Improving Societal Outcomes in the Organ Donation Value Chain

Priyank Arora  
Ravi Subramanian  
Scheller College of Business, Georgia Institute of Technology, Atlanta, GA 30308,  
priyank.arora@scheller.gatech.edu, ravi.subramanian@scheller.gatech.edu

In the context of the mismatch between demand and supply of organs, a statistic worth noting is that across the donor service areas in the US, the average (median) percentage of eligible potential donors from whom no organs are recovered is 26.6% (28%). The significant socioeconomic costs arising from the suboptimal quantity and/or quality of recovered organs form the context of our study, which takes the perspective of the social planner that has an overall quality-adjusted-life-year (QALY) improvement objective. We model the operational decisions of the two key supply-side entities in a cadaver ODVC, namely, the organ procurement organization (OPO) and hospital (trauma center). Specifically, we consider the OPO’s effort level in seeking authorization for organ donation, the hospital’s effort level in identifying and referring potential donors, and the hospital’s priority scheme for scheduling organ recovery and other surgeries in its operating room (OR). The main contributions of our work are two-fold: First, we develop an analytical model to study the effects of contextual parameters and decisions of the OPO and the hospital on their respective payoffs and on societal outcomes. This model interrelates key components, including donor heterogeneity, organ recovery reimbursement rates for the hospital, cost to the hospital from the wait times experienced by its other patients, shared OR capacity between organ recovery and other procedures, and QALY increments for organ recipients and the hospital’s other patients. Our analysis identifies current misalignments in the objectives of the OPO, the hospital, and the social planner, that lead to socially suboptimal fractions of organs recovered from different types of potential donors. Second, we recommend an administratively feasible and Pareto-improving contract that the social planner can use to help the ODVC achieve socially-optimal performance.

Key words: healthcare operations; organ donation; organ recovery; contracts

History: This version: June 2017

1. Introduction

The mismatch between demand and supply of organs for transplantation is significant: as of December 2016, approximately 120,000 individuals in the US were registered on the waitlist to receive an organ. This number is substantially large compared to supply-side figures: in the year
2016, there were 15,948 donors who accounted for 33,611 transplants (HRSA 2016). Despite the creation of donor registries, the use of organs from expanded-criteria donors, and improvements in surgical techniques, organ preservation, and immunosuppressant drugs, there is still a significant gap between demand and supply of organs, which results in significant socioeconomic costs. For instance, according to the US Renal Data System’s report (USRDS 2013), the cost to an end-stage renal disease patient for dialysis is $89,000/year as compared to approximately $32,000 for renal transplant surgery and $25,000/year for post-surgery care. The US Medicare’s budget absorbs about 80% of dialysis costs and about 53% of renal transplantation surgery and post-surgery care costs.

In the context of the mismatch between demand and supply of organs, a statistic worth noting is that across the donor service areas\(^1\) (DSAs) in the US, the average (median) percentage of eligible\(^2\) potential donors from whom no organs are recovered is 26.6% (28%). Figure 1 highlights the significant gaps in the numbers of donors from whom at least one organ is recovered and the numbers of eligible deaths (i.e., deceased individuals who meet the criteria for organ donation), across the 58 DSAs in the US.

While the healthcare operations management literature has largely addressed the issue of optimal allocation of organs to waiting patients once an organ becomes available for transplantation (Alagoz et al. 2004; 2007, Ata et al. 2016, Kreke et al. 2002, Su and Zenios 2006, Zenios et al. 2000, Zenios 2002), our work addresses the significant supply-side gap between the numbers of eligible potential donors and eventual donors. We model the operational decisions of the

---

\(^1\) The Centers for Medicare and Medicaid Services has divided the US into 11 regions, which are further divided into 58 donor service areas for planning and managing organ recovery and allocation efforts.

\(^2\) According to Federal regulations, an eligible potential donor is an individual 75 years or younger whose death meets American Academy of Neurology Practice’s (AANP’s) neurological criteria for determining brain death, and who does not exhibit one or more active infections that are contraindications for organ donation such as tuberculosis, rabies, meningitis, malaria, etc. (OPTN 2017; p. 4–5).
Note: This figure shows, for each of the 58 DSAs in the US, the number of eligible potential donors from whom no organ is recovered as a percentage of the number of eligible deaths at hospitals in the DSA. The plots are based on data for the year 2016 published by the US Scientific Registry of Transplant Recipients (SRTR 2016). The first two letters in the horizontal axis marker indicate the State to which the DSA belongs.

Figure 1  Eligible potential donors from whom no organ is recovered as a percentage of the number of eligible deaths at hospitals, by donor service area

two key supply-side entities in a cadaver organ donation value chain (ODVC) within a DSA – namely, the organ procurement organization (OPO) and hospital (typically the main trauma center within the DSA) – that determine organ recovery outcomes. The OPO is a non-profit entity designated by the Centers for Medicare and Medicaid Services (CMS) to carry out the following main tasks in an ODVC: collaborate with hospitals in its DSA, educate hospital staff on standard procedures for organ donation, review medical records of identified potential organ donors, request family authorization for organ donation, and coordinate recovery, matching, and transportation of recovered organs. From the standpoint of the hospital, one of the key provisions of the revised Medicare conditions for organ donation (Federal Ruling 42 CFR 482 on June 22, 1998), also termed as the “donation rule,” is that the hospital is required to contact its affiliated OPO in a timely manner about individuals whose death is imminent or who have
died while in the hospital’s care. Traino et al. (2012) underscore the importance of coordinating the operational actions of these two supply-side entities. Their study emphasizes the need for strong OPO–hospital relationships in the form of effective channels of communication and frequent contact in order to ensure timely identification and referral of potential donors by hospitals to OPOs.

However, several studies lament the lack of coordination between hospitals and OPOs as one of the main reasons for suboptimal levels of organ recovery from eligible deceased donors (Brown 2000, Graham et al. 2009). Shafer et al. (2003; 2004) point out that OPOs are resource-constrained organizations with limited budgets and have to be selective in the number and type of staff on their payroll, which has implications for organ recovery outcomes. For example, Shafer et al. (2004) show that although it requires a considerable investment of OPO resources, having a full-time OPO staff member present at the donor hospital can significantly improve organ recovery outcomes.

Further, studies also show that suboptimal organ recovery outcomes result from hospitals’ priority to providing care to living patients as compared to deceased donors, health care professionals’ (HCPs’) lack of knowledge regarding the concept of brain death (BD), and delays by hospital staff in referring potential donors to the OPO (Opdam and Silvester 2004, Sheehy et al. 2003, Walters 2009). Doyle et al. (2014) and Walker (1998) highlight that deceased donors often receive lower priority for care than other patients in the hospital, resulting in delays in access to the operating room (OR) for surgical recovery of organs. Jendrisak et al. (2002, p. 1), Jensen (2011, p. 165–166), Moazami et al. (2007), and Zaroff et al. (2002) point out that, of the logistical issues involved in the organ recovery process, timely access to the hospital’s OR for surgical recovery of organs can be one of the most problematic. From a socioeconomic standpoint, delays in the organ recovery process negatively influence the quality of the recovered organ and, hence, graft survival and quality of life of the recipient post-transplantation (Cantin et al. 2003, Schnitzler et al. 2003).
The significant socioeconomic costs arising from the suboptimal quantity and/or quality of recovered organs form the context of our study, which focuses on the following operational actions for organ recovery by the OPO and the hospital: the OPO’s effort level in seeking authorization for organ donation, the hospital’s effort level in identifying and referring potential donors, and the hospital’s priority scheme for scheduling organ recovery and other surgeries in its OR. In our analysis, we take the perspective of the social planner (Department of Health and Human Services, \(^3\) DHHS), that uses quality-adjusted life years (QALYs) as a measure of health outcomes. Introduced by Klarman and Rosenthal (1968), QALY was developed as a way to combine length and quality of life into a single measure. The US Panel on Cost-Effectiveness in Health and Medicine recommends that QALY be used as the principal measure of health outcomes (Siegel et al. 1996). This measure has also been used in the operations management literature that focuses on the demand side of the ODVC to examine trade-offs involved in organ allocation policies (e.g., Ahn and Hornberger 1996, Bertsimas et al. 2013, Dai et al. 2017, Zenios et al. 2000, Zenios 2002).

We aim to answer the following research questions: (i) How do the operational actions of the supply-side ODVC players impact their individual payoffs and societal outcomes? (ii) What mechanisms can be implemented by the social planner to improve societal outcomes in the ODVC as well as ensure that no player is worse-off? The main contributions of our work are two-fold. First, we develop a model that enables us to study the effects of the operational decisions of the supply-side players (OPO and hospital) in an ODVC, on their respective payoffs and on societal outcomes. The model reflects key contextual components, including donor heterogeneity, organ recovery reimbursement rates for the hospital, cost to the hospital from the wait times experienced by its other patients, shared OR capacity between organ recovery

\(^3\) The Department of Health and Human Services is a federal agency tasked with protecting the health of all US nationals and providing essential human services.
and other procedures, and increments in QALYs for organ recipients and the hospital’s other patients. Our analysis identifies current misalignments in the objectives of the OPO, the hospital, and the social planner, that lead to socially suboptimal fractions of organs recovered from different types of potential donors. Second, we recommend an administratively feasible and Pareto-improving contract that the social planner can use to help the ODVC achieve socially-optimal performance.

2. Literature Review

Our work draws on and contributes to three literature streams: (i) The operations management (OM) and healthcare economics literatures that address the demand and supply of organs for transplantation; (ii) The medical literature on organ donation and recovery that presents the findings and perspectives of medical researchers and healthcare professionals; and (iii) The literature on non-profit operations management and humanitarian supply chains that considers the distinctive features of non-profit operations.

The majority of the related OM literature focuses on the demand side of an ODVC. The demand for organs is known in the form of the national waitlist maintained by the United Network for Organ Sharing (UNOS). Several research studies have addressed the issue of optimal allocation of organs to waiting patients once an organ becomes available for transplantation. These studies have recommended various organ allocation policies to maximize social or individual welfare, with organ supply assumed to be exogenous (Alagoz et al. 2004; 2007, Su and Zenios 2006, Zenios et al. 2000). Ata et al. (2016), Ata et al. (2017), Kreke et al. (2002), and Zenios et al. (2000) use simulation models and the UNOS database of previous organ transplantations to study the potential effects of changes to existing organ allocation policies. Recent studies in healthcare economics have examined ways to improve the supply of organs for transplantation, e.g., donations by living donors. For instance, Ashlagi and Roth (2012) and Roth et al. (2007) show that two-way and three-way kidney exchanges involving patients for whom a living donor
is willing to donate an organ but is incompatible (e.g., blood- or tissue-type incompatibility between the patient and the living donor) could lead to a substantial increase in the number of transplants.

To our knowledge, the paper by Arikan et al. (2017) is the only other paper in the OM literature that discusses the supply side of organ transplantation from cadaveric donors. They perform an empirical study to understand drivers of geographical differences in organ procurement rates and propose that low-quality organs should immediately be made available more widely rather than sticking to geographical constraints that currently apply to all organ quality types. Our work complements this paper by studying the operational actions of the supply-side players (OPO and hospital) in a cadaver ODVC.

We also draw on the medical literature to consider the findings and views of the medical community on ways to alleviate the wide gap between the demand and supply of organs for transplantation. Traino et al. (2012) implemented a national test of the Rapid Assessment of hospital Procurement barriers in Donation (a tool for identifying barriers to donor identification, donor referral, and requesting families for authorization) and found that the hospital–OPO relationship has inherent tensions. This is because the HCPs caring for potential donors and trying to keep them alive are the same professionals on whom the system relies for referrals. Our work builds on their major recommendation – as also reiterated by other researchers (see Goodman et al. 2003) – that special efforts, such as interventions to improve donor identification and periodic meetings among ODVC members, are needed to ensure that OPOs and hospitals coordinate their efforts to ensure the successful conduct of organ recovery activities.

Ranjan et al. (2006) conducted a retrospective financial analysis of the management of potential donors showed that organ recovery for Medicare-approved transplant hospitals can in fact be financially attractive for those hospitals. OPOs typically reimburse hospitals in a timely manner and at Medicare-prevailing rates, which are typically higher than third-party insurance
rates. Thus, the financially attractive and timely reimbursements are at odds with the observation of low or insufficient priority for organ recovery by hospitals (Jensen 2011; p. 165–166). However, in their analysis, Ranjan et al. (2006) only include the direct costs of maintaining a potential donor and do not take into account indirect costs (e.g., disutility associated with dealing with the subject of organ donation or the effect of organ recovery activities on delays experienced by the hospital’s other patients). Domingos et al. (2012) emphasize that organ recovery outcomes must be considered as a component in the measure of the quality of care provided by hospitals in order to drive improvements in organ recovery outcomes. They propose that not only OPOs but also intensive care units at hospitals should be evaluated on the conversion rate of potential to actual donors (i.e., donors from whom organs are eventually recovered). Through our analytical model, we attempt to explore the possibility of introducing contractual levers in order to better align the objectives of the OPO and the hospital with the societal objective of the social planner.

There is an emerging stream of literature in OM that recognizes the unique challenges associated with non-profit operations within supply chains because of the distinct sets of objectives targeted by the for-profit versus the non-profit organizations (Feng and Shanthikumar 2016, Tomasini and Van Wassenhove 2009). Bhattacharya et al. (2014) and Holguin-Veras et al. (2012) discuss the differences between humanitarian supply chains and traditional supply chains, including supply chain design requirements and coordination mechanisms. Berenguer et al. (2014) highlight the lack of inter-organization collaboration as a recurring challenge in non-profit operations, which is true of our context as well; the divergence between individual and societal objectives in an ODVC leads to unique interrelationships and coordination challenges. Ergun et al. (2014) utilize a cooperative game theoretic model to explore improvements in humanitarian operations through collaboration among the different parties involved, including governmental, private, and non-governmental organizations. Similar to their work, we utilize a
game theoretic model to analyze the interrelationships between the operational decisions and the payoffs of the supply-side players in an ODVC.

The preceding review of the related literature underscores the importance of coordinating the operational actions of the supply-side ODVC players in the form of relevant efforts, timely referrals, and timely organ recovery; yet interactions among contextual parameters and the decisions of these players have not been well-studied. We address this important literature gap with our analysis that takes the perspective of the social planner and considers the privately- and socially-optimal operational actions of the OPO and the hospital operating within an ODVC.

3. Model (Current Scenario)

The study by Sheehy et al. (2003) shows that larger hospitals (which are typically trauma centers as well) have significantly greater numbers of potential and actual donors – specifically, 19% of hospitals account for 80% of the donations. We therefore focus our analysis on a large hospital in the OPO’s DSA. Based on the organ recovery process and the sequence of multiple interactions between the OPO and the hospital, we develop an analytical model to represent the current scenario for a focal organ and then, through our analysis, propose a contractual mechanism to improve the societal outcome of the ODVC. Table 1 lists the abbreviations that we use and Table 2 summarizes our notation.

Several research studies (e.g., Schold et al. 2005, Swanson et al. 2002) highlight that a combination of factors, such as age and medical history of the potential donor, leads to variation in the quality of organs that can be recovered from potential donors. Ojo (2005) reports that graft survival for “expanded-criteria” or marginal-quality cadaveric donor kidneys is 8% lower at 1-year, and 15-20% lower at 3-5 years post-transplantation compared to non-marginal kidneys (an expanded-criteria-donor for kidneys is any donor above the age of 60 years, or a donor above the age of 50 years with two of the following: history of high blood pressure, creatinine
level greater than or equal to 1.5 milligrams per deciliter, or death resulting from a stroke). Therefore, since organ quality determines graft survival as well as the quality of life of the recipient post-transplantation (Cantin et al. 2003, Schnitzler et al. 2003), the quality of the recovered organ is an important consideration for the social planner (DHHS).

To allow for this heterogeneity, we consider the arrival of two types of medically suitable potential donors who meet the criteria for imminent BD – *type 1*, from whom the focal organ recovered would be of higher quality (known *a priori* based on health history and tests early in the process), and *type 2*, from whom the organ recovered would be of lower quality. We assume that the arrival rates of potential donors of each type are independent. Let $\lambda_{p1}$ and $\lambda_{p2}$ denote the (Poisson) arrival rates of type 1 and type 2 potential donors, respectively, such that $\lambda_{p1} + \lambda_{p2} = \lambda_p$.

However, current federal regulations and donor referral and donor management practices do not require the hospital or the OPO to differentiate their operational actions based on donor heterogeneity. For instance, the reporting by the OPO to CMS of numbers of eligible deaths and eventual donors (the data that is presented in Figure 1), does not reflect the heterogeneity in the quality of organs that can be recovered from potential donors or in the quality of organs that are recovered from authorized donors. Further, OPOs provide hospital staff with criteria or “triggers” pertaining to the health condition of the patient which, when met, indicate that the patient may be a potential organ donor and that the OPO should be contacted (see Appendix A for an example of a “trigger card” provided by an OPO to HCPs). Neither the donation rule (discussed in Section 1) nor do the descriptions of triggers reflect the heterogeneity in the quality of organs that can be recovered from potential donors. Because OPOs and hospitals currently make their operational decisions without explicitly taking into account the quality of organs that can be recovered, for the model that captures the current scenario, incoming potential donors are regarded as a single pool with overall arrival rate $\lambda_p$. 
3.1. OPO and Hospital Actions

We model the interaction between the OPO and the hospital within an ODVC as a Stackelberg game. The hospital, because of its say in the referral of potential donors to the OPO and in OR scheduling, is the Stackelberg leader. The hospital makes two operational decisions – the level of effort to commit to organ recovery activities, and the OR priority assigned to organ recovery as compared to other procedures. The OPO’s decision is the level of effort to commit towards interacting with potential donors’ families and seeking authorization for organ donation. We discuss these decisions further in Sections 3.2, 3.3, and 3.4.

3.2. Hospital’s Effort Level

The hospital decides its level of effort ($\xi_h$) to commit to organ recovery activities – specifically, personnel training, time, and effort involved in identifying and referring potential donors. Based on an audit of 5,551 deaths in 12 hospitals, Opdam and Silvester (2004) report that hospitals withdrew medical support for 38% of potential donors who met eligibility criteria for organ donation. Using data on 3,329 potential heart-beating donors across 284 hospitals, Walters (2009) reports that 23% of these donors did not have BD tests carried out. Several studies reveal that HCPs find it difficult to broach and discuss the subject of organ donation with the families of potential donors (Chernenko et al. 2005, Regehr et al. 2004). In particular, Chernenko et al. (2005) estimated that 77% of registered nurses and 44% of doctors found it difficult to communicate and explain the concept of brain death to families. Training-related reasons for HCPs not referring potential donors include lack of knowledge regarding: the concept of brain death; the process of referring potential organ donors to the OPO; and the life-changing outcomes for organ recipients (Molzahn et al. 2003). Also, participating in the process of organ donation is only one aspect of their complex jobs.

We normalize the hospital's effort level such that $\xi_h \in [0, 1]$ translates into the fraction ($f_h$) of potential donors who end up as “referred” donors. Missed referral opportunities by the hospital or referrals made after pronouncement of BD are collectively termed as “missed” referrals (Federal Ruling 42 CFR 482 on June 22, 1998).
3.3. OPO’s Effort Level

Once a potential donor is referred to the OPO, the OPO appoints a family care coordinator to the case, who is responsible for following up with the hospital regarding the health condition of the potential donor and attending to the needs of the potential donor’s family members facing the stressful situation.

Siminoff et al. (2009) conducted an event study to show that a training workshop on the use of “effective relational and affective communication techniques,” for OPO staff responsible for seeking authorization from the nearest-of-kin for organ donation, led to a significant increase in the likelihood of authorization. The study by Frutos et al. (2002) found that 20.3% of families who initially refused authorization and 65.8% of families who were undecided at first, eventually consented when approached again. Thus, the OPO’s effort level $\xi_o \in [0, \theta]$ in the form of interactions with the potential donor’s family members, influences the fraction of referred donors who are “authorized” for organ donation after pronouncement of BD. $0 < \theta \leq 1$ captures the resource-constrained environments in which OPOs operate (as discussed in Section 1; Shafer et al. 2003; 2004).

We denote the fraction of referred donors who end up as authorized donors by $f_o$ and assume $f_o = \tau \xi_o$, where $\tau$ ($0 < \tau \leq 1$) captures donors’ or their families’ (exogenous) affinities towards organ donation. Thus, the rate of authorized donors from whom organs are recovered is $\lambda_a = f \lambda_p$, where $f = f_h f_o = \tau \xi_h \xi_o$ (see Figure 2). As discussed before, in the current scenario, both the OPO and the hospital regard potential donors as a single pool with overall arrival rate $\lambda_p$. 

\[ \begin{align*}
\text{Potential donors} & \quad \lambda_p = \lambda_{p1} + \lambda_{p2} \\
\text{Family refusals} & \quad (1 - f_o) f_h \lambda_p \\
\text{Missed referrals} & \quad (1 - f_h) \lambda_p \\
\text{Authorized donors} & \quad \lambda_a = f_o f_h \lambda_p \\
\text{Other patients} & \quad \lambda_h \\
\text{Operating Room (OR)} & \quad \lambda_a \\
\end{align*} \]
Accordingly, in the current scenario, the fractions of referred donors who end up as authorized donors are the same for both donor types. We assume that the flows in Figure 2 are such that we can use the Partition Theorem for Poisson Processes (see Cramér and Leadbetter 1967). Also, for simplicity, we assume that the organ recipient is identified from the waitlist immediately after the potential donor becomes an authorized donor, or, that authorized donors can proceed immediately to organ recovery surgery once the hospital’s OR becomes available.

3.4. Hospital’s OR Scheduling Policy

The OR is often a hospital’s most constrained resource because of the expensive medical equipment and specially trained staff that are required (Cardoen et al. 2010, Van Houdenhoven et al. 2007). Within the context of ODVCs, Walters (2009) states that a hospital’s resource constraints, including limited OR capacity, may result in suboptimal organ recovery decisions. Jendrisak et al. (2002), Jensen (2011), Moazami et al. (2007), and Zaroff et al. (2002) point out that of the logistical issues involved in the organ recovery process, timely access to the hospital’s OR for surgical recovery of organs can be one of the most problematic. Based on our conversations with ODVC members and reviews of the medical literature, potential reasons for organ recovery surgeries not being accorded requisite OR priority are: (i) Hospitals (implicitly or explicitly) favor living patients over deceased donors in decisions involving allocation or prioritization of use of hospital resources (Doyle et al. 2014, Walker 1998), and (ii) The transplant team members tasked with organ recovery are typically unknown to the hospital’s OR staff as those members come from the transplant hospital, rather than the donor hospital (Regehr et al. 2004).

We model the hospital’s OR as an $M/G/1/[2]$ queueing system. The two classes of patients competing for OR access are: authorized donors ($a$), and other hospital patients ($h$) who need OR care, with respective Poisson arrival rates $\lambda_a = f \lambda_p$ and $\lambda_h$ (see Figure 2). Patients in each class require a random amount of time in OR care, with mean $\frac{1}{\mu_x}$ and second moment $\nu_x$, where
\( x \in \{a, h\} \). Denote \( \rho_x = \frac{\lambda_x}{\mu_x} \). Let \( \mathcal{S} \) be the set of possible OR scheduling policies. For a chosen policy \( \chi \in \mathcal{S} \), let \( w_a \) and \( w_h \) denote the resulting average wait times in queue for the patients of each class. We assume \( \rho_a + \rho_h < 1 \), non-preemptive priority (i.e., the patient under consideration is processed completely before the next patient) and first-come first-serve policy within each of the two classes of patients. Based on our interviews with ODVC members, including a hospital trauma care manager, we make the realistic assumption that the mean service time for the organ recovery surgery in the OR is relatively low compared to the mean interarrival time of potential donors to the process.

**Assumption A1.** The mean service time for organ recovery surgery in the OR is low compared to the mean interarrival time of potential donors, i.e., \( \frac{1}{\mu_a} \ll \frac{1}{\lambda_p} \).

### 3.5. OPO and Hospital Objectives

The OPO has to meet the volume-based standard (specifically, conversion rate of potential donors to actual donors) set by CMS in order to maintain its Medicare certification. Also, a review of the mission statements of various OPOs, including Carolina Donor Services, Lifelink of Georgia, Living Legacy Foundation of Maryland, Gift of Life Michigan, and Nevada Donor Network, among others, indicates that OPOs aim to maximize the volume of authorized donors in their respective DSAs. Thus, the OPO currently maximizes \( \pi_{opo} = \lambda_a = f\lambda_p \), the rate of authorized donors. We denote the OPO’s effort level that maximizes \( \pi_{opo} \) as \( \xi_o^* \).

For its services related to caring for authorized donors, the hospital receives a reimbursement from the OPO, irrespective of organ recovery surgery outcomes (e.g., if the organ is deemed to be unviable by the recipient’s transplant surgeon during organ recovery surgery; CMS Ruling CMS-1543-R, p. 6–8). This reimbursement consists of two components: a fixed reimbursement rate per authorized donor \( (R_{af}) \) that includes charges for laboratory tests, medical equipment use, and anesthesiology consultation, among others, and a variable reimbursement rate per authorized donor per unit care time \( (R_{av}) \), which depends on the wait time \( (w_a) \) of the authorized donors.
in the intensive care unit before organ recovery surgery in the OR. The emotional stresses and discomfort for the hospital staff involved in organ recovery efforts (Chernenko et al. 2005, Regehr et al. 2004) results in a non-linear cost $C_e$ to the hospital; we assume $C_e = \frac{c_e}{2}\xi^2 h$.

We denote the average per-patient reimbursement rate associated with the hospital’s other patients by $R_h$. Let $C_h$ denote the net cost to the hospital in the form of customer dissatisfaction and reputation loss arising from the wait times experienced by the hospital’s other patients (Guinet and Chaabane 2003, Hall 2006). For exposition, we assume that $C_h > 0$ subsumes the variable reimbursement rate (per patient per unit time) applicable to the care time of the hospital’s other patients. As in Green et al. (2006) and Guinet and Chaabane (2003), we assume $C_h$ to be linear in the wait time $w_h$, i.e., $C_h = c_h w_h$. Thus, the hospital’s payoff rate is:

$$\pi_h = f\lambda_p R_{af} + f\lambda_p R_{av} w_a + \lambda_h R_h - C_h - C_e.$$

(1)

Friedman and Pauly (1983) find that even not-for-profit hospitals exhibit a profit-maximizing response to reimbursement terms. Within the context of ODVCs, Rios-Diaz et al. (2017) find that the hospital’s ownership-type (for-profit versus not-for-profit) is not associated with significantly different conversion rates of potential donors to actual donors. Furthermore, studies also find that for-profit and not-for-profit hospitals behave in a similar manner when responding to policy changes (Deneffe and Masson 2002, Duggan 2000). Therefore, in the ODVC context, the hospital’s payoff function is likely to be structurally similar if it were not for-profit. However, in Section 6, we reflect on the implications for our results if the hospital’s objective function had a volume-of-care component.

### 3.6. Social Planner’s Objective

We represent the social planner’s objective in terms of the quality-adjusted life years (QALYs) added to patients post-care. Introduced by Klarman and Rosenthal (1968), QALY was developed as a way to combine length and quality of life into a single measure. QALY calculations
are based on the idea that individuals transition through health states over time and that each health state has a utility attached to it (in terms of life-years weighted by their qualities). The utility of a health state is measured on a cardinal scale of 0–1, where 0 indicates death and 1 indicates full health. Standardized survey instruments (such as the EQ-5D survey) are used to estimate transition probabilities for the population. Responses to the surveys are elicited either from samples of the general population or from groups of patients. Finally, to obtain the QALYs added, the discounted utilities of health states over time are summed (for a more detailed explanation, see Weinstein et al. 2009). The US Panel on Cost-Effectiveness in Health and Medicine recommends that QALY be used as the principal measure of health outcomes (Siegel et al. 1996).

From a socioeconomic standpoint, delays in the organ recovery process negatively influence the quality of the recovered organ and, hence, graft survival and quality of life of the recipient post-transplantation (Cantin et al. 2003, Schnitzler et al. 2003). Jendrisak et al. (2005) tested the outcomes of a novel program whereby authorized donors were transported from the donor hospital to an independent facility housing a dedicated OR for organ recovery surgery. The results of their study demonstrated the benefits of timely OR access for organ recovery surgery, including improved viability of recovered organs. Prolonged waiting for the OR after declaration of BD can lead to an increased likelihood of rejection of the transplanted organ by the recipient’s body or failure of the organ post-transplantation (Blasco et al. 2007, Kunzendorf et al. 2002, Van Der Hoeven et al. 2003).

In addition to delays in the organ recovery process, the variation in organ quality stemming from donor heterogeneity too has a bearing on graft survival as well as quality of life of the recipient post-transplantation. For instance, Ojo (2005) reports that graft survival for expanded-criteria or marginal-quality cadaveric donor kidneys is 8% lower at 1-year, and 15-20% lower at 3-5 years post-transplantation compared to non-marginal kidneys (an expanded-criteria-donor
for kidneys is any donor above the age of 60 years, or a donor above the age of 50 years with two of the following: history of high blood pressure, creatinine level greater than or equal to 1.5 milligrams per deciliter, or death resulting from a stroke). Thus, unlike the current objectives of the OPO and the hospital, the social planner’s payoff does depend on the respective fractions of the different types of potential donors who end up as authorized donors. Let \( f_i \) denote the fraction of type \( i \in \{1, 2\} \) potential donors who end up as authorized donors, and \( w_{ai} \) denote the average wait time experienced by type \( i \in \{1, 2\} \) authorized donors while waiting for the OR to become available.

We assume the QALYs added for a recipient of the focal organ recovered from a type \( i \) potential donor to be of the form \( Q_{ai}(1 - q_aw_{ai}) \), whereby the value is a decreasing function of the delay \( (w_{ai}) \) experienced by the donor in the OR queue (i.e., post-BD), with an upper limit on the delay \( \left( \frac{1}{q_a} \right) \) beyond which there is no QALY addition (Blasco et al. 2007). The parameter \( q_a \) captures the sensitivity of the QALYs added to the delay experienced by the donor while waiting for the OR to become available.

Further, empirical studies find that delays experienced by the hospital’s other patients while waiting for the OR to become available, lead to increases in postoperative complications and mortality rates (Moran et al. 2005, Shiga et al. 2008). Therefore, we similarly assume the QALYs added for the hospital’s other patients who access the OR to be of the form \( Q_h(1 - q_hw_h) \), where \( w_h \) is the average delay experienced by these patients in the OR queue, and \( q_h \) is the sensitivity of QALYs added for the hospital’s other patients to the delay experienced by them. Thus, the social planner’s payoff rate is:

\[
\pi_S = \lambda_h Q_h(1 - q_hw_h) + f_1\lambda_{p1}Q_{a1}(1 - q_aw_{a1}) + f_2\lambda_{p2}Q_{a2}(1 - q_aw_{a2})
\] (2)

Note that \( Q_{a1} > Q_{a2} \); for example, transplants of expanded-criteria or marginal-quality kidneys add about 15% fewer QALYs in contrast to non-expanded-criteria kidneys (Ojo 2005).
4. Analysis

We first present the equilibrium operational decisions of the OPO and the hospital in the current scenario and discuss their divergence from socially-optimal decisions. We then supplement the analytical results with a numerical illustration that is grounded in practice. Thereafter, in Section 5, we discuss the design of contracts that yield socially-optimal outcomes. We restrict our analysis to parameter settings where all constraints on the model’s variables and decisions are met.

4.1. Current (Uncoordinated) Scenario: OPO and Hospital Equilibrium Decisions

**OPO:** Recall that the OPO is the Stackelberg follower, with the objective of maximizing the volume of care. It is therefore intuitive that the OPO will exert its highest effort level in order to maximize the rate of authorized donors. Hence, \( \xi_o^* = \theta \).

**Hospital:** As discussed earlier, the hospital’s chosen OR scheduling policy \( \chi \in \mathcal{A} \) will impact its payoff through the resulting average wait times in the OR queue for authorized donors and the hospital’s other patients. We transform the control problem of the hospital’s choice of OR scheduling policy into an optimization problem of choosing the vector of resulting average wait times \( \{w_a, w_h\} \) in the corresponding achievable region for work-conserving scheduling policies. In order to obtain \( w_a \) and \( w_h \), we use the results from Gelenbe and Mitrani (1980) that characterize the average wait times in queue achievable in a multiclass queueing system.

\[
    w_x \geq \frac{\lambda_a \nu_a + \lambda_h \nu_h}{2(1 - \rho_x)}, \quad x \in \{a, h\} \tag{3}
\]

\[
    \rho_a w_a + \rho_h w_h = (\rho_a + \rho_h) \frac{\lambda_a \nu_a + \lambda_h \nu_h}{2(1 - \rho_a - \rho_h)} \tag{4}
\]

We denote the work-conserving scheduling policy that gives absolute priority to authorized donors by \( I \) and the policy that gives absolute priority to the hospital’s other patients by \( II \). The constraint given by (3) is held at equality when \( x = a \) for \( \chi = I \), and when \( x = h \) for \( \chi = II \). Under Assumption A1, for both \( \chi = I \) and \( \chi = II \), \( w_a \) and \( w_h \) are linear increasing in
the fraction \( f \) of potential donors who end up as authorized donors. Because \( f \) increases in \( \xi \) and in \( \xi_o \), a consequence of increased effort by either the hospital or the OPO is greater OR congestion, leading to longer wait times for both authorized donors as well as the hospital’s other patients.

In maximizing the hospital’s payoff under the constraints given by (3) and (4), we obtain that it is optimal for the hospital to always prioritize other patients for service over authorized donors, i.e., \( \chi^*_h = II \). This is because a higher priority for authorized donors results in: (i) Longer wait times for the hospital’s other patients, resulting in a higher net cost to the hospital \( C_h \), (ii) Shorter wait times for authorized donors, leading to a lower revenue from the variable component \( R_{av} \) in the hospital’s payoff in (1). In terms of practice, the report by Cantin et al. (2003) reflects that hospitals typically delay organ recovery surgeries until other procedures have been completed. Furthermore, our interviews with OPO officials consistently revealed that organ recovery is currently not accorded sufficient priority by hospitals in OR scheduling. Proposition 1 characterizes the hospital’s equilibrium effort level \( \xi^*_h \) in the current scenario (all proofs are included in Appendix B).

**Proposition 1.** \( \exists \bar{c}_e > 0, \bar{c}_h > 0, \text{and } \bar{c}_h(c_e) > \bar{c}_h > 0, \text{such that the following four cases characterize the hospital’s equilibrium effort level } \xi^*_h: \)

1. \( c_e > \bar{c}_e, c_h \geq \bar{c}_h: \pi_h \text{ is decreasing in } \xi_h, \text{i.e., } \xi^*_h = 0; \)
2. \( c_e > \bar{c}_e, c_h < \bar{c}_h: \pi_h \text{ is concave in } \xi_h \text{ and either:} \)
   
   (a) unimodal, or (b) increasing; i.e., \( 0 < \xi^*_h \leq 1; \)
3. \( c_e < \bar{c}_e, c_h < \bar{c}_h: \pi_h \text{ is increasing in } \xi_h, \text{i.e., } \xi^*_h = 1; \)
4. \( c_e < \bar{c}_e, c_h \geq \bar{c}_h: \pi_h \text{ is convex in } \xi_h, \text{ and:} \)
   
   (a) \( \xi^*_h = 1 \text{ if } \bar{c}_h \leq c_h < \bar{c}_h(c_e), \text{ or (b) } \xi^*_h = 0 \text{ if } c_h \geq \bar{c}_h(c_e). \)

Denote \( \pi^*_h = \pi_h(\xi^*_h, \chi^*_h) \). Recall that \( c_e \) is the coefficient in the cost to the hospital for its efforts towards organ recovery activities, and \( c_h \) is the coefficient in the cost to the hospital from the
wait times experienced by the hospital’s other patients. Intuitively, when both $c_e$ and $c_h$ are sufficiently low (cases (iii) and (iv)a of Proposition 1), it is optimal for the hospital to exert its highest effort level (i.e., $\xi^*_h = 1$). On the other hand, when $c_e$ and/or $c_h$ are sufficiently high (cases (i) and (iv)b of Proposition 1), it is optimal for the hospital to not exert any effort (i.e., $\xi^*_h = 0$). However, when $c_e$ is sufficiently high and $c_h$ is sufficiently low (case (ii) of Proposition 1), the hospital’s payoff is concave in its effort level and the optimal level of effort for the hospital could be an interior value (i.e., $0 < \xi^*_h \leq 1$).

So that the OPO’s objective of volume of care is maximized, it is in the OPO’s private interest for the hospital to exert the highest effort level. However, Proposition 1 points out several instances where the private incentives of the OPO and the hospital are misaligned. Restricting our focus to the nontrivial cases where $\xi^*_h > 0$, let $\Theta_1$ denote the set of parameter conditions such that $\xi^*_h = 1$, and let $\Theta_2$ denote the set of parameter conditions such that $0 < \xi^*_h < 1$.

Also, let $\tilde{f}$ denote the equilibrium overall fraction of potential donors who end up as authorized donors in the current scenario, i.e. $\tilde{f} = \tau \xi^*_h \xi^*_o$. Thus, $\tilde{f}$ is either equal to $\tau \theta$ (under $\Theta_1$) or strictly less than $\tau \theta$ (under $\Theta_2$). Note that in the current scenario, the equilibrium fractions of both types of potential donors who end up as authorized donors, are the same; i.e., $\tilde{f}_1 = \tilde{f}_2 = \tilde{f} (= \tau \xi^*_h \xi^*_o)$. Next, we characterize our proposed centralized scenario wherein we examine the social planner’s objective and its implications for socially-optimal actions by the hospital and the OPO.

### 4.2. Proposed Centralized Scenario: Socially-Optimal Actions

The operational actions of the OPO and the hospital: their respective effort levels and the OR scheduling policy chosen by the hospital, affect the delays experienced by authorized donors and the hospital’s other patients while waiting for the OR to become available and, thus, the social planner’s payoff $\pi_S$. Note that, for a given OR scheduling policy $\chi$, the social planner’s payoff depends only on the effective fraction ($f_i$) of type $i \in \{1, 2\}$ potential donors who end up
as authorized donors, and not the individual values of $f_h$ and $f_o$ (recall that $f_h$ is the fraction of potential donors who end up as referred donors and $f_o$ is the fraction of referred donors who end up as authorized donors). The social planner’s payoff in (2) can be rewritten as:

$$\pi_S(f_1, f_2, \chi) = \lambda_h Q_h [1 - q_h w_h(f_1, f_2, \chi)] + f_1 \lambda_{p1} Q_{a1} [1 - q_a w_{a1}(f_1, f_2, \chi)] + f_2 \lambda_{p2} Q_{a2} [1 - q_a w_{a2}(f_1, f_2, \chi)].$$

Under the proposed centralized scenario, the hospital will be required to prioritize among three classes of patients competing for OR access, namely, type 1 authorized donors ($a_1$), type 2 authorized donors ($a_2$), and the hospital’s other patients ($h$). Let $\mathcal{A}_S$ denote the set of possible OR scheduling policies for these three classes of patients. The vector $\{w_{a1}, w_{a2}, w_h\}$ of average wait times experienced by the three classes of patients are characterized by Equations (5) and (6) (Gelenbe and Mitrani 1980). We reasonably assume that the mean ($\frac{1}{\mu_a}$), and second moment ($\nu_a$) of the service time in the OR is same for both classes ($a_1$ and $a_2$) of authorized donors (e.g., Brockmann et al. 2006). Denote $\rho_{ai} = \frac{f_i \lambda_{pi}}{\mu_a}$, where $i \in \{1, 2\}$.

$$w_y \geq \frac{f_1 \lambda_{p1} \nu_a + f_2 \lambda_{p2} \nu_a + \lambda_h \nu_h}{2(1 - \rho_y)}, \ y \in \{a1, a2, h\}$$

$$\rho_{a1} w_{a1} + \rho_{a2} w_{a2} + \rho_h w_h = (\rho_{a1} + \rho_{a2} + \rho_h) \frac{f_1 \lambda_{p1} \nu_a + f_2 \lambda_{p2} \nu_a + \lambda_h \nu_h}{2(1 - \rho_{a1} - \rho_{a2} - \rho_h)}$$

Under the proposed centralized scenario, $\chi = I$ denotes that the hospital gives priority to class $a_1$ and $a_2$ patients over class $h$ patients. Under Assumption A1, given the hospital’s choice of OR priority between authorized donors and the hospital’s other patients, it turns out that the relative priority between the two classes of authorized donors has a negligible impact on their wait times in the OR queue because $w_{a1}(f_1, f_2, \chi) \approx w_{a2}(f_1, f_2, \chi)$. Furthermore, since $w_{ai}(f_1, f_2, \chi) > w_{ai}(f_1, f_2, I) \ \forall \ \chi \in \mathcal{A}_S \setminus \{I\}, i \in \{1, 2\}$, the social planner’s payoff is higher if the hospital prioritizes authorized donors over its other patients. In other words, the socially-optimal OR scheduling policy (denoted by $\chi^S$) is $I$, which is at odds with the hospital’s privately-optimal OR scheduling policy ($\chi^h = II$; see Section 4.1).

The social planner is faced with a quantity-quality trade-off. A larger fraction of either type of potential donors converted to authorized donors (i.e., higher quantity) also leads to greater
OR congestion, which results in longer wait times in the OR queue for both types of authorized donors as well as the hospital’s other patients. The longer wait times faced by authorized donors in the OR queue adversely impact QALY outcomes (i.e., lower quality). Thus, from the social planner’s perspective, it is valuable to understand how efforts should be allocated between the two types of potential donors. Proposition 2 characterizes the socially-optimal fractions of type 1 and type 2 potential donors who should end up as authorized donors, based on the characteristic $q_a$ of the focal organ. Recall that $q_a$ is the sensitivity of QALYs added for the organ recipient, to the delay experienced by the organ donor while waiting in the OR queue.

**Proposition 2.** With a higher OR priority for authorized donors over the hospital’s other patients, $\exists \bar{q}_a > 0$ and $\bar{q}_a > 0$, such that the following three cases characterize the socially-optimal fractions, $f_1^S$ and $f_2^S$, respectively, of type 1 and type 2 potential donors who should end up as authorized donors:

(i) $q_a < \bar{q}_a$: $f_1^S = \tau \theta$, $f_2^S = \tau \theta$;

(ii) $\bar{q}_a \leq q_a < \bar{q}_a$: $f_1^S = \tau \theta$, $f_2^S = \max \left\{ \frac{(1-\rho_h)[Q_a1(2-q_a\lambda_h\nu_h-q_a\tau \theta \lambda_{p1}\nu_{\theta})-Q_a2q_a\tau \theta \lambda_{p1}\nu_{\theta}]-Q_hq_h\lambda_h\nu_a}{2Q_a1(1-\rho_h)q_a\lambda_{p2}\nu_a}, 0 \right\}$;

(iii) $q_a \geq \bar{q}_a$: $f_1^S = \min \left\{ \frac{(1-\rho_h)[Q_a1(2-q_a\lambda_h\nu_h)-Q_hq_h\lambda_h\nu_a]}{2Q_a1(1-\rho_h)q_a\lambda_{p2}\nu_a}, \tau \theta \right\}$, $f_2^S = 0$.

Denote $\pi_S^S = \pi_S(f_1^S, f_2^S, \chi^S)$. From Proposition 2, we observe that it is always optimal for the social planner to first allocate ODVC resources to convert type 1 potential donors to authorized donors, and then to allocate any remaining resources to convert type 2 potential donors to authorized donors. This result is intuitive given that the focal organ recovered from a type 1 potential donor adds more QALYs for the organ recipient as compared to the focal organ recovered from a type 2 potential donor.

Note that the upper limit on the OPO’s effort level, in essence, places an upper limit ($= \tau \theta \lambda_p$) on the overall rate of authorized donors that can be achieved by the ODVC. Proposition 2 points out cases where it is not socially optimal to exhaust the available resources of the supply-side ODVC players; i.e., under certain conditions, $f_1^S \lambda_{p1} + f_2^S \lambda_{p2}$ is strictly less than $\tau \theta \lambda_p$. 
For instance, when the QALYs added to the organ recipient is highly sensitive to the delay experienced by the organ donor in the OR queue (i.e., $q_a$ is higher), it is socially optimal for all type 1 potential donors to end up as authorized donors; however, it is not socially optimal to exhaust the remaining resources to convert all possible type 2 potential donors to authorized donors (i.e., cases (ii) and (iii) in Proposition 2). This is because a larger fraction of type 2 authorized donors increases OR congestion and, hence, adversely affects the QALYs added for the recipients of organs from both types of potential donors – including the recipients of type 1 organs. This observation is pertinent because, despite the scarcity of organs, 10% of livers and similarly high percentages of other types of organs are turned down by transplant surgeons during or after organ recovery surgery in the OR because of concerns regarding organ quality (Colpart et al. 1999). In other words, there exists opportunity for improvement in the allocation of scarce resources based on potential-donor types.

Proposition 2 points out several instances where $f_{1i}^S \leq \tilde{f}_i^1$ or $f_{2i}^S \leq \tilde{f}_2$, i.e., that the equilibrium fractions that result in the current scenario are not socially optimal. Consider the following two cases: (i) $\tilde{f}_i > f_{1i}^S \forall i \in \{1, 2\}$. For example, a sufficiently large $q_a$ would lead to this relationship. In this case, it is socially optimal for the hospital or the OPO to exert a lower effort level than in the current scenario due to the adverse effect on OR congestion and, hence, on QALY outcomes. However, this would conflict with the OPO’s volume-of-care objective, and, possibly, the hospital’s objective. (ii) $\tilde{f}_i < f_{1i}^S \forall i \in \{1, 2\}$. For example, a sufficiently small $q_a$ would lead to this relationship. Since the OPO always exerts its maximum effort level (i.e., $\xi_o^* = \theta$), in this case, it is socially optimal for the hospital to exert greater effort than in the current scenario in order to increase the rate of authorized donors, although this would not be privately optimal for the hospital.

The problem at hand can be viewed as a principal-agent problem, wherein a contract needs to be designed if the social planner (principal) intends for the hospital and the OPO (agents) to
make operational decisions that would lead to the socially-optimal fractions of type 1 and type 2 potential donors being converted to authorized donors. Contracts can be designed that either: (i) directly specify the OPO’s level of effort, and the hospital’s level of effort and OR scheduling policy, or (ii) alter the objectives of the OPO and the hospital to induce the desired operational actions and, thus, organ recovery outcomes. We focus on the latter because OPOs operate under a federal mandate that requires them to collect data on missed referrals by hospitals and time stamps related to key organ recovery milestones (e.g., administration of various medical tests and drugs, intra-hospital transfers, etc.). Before discussing the design of socially-optimal contracts, we present a numerical illustration that is grounded in practice, in order to reinforce the context of our analysis thus far.

4.3. Numerical Illustration

Using numerical values that are grounded in practice, we illustrate current misalignments in the objectives of the social planner, the OPO, and the hospital. We use the following representative values for our model parameters based on the medical literature and interviews with OPO staff:

(i) $\lambda_h = 400$ patients per month (based on the number of OR procedures at a large donor hospital in the OPO’s DSA; AHD 2016), and $\rho_h = 0.8$ (based on estimates reported in the set of studies reviewed by Cardoen et al. 2010, p. 924). \footnote{We present another numerical illustration in Appendix C with a lower $\rho_h$, and show that our findings are qualitatively similar.}

(ii) According to the medical literature, kidney transplants add about 4.7 QALYs for recipients, as compared to about 0.8 QALYs added by hip arthroplasty and 0.66 QALYs added by bypass surgery (CEAR 2017, Held et al. 2016). Accordingly, we use $\frac{Q_{a1}}{Q_h} = 6$. Also, since transplants of expanded-criteria or marginal-quality kidneys add about 15% fewer QALYs in contrast to non-expanded-criteria kidneys (Ojo 2005), we use $\frac{Q_{a2}}{Q_{a1}} = 0.85$.

(iii) Based on data provided by OPO staff for kidneys, $\lambda_p = 12$ per month, $R_{af} = 200,000$ USD per donor, and $R_{av} = 2,000$ USD per donor per day. Also, we use $\lambda_p2 = 4$ per month (on
average, marginal donors constitute approximately 30-34% of all kidney donors; Klein et al. 2010), and \( \frac{1}{\mu_a} = 4 \) hours (LifeSource 2016).

We assume that \( \theta = 1 \) and \( \tau = 0.8 \), and choose appropriate values for the remaining model parameters \((c_h, c_e, \nu_a, \text{and } \nu_h)\) in order to examine the conditions discussed in Section 4.1: (a) Conditions \( \Theta_1 \): A sufficiently low \( c_e \) and a sufficiently low \( c_h \) lead to \( \xi^*_h = 1 \), in which case, \( \tilde{f}_1 = \tilde{f}_2 = \tilde{f} = \tau \xi^*_h \xi^*_e = 0.8 \); and (b) Conditions \( \Theta_2 \): A sufficiently high \( c_e \) and a sufficiently low \( c_h \) lead to \( 0 < \xi^*_h < 1 \). Here, we choose parameter values such that \( \xi^*_h = 0.80 \), in order to match the resulting value of \( f_h \) with the average referral rate reported in Sheehy et al. (2003). Thus, in this case, \( \tilde{f}_1 = \tilde{f}_2 = \tilde{f} = \tau \xi^*_h \xi^*_e = 0.64 \).

The socially optimal fractions of type 1 and type 2 potential donors who should end up as authorized donors, or \( f_1^S \) and \( f_2^S \), take values according to Proposition 2 in Section 4.2. Based on the model parameters outlined above, we obtain \( \frac{1}{\tilde{q}_a} \approx 25.4 \) hours and \( \frac{1}{\tilde{q}_a} \approx 16.9 \) hours. These values are in reasonable agreement with the medical literature that recommends recovery of kidneys from cadaver donors within about 24 hours of declaration of BD (e.g., Blasco et al.
Figure 3 includes plots of the equilibrium fractions ($\tilde{f}_1 = \tilde{f}_2 = \tilde{f}$) of potential donors who currently end up as authorized donors and the socially-optimal fractions ($f_1^S$ and $f_2^S$) who should end up as authorized donors, with respect to the sensitivity ($q_a$) of QALYs added for the organ recipient to the delay experienced by the organ donor while waiting in the OR queue.

The plots in Figure 3 show that, under both sets of conditions ($\Theta_1$ and $\Theta_2$), there exist regions wherein the equilibrium fractions in the current scenario are not socially optimal. Under $\Theta_1$, $f_1^S < \tilde{f}_1$ for $q_a > \tilde{q}_a$ and $f_2^S < \tilde{f}_2$ for $q_a > \tilde{q}_a$ (see Figure 3(a)); and under $\Theta_2$, $f_1^S > \tilde{f}_1$ for $q_a < \tilde{q}_a$ and $f_2^S > \tilde{f}_2$ for $q_a < \tilde{q}_a$ (see Figure 3(b)). Thus, the numerical illustration highlights the realistic possibility of the incentive misalignments discussed in Section 4.2.

5. Socially Optimal Contracts

We propose a multiparameter contract to address the misalignments between the social planner’s objective, and the OPO’s and hospital’s objectives. With regard to the OPO, the social planner needs to incentivize the OPO to also take into account donor heterogeneity and the adverse impact of OR congestion on the hospital’s other patients. Accordingly, we recommend that the social planner revise the OPO’s objective function ($\hat{\pi}_{opo}$) to be based on composite criteria. Specifically, the OPO’s revised objective function in (7) below comprises: (i) QALY-weighted volume components corresponding to each type of potential donors, and (ii) a suitably-weighted ($\alpha > 0$) volume-based component that addresses the adverse impact on QALYs added for the hospital’s other patients.

$$\hat{\pi}_{opo} = f_1 \lambda_{p1} Q_{a1} (1 - q_a w_{a1}) + f_2 \lambda_{p2} Q_{a2} (1 - q_a w_{a2}) - \alpha (f_1 \lambda_{p1} + f_2 \lambda_{p2})$$  (7)

With the revised objective function, the OPO would be required to differentiate its operational actions based on donor heterogeneity; therefore, $f_i = \tau_i \xi_h \xi_{oi}$, where $\xi_{oi} \in [0, \theta_i]$ and $0 < \theta_i \leq 1$; $i \in \{1, 2\}$. Per CMS Ruling 45 CFR §164.512(h), once a referral call is made to the OPO, the assigned OPO personnel can access medical records (e.g., health history and test results) of
the potential donor. Based on this information, the OPO can a priori determine the potential donor’s type. Let \( \hat{\xi}_{oi} \) denote the OPO’s effort level for type \( i \) potential donors referred by the hospital, that maximizes \( \hat{\pi}_{opo} \).

Next, we suggest that the social planner levy the following two penalties on the hospital: (i) a penalty (at rate \( p_m \)) on the hospital for each missed referral, and (ii) a penalty (at rate \( p_d \) per unit wait time) on the hospital for the average wait time experienced by type 1 authorized donors in excess of the average wait time for them under the proposed centralized scenario (see Section 4.2). With these penalties, the hospital’s revised payoff is:

\[
\hat{\pi}_h = \lambda_a R_{af} + (f_1 \lambda_{p1} w_{a1} + f_2 \lambda_{p2} w_{a2}) R_{av} + \lambda_h R_h - c_h w_h - \frac{c_e}{2} \hat{\xi}_h^2 - \lambda_p (1 - f_h) p_m - (w_{a1} - w_{a1}^S) p_d \tag{8}
\]

Where, \( \lambda_a = \tau \xi_h (\xi_{o1} \lambda_{p1} + \xi_{o2} \lambda_{p2}) \) and \( w_{a1}^S := w_{a1} (f^S_1, f^S_2, \chi^S) \). We denote the hospital’s optimal effort level and OR scheduling policy under the contract by \( \hat{\xi}_h^* \) and \( \hat{\chi}_h^* \), respectively. Denote \( \hat{\pi}_h^* = \hat{\pi}_h(\hat{\xi}_h^*, \hat{\chi}_h^*) \).

In practice, hospitals do not bear any financial penalties for adverse ODVC performance outcomes, except maybe for annotations in accreditation reports prepared by federal agencies such as the Joint Commission on Accreditation of Healthcare Organizations, for choosing suboptimal organ recovery effort levels that lead to missed referrals (see the DHHS report by Inspector General Brown 2000; p. 19–20). This DHHS report recommends that some kind of mechanism (incentive or disincentive) be implemented based on data that is already being collected, in order to improve hospitals’ compliance with the donation rule. However, no such mechanism exists in practice. OPOs already audit patient records of hospitals in their respective DSA at regular intervals (monthly or quarterly) to collect data on missed referrals and time stamps related to key organ recovery milestones, so there would be limited additional administration costs for implementing these penalties based on missed referrals and wait times experienced by type 1 authorized donors. There are other examples of penalties that have been introduced to
influence the operational actions of hospitals and improve societal outcomes. One example is the Readmission Penalty that was introduced by CMS in 2012 to curb the loss in quality of care and substantial costs incurred due to avoidable re-hospitalizations of Medicare beneficiaries (Berenson et al. 2012, CMS 2012). A few recent studies have analyzed the effects of this penalty on hospitals’ operational actions such as readmission-reduction efforts (Andritsos and Tang 2015, Zhang et al. 2016). Another example is the Hospital-Acquired Condition (HAC) Reduction Program introduced by CMS in 2014, that penalizes hospitals that perform poorly on HAC quality measures (CMS 2014).

The set of contractual levers \{\alpha, p_m, p_d\} outlined above can help the ODVC attain the socially optimal payoff \(\pi^S\) (see Section 4.2) and, when appropriately specified, can even ensure that both the OPO and the hospital are strictly better-off with the implementation of the contract. For the hospital, strictly better-off implies that its equilibrium payoff under the contract attains a larger value compared to the current (uncoordinated) scenario. For the OPO, strictly better-off means that its total equilibrium effort under the contract (= \(\hat{\xi}^*_o + \hat{\xi}^*_o\)) is less than \(2\xi^*_o\), where \(\xi^*_o\) is the OPO’s equilibrium effort for each type of potential donors under the current scenario (see Section 4.1). Note that the factor of 2 in the aforementioned comparison arises because, under the current scenario we have \(f_1 = f_2 = \tau\xi_h\xi_o\), and under the contract we have \(f_1 = \tau\xi_h\xi_{o1}\) and \(f_2 = \tau\xi_h\xi_{o2}\).

We focus on Pareto-improving contracts for their promise of acceptability. In our context, Pareto improvement implies that none of the three entities, namely, the OPO, the hospital, or the social planner is worse-off, and at least one of them is strictly better-off with the implementation of the contract. The suggested contractual levers in Proposition 3 not only influence the operational actions of the hospital and the OPO to help the ODVC achieve socially-optimal performance, but also achieve strict Pareto improvement for the OPO and the hospital.
Proposition 3. \( \exists \bar{p}_m > 0 \) and \( \bar{p}_d > 0 \) s.t. \( \forall p_m > \bar{p}_m, p_d > \bar{p}_d, \) and \( \alpha = \frac{Q_h q_h \lambda_h v_a}{2(1-p_m)} \), the set of contractual levers \( \{\alpha, p_m, p_d\} \) ensures: (i) Social optimality, i.e., \( \hat{\pi}_S(\alpha, p_m, p_d) = \pi^S \) and, (ii) Strict Pareto improvement for the hospital and the OPO, i.e., \( \hat{\pi}^*_h > \pi^*_h \) and \( \hat{\xi}^*_o + \hat{\xi}^*_o < 2 \xi^*_o \).

As stated in Section 4.2 (proposed centralized scenario), the OR scheduling policies that are respectively optimal for the social planner and the hospital, are divergent (\( \chi^S = I \) and \( \chi^*_h = II \)). Under the revised objective function for the hospital, choosing \( p_d > \bar{p}_d \) ensures that the hospital accords absolute priority to authorized donors over its other patients, i.e., \( \hat{\chi}^*_h = I \) (see Proof of Proposition 3). Recall that the OPO is the Stackelberg follower. An appropriately-chosen weight \( \alpha \) in the OPO’s revised objective function ensures that the OPO adjusts its effort levels for both types of referred donors in response to the hospital’s chosen effort level, thereby balancing QALY outcomes across organ recipients and the hospital’s other patients. The social planner can always choose a value of \( p_m \) such that \( \hat{\xi}^*_o + \hat{\xi}^*_o < 2 \xi^*_o \), so that the OPO is strictly better-off under the contractual mechanism in Proposition 3 as compared to the current scenario.

Setting \( \alpha \) and \( p_d \) as in Proposition 3, effectively fixes the overall fraction \( (f) \) of potential donors who end up as authorized donors because \( \lambda_a = f_1^S \lambda_{p1} + f_2^S \lambda_{p2} \), which implies that \( f = \frac{f_1^S \lambda_{p1} + f_2^S \lambda_{p2}}{\lambda_{p1} + \lambda_{p2}} \). If the penalty for missed referrals \( (p_m) \) were zero, the hospital’s payoff would be decreasing in \( \xi_h \), implying that the equilibrium effort level of the hospital under the contract would be zero. By levying a penalty \( (p_m > 0) \) on the hospital for missed referrals, the social planner induces an interesting dynamic between the OPO and the hospital. In the presence of a non-zero penalty for missed referrals, it becomes costly for the hospital to not exert effort towards organ recovery activities. However, because of the presence of the QALY-based components in the OPO’s revised objective, the OPO (Stackelberg follower) responds to an increased effort level by the hospital by substantially reducing its own effort levels for type 1 and/or type 2 potential donors referred by the hospital. The resulting equilibrium decisions of the OPO and the hospital end up being beneficial for the hospital because of decreased OR congestion.
and, hence, lower wait times experienced by the hospital’s other patients. The reduced OR congestion also favorably impacts QALY outcomes and helps the ODVC achieve socially-optimal performance.

6. Conclusion

While the majority of the healthcare operations management literature focuses on the demand side of the ODVC, we develop an analytical model to study the effects of contextual parameters and operational actions of the supply-side entities (OPO and hospital) on their respective payoffs and on societal outcomes. Our analysis of the current (uncoordinated) and proposed centralized scenarios reveals several key findings. First, we show that higher effort levels towards organ recovery activities chosen by either the OPO or the hospital have counteracting effects on the societal outcome (assessed based on quality-adjusted life years, or QALYs). Although higher effort levels lead to larger fractions of conversions of potential donors to authorized donors, another resulting effect is greater congestion at the OR. Consequently, authorized donors as well as the hospital’s other patients experience longer wait times in the OR queue, resulting in adverse impacts on QALY outcomes.

Second, in contrast to current ODVC policies and practices that do not require the hospital or the OPO to differentiate their operational actions based on the quality of organs that can be recovered from potential donors, our analysis shows that the social planner’s objective may be non-monotonic in the respective fractions of the different types of potential donors who end up as authorized donors. Using a numerical illustration that is grounded in practice, we demonstrate the realistic possibility of incentive misalignments between the social planner, and the OPO and hospital. These misalignments lead to socially suboptimal fractions of organs recovered from the different types of potential donors.

Third, we recommend a multiparameter contract to address the misalignments between the social planner’s objective, and the OPO’s and hospital’s objectives. We show that by: (i)
including QALY-based components corresponding to each type of potential donors and an appropriately-weighted volume-based component in the OPO objective (i.e., revising the OPO’s objective to reflect composite criteria); and (ii) levying appropriately-chosen penalties on the hospital for missed referrals and for socially suboptimal delays experienced by authorized donors while waiting for the OR to become available; the social planner can help the ODVC achieve not only socially-optimal performance, but also strict Pareto improvement for the OPO and the hospital. This contractual mechanism is administratively feasible since OPOs already collect data on missed referrals by hospitals and time stamps related to key organ recovery milestones, and it also has the promise of acceptability since no player would be worse-off.

Under our recommended contractual mechanism, it may be possible for the resource-constrained OPO to increase the allocation of its efforts towards accomplishing other important tasks, such as fundraising, public awareness campaigns, and community engagement. Also, the OPO may be able to allocate more effort towards assisting hospital staff in lowering their discomfort associated with organ donation activities (e.g., through training programs on cultural sensitivities towards organ donation, psychological support skills to enhance end-of-life care, etc.; DOT Grant Program Report 2005). As an example, New England Organ Bank, the OPO for the six New England states, has instituted educational programs for HCPs on psychological support skills in discussions surrounding death and dying, that have led to a significant increase in the number of referrals of potential donors by hospitals in its DSA.

Consistent with studies that consider principal-agent problems in the non-profit and healthcare contexts (e.g., Devalkar et al. 2016, Gupta and Mehrotra 2015), we consider that the social planner (the principal), strives to optimize the benefit delivered to organ recipients and the hospital’s other patients. Thus, in our setting, the social planner’s payoff is different from the sum of the payoffs of the individual players. However, our results remain structurally similar if we add the hospital’s and OPO’s objectives within the social planner’s payoff. Also, in our work,
we consider the hospital to be a profit-maximizing entity. Although empirical studies (Deneffe and Masson 2002, Duggan 2000) find that not-for-profit hospitals behave in a similar manner as profit-maximizing entities, including in the context of ODVCs (Rios-Diaz et al. 2017), if the hospital too had a volume-of-care component in its objective, the incentive misalignments would be less severe and the interventions from the social planner for both the OPO and the hospital would be in the form of composite criteria (i.e., QALY-based components added to the existing objective). Apart from contractual levers, our model points to other potentially viable operational interventions for improving organ recovery outcomes: expanding OR capacity (increasing $\mu$), improving the waiting experience in the OR queue (decreasing $c_h$), and emphasizing HCP training to reduce the discomfort associated with organ recovery activities (decreasing $c_e$).

With regard to limitations of our work and avenues for future research, we note that our model focuses on one focal organ whereas multiple organs can potentially be recovered from a cadaveric donor. It will be interesting to study the privately- and socially-optimal levels of efforts by the OPO and the hospital, as well as the OR priorities accorded by the hospital in the presence of heterogeneity in organ yields across potential donors or heterogeneity in the characteristics of organs (e.g., sensitivities of different organs to delays experienced by the donors). Also, as discussed earlier, the existing OM literature focuses on the demand side of the ODVC whereas our paper adds to the sparse literature on the supply side. We believe that there will be value in capturing the interplay between demand- and supply-side actions in order to further improve societal outcomes. For instance, it will be valuable to study the interplay between the operational actions in the organ recovery process and the trade-offs involved in the allocation of organs to patients on the waitlist.

**Acknowledgments**

The authors gratefully acknowledge Dr. Shri Bharadwaj (Lifelink of Georgia), Dr. Sundaresan Jayaraman (Georgia Tech) and Dr. Ram Subramanian (Emory University) for their collaboration on a National Institute of Health (NIH) proposal submitted in 2012 along with the second author. The proposal underscored the need to design and develop a framework for improving the performance of the ODVC for kidneys.
References


Appendix A: Example of a Trigger Card issued by an OPO

```
ORGAN DONATION
CLINICAL TRIGGERS

Your patient is intubated and shows evidence of the following:
- Coma
- Stroke
- Hypoxia
- Brain Tumor
- Cerebral Injury
- Near-Drowning
- Cerebral Edema
- Cerebral Hemorrhage

AND

Any of the following criteria are met:
- GCS≤5, not sedated
- Unresponsive or posturing
- No pupillary or corneal reflex
- No cough or gag
- No spontaneous respiration
- Discussion of DNR or withdrawal of support
```

Appendix B: Proofs

Let \( w^x \) denote the average wait time in OR queue for each patient class, where: \( x \in \{a, h\} \) and priority \( \chi \in \mathcal{A} \) in the current scenario; \( x \in \{a1, a2, h\} \) and priority \( \chi \in \mathcal{A}_S \) in the proposed centralized scenario.

**PROOF OF PROPOSITION 1.** From the discussion following (3) and (4), we have:

\[
w^{I}_a = \frac{(\lambda_a \nu_a + \lambda_h \nu_h)}{2(1 - \rho_a)}, \quad w^{I}_h = \frac{(\lambda_a \nu_a + \lambda_h \nu_h)}{2(1 - \rho_a - \rho_h)}, \quad w^{II}_a = \frac{(\lambda_a \nu_a + \lambda_h \nu_h)}{2(1 - \rho_h)}, \quad \text{and} \quad w^{II}_h = \frac{(\lambda_a \nu_a + \lambda_h \nu_h)}{2(1 - \rho_h)}. \]

Under Assumption A1, \( w^x_h \equiv \frac{(\lambda_a \nu_a + \lambda_h \nu_h)}{2(1 - \rho_h)} \) \( \forall \chi \in \mathcal{A} \). Substituting this expression in (3) and (4), we obtain the following relationship:

\[
w^a(f, \chi = I) < w_a(f, \chi) < w^a(f, \chi = II) \forall \chi \in \mathcal{A}_S \setminus \{I, II\}. \]

Since \( \pi_h \) is increasing in \( w_a \), we obtain \( \chi^*_h = II \). Substituting the expression for \( w^II_a \) in place of \( w^a \), and substituting \( f = \tau \xi_a \xi_h = \tau \theta \xi_h \) in (1), we obtain:

\[
\frac{\partial \pi_h}{\partial \xi_h} = \left( R_a f + R_a \frac{\lambda_h \nu_h}{2(1 - \rho_h)} - c_h \frac{\nu_a}{2(1 - \rho_h)} \right) \tau \theta \lambda_p + \xi_h \left( R_a \frac{\lambda_h \nu_a}{2(1 - \rho_h)} - c_e \right) =: B + \xi_h A
\]

where \( A = R_a \frac{\lambda_h \nu_a}{2(1 - \rho_h)} - c_e \), and \( B = \left( R_a f + R_a \frac{\lambda_h \nu_h}{2(1 - \rho_h)} - c_h \frac{\nu_a}{2(1 - \rho_h)} \right) \tau \theta \lambda_p \). Denote \( \bar{c}_e := R_a \frac{\lambda_h \nu_a}{2(1 - \rho_h)} \), and \( \bar{c}_h := R_a \frac{\nu_a}{\nu_a} + R_a \frac{\lambda_h \nu_h}{\nu_a} \). Note that \( c_e \leq \bar{c}_e \Leftrightarrow A \geq 0 \), and \( c_h \leq \bar{c}_h \Leftrightarrow B \geq 0 \).

When \( A < 0 \) (or \( c_e > \bar{c}_e \)), \( \pi_h \) is concave in \( \xi_h \), otherwise it is convex (because \( \frac{\partial^2 \pi_h}{\partial \xi_h^2} = A \)).

Therefore, the hospital’s objective and optimal effort level (\( \xi^*_h \)) can be characterized by the following four exhaustive cases:
(i) \( c_e > \bar{c}_e \) and \( c_h \geq \bar{c}_h \), or \( A < 0 \) and \( B \leq 0 \) \( \Rightarrow \frac{\partial \pi_h}{\partial \xi_h} < 0 \) \( \Rightarrow \xi_h^* = 0 \).

(ii) \( c_e > \bar{c}_e \) and \( c_h < \bar{c}_h \), or \( A < 0 < B \): (a) \( \frac{B}{A} < 1 \) \( \Rightarrow \pi_h \) is concave (unimodal) and \( \xi_h^* = \frac{B}{A} < 1 \). (b) \( \frac{B}{A} > 1 \) \( \Rightarrow \frac{\partial \pi_h}{\partial \xi_h} \geq 0 \) \( \Rightarrow \pi_h \) is concave increasing and \( \xi_h^* = 1 \).

(iii) \( c_e < \bar{c}_e \) and \( c_h < \bar{c}_h \), or \( A > 0 \) and \( B > 0 \) \( \Rightarrow \frac{\partial \pi_h}{\partial \xi_h} > 0 \) \( \Rightarrow \xi_h^* = 1 \).

(iv) \( c_e < \bar{c}_e \) and \( c_h \geq \bar{c}_h \), or \( B \leq 0 < A \): \( A > 0 \) implies that \( \pi_h \) is convex in \( \xi_h \). The following two sub-cases arise from comparing \( \pi_h\mid_{\xi_h=1} \) and \( \pi_h\mid_{\xi_h=0} \): Denote \( \bar{c}_h(c_e) = \bar{c}_h + (\bar{c}_e - c_e) \left( \frac{1-\rho_h}{\lambda_p \theta^2 \nu_a} \right) \);

Case (a): For \( \bar{c}_h \leq c_h < \bar{c}_h(c_e) \), \( \pi_h\mid_{\xi_h=1} - \pi_h\mid_{\xi_h=0} = \frac{A}{2} + B \geq 0 \) \( \Rightarrow \xi_h^* = 1 \); and, Case (b): For \( c_h \geq \bar{c}_h(c_e) \), \( \pi_h\mid_{\xi_h=1} - \pi_h\mid_{\xi_h=0} = \frac{A}{2} + B \leq 0 \) \( \Rightarrow \xi_h^* = 0 \). Note that if \( B = -\frac{A}{2} \), the hospital is indifferent between \( \xi_h^* = 1 \) and \( \xi_h^* = 0 \).

PROOF OF PROPOSITION 2. We first show that \( \chi^S = I \). Note that, since \( w^\chi_{a_1} \equiv w^\chi_{a_2} \forall \chi \in \mathcal{A}_S \), we drop 1 or 2 from the subscript such that \( w^\chi_a \) denotes the average wait time experienced by authorized donors of either class. Under Assumption A1, \( w^\chi_h \equiv \frac{(\lambda_a \nu_a + \lambda_h \nu_h)}{2(1-\rho_h)} \forall \chi \in \mathcal{A} \). Since \( \pi_S \) decreases in \( w_a \), and \( w_a(f_1, f_2, \chi = II) > w_a(f_1, f_2, \bar{\chi}) > w_a(f_1, f_2, \chi = I) \forall \bar{\chi} \in \mathcal{A}_S \setminus \{I, II\} \), we have \( \chi^S = I \).

From the discussion following (5) and (6), we obtain \( w^I_a \equiv \frac{f_1 \lambda_p \nu_a + f_2 \lambda_p \nu_a + \lambda_h \nu_h}{2} \). Substituting the expression for \( w^I_a \) in place of \( w_{a_1} \) and \( w_{a_2} \) in (2), we get:

\[
\frac{\partial \pi_S}{\partial f_1} = \lambda_p 1 \left( Q_{a_1} - Q_{a_1} f_1 q_a \lambda_p \nu_a - (Q_{a_1} + Q_{a_2}) f_2 q_a \nu_a \right),
\end{equation}

\[
\frac{\partial \pi_S}{\partial f_2} = \lambda_p 2 \left( Q_{a_1} - Q_{a_1} f_2 q_a \lambda_p \nu_a - (Q_{a_1} + Q_{a_2}) f_1 q_a \nu_a \right)
\]

Denoting \( \lambda_p := \min\{\lambda_p, \lambda_p, \tilde{\lambda}_p := \max\{\lambda_p, \lambda_p, \tilde{\lambda}_p \} \), we consider the three cases in the statement of Proposition 2:

(i) \( q_a < \bar{q}_a := \frac{2Q_{a_1}(1-\rho_h) - Q_{a_1}q_a \lambda_h \nu_h}{(1-\rho_h)[Q_{a_1}(\lambda_h \rho_h + \theta \lambda_p \nu_a + 2\theta \lambda_p \nu_a) + Q_{a_2} \theta \lambda_p \nu_a + Q_{a_2} \theta \lambda_p \nu_a]} \Rightarrow \frac{\partial \pi_S}{\partial f_1} > 0 \) and \( \frac{\partial \pi_S}{\partial f_2} > 0 \) \( \forall f_1 \in [0, \tau \theta] \) and \( f_2 \in [0, \tau \theta] \), i.e., \( \pi_S \) is maximized when \( f_1^S = \tau \theta \) and \( f_2^S = \tau \theta \).

(ii) \( \bar{q}_a \leq q_a < \bar{q}_a := \frac{2Q_{a_1}(1-\rho_h) - Q_{a_1}q_a \lambda_h \nu_h}{(1-\rho_h)[Q_{a_1}(\lambda_h \rho_h + \theta \lambda_p \nu_a + 2\theta \lambda_p \nu_a) + Q_{a_2} \theta \lambda_p \nu_a + Q_{a_2} \theta \lambda_p \nu_a]} \Rightarrow \frac{\partial \pi_S}{\partial f_1} > 0 \) \( \forall f_1 \in [0, \tau \theta] \) and \( f_2 \in [0, \tau \theta] \Rightarrow f_1^S = \tau \theta \). Also, \( \frac{\partial \pi_S}{\partial f_2} \bigg|_{f_1=\tau \theta} \) decreases linearly in \( f_2 \) such that \( \frac{\partial \pi_S}{\partial f_2} \bigg|_{f_1=\tau \theta, f_2=0} > 0 \).
and \( \frac{\partial \pi_S}{\partial f_1} |_{f_1=\tau, f_2=\tau} < 0 \Rightarrow \pi_S(f_1 = \tau) \) is unimodal in \( f_2 \) and \( f_2^S \) can be obtained by equating \( \frac{\partial \pi_S}{\partial f_2} |_{f_1=\tau} \) to 0. Therefore, \( f_2^S = \max \left\{ \frac{(1-\rho_b)[Qa_1(2-qa_1\lambda_h\nu_h-q\tau\theta\lambda_p\nu_a)-Qa_2qa_2\theta\lambda_p\nu_a]-Qh, qh, \lambda_h\nu_h}{2Qa_1(1-\rho_b)qa_2\lambda_p\nu_a}, 0 \right\} \).

(iii) Analogous to (ii) above, when \( qa \geq \bar{q}_a \), we have that \( \pi_S \) is unimodal in \( f_1 \forall f_2 \in [0, \tau] \). Equating \( \frac{\partial \pi_S}{\partial f_1} \) to 0, we have \( f_1(f_2) = \frac{(1-\rho_b)[Qa_1(2-qa_1\lambda_h\nu_h-q_2f_2\lambda_p\nu_a)-Qa_2qa_2\lambda_p\nu_a]-Qh, qh, \lambda_h\nu_h}{2Qa_1(1-\rho_b)qa_2\lambda_p\nu_a} \). Substituting this expression in place of \( f_1 \) in (B2), we have \( \frac{\partial \pi_S}{\partial f_2} < 0 \Rightarrow f_2^S = 0 \). Replacing \( f_2 \) with 0 in the expression for \( f_1^S \), we obtain \( f_1^S = \min \left\{ \frac{(1-\rho_b)[Qa_1(2-qa_1\lambda_h\nu_h-q_2f_2\lambda_p\nu_a)-Qh, qh, \lambda_h\nu_h]}{2Qa_1(1-\rho_b)qa_2\lambda_p\nu_a}, \tau \right\} \). ■

**PROOF OF PROPOSITION 3.**

(i) Social optimality of the recommended contractual mechanism: Replacing \( w_{a_1}^S \equiv \frac{f_1^S\lambda_p\nu_a+f_2^S\lambda_p^2\nu_a+\lambda_h\nu_h}{2} \) in (8), we have that \( \hat{\pi}_h \big|_{\chi_h=I} > \hat{\pi}_h \big|_{\chi_h=II} \) iff \( pd > R_{av}(f_1^S\lambda_p + f_2^S\lambda_p^2) \). Denote \( \tilde{p}_d = R_{av}(f_1^S\lambda_p + f_2^S\lambda_p^2) \). Therefore, under the contract, it follows that \( pd > \tilde{p}_d \Rightarrow \hat{\chi}_h^* = I \).

Under Assumption A1, \( w_{a_1}^\chi \equiv w_{a_2}^\chi \forall \chi \in \mathcal{A}_S \). After substituting \( w_a^I = \frac{f_1\lambda_p\nu_a+f_2\lambda_p^2\nu_a+\lambda_h\nu_h}{2} \) for \( w_{a_1} \) and \( w_{a_2} \) in (7), we equate the expressions for \( \frac{\partial \pi_{opo}}{\partial f_1} \) and \( \frac{\partial \pi_{opo}}{\partial f_2} \) to the right hand sides of (B1) and (B2), respectively, to ensure that the values of \( f_1 \) and \( f_2 \) that maximize \( \hat{\pi}_{opo} \) are also socially optimal. Thus, we obtain \( \alpha = \frac{Qh, qh, \lambda_h\nu_h}{2(1-\rho_b)} \) and we have \( \pi_S \left( \alpha = \frac{Qh, qh, \lambda_h\nu_h}{2(1-\rho_b)}, pm, pd > \tilde{p}_d \right) = \pi_S^* \).

(ii) Strict Pareto improvement for the hospital and the OPO: Under the given contractual mechanism, \( \frac{\partial \hat{\pi}_h}{\partial \chi_h} = -c_e\lambda_h + pm\lambda_p \). \( \frac{\partial \hat{\pi}_h}{\partial \chi_h} = -c_e < 0 \Rightarrow \hat{\pi}_h \) is maximized at \( \hat{\chi}_h^* = \frac{pm\lambda_p}{c_e} \). Under the contract, let \( f_S := \frac{f_1^S\lambda_p+f_2^S\lambda_p^2}{\lambda_p+f_p^2} \) denote the overall fraction of potential donors who end up as authorized donors. Substituting \( f_S \) and \( \hat{\chi}_h^* \) in (8) after expanding, we specify the conditions that ensure that the hospital is strictly better-off under the contract (i.e., \( \hat{\pi}_h^* - \pi_h^* > 0 \)) in the following three exhaustive cases:

(a) Conditions \( \Theta_2 \) (i.e., where \( 0 < \chi_h^* = \frac{B}{1} < 1 \) and \( \hat{f} < f_S(1-\rho_h) \)): Here, it is sufficient to show that the following inequality holds: \( ap_m^2 + bp_m + c \geq 0 \), where \( a = \lambda_p^2, b = -2c_e\lambda_p, \) and \( c = c_e B^2 - c_e \frac{\alpha}{1-\rho_h} \left[ f_S(1-\rho_h) - (\tau\theta B A) \right] \lambda_p\nu_a + 2c_e R_{av} f_p \lambda_p \left[ f_S - (\tau\theta B A) \right] \). The discriminant of this quadratic equation is always \( \geq 0 \). This implies that there exist two real roots. Since the sum of
the two roots $\frac{-b}{a} = \frac{2c}{\lambda_p} > 0$, it follows that there exists at least one root (say, $p_1$) that is $> 0 \Rightarrow ap^2_m + bp_m + c > 0 \forall p_m > p_1$.

(b) Conditions $\Theta_2$ and $\tilde{f} > f^S(1 - \rho_h)$: In this case, the condition that ensures $\hat{\pi}_h^* - \pi_h^* > 0$ can be written as $qp^2_m + rp_m + s > 0$, where $q = \lambda_p^2$, $r = -2c_e\lambda_p$, $s = -2\frac{c_eB^2}{A} - c_eR + c_e\frac{\epsilon_h}{1-\rho_h} (\tau\theta B - f^S)\lambda_p\nu_a$, and $R = R_{af} (\tau\theta B - f^S) + R_{av} \frac{\lambda_p}{1-\rho_h} \left[ \left( \frac{\theta^2B^2}{A} - (f^S)^2 (1 - \rho_h) \right) \lambda_p\nu_a + (\tau\theta B - f^S (1 - \rho_h))\lambda_h\nu_h \right]$. By following the same line of reasoning as in case (a) above, we can show that there exists at least one root (say, $p_2$) that is $> 0 \Rightarrow qp^2_m + rp_m + s > 0 \forall p_m > p_2$.

(c) Conditions $\Theta_1$ (i.e., where $\xi_h^* = 1$) and $\tilde{f} > f^S(1 - \rho_h)$: Note that conditions $\Theta_1$ imply $\tilde{f} > f^S(1 - \rho_h)$. The proof of case (c) is similar to that of case (b) above, by replacing $\frac{B}{A}$ with 1. For case (c), let $p_3$ denote the threshold such that $\hat{\pi}_h^* - \pi_h^* > 0 \forall p_m > p_3$.

Rearranging the terms in the relationships $f^S_1 = \tau\hat{\xi}_h^*\hat{\pi}_h^*$ and $f^S_2 = \tau\hat{\xi}_h^*\hat{\pi}_h^*$, we have $\hat{\xi}_o1 = \frac{f^S_1}{\pi_h^*}$ and $\hat{\xi}_o2 = \frac{f^S_2}{\pi_h^*}$. For each of the cases in Proposition 2, it is straightforward to find the corresponding values of the penalty (say, $p_4$) such that $\hat{\xi}_o1^* + \hat{\xi}_o2^* < 2\xi_o^* = 2\theta$, or that the OPO is strictly better-off under the contract $\forall p_m > p_4$. Strict Pareto improvement for the hospital and the OPO follows by setting $\bar{p}_m = \max\{p_1, p_2, p_3, p_4\}$.

### Appendix C: Numerical Illustration (with a lower $\rho_h$)

In this section, we present a numerical illustration that considers a lower value of $\rho_h$ than in Section 4.3 (specifically, a higher $\mu_h$ for a given $\lambda_h$). Using the same representative values for our model parameters as in Section 4.3, and changing $\rho_h$ from 0.8 to 0.6, we obtain plots of the equilibrium fractions ($\hat{f}_1 = \hat{f}_2 = \hat{f}$) of potential donors who currently end up as authorized donors and the socially-optimal fractions ($f^S_1$ and $f^S_2$), with respect to the sensitivity ($q_a$) of QALYs added for the organ recipient to the delay experienced by the organ donor while waiting in the OR queue (see Figure C1). With these parameter choices, the obtained threshold values ($\bar{q}_a$ and $\bar{q}_a$) are higher than those in Section 4.3, implying that cases (i) and (ii) of Proposition
Figure C1  Numerical Illustration (with a lower $\rho_h$)

2 are more likely (or, that it would more likely be socially optimal for the OPO and the hospital to exert greater effort levels when $\rho_h$ is lower).

Similar to the plots in Section 4.3, the plots in Figure C1 show that, under both sets of conditions ($\Theta_1$ and $\Theta_2$), there exist regions where the equilibrium fractions in the current scenario are not socially optimal. For instance, as seen in Figure 3(b), under $\Theta_2$, $f_1^S > \tilde{f}_1$ when $q_a < \bar{q}_a$ and $f_2^S > \tilde{f}_2$ when $q_a < \bar{q}_a$. 

Note: This figure includes plots of the equilibrium fraction ($\tilde{f}_1 = \tilde{f}_2 = \tilde{f}$) of potential donors who currently end up as authorized donors, and the socially-optimal fractions ($f_1^S$ and $f_2^S$) who should end up as authorized donors, with respect to the sensitivity ($q_a$) of QALYs added for the organ recipient to the delay experienced by the organ donor while waiting in the OR queue.
### Table 1: Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BD</td>
<td>Brain Death</td>
</tr>
<tr>
<td>CMS</td>
<td>Centers for Medicare and Medicaid Services</td>
</tr>
<tr>
<td>DHHS</td>
<td>Department of Health and Human Services</td>
</tr>
<tr>
<td>DSA</td>
<td>Donor Service Area</td>
</tr>
<tr>
<td>HCP</td>
<td>Health Care Professional</td>
</tr>
<tr>
<td>ODVC</td>
<td>Organ Donation Value Chain</td>
</tr>
<tr>
<td>OPO</td>
<td>Organ Procurement Organization</td>
</tr>
<tr>
<td>OR</td>
<td>Operating Room</td>
</tr>
<tr>
<td>QALY</td>
<td>Quality-Adjusted Life Year</td>
</tr>
</tbody>
</table>

### Table 2: Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Weight of the volume-based component in the OPO’s revised objective</td>
</tr>
<tr>
<td>$C_e$</td>
<td>Cost of hospital’s efforts towards organ recovery activities</td>
</tr>
<tr>
<td>$c_e$</td>
<td>Coefficient in the cost to the hospital for its efforts towards organ recovery activities ($C_e$)</td>
</tr>
<tr>
<td>$C_h$</td>
<td>Cost to the hospital from the wait times experienced by the hospital’s other patients</td>
</tr>
<tr>
<td>$c_h$</td>
<td>Coefficient in the cost to the hospital from the wait times experienced by the hospital’s other patients ($C_h$)</td>
</tr>
<tr>
<td>$\chi$</td>
<td>Hospital’s OR scheduling policy</td>
</tr>
<tr>
<td>$f$</td>
<td>Overall fraction of potential donors who end up as authorized donors in the current scenario</td>
</tr>
<tr>
<td>$f_h$</td>
<td>Fraction of potential donors who end up as referred donors</td>
</tr>
<tr>
<td>$f_i$</td>
<td>Fraction of type $i$ potential donors who end up as authorized donors</td>
</tr>
<tr>
<td>$f_o$</td>
<td>Fraction of referred donors who end up as authorized donors</td>
</tr>
<tr>
<td>$f^o_i$</td>
<td>Socially-optimal fraction of type $i$ potential donors who should end up as authorized donors</td>
</tr>
<tr>
<td>$\bar{f}$</td>
<td>Equilibrium fraction of potential donors who end up as authorized donors (current scenario)</td>
</tr>
<tr>
<td>$\tilde{f}_i$</td>
<td>Equilibrium fraction of type $i$ potential donors who end up as authorized donors (current scenario)</td>
</tr>
<tr>
<td>$\lambda_x$</td>
<td>Arrival rate of class $x$ patients</td>
</tr>
<tr>
<td>$\mu_y$</td>
<td>Service rate (in the OR) for class $y$ patients</td>
</tr>
<tr>
<td>$\nu_y$</td>
<td>Second moment of the service time required in the OR for class $y$ patients</td>
</tr>
<tr>
<td>$p_d$</td>
<td>Penalty rate levied by the social planner on the hospital per unit wait time for type 1 authorized donors in excess of their average wait time under the proposed centralized scenario</td>
</tr>
<tr>
<td>$p_m$</td>
<td>Penalty rate levied by the social planner on the hospital for each missed referral</td>
</tr>
</tbody>
</table>

*Continued on next page*
**Notation – Continued from previous page**

- $Q_{ai}$: Average QALYs added to a recipient of the focal organ recovered from type $i$ potential donor.
- $q_a$: Sensitivity of QALYs ($Q_{ai}$) added to the delay experienced by the donor while waiting in the OR queue.
- $Q_h$: Average QALYs added to the hospital’s other patients who access the OR.
- $q_h$: Sensitivity of QALYs ($Q_h$) added to the delay experienced by the hospital’s other patients while waiting in the OR queue.
- $R_{af}$: Average fixed reimbursement rate per authorized donor.
- $R_{av}$: Average variable reimbursement rate per authorized donor per unit care time.
- $R_h$: Average fixed reimbursement rate (per-patient) associated with the hospital’s other patients.
- $\tau$: Donors’ or their families’ (exogenous) affinities towards organ donation.
- $\theta$: Upper limit on the OPO’s effort level.
- $w_y$: Average wait time in the OR queue for patients of class $y$.
- $\xi_h$: Hospital’s effort level towards organ recovery activities.
- $\xi_o$: OPO’s effort level in the current (uncoordinated) scenario.
- $\xi_{oi}$: OPO’s effort level for type $i$ potential donors under the contract.

**Note:** Different classes of patients are: authorized donors converted from the overall pool of potential donors ($a$); type $i$ authorized donors ($ai$), $i \in \{1, 2\}$; the hospital’s other patients ($h$); the overall pool of potential donors ($p$); and type $i$ potential donors ($pi$), $i \in \{1, 2\}$. Also, $x \in \{a, h, p, p1, p2\}$; $y \in \{a, a1, a2, h\}$. 