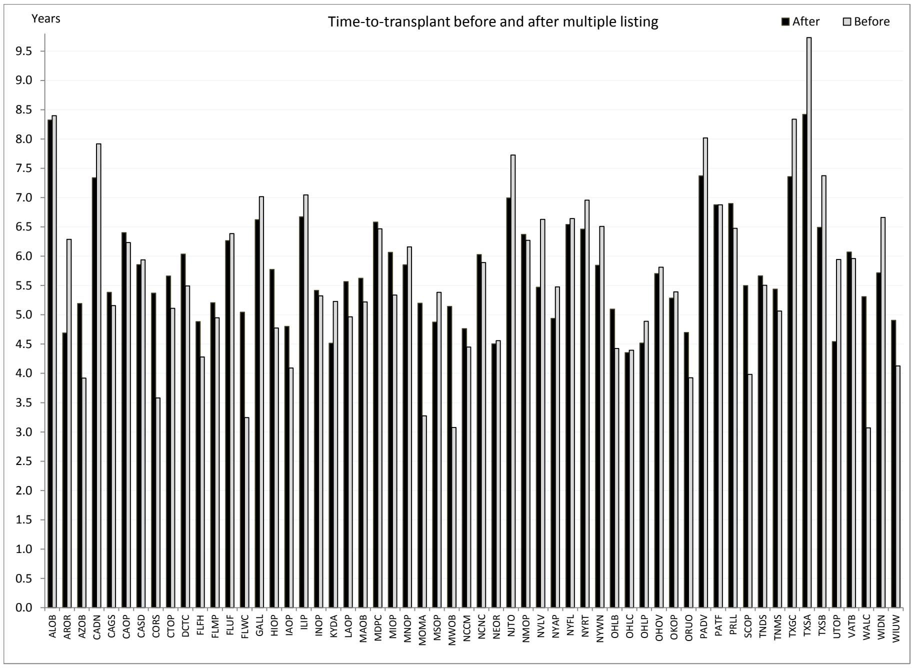


Figure 9: Waiting times at 58 DSAs (based on patient and organ arrival data from 2009–2013 provided by UNOS) for blood type B, before and after multiple listing for $d = 1200$ and $\pi = 0.25$.

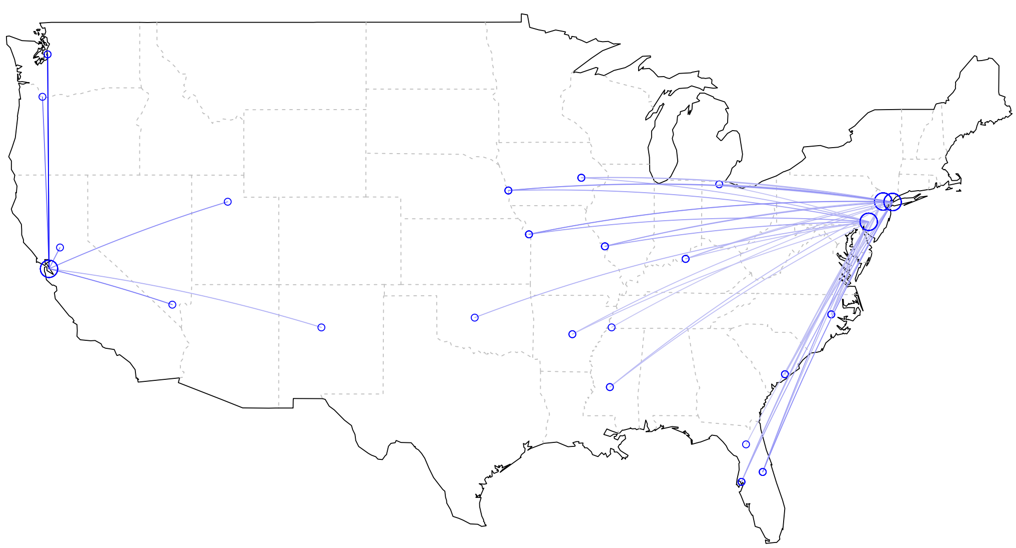


Figure 10: Equilibrium multiple listing rates for the partial solution for blood type O patients ($\pi = 0.25$ and $d = 1200$). In this illustrative scenario, OrganJet offers its services only in certain DSAs: Oakland, CA, New York, NY, New Providence, NJ and Philadelphia, PA.

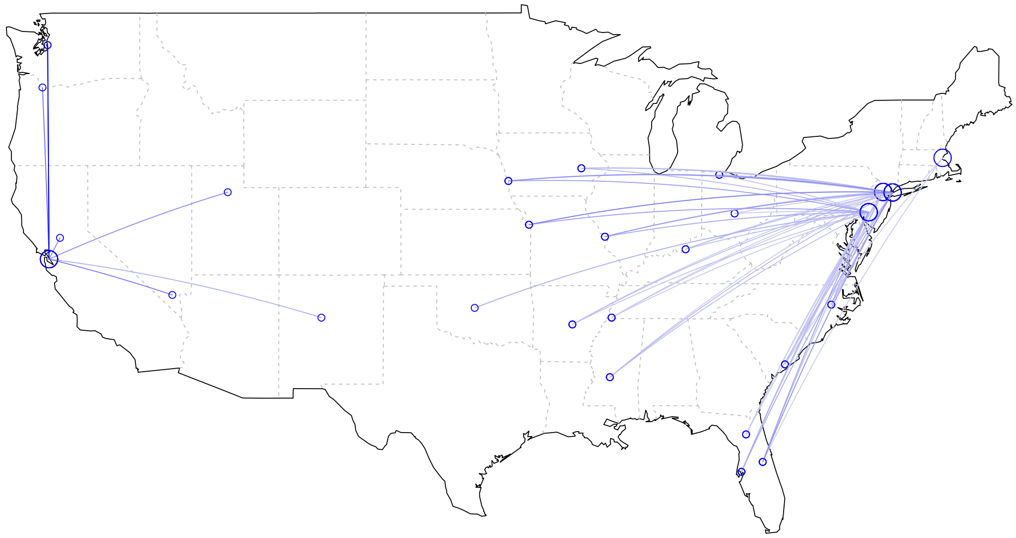


Figure 11: Equilibrium multiple listing rates for the partial solution for blood type O patients ($\pi = 0.25$ and $d = 1200$). In this illustrative scenario, OrganJet offers its services only in certain DSAs: Oakland, CA, New York, NY, New Providence, NJ, Philadelphia, PA and Waltham, MA.

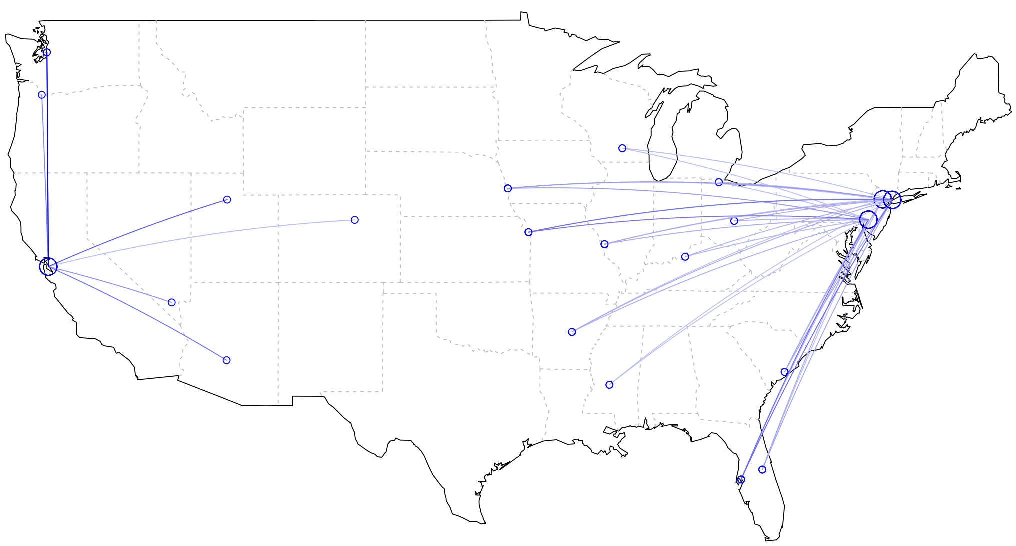


Figure 12: Equilibrium multiple listing rates for the partial solution for blood type B patients ($\pi = 0.25$ and $d = 1200$). In this illustrative scenario, OrganJet offers its services only in certain DSAs: Oakland, CA, New York, NY, New Providence, NJ and Philadelphia, PA.

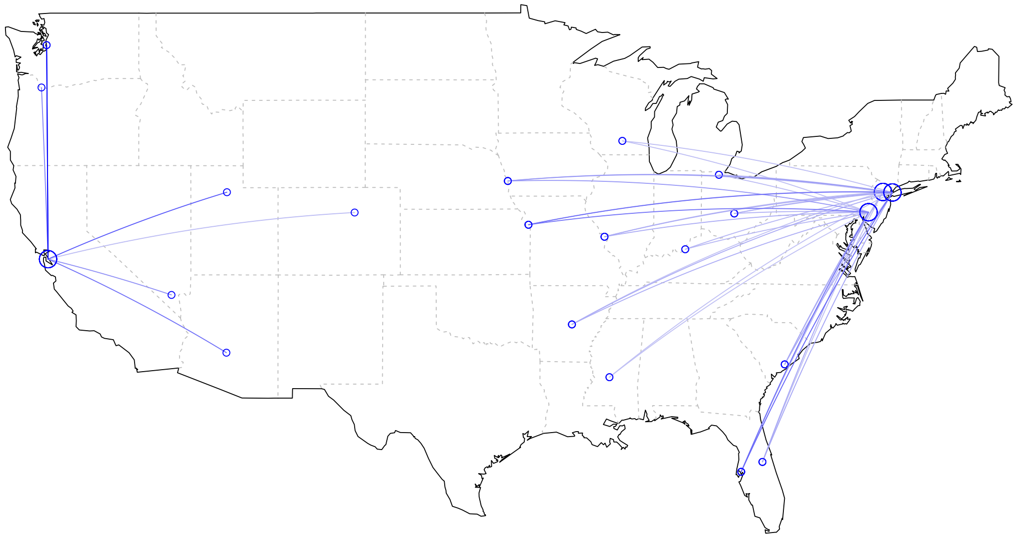


Figure 13: Equilibrium multiple listing rates for the partial solution for blood type B patients ($\pi = 0.25$ and $d = 1200$). In this illustrative scenario, OrganJet offers its services only in certain DSAs: Oakland, CA, New York, NY, New Providence, NJ, Philadelphia, PA and Waltham, MA.

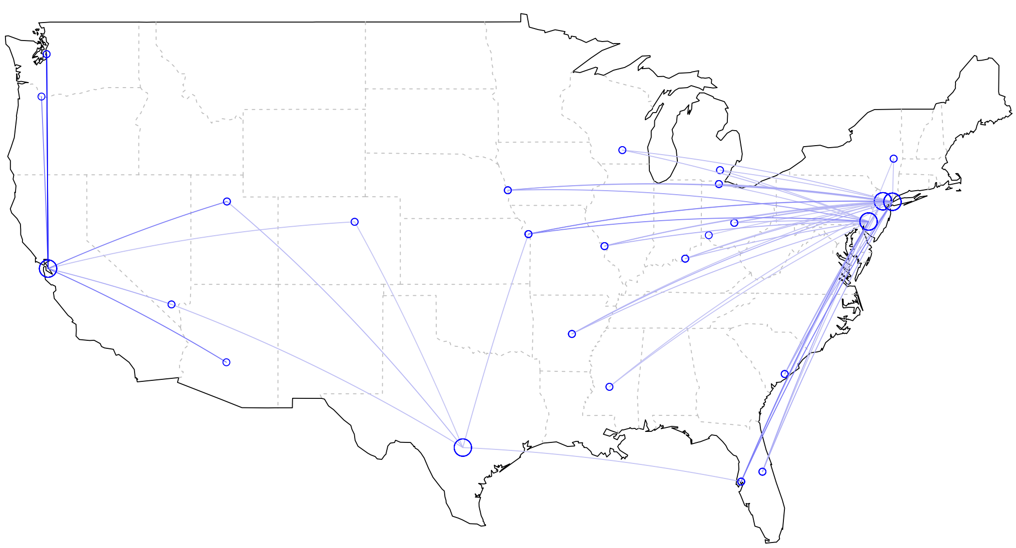


Figure 14: Equilibrium multiple listing rates for the partial solution for blood type B patients ($\pi = 0.25$ and $d = 1200$). In this illustrative scenario, OrganJet offers its services only in certain DSAs: Oakland, CA, New York, NY, New Providence, NJ, Philadelphia, PA, Waltham, MA and San Antonio, TX.

H Simulation Study Details

H.1 Finding Patient Arrival Rates (See Excel File: Patient Arrival Rates 01-04)

UNOS provides reports based on its transplant data on the following web site:
http://optn.transplant.hrsa.gov/latestData/advancedData.asp.
Making the selections shown in Figure 15 gives the number of 'registrations' in a particular year to waiting lists for kidney transplants in each DSA for each blood type. Part of the report that is built via these selections is presented in Figure 16 as an example. On the same web site, entering the selections shown in Figure 17a gives the number of kidney 'transplants from living donors' in a particular year in each DSA for each blood type. Subtracting 'Transplants from living donors' from 'new registrations' for each DSA and blood type gives annual patient arrival rates. Average rates for a range of years can be found by averaging annual rates.

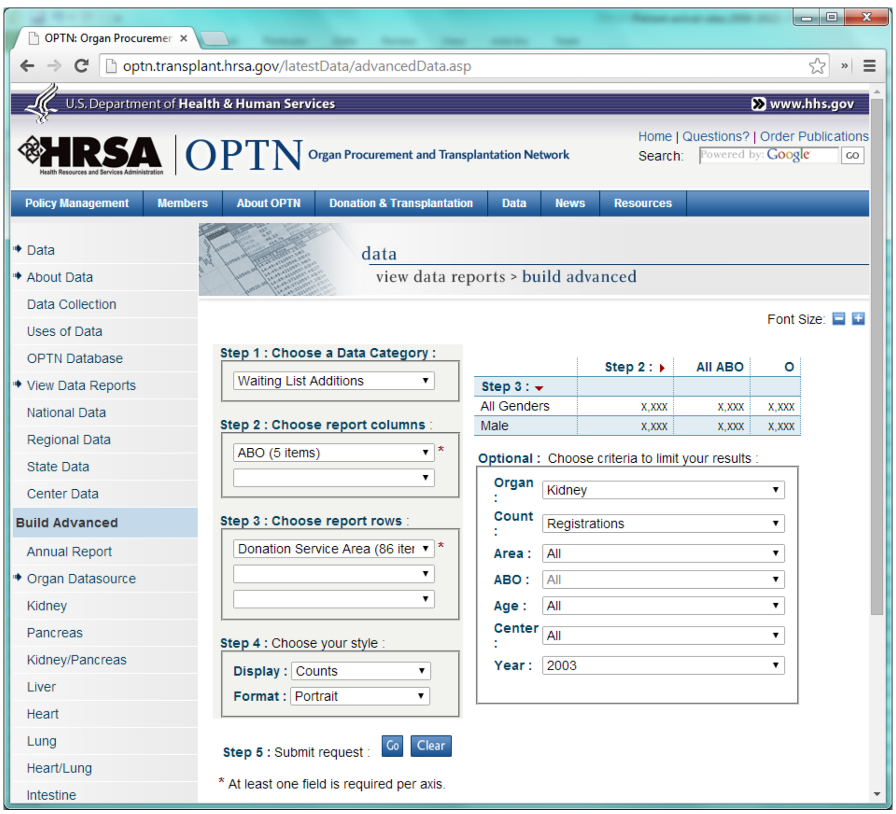


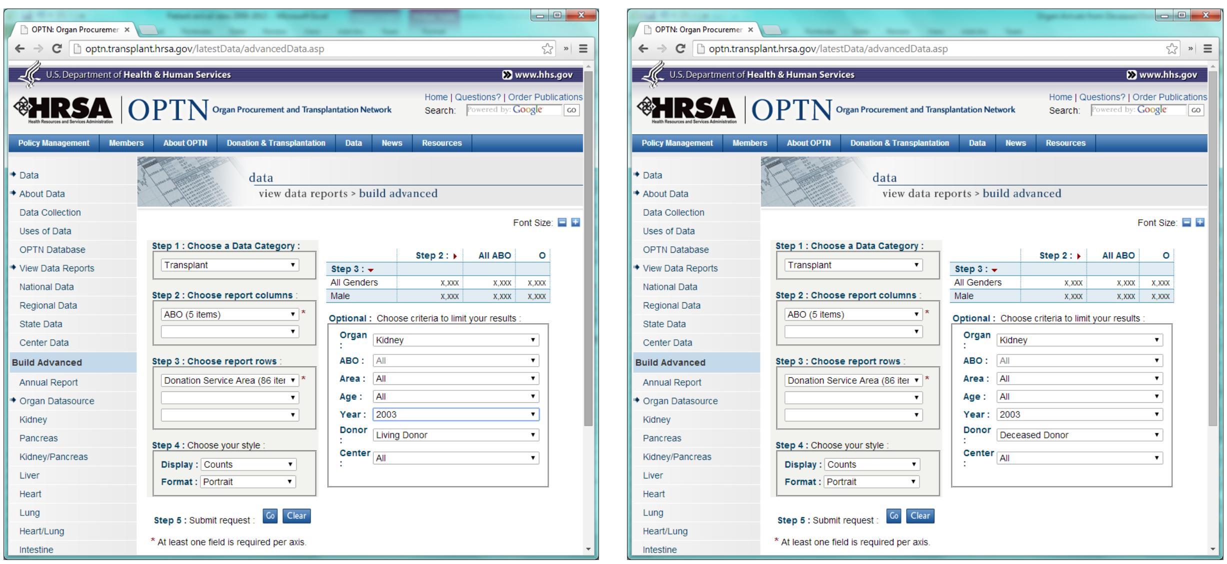
Figure 15: Selections entered at "http://optn.transplant.hrsa.gov/latestData/advancedData.asp" to find out annual number of registrations for kidney per blood type per DSA in a particular year.

H.2 Finding Organ Arrival Rates (See Excel File: Organ Arrivals from Deceased Donors 01-04)

'Kidney transplants from deceased donors' represent the number of patients (of each blood type) that received transplants. This data report can be created from the UNOS data through the selections shown in Figure 17b. Let us refer to this data as 'transplants categorized by patient blood type'. The number of transplants is not the appropriate input for the simulation model.

Number of New Registrations per year	2003 All ABO	O	A	B	AB
All Donation Service Areas	25,579	12,429	8,476	3,704	970
ALOB-OP1 Alabama Organ Center	775	392	238	123	22
AROR-OP1 Arkansas Reg. Organ Recovery Agency	106	56	38	10	2
AZOB-OP1 Donor Network of Arizona	364	204	110	38	12
CADN-OP1 CA Transplant Donor Network	1,602	770	504	255	73
CAGS-OP1 Sierra Donor Services	232	108	70	45	9
CAOP-OP1 OneLegacy	2,268	1,101	733	341	93
CASD-IO1 Lifesharing - A Donate Life Org.	381	193	115	63	10
CORS-OP1 Donor Alliance	342	165	124	39	14
CTOP-OP1 LifeChoice Donor Services	97	60	25	11	1
DCTC-OP1 Washington Reg Transplant Community	566	275	165	102	24
FLFH-IO1 TransLife	252	136	63	46	7
FLMP-OP1 Life Alliance Organ Recovery Agency	284	136	100	39	9
FLUF-IO1 LifeQuest Organ Recovery Services	368	175	115	60	18
FLWC-OP1 LifeLink of Florida	168	82	53	28	5
GALL-OP1 LifeLink of Georgia	618	302	188	104	24
HIOP-OP1 Legacy of Life Hawaii	96	36	32	24	4
IAOP-OP1 Iowa Donor Network	162	71	66	21	4
ILIP-OP1 Gift of Hope	1,358	645	488	171	54
INOP-OP1 Indiana OPO	279	133	98	39	9
KYDA-OP1 KY Organ Donor Affiliates	159	92	42	18	7

Figure 16: Report generated by selections shown in Figure 15.



(a) living donors (b) deceased donors

Figure 17: Number of kidney transplants per DSA report can be built for a particular year.

The number of kidneys offered to patients is needed (including the ones that are wasted). We can also generate reports of the number of kidneys that were 'recovered' from deceased donors for a particular year and those that were 'transplanted' to patients categorized by donor blood type (using selections shown in Figure 18). Since it includes kidneys that are wasted, 'kidneys recovered' data may seem appropriate as simulation input. However, patients do not necessarily receive organs from donors with the same blood type. That is, data categorized by donor blood type does not give the true account of how many patients of each blood type get kidneys. Therefore we used the ratio of kidneys recovered to kidneys transplanted (data categorized by donor blood type) to scale up the numbers in the data we labeled above as 'transplants categorized by patient blood type'. This way, at a given DSA for a particular blood type, we can estimate the number of organs that were offered and use this in the simulation as organ arrival rate.

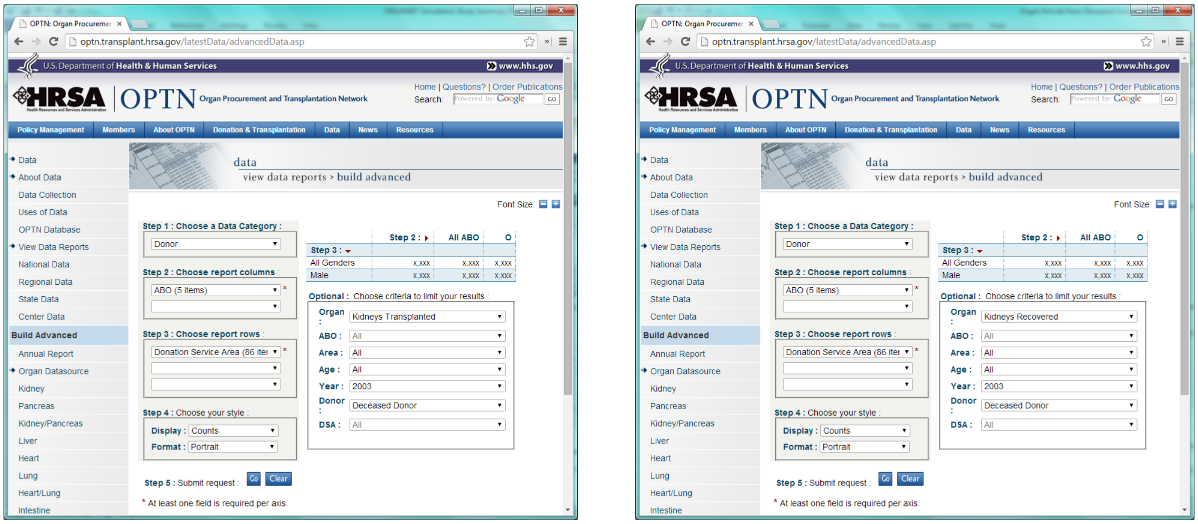


Figure 18: ‘Kidneys Transplanted’ and ‘Kidneys Recovered’ reports can be built for a given year.

H.3 Calculating GCVs

The geographical coefficient of variation for waiting times and for access to transplant, respectively, are given by

$$GCV_w = \sqrt{\sum_k \mu_k^2 (w_k - \bar{w})^2 \Big/ \sum_k \mu_k^2} \Big/ \bar{w} \quad \text{and} \quad GCV_\phi = \sqrt{\sum_k \lambda_k^2 (\phi_k - \bar{\phi})^2 \Big/ \sum_k \lambda_k^2} \Big/ \bar{\phi}.$$

where

$$\bar{w} = \frac{\sum_k \mu_k w_k}{\mu_k} \quad \text{and} \quad \bar{\phi} = \frac{\sum_k \lambda_k \phi_k}{\lambda_k}.$$

Analytical calculations (and definitions):

$w_k = \log(\lambda_k/\mu_k) * 365.25/\gamma$.

$\phi_k = \mu_k/\lambda_k$.

λ_k is the patient arrival rate to DSA k after multiple-listing (assuming multiple listing patient only arrived to the DSA it multiple-listed to).

μ_k is the organ arrival rate to DSA k.

$1/\gamma$ is the average time-to-death without a transplant. γ is 0.1694 and 0.1719, respectively, for blood types B and O.

Simulation output (and definitions):

w_k is the average waiting time of patients who received an organ at DSA k and μ_k is the number of transplants at DSA k. ϕ_k, λ_k and average time-to-death without a transplant are defined as described above under analytical calculations.

H.4 Acceptance Probabilities

The following formula is used to calculate the acceptance probability for DSA k

$$1 - \exp^{\ln(1-a_k)/N} \tag{113}$$

Where a_k is the percentage of offers accepted and N is the maximum number of times each organ is offered[47].

H.5 Median Waiting Times reported on UNOS web site.

Median waiting times for patients registered between 1999 and 2004 could be produced using the following UNOS web site : http://optn.transplant.hrsa.gov/latestData/step2.asp The category and organ selections are, respectively "Median Waiting Time" and "Kidney" and the report is titled "Waiting Time by Blood Type".

References

Billingsley, P. (1999). *Convergence of Probability Measures* (second ed.). Wiley.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.

Cain, M. (1994). The moment-generating fuction of the minimum of bivariate normal random variables. *The American Statistician 48*(2), 124–125.

Clark, C. E. (1961). The greatest of a finite set of random variables. *Operations Research 9*, 145–162.

Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein (2009). *Introduction to Algorithms* (third ed.). The MIT Press, Cambridge, MA.

Cramer, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Horowitz, J. (1980). Extreme values from a nonstationary stochastic process: An application to air quality analysis. *Technometrics 22*, 469–478.

Jennings, O. B. and J. E. Reed (2012). An overloaded multiclass FIFO queue with abandonments. *Forthcoming in Operations Research*.

Kella, O. (1986). On the distribution of the maximum of bivariate normal random variables with general means and variances. *Communications in Statistics-Theory and Methods 15*(11), 3265–3276.

Liu, Y. and W. Whitt (2011a). The G_t/GI/S_t+GI many-server fluid queue. *Working Paper*.

Liu, Y. and W. Whitt (2011b, July-August). A network of time-varying many-server fluid queues with customer abandonment. *Operations Research 59*(4), 835–846.

Marshall, A., I. Olkin, and B. Arnold (2009). *Inequalities: Theory of Majorization and Its Applications* (second ed.). Springer.

Mas-Colell, A. (1984). On a theorem of Schmeidler. *Journal of Mathematical Economics 13*(3), 201–206.

[47] Data of total offers vs total accepts for each DSA from 7/1/2000 to 8/31/2010, from which we compute percentage of offers accepted.

Talreja, R. and W. Whitt (2008). Fluid models for overloaded multiclass many-server queueing systems with first-come, first-served routing. *Management Science 54*(8), 1513–1527.

Ward, A. (2012, January). Asymptotic analysis of queueing systems with reneging: A survey of results for fifo, single class models. *Surveys in Operations Research and Management Science 17*(1), 1–14.

Whitt, W. (2002). *Stochastic-Process Limits.* Springer.

Whitt, W. (2004, October). Efficiency-driven heavy-traffic approximations for many-server queues with abandonments. *Management Science 50*(10), 1449–1461.