Characterizing Rational Transplant Program Response to Outcome-Based Regulation

David Mildebrath[1], Taewoo Lee[2], Saumya Sinha[1,3], Andrew J. Schaefer[1], and A. Osama Gaber[3]

[1]Department of Computational and Applied Mathematics, Rice University, Houston, TX

[2]Department of Industrial Engineering, University of Houston, Houston, TX

[3]Department of Surgery, Houston Methodist Hospital, Houston, TX

October 14, 2021

Abstract

Organ transplantation is an increasingly common therapy for many types of end-stage organ failure, including lungs, hearts, kidneys and livers. The past twenty years have seen increased scrutiny of post-transplant outcomes in the United States, in order to ensure the efficient utilization of the scarce organ supply. Under regulations by the Organ Procurement Transplantation Network (OPTN) and Centers for Medicare and Medicaid (CMS), the United States has seen a rise in risk-averse patient selection among transplant programs, resulting in decreased transplantation volume for some programs. Despite this observed decrease, the response of transplant programs to OPTN/CMS regulations remains poorly understood. In this work, we consider the perspective of a transplant program which seeks to simultaneously maximize transplant volume and control the risk of OPTN/CMS penalization. Using a chance-constrained mixed-integer programming model, we demonstrate that under certain conditions, it may be rational for a transplant program to curtail its transplant volume in order to avoid penalization under OPTN/CMS regulations. This finding, which confirms empirical results observed in the clinical literature, suggests that such regulations may be inherently unsuitable for use in incentivizing improved program performance. We also highlight other structural shortcomings of OPTN/CMS regulations which have not been observed previously in the literature.

1 Introduction

Organ transplantation is the most effective, and often the only viable, therapy for end-stage organ failure such as chronic obstructive lung disease, end-stage liver disease, and end-stage renal disease. Unfortunately, the supply of transplantable organs has not kept pace with the demand, leading to a significant waiting list for transplantation—as of early 2021, there are more than 100,000 patients on various transplant waiting lists in the United States (OPTN Database 2021). The limited access to transplant necessitates the efficient utilization of the available supply of organs. However, not all transplant programs are equally proficient at turning organs into successful transplants (Buccini et al. 2014, Nassir et al. 2015). In order to ensure that available organs are being effectively used for treatment, the last twenty years have seen increased scrutiny of transplant program operations and outcomes. (In keeping with recently issued federal guidelines, we use the term *program* to refer to the transplantation of a single organ. A hospital that performs both lung and kidney transplants contains two transplant programs.)

This scrutiny has led to the implementation of two closely-related sets of requirements on post-transplant outcomes (McDiarmid et al. 2008, Ho et al. 2015). The first of these is overseen by the Organ Procurement and Transplantation Network (OPTN), a public-private partnership contracted by the United States Department of Health and Human Services (HHS), whose goal is to increase access to transplantation and improve post-transplant survival (OPTN Database 2021). The second set of regulations was overseen by the Centers for Medicare and Medicaid Services (CMS), an agency contained within HHS. The CMS regulations were in effect from June 30, 2007 until November 29, 2019 (the OPTN regulations are still in effect).

Broadly speaking, both sets of regulations function as follows: a regulatory agency (OPTN or CMS) considers all of the transplants performed by a program over a period of two and a half years. Regulators compare the actual number of transplants which resulted in a graft failure within one year of the transplant date and the *expected* number of one-year post-transplant graft failures based on the experience of similar programs nationwide. If the observed number of failures "far exceeds" the expected number, then the transplant program is subject to penalties.

Despite their stated goals of improving transplant outcomes, CMS regulations in particular led to adverse unintended consequences (Schold et al. 2010, 2013, Massie and Segev 2013). Because these regulations focused only on post-transplant outcomes for evaluation, many clinicians and stakeholders expressed concern that programs might have an incentive to be more selective when

determining their patient mix, potentially rejecting some patients that seem risky (i.e., not likely to survive after transplantation) (Abecassis et al. 2009, Kasiske et al. 2012, Schold et al. 2013, Schold and Axelrod 2014, Dolgin et al. 2016). Indeed, there is evidence that the number of patients accepted at certain programs declined—sometimes substantially—exacerbating the limited access to transplant (Schold et al. 2013, White et al. 2015, Dolgin et al. 2016).

In this paper, we study the problem of regulatory-induced risk aversion from the perspective of the transplant program. As we will discuss in detail in Section 3, there is debate in the clinical literature over the impact of these regulations on programs' behavior. What is missing from this debate is a quantitative understanding of the rational response of transplant programs in the face of these regulations. Consequently, the goal of this work is to characterize the response of a rational program to outcome-based regulations such as those used by OPTN and CMS. Our results are particularly significant to regulators, as they provide rigorous theoretical evidence for the existence of several unintended consequences and misaligned incentives of the regulations which had previously been observed only empirically.

The main contributions of this paper are as follows:

- We develop the first optimization framework to model organ transplantation under outcome-based regulation from the perspective of a transplant program. Our work is the first in the operations research literature to study the organ transplantation process from the perspective of the transplant program itself, rather than the perspective of regulators or individual patients.

- Using data provided by Houston Methodist Hospital, our numerical experiments describe realistic conditions under which it is optimal for a program to decrease the number of patients it adds to its waitlist in order to control the risk of regulatory penalization. However, we also demonstrate that under certain conditions, decreasing the number of patients accepted to the waitlist may *increase* the probability of penalization. This finding suggests that controlling penalization risk through adverse patient selection could be extremely difficult in practice, and may indeed increase penalization risk if performed poorly.

- We identify several shortcomings of CMS/OPTN regulations. In particular, we demonstrate that the regulations can lead to increased selectivity by programs—likely exacerbating the already significant mismatch between organ supply and demand—even when the risk adjustment used to model expected outcomes is adequate. We also demonstrate that the regula-

tions may unfairly penalize medium-sized programs, a result which has not been previously observed. Moreover, our model quantifies which patients may be most at risk of adverse selection from programs.

- Our results provide insights to policymakers by quantitatively characterizing the response of rational programs to outcome-based regulations. Our model can be used to explore the effects of potential changes to the regulations, which we demonstrate in the numerical experiments.

For the remainder of the paper, we focus exclusively on lung transplantation programs. Lungs are the fourth most transplanted organ in the United States, and the total number of lung transplants has increased steadily in the past decade (OPTN Database 2021). However, lungs have received very little attention in the operations research literature, which has focused mostly on kidney and liver transplants. Moreover, as we discuss in the following sections, lung transplantation is somewhat unique compared to other organs—for example, wait times to receive a lung transplant in the US are far shorter than for other organs, and donor organ quality is less variable than other organs. These features provide a unique modeling advantage, allowing us to focus solely on the effects of the regulations, without considering complicating factors like waitlist dynamics or donor matching.

The remainder of this paper is organized as follows. Section 2 briefly outlines the history and structure of the OPTN and CMS regulations. Section 3 reviews relevent literature from both the operations research and clinical communities. In Section 4, we introduce a model of rational program behavior in response to OPTN/CMS regulation, which we analyze in Section 5. Section 6 contains a series of computational experiments. Concluding remarks are given in Section 7. Omitted proofs and additional technical details can be found in the appendix.

2 Overview and History of OPTN and CMS Regulations

In 2000, HHS granted OPTN increased authority to monitor the performance of transplant programs. Soon thereafter, OPTN implemented a series of regulations designed to help programs improve their post-transplant outcomes (McDiarmid et al. 2008). These requirements have remained in place with relatively little modification since. Following a series of high-profile failures in organ transplantation (Abecassis et al. 2008, Table 1), CMS proposed a set of outcome-based regulations similar to those used by OPTN in order to identify underperforming programs in 2005. These regulations took effect beginning on June 30, 2007. An important factor in CMS regulation

is that CMS has the authority to cease Medicare and Medicaid reimbursement for underperforming programs. According to SRTR data, in 2018 more than two-thirds of all adult lung transplant programs received CMS reimbursement for at least 50% of their patients. For these programs, a loss of CMS reimbursement would likely have profound effects, potentially leading to program insolvency. In contrast, the penalties imposed by OPTN are perceived to be milder, and are typically focused on remediation rather than punitive measures(UNOS 2018, Schold 2020)

Due to the severity of the penalties, the CMS regulations have received far more attention in the medical community than their OPTN counterpart. Widespread criticism from the transplant community, as well as mounting evidence that the regulations were correlated with a decrease in transplant volume for some patients, led to the repeal of the CMS regulations, effective November 29, 2019 (see Section 3).

In spite of the recent repeal of the CMS regulations, the study of CMS/OPTN outcome-based regulations remains relevant for several reasons. First, as noted by Schold (2020), "[t]he effects of [OPTN] oversight [...] may still yield risk averse behaviors even if the ramifications of low performance evaluations are not as severe." Flagging by OPTN requires the program to undergo a review process, resulting in increased administrative burden, and may also result in negative media attention, depending on the nature of the failure. Moreover, CMS requires all programs to maintain OPTN membership in order to received Medicare/Medicaid reimbursement, and this requirement is unchanged by the repeal enacted in 2019. Under current rules, OPTN has the authority to rescind the membership of flagged programs (OPTN 2019, Appendix L). Consequently, the current regulatory configuration still allows for the possibility that a program which is flagged by OPTN can lose Medicare/Medicaid reimbursement (Andreoni 2020).

Second, the CMS requirements remain an important example of a large-scale pay-for-performance scheme in the United States. As such schemes proliferate in the United States (see, e.g., CMS (2020)), it is increasingly important to gain a quantitative understanding of the incentives such structures create for healthcare providers. Finally, CMS continues to make each program's post-transplant survival outcome data available to the public. Hence, programs still have an incentive to meet certain post-transplant outcome benchmarks, in order to avoid losing patients to other, higher-performing programs (Howard and Kaplan 2006). For these reasons, we consider both OPTN and CMS regulations in this paper.

We conclude this section by summarizing the details of each set of regulations. For clarity, in this section we refer to both OPTN and CMS regulations in the present tense, in spite of the recent

changes to the CMS requirements. Both CMS and OPTN regulations are based on data published twice-yearly by the Scientific Registry for Transplant Recipients (SRTR). These data, released in a document called a Program-Specific Report (PSR), are publicly available on the SRTR website (`www.srtr.org`), and contain a large amount of information about each transplant program in the United States, broken down by organ type. Each PSR incorporates data from all patients transplanted over a 2.5 year period, called an *evaluation window*, and is published one year after the end of the window. A new evaluation window begins every six months; consequently, there are five "active" evaluation windows at any given time (Figure 1).

Both CMS and OPTN regulations function by considering two numbers reported in each PSR: the number of *observed* graft failures (including deaths) during the first year after transplant (denoted O) and the number of *expected* graft failures during the first year after transplant (denoted E). The expected number of graft failures E is computed by the SRTR using a Cox Proportional Hazards Model (Cox 1972), which is normalized to national transplant outcomes in order to incentivize better performance by transplant programs nationally. Unfortunately, the Cox models used by the SRTR have been widely criticized within the transplant community for failing to adequately capture important medical characteristics (e.g., cardiac risk criteria (Pelletier et al. 2014)) that surgeons believe may affect survival outcomes (Abecassis et al. 2008, 2009, Howard et al. 2009, Kasiske et al. 2012). (For a more complete guide to the methods used to create the PSRs, see Dickinson et al. (2008) and SRTR (2019b).)

For each evaluation window, both OPTN and CMS compare the values O and E reported in the PSR. Based on these values, OPTN/CMS use specific criteria to decide whether a program should be subject to additional scrutiny or penalization, a situation known as "being flagged". The primary difference between OPTN and CMS regulations is how these values are compared, and the penalties associated with being flagged.

The OPTN criteria, derived using Bayesian methods (Salkowski et al. 2014a,b), are outlined in §D11.A of the OPTN bylaws (OPTN 2019). Under these criteria, a program is flagged if either $F_{O,E}(1.2) < 0.25$, or $F_{O,E}(2.5) < 0.9$ (or both), where $F_{O,E}$ is the cumulative distribution function of the Gamma distribution with mean and variance equal to $(O+2)/(E+2)$ and $(O+2)/(E+2)^2$, respectively. These criteria may be equivalently re-written as $O \leq g^{\text{OPTN}}(E)$ where $g^{\text{OPTN}}(x) = \min\{h^{1.2}(x), h^{2.5}(x)\}$, and the functions $h^{1.2}$ and $h^{2.5}$ are defined implicitly by the relations $F_{O,E}(1.2) = 0.25$ and $F_{O,E}(2.5) = 0.9$, respectively (Kremers 2014).

Until November 2019, a program was flagged under CMS regulations (outlined in 42 CFR

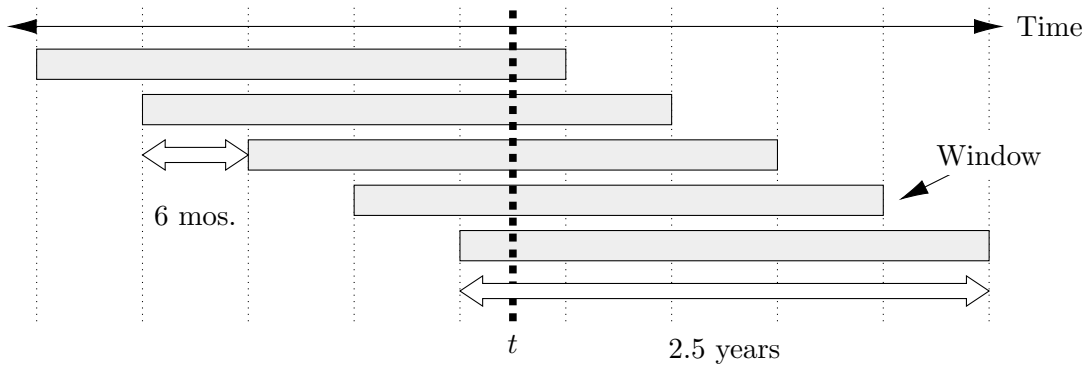

Figure 1: Timeline of rolling evaluation windows, used by both CMS and OPTN. Each evaluation window lasts 2.5 years, and a new window begins every six months. Consequently, at any given time t, there are five "active" evaluation windows. Put another way, any single transplant contributes to five different SRTR reports. Note that the PSR for a single evaluation window is published one year after the conclusion of the evaluation window.

§482.82) if all three of the following occurred: $O > E + 3$, $O > 1.5E$, and the one-sided p-value is less than 0.05, or, equivalently (SRTR 2019b),

$$f(O) \equiv O\left(1 - \frac{1}{9 \cdot O} - \frac{1.96}{3\sqrt{O}}\right)^3 > E.$$

More compactly, CMS required $O \leq g^{\text{CMS}}(E)$, where $g^{\text{CMS}}(x) = \max\{x + 3, 1.5x, f^{-1}(x)\}$ (see Figure 2).

3 Literature Review

This work contributes to two different streams of literature. The first is the clinical literature surrounding the impact of CMS and OPTN outcome-based regulation on transplant programs, insurers, and patients. The second is the operations research literature related to decision-making in organ transplantation. Below, we outline each stream and its relation to the present work.

Since their introduction in 2007, the CMS regulations have received significant attention from the transplant community. Soon after their enactment, many clinicians expressed concerns that the regulations could restrict access to transplant for patients deemed "high-risk" (Abecassis et al. 2008, 2009, Howard et al. 2009, Weinhandl et al. 2009). Indeed, a survey of 63 transplantation personnel conducted by Schold et al. (2010) found that respondents from programs flagged by CMS were significantly more likely to indicate that they had become more restrictive in selecting both transplant candidates and donors.

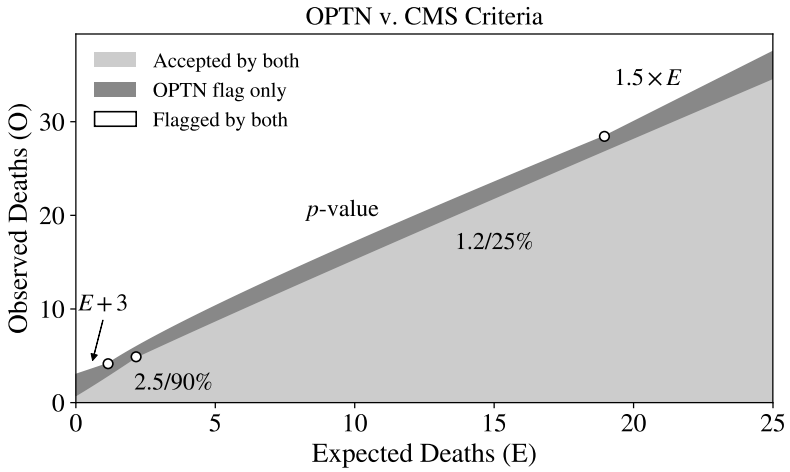

Figure 2: Comparison of OPTN and CMS evaluation criteria in the O/E space. The regulations divided the non-negative quadrant into three regions: points where a program is flagged under both criteria (white coloring), under the OPTN criteria only (dark gray coloring) and under neither criteria (light gray coloring). The boundary between white and light gray is divided into three regions (indicated by white circles), corresponding to the three criteria enforced by CMS. Similarly, the boundary between the light and dark gray regions is divided into two regions, corresponding to the two criteria enforced by OPTN.

As data became available, researchers observed that programs that were flagged under the CMS criteria exhibited a decline in transplant volume, sometimes as high as 38% (White et al. 2015). However, the *cause* of this decline has been debated. White et al. (2015) found that flagged programs reduced the wait-listing of candidates from low-income neighborhoods, which they note "is likely to be due at least in part to more conservative candidate selection criteria." The authors also found that the aggregate risk of one-year post-transplant graft failure is lower among programs that were never flagged, which "may suggest selection bias on the part of these programs." Regulators themselves attributed the decline to "the tendency of [flagged programs] to reduce volume as they regroup to improve their outcomes" (Hamilton 2013). Schold et al. (2013) do not find that the observed decline in volume could be attributed to increased risk aversion. Indeed, the authors question whether the observed "reduction in transplant volume is [...] rational from the [program] perspective." Kasiske et al. (2016) note that "[a]voiding suitable but high-risk transplants to avoid regulatory scrutiny is [...] ineffective, assuming the risk is measured and adjusted for in the models."

In this work, we address these questions by demonstrating that it may in fact be rational for programs to reduce their transplant volume in response to outcome-based regulations, even when the risk adjustment used to assess expected outcomes is adequate. While our findings do

not definitively demonstrate that the observed decline in volume at flagged programs is due to increased risk aversion, they do answer the question of whether it can be rational for programs to reduce volume, and suggest that the observed decline may be at least partially attributable to this increased risk aversion.

Our work also contributes to a stream of papers which apply the tools of operations research to problems in organ transplantation. Historically, this literature has fallen into two broad categories: "organ allocation decisions" and "organ acceptance decisions" (Su and Zenios 2006), which we refer to as the societal perspective and the patient perspective, respectively. The former category refers to decisions made from the perspective of the government, regulators, and other national and regional players in the organ transplantation process (e.g., design of an allocation policy, transportation networks, and organ procurement regions). Such problems have been addressed using a combination of both optimization (e.g., Zenios et al. 2000, Segev et al. 2005b, Stahl et al. 2005, Kong et al. 2010, Akan et al. 2012, Zahiri et al. 2014, Beliën et al. 2013, Davis et al. 2015, Kilambi and Mehrotra 2017) and simulation (e.g., Thompson et al. 2004, Segev et al. 2005a, Davis et al. 2013, Lehr et al. 2019). The latter category, on the other hand, mainly focuses on decisions made by individual patients (e.g., whether or not to accept a donor organ). This area has primarily employed Markov decision processes to model problems like optimal acceptance of livers (e.g., Alagoz et al. 2007a, Sandıkçı et al. 2008), and general models of the organ waitlist from the patient perspective (e.g., Alagoz et al. 2007b, Sandıkçı et al. 2013). To our knowledge, this is the first work in the operations research literature to consider the problem of organ transplantation from the perspective of the transplant program, i.e., to explicitly consider the choices made by the transplant program itself. We note that this perspective lies somewhere between the "societal perspective" and the "patient perspective" discussed above, as programs must balance both the regulatory pressures from society and the welfare of their patients (Figure 3). However, in this paper we consider programs as "rule takers," and study their response under regulatory pressure.

4 Model

In this section, we develop a program-level patient-mix decision model that determines which patients to add to the waitlist so as to maximize expected transplant volume while keeping the probability of violating the regulatory criteria (either CMS or OPTN) below a pre-specified risk threshold. In practice, programs hold periodic meetings to deliberate on which potential transplant

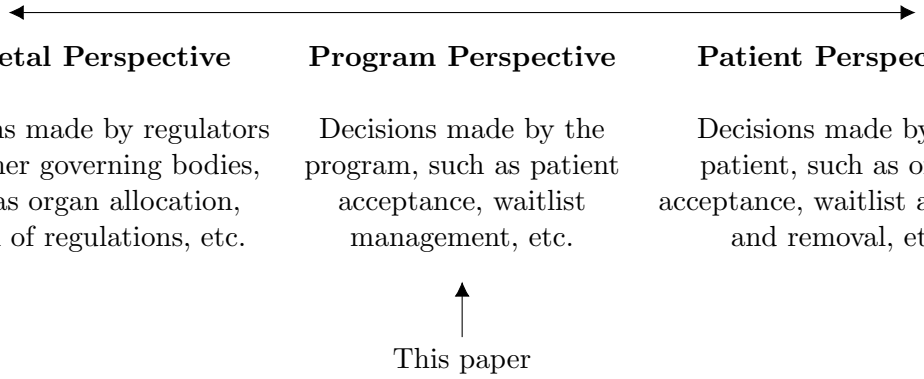

Figure 3: Relationship between the societal perspective and patient perspective, and the current work. Ours is among the first works in the operations research literature to consider organ transplantation from the perspective of the program itself, which lies in between societal decisions and patient decisions.

candidates to add to their waitlists. These meetings typically have many parties present, including surgeons and physicians, financial coordinators, social workers, psychiatrists, and others. These decisions are often made with some degree of subjectivity, with little or no quantitative insight. Our model addresses how a self-interested program may modify this step of the decision process in order to control their risk of penalization.

Throughout the remainder of the paper, we assume that all graft failures are due to patient deaths. Recent SRTR data indicate that this assumption holds true in over 90% of lung transplantation cases in the United States (SRTR 2019a,b). For this reason, we refer to the quantity O as the "observed number of deaths," and refer generically to patient survival, rather than graft survival. Table 4 and Table 5 in the appendix summarize our notation.

Under both CMS and OPTN criteria, patients are characterized by their post-transplant survival probabilities as assessed by both the SRTR risk model and the program's own experience with the specific patient subgroup. Let $\mathcal{I} = \{1, \ldots, I\}$ denote the set of patient classes based on health characteristics. With each patient class $i \in \mathcal{I}$, we associate two numbers: $c_i \in (0, 1)$ denoting the program's expected one-year post-transplant death probability for patients of type i, and $e_i \in (0, 1)$ denoting the SRTR expected one-year post-transplant death probability for patients of type i. In practice, the values for c_i are determined by a program-specific risk model developed based on the program's own experience. Such models are common across many types of solid organ transplantation: some examples include kidneys (Pieloch et al. 2015), hearts (Joyce et al. 2018) and lungs (Russo et al. 2009, Chan et al. 2019). We assume for simplicity that c_i and e_i are time homogeneous, though this assumption may be relaxed. We also assume that c_i and e_i depend

only on the patient's health, not on the quality of the donated graft. Although an exceptionally low-quality graft may adversely affect transplant outcomes, a large majority of the clinical factors associated with graft failure in lung transplantation depend mainly on the physiology of the recipient (Diamond et al. 2013). For example, as of the July 2019 PSRs, 13 of the 16 factors included in the SRTR risk model for graft survival depend only on the recipient, not the donated organ (SRTR 2019a).

The set of decision epochs is given by $\mathcal{T} \cup \{0\}$, where $\mathcal{T} \equiv \{1, \dots, T\}$, and $T \in \mathbb{N}$ is some finite planning horizon. At Houston Methodist, transplant decisions are made on a weekly basis, and thus for convenience we refer to elements of $\mathcal{T}$ as weeks, though this choice can be easily modified. Let $\mathcal{P}$ denote the set of patients who have arrived to the program seeking transplant in the current week, indexed by $t = 0$. For each $j \in \mathcal{P}$, we introduce the control variable $z_j \in \{0, 1\}$, where $z_j = 1$ if patient j is added to the waitlist and 0 otherwise. Let $w = 1, \dots, 5$ index the five evaluation windows active at $t = 0$ (cf. Figure 1) and $w = 6, \dots, W$ index the remaining evaluation windows that begin after $t = 0$ but end on or before $t = T$ (we assume that T is chosen sufficiently large so that at least one such window exists). For each window $w \in \mathcal{W} \equiv \{1, \dots, W\}$, let $\mathcal{T}_w$ denote the set of weeks in $\mathcal{T}$ contained in window w.

For each $i \in \mathcal{I}$ and $t \in \mathcal{T}$, let the $\mathbb{Z}_+$-valued random variable Z_{it} denote the number of patients of type i to arrive seeking transplant in week t. For each $j = 1, \dots, Z_{it}$, let A_{it}^j be a Bernoulli random variable indicating whether the j^{th} patient of type i to arrive in week t is added to the waitlist. The mean of A_{it}^j is the control variable $u_{it} \in [0, 1]$ for all j. The number of patients Y_{it} of type i added to the waitlist in week t is given by $Y_{it} = \sum_{j=1}^{Z_{it}} A_{it}^j$, where $Y_{it} = 0$ if $Z_{it} = 0$.

We make the following assumptions:

Assumption 1. *The random variables* $\left\{A_{it}^j \mid (i,t) \in \mathcal{I} \times \mathcal{T}, j \in \{1, \dots, Z_{it}\}\right\}$ *are independent. For all* $i \in \mathcal{I}$, $t \in \mathcal{T}$, $A_{it}^j \sim \text{Bernoulli}(u_{it})$ *for all* $j \in \{1, \dots, Z_{it}\}$. *That is, the addition of patients to the waitlist is governed by independent Bernoulli random variables.*

Assumption 2. *Patients accepted for transplant are immediately matched with a graft.*

Assumption 2 is reasonable for lung transplantation, where the median wait time to transplant is less than 6 months, and more than one third of patients receive a transplant in less than 90 days. By comparison, only 18% of kidney transplant candidates received a transplant within 6 months (OPTN 2019). Note that Assumption 2 implies that patient health status is static pre-transplant. We make the following assumptions regarding patient arrivals Z_{it}:

Assumption 3. *The expected number of patients of type i to arrive is constant across all weeks t. That is, the expected values $\lambda_i = \mathbb{E}[Z_{it}]$ for $i \in \mathcal{I}$ are time-homogeneous.*

Assumption 4. *The total number of patients to arrive each week is governed by a stationary Poisson process with mean $\lambda = \sum_{i\in\mathcal{I}} \lambda_i$. That is, $\sum_{i\in\mathcal{I}} Z_{it} \sim \text{Poisson}(\lambda)$ for all $t \in \mathcal{T}$.*

Using the Thinning Theorem for Poisson point processes (see, e.g., Dobrow (2016) Section 6.4), one can show that under Assumption 4, for each $i \in \mathcal{I}$, the random process $\{Z_{it}\}_{t\in\mathcal{T}}$ is a Poisson point process with mean λ_i, and that these processes are independent for each $i \in \mathcal{I}$.

Assumption 4 is common in the literature; see, e.g., Zenios et al. (2000) or Shechter et al. (2005). To test this assumption, we performed Pearson's chi-squared test on all 995 patient arrivals to the Houston Methodist Hospital lung transplant program between January 2014 and January 2019. Typically, Houston Methodist does not evaluate transplant candidates (i.e. there are no patient arrivals) during certain weeks of the year—for example, around the Christmas holidays. When controlling for these "empty" weeks, the chi-squared test fails to reject the hypothesis of Poisson arrivals ($p = 0.619$). Moreover, the same data indicate that the mean number of patient arrivals is stationary over the period considered. We explore the sensitivity of our model to Assumption 4 in the appendix.

For each window $w \in \{1, \ldots, 5\}$, let O_w^{init} denote the number of patients who were transplanted in window w who did not reach the one-year post-transplant survival mark, and died before $t = 0$. Let the $\mathbb{Z}_+$-valued random variable $\widehat{O}_w$ denote the number of patients who were transplanted in window w before $t = 0$ who are still alive at $t = 0$, but die before the one-year post-transplant mark. Then, under Assumption 2, the observed number of deaths O_w for any window $w \in \mathcal{W}$ is given by

$$O_w = \sum_{t\in\mathcal{T}_w} \sum_{i\in\mathcal{I}} \sum_{j=1}^{Y_{it}} X_{it}^j + \begin{cases} O_w^{\text{init}} + \widehat{O}_w + \displaystyle\sum_{j\in\mathcal{P}} X_0^j z_j, & w \in \{1, \ldots, 5\}, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where $X_0^j \sim \text{Bernoulli}(c_0^j)$ indicates whether patient $j \in \mathcal{P}$ dies within a year of transplant, $c_0^j \in [0, 1]$ denotes the program-expected one-year death probability for the patient, and $X_{it}^j \sim \text{Bernoulli}(c_i)$ indicates whether the j^{th} patient of type i added to the waitlist (and by Assumption 2, transplanted) in week t fails to reach the one-year post-transplant survival mark.

Similarly, for each $w \in \{1, \ldots, 5\}$, let E_w^{init} denote the SRTR-expected number of deaths for patients transplant in window w before $t = 0$. Then, again using Assumption 2, the SRTR-expected

number of deaths E_w for patients transplanted in window $w \in \mathcal{W}$ is given by

$$E_w = \sum_{t \in \mathcal{T}_w} \sum_{i \in \mathcal{I}} e_i Y_{it} + \begin{cases} E_w^{\text{init}} + \sum_{j \in \mathcal{P}} e_0^j z_j, & w \in \{1, \ldots, 5\}, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where e_0^j is the SRTR-expected one-year death probability for patient $j \in \mathcal{P}$.

As discussed in Section 1, both OPTN and CMS regulations can be written in the form $O_w \leq g(E_w)$ for some function g. Hence, for each window $w \in \mathcal{W}$, we enforce the constraint $\mathbb{P}\big[O_w \leq g(E_w)\big] \geq 1 - \alpha_w$, where $\alpha_w \in [0, 1]$ is a pre-selected risk parameter for each $w \in \mathcal{W}$. We use chance constraints due to their simple interpretability (which is important to clinicians and healthcare policy makers), as well as the tractability of the resulting model.

We additionally enforce a variable lower bound on the expected number of transplants performed in each window w, in order to prevent feasible solutions that require dramatic decreases in transplant volume. In order to enforce these constraints, let the $\mathbb{Z}_+$-valued random variable N_w represent the number of patients added to the waitlist in window $w \in \mathcal{W}$, given by

$$N_w = \sum_{t \in \mathcal{T}_w} \sum_{i \in \mathcal{I}} Y_{it} + \begin{cases} N_w^{\text{init}} + \sum_{j \in \mathcal{P}} z_j, & w \in \{1, \ldots, 5\}, \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

where N_w^{init} is the (constant) number of patients added to the waitlist in window w before $t = 0$. For each $w \in \mathcal{W}$, let U_w be a valid upper bound for $\mathbb{E}[N_w]$. Let $L_w \geq 0$ be a parameter selected by the program, representing the minimum expected number of transplants that a program wishes to perform to remain open. We enforce the constraint that $L_w y_w \leq \mathbb{E}\big[N_w\big] \leq U_w y_w$, where $y_w \in \{0, 1\}$ is a binary variable indicating whether or not a program performs any transplants in window w.

Finally, we specify the objective function of our optimization model. The total number of patients added to the waitlist in the weeks $t \in \mathcal{T} \cup \{0\}$ is given by $\sum_{j \in \mathcal{P}} z_i + \sum_{(i,t) \in \mathcal{I} \times \mathcal{T}} Y_{it}$. We seek to maximize the expected number of patients added to the waitlist, and thus, by Assumption 2, the number of patients who receive a transplant. Hence, our objective is to maximize

$$\sum_{j \in \mathcal{P}} z_j + \mathbb{E}\left[\sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} Y_{it}\right] = \sum_{j \in \mathcal{P}} z_j + \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} \lambda_i u_{it},$$

where the last step follows from the definition of Y_{it} and Wald's lemma (Wald 1944). While other

choices for the objective function are possible, e.g., maximizing revenue from transplantation or maximizing quality-adjusted life-years (Akan et al. 2012, Zenios et al. 2000), we focus on the choice of maximum expected volume for simplicity and note that expected transplant volume is highly correlated with these other metrics. The resulting chance-constrained optimization model is given by:

$$\max \sum_{j\in\mathcal{P}} z_j + \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}} \lambda_i u_{it}, \tag{M0-a}$$

$$\text{s.t. } \mathbb{P}\big[O_w \le g(E_w)\big] \ge 1-\alpha_w \text{ for all } w\in\mathcal{W}, \tag{M0-b}$$

$$L_w y_w \le \mathbb{E}\big[N_w\big] \le U_w y_w \text{ for all } w\in\mathcal{W}, \tag{M0-c}$$

$$y_w \in \{0,1\} \text{ for all } w\in\mathcal{W}, \tag{M0-d}$$

$$z_j \in \{0,1\} \text{ for all } j\in\mathcal{P}, \tag{M0-e}$$

$$0 \le u_{it} \le 1 \text{ for all } (i,t)\in\mathcal{I}\times\mathcal{T}, \tag{M0-f}$$

where O_w and E_w are given by (1) and (2), respectively, and $\mathbb{E}\big[N_w\big]$ is given by the linear expression

$$\mathbb{E}\big[N_w\big] = \sum_{t\in\mathcal{T}_w}\sum_{i\in\mathcal{I}} \lambda_i u_{it} + \begin{cases} N_w^{\text{init}} + \sum_{j\in\mathcal{P}} z_j, & w\in\{1,\dots,5\}, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

5 Structural Properties

We now turn our attention to analyzing the solution structure of the model (M0). Throughout this section we make the very mild assumption that $\alpha < 1/2$, i.e., no program is willing to tolerate a risk of being flagged greater than 50%. Our analysis is divided into two parts. Section 5.1 analyzes the steady-state structure of the model and presents necessary conditions for high-volume transplant programs to remain open. In Section 5.2, we present and analyze a computationally tractable approximation to the model (M0), which we use for the numerical experiments in Section 6. All proofs are available in the appendix. Table 1 provides a summary of the models introduced in this section.

5.1 Steady-State Analysis

To simplify our analysis, we replace the function g in (M0-b) with a convex, piecewise linear (PWL) approximation. The choice of this approximation for both CMS and OPTN criteria is discussed in the appendix. Additionally, we will take $\mathcal{P} = \varnothing$ (eliminating the binary variables z_j) and $L_w = 0$ for all $w \in \mathcal{W}$ (eliminating the binary variables y_w). With these modifications, the model (M0) becomes

$$\max \sum_{t\in\mathcal{T}} \sum_{i\in\mathcal{I}} \lambda_i u_{it}, \tag{M1-a}$$

$$\text{s.t. } \mathbb{P}\big[O_w \le \max_{\ell\in\{1,\dots,L\}} \{m_\ell E_w + b_\ell\}\big] \ge 1 - \alpha_w \text{ for all } w \in \mathcal{W}, \tag{M1-b}$$

$$0 \le u_{it} \le 1 \text{ for all } (i,t) \in \mathcal{I} \times \mathcal{T}, \tag{M1-c}$$

for some vectors $m, b \in \mathbb{R}^L$, where the random variables O_w and E_w are given by (1) and (2).

Assumption 3 and Assumption 4 allow us to prove the existence of a stationary solution for a single-window relaxation of the model (M1):

Proposition 1. *Consider the model (M1) for a single evaluation window w satisfying $w \ge 6$ (cf. the indexing convention in Section 4). Under Assumption 3 and Assumption 4, there exists a stationary optimal solution u^* to (M1) (that is, $u^*_{it_1} = u^*_{it_2}$ for all $i \in \mathcal{I}$, $t_1, t_2 \in \mathcal{T}_w$).*

Proposition 1 inspires the model

$$\max D \sum_{i\in\mathcal{I}} \lambda_i u_i, \tag{M2-a}$$

$$\text{s.t. } \mathbb{P}\big[O_w \le \max_{\ell\in\{1,\dots,L\}} \{m_\ell E_w + b_\ell\}\big] \ge 1 - \alpha, \tag{M2-b}$$

$$0 \le u_i \le 1 \text{ for all } i \in \mathcal{I}, \tag{M2-c}$$

for a single window $w \ge 6$, with decision variables $u \in [0,1]^I$. The factor D appearing in the objective function is the number of decision epochs in an evaluation window (so $D = 130$ if decisions are made weekly). The following result shows that the optimal value of (M2) serves as a tight upper bound on the optimal objective value of the model (M1). We let $z^*_T(\alpha)$ denote the optimal objective value of (M1) with risk parameters $\alpha \in [0,1]^W$ and horizon T. Note that $z^*_T(\alpha)$ depends on the initial data $\{(E^{\text{init}}_w, O^{\text{init}}_w, \widehat{O}_w)\}_{w=1}^5$, though we suppress this dependence for brevity.

Theorem 1. *Let $\widetilde{u}$ be optimal for (M2) with risk parameter $\alpha_0 \in [0,1]$. Suppose that the planning horizon T is divisible by D, and that window $w = 6$ begins in week $t = 1$. Then for any initial data $\{(E_w^{\text{init}}, O_w^{\text{init}}, \widehat{O}_w)\}_{w=1}^5$ and risk parameters $\{\alpha_w\}_{w=1}^5$,*

$$\frac{z_T^*(\widetilde{\alpha})}{T} \leq \sum_{i \in \mathcal{I}} \lambda_i \widetilde{u}_i,$$

where $\widetilde{\alpha}_w = \alpha_w$ for $w = 1, \ldots, 5$ and $\widetilde{\alpha}_w = \alpha_0$ otherwise. Moreover, for any initial data there exist risk parameters $\{\alpha_w\}_{w=1}^5$ for which this bound holds with equality.

Theorem 1 indicates that the average number of patients accepted to the waitlist per week (according to (M1)) is bounded above by the model (M2), and that this bound is tight. In fact, we observe empirically that, as T grows sufficiently large, the slack in the bound of Theorem 1 becomes arbitrarily small, i.e., the average number of patients added to the waitlist per week converges to $\sum_i \lambda_i \widetilde{u}_i$ (see Section 6). This result combined with our empirical observations motivate further study of the model (M2).

To this end, we consider a model that covers only a single window $w \geq 6$ whose decision epochs (weeks) are indexed by $t = 1, \ldots, D$. We drop subscripts w on the random variables O_w and E_w for the remainder of the section. Under Assumption 1, Assumption 3 and Assumption 4, the mean and variance of the random variables $O - m_\ell E - b_\ell$ for $\ell = 1 \ldots, L$ are given by

$$\mu_\ell = \mathbb{E}[O - m_\ell E - b_\ell] = -b_\ell + D \sum_{i \in \mathcal{I}} (c_i - m_\ell e_i) \lambda_i u_i, \tag{5a}$$

$$\sigma_\ell^2 = \text{Var}[O - m_\ell E - b_\ell] = D \sum_{i \in \mathcal{I}} \left[(c_i - m_\ell e_i)^2 + c_i(1 - c_i) \right] \lambda_i u_i, \tag{5b}$$

respectively. Note that the variance (5b) is linear in the decision variables u. This helpful property is a consequence of Assumption 4—if the variables Z_{it} followed other distributions, the variance σ_ℓ^2 would, in general, be a non-convex function of u.

In order to analyze the quantity $\mathbb{P}\big[\min_\ell \{O - m_\ell E - b_\ell\} \leq 0\big]$ which appears implicitly in (M2-b), we perform two simplifications. First, by the countable additivity of measure

$$\mathbb{P}\left[\min_{\ell \in \{1,\ldots,L\}} \{O - m_\ell E - b_\ell\} \leq 0\right] \geq \max_{\ell \in \{1,\ldots,L\}} \left\{ \mathbb{P}\big[O - m_\ell E - b_\ell \leq 0\big] \right\}. \tag{6}$$

Hence we may approximate the constraint (M2-b) with the constraint

$$\max_{\ell\in\{1,\dots,L\}}\left\{\mathbb{P}\big[O-m_\ell E-b_\ell\le 0\big]\right\}\ge 1-\alpha. \tag{7}$$

Second, we make the following assumption:

Assumption 5. *The random variables $O-m_\ell E-b_\ell$ for $\ell=1,\dots,L$ are normally distributed, with mean and variance given by (5a) and (5b), respectively.*

From the definitions of O_w and E_w ((1) and (2), respectively) for $w\ge 6$, we see that the random variable $O-m_\ell E-b_\ell$ is a sum of $D=|\mathcal{T}_w|$ independent, identically distributed (iid) random variables with finite variance. By the classical central limit theorem, the sum of D iid random variables with finite variance converges in distribution to a normal random variable as $D\to\infty$. For all results in this paper we take $D=130$, so that $O-m_\ell E-b_\ell$ is the sum of 130 iid random variables, and is well-approximated by a normal distribution.

Under Assumption 5, we may replace (7) with its deterministic equivalent (Charnes et al. 1958):

$$\min_{\ell\in\{1,\dots,L\}}\left\{\mu_\ell+\Phi^{-1}(1-\alpha)\sigma_\ell\right\}\le 0, \tag{8}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Replacing (M2-b) with (8), we obtain the model:

$$\max\ D\sum_{i\in\mathcal{I}}\lambda_i u_i, \tag{M3-a}$$

$$\text{s.t.}\ \min_{\ell\in\{1,\dots,L\}}\left\{\mu_\ell+\Phi^{-1}(1-\alpha)\sigma_\ell\right\}\le 0, \tag{M3-b}$$

$$\mu_\ell=-b_\ell+D\sum_{i\in\mathcal{I}}(c_i-m_\ell e_i)\lambda_i u_i\ \text{for}\ \ell\in\{1,\dots,L\}, \tag{M3-c}$$

$$\sigma_\ell^2=D\sum_{i\in\mathcal{I}}\left[(c_i-m_\ell e_i)^2+c_i(1-c_i)\right]\lambda_i u_i\ \text{for}\ \ell\in\{1,\dots,L\}, \tag{M3-d}$$

$$0\le u_i\le 1\ \text{for all}\ i\in\mathcal{I}. \tag{M3-e}$$

Remark 1. *In addition to being non-convex and potentially disconnected (Hillestad and Jacobsen 1980), the feasible region of (M3) has a particularly notable structure: there are situations in which decreasing transplant volume may increase penalization risk. This counter-intuitive fact is explained by the quadratic structure of the constraints $\mu_\ell+\Phi^{-1}(1-\alpha)\sigma_\ell\le 0$. Decreasing u strictly decreases*

σ_ℓ, but may increase μ_ℓ (making it less negative). This may eventually result in the infeasibility of the model, i.e., unacceptably high risk of penalization. This result has significant implications for programs: even if a program wishes to attempt to control penalization risk through adverse patient selection, it is likely that— unless done with extreme care, with accurate data calibration and forecasting—the program could actually be increasing its penalization risk.

Theorem 2 highlights another important structural feature of (M3):

Theorem 2. *There exists an optimal solution u^* to the normal single-window model (M3) such that at most one patient class is accepted at a fractional rate, with all other classes either uniformly accepted or uniformly declined.*

Although our models theoretically allow for stochastic acceptance of patients (through the Bernoulli variables A_{it}^j), Theorem 2 demonstrates that, at equilibrium, the optimal solution is almost entirely deterministic—the probability of acceptance for most patients is either 0 or 1.

To conclude this section, we provide necessary conditions on the risk threshold α such that the optimal transplant volume of a program is non-zero. These conditions provide quantitative insight about which kinds of patients are the most favorable *vis-à-vis* CMS and OPTN regulations. Any future regulations should seek to minimize this bound—that is, future regulations should seek to minimize the flagging rate a program must accept in order to remain operational. We state the necessary conditions for high-volume transplant programs under the CMS criteria. Although these assumptions lead to a slight loss of generality, they cover a very important case, as high-volume programs naturally account for a majority of transplant volume (as of the July 2019 PSRs, just 10 programs accounted for almost 40% of the lung transplant volume in the US (SRTR 2019a)). We expect qualitatively similar results for smaller transplant programs and the OPTN criteria. In this context, we define a high-volume transplant program to be a program for which the most restrictive CMS constraint is the $O \leq 1.5E$ constraint (Dickinson et al. 2006, Figure 2). For such programs, we take $L = 1$, with $b_L = 0$ and $m_L = 1.5$, to obtain the following model:

Table 1: Hierarchy of models presented in this section. Each model Mi represents a simplification of the model $M(i-1)$, with M0 being the most general.

Model	Contains y_w, z_j binaries	Contains multiple windows	Chance constraint (CC)	Approximation to function g	Structural results
M0	Yes	Yes	Full CC	None	
M1	No	Yes	Full CC	PWL	Proposition 1
M2	No	No	Full CC	PWL	Theorem 1
M3	No	No	Normal approx	PWL	Theorem 2
M4	No	No	Normal approx	Linear	Theorem 3

$$\max\ D\sum_{i\in\mathcal{I}}\lambda_i u_i, \tag{M4-a}$$

$$\text{s.t. } \mu+\Phi^{-1}(1-\alpha)\sigma\le 0, \tag{M4-b}$$

$$\mu = D\sum_{i\in\mathcal{I}}(c_i-1.5e_i)\lambda_i u_i, \tag{M4-c}$$

$$\sigma^2 = D\sum_{i\in\mathcal{I}}\left[(c_i-1.5e_i)^2+c_i(1-c_i)\right]\lambda_i u_i, \tag{M4-d}$$

$$0\le u_i\le 1 \text{ for all } i\in\mathcal{I}. \tag{M4-e}$$

For all i such that $c_i < 1.5e_i$, let $\beta_i := \sqrt{c_i(1-c_i)/(c_i-1.5e_i)^2}$ denote the coefficient of variation of the random variable $X - 1.5e_i$, where $X \sim \text{Bernoulli}(c_i)$ indicates whether a patient of class i survives one year post-transplant. Then the desired necessary conditions, derived via the KKT conditions, are given as follows:

Theorem 3. *If the optimal solution u to (M4) is non-zero, then $\alpha \ge 1-\Phi(\sqrt{D\sum_{i\in\mathcal{I}^-}\lambda_i/(1+\beta_i^2)})$, where $\mathcal{I}^- := \{i\in\mathcal{I} : c_i < 1.5e_i\}$.*

Theorem 3 indicates that patient classes that are most likely to help a program to remain open are those for which the ratio c_i/e_i is favorable (less than 1.5), the expected number of patients λ_i is large, and the variance in the probability of death (as measured by β_i) is small.

5.2 Tractable Approximation

We now derive a tractable approximation of the model (M0). As in Section 5.1, we begin by replacing the nonlinear function g in (M0-b) with a convex, PWL approximation of the form

$g^{\text{PWL}}(x) = \max_{\ell=1,\ldots,L}\{m_\ell x + b_\ell\}$. Next, following the derivation of (8), we use Assumption 5 to replace (M0-b) with its deterministic equivalent

$$\min_{\ell\in\{1,\ldots,L\}} \left\{ \mu_{w\ell} + \Phi^{-1}(1-\alpha_w)\sigma_{w\ell} \right\} \leq 0 \tag{9}$$

for each $w \in \mathcal{W}$. The values $\mu_{w\ell}$ and $\sigma^2_{w\ell}$ are given by the linear expressions

$$\mu_{w\ell} = -b_\ell + \sum_{t\in\mathcal{T}_w}\sum_{i\in\mathcal{I}}(c_i - m_\ell e_i)\lambda_i u_{it} + \begin{cases} \widehat{\mu}_w + \left(O^{\text{init}}_w - m_\ell E^{\text{init}}_w\right) + \sum_{j\in\mathcal{P}}\left(c^j_0 - m_\ell e^j_0\right)z_j, & w = 1,\ldots,5 \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

and

$$\sigma^2_{w\ell} = \sum_{t\in\mathcal{T}_w}\sum_{i\in\mathcal{I}}\left[(c_i - m_\ell e_i)^2 + c_i(1-c_i)\right]\lambda_i u_{it} + \begin{cases} \widehat{\sigma}^2_w + \sum_{j\in\mathcal{P}} c^j_0(1-c^j_0)z_j, & w = 1,\ldots,5, \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where $\widehat{\mu}_w$ and $\widehat{\sigma}^2_w$ are the mean and variance, respectively, of the random variable $\widehat{O}_w$ which appears in (1), and c^j_0 (*resp.* e^j_0) is the program's (*resp.* SRTR's) expected probability of death for patient $j \in \mathcal{P}$. We replace the square root function that appears implicitly in (9) with a piecewise linear approximation. Because the square root function is concave, for any $\varepsilon > 0$ there exists $K \in \mathbb{N}$ and scalars $\beta_1,\ldots,\beta_K$, $\gamma_1,\ldots,\gamma_K$ such that

$$\left|\sqrt{\sigma^2_{w\ell}} - \min_{k=1,\ldots,K}\{\beta_k\sigma^2_{w\ell} + \gamma_k\}\right| < \varepsilon,$$

for any $\sigma^2_{w\ell} \in [0, U]$, where U is a valid upper bound on the quantity (11). Finally, in order to prevent infeasibility, we introduce a binary variable x_w for each window $w \in \mathcal{W}$, where $x_w = 0$ if and only if the chance constraint for window w is satisfied. We then replace the right-hand side of (9) with $M_w x_w$ (where M_w is a sufficiently large constant), and add a term $-\rho_w x_w$ to the objective function, where $\rho_w > 0$ is a penalty term associated with violating the chance constraint in window w.

With these modifications, we obtain the following *approximation model*:

$$\max \sum_{j\in\mathcal{P}} z_j + \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}} \lambda_i u_{it} - \sum_{w\in\mathcal{W}} \rho_w x_w, \tag{MA-a}$$

$$\text{s.t.} \min_{\ell=1,\dots,L} \left\{\mu_{w\ell} + \Phi^{-1}(1-\alpha_w)\widetilde{\sigma}_{w\ell}\right\} \leq M_w x_w \text{ for all } w \in \mathcal{W}, \tag{MA-b}$$

$$\widetilde{\sigma}_{w\ell} = \min_{k=1,\dots,K} \{\beta_k \sigma_{w\ell}^2 + \gamma_k\} \text{ for all } w \in \mathcal{W} \text{ and } \ell \in \{1,\dots,L\}, \tag{MA-c}$$

$$L_w y_w \leq \mathbb{E}\big[N_w\big] \leq U_w y_w \text{ for all } w \in \mathcal{W}, \tag{MA-d}$$

$$x_w, y_w \in \{0,1\} \text{ for all } w \in \mathcal{W}, \tag{MA-e}$$

$$z_j \in \{0,1\} \text{ for all } j \in \mathcal{P}, \tag{MA-f}$$

$$0 \leq u_{it} \leq 1 \text{ for all } (i,t) \in \mathcal{I}\times\mathcal{T}, \tag{MA-g}$$

where $\mathbb{E}[N_w]$, μ_{wl}, and $\sigma_{w\ell}^2$ are given by the linear expressions (4), (10) and (11), respectively. Because the $\min\{\cdot\}$ expressions appearing in (MA) can be computed by introducing additional binary variables and linear constraints (see, e.g., Vielma and Nemhauser (2011)), the model (MA) is a mixed-integer (linear) program, and may be solved with a commerical mixed-integer programming solver. The original model (M0) and the approximation (MA) are related via the following proposition:

Proposition 2. *Suppose that all of the following hold:*

1. *The piecewise linear approximation of the square root function satisfies $\widetilde{\sigma}_{w\ell} \geq \sigma_{w\ell}$ for all $\ell \in \{1,\dots,L\}$ and $w \in \mathcal{W}$.*
2. *The function g^{PWL} is chosen so that $g^{\text{PWL}}(E_w) \leq g(E_w)$ for all $E_w \geq 0$.*
3. *The random variables $O_w - m_\ell E_w - b_\ell$ are normally distributed with mean $\mu_{w\ell}$ and variance $\sigma_{w\ell}^2$ for all $\ell \in \{1,\dots,L\}$ and $w \in \mathcal{W}$.*

Then any optimal solution to the approximation model (MA) with $x_w = 0$ for all $w \in \mathcal{W}$ is a feasible solution to the model (M0).

6 Numerical Experiments

We now perform a series of numerical experiments to illustrate several significant results that follow from our model. We begin by describing a method for generating the patient classes $\mathcal{I}$ which we use

in our experiments. All experiments were performed using the Gurobi Optimizer, v9.0.0 (Gurobi Optimization, LLC 2019).

6.1 Selecting Patient Classes

We generated a set of patient classes $\mathcal{I}$ using the data from all patients who were either added to the lung transplant waitlist or received a lung transplant at Houston Methodist Hospital (HMH) between January 1, 2014 and December 31, 2018 (multi-organ transplants were excluded). We emphasize that the technique used in this section is only one possible choice of patient class selection; other possibilities could be explored in future work. Moreover, the data used to create these patient classes do not include any patients who sought transplant but were declined to be listed. We discuss the effect of this missing proprietary data in Section 7.

Of the $n = 474$ patients in the cohort, 374 underwent transplant, of which we exclude five due to either unusual medical conditions or data reporting errors. For each of the remaining 469 patients in the cohort, we computed e, the SRTR-estimated one-year post-transplant death probability, using the Cox model provided by the SRTR in its July 2019 PSRs, which includes 16 covariates (SRTR 2019a). Three of the 16 covariates require information about the donor—for the 100 patients who did not receive transplant, we treated these three covariates as missing values. Similarly, for each of the 469 patients, we computed c, the program-estimated one-year post-transplant death probability, using the survival model of Chan et al. (2019). This model uses eight covariates, none of which require donor information.

Given a pair (e, c) for all 469 patients in the cohort, we created patient classes using a two-dimensional histogram. We discretized the unit square into square bins of edge length 0.08 (8%)—we chose this bin size because the program survival model of Chan et al. (2019) has been shown to discriminate patient classes with survival percentages approximately 8 percentage points apart on validation data. For each bin, we created one patient class, whose e_i and c_i values are given by the average of all the (e, c) pairs of patients contained in the bin. The arrival rate λ_i was computed proportional to the number of patients in the bin, scaled so that the total number of arrivals per evaluation window, $130 \times \sum_{i \in \mathcal{I}} \lambda_i$, equals 225, the observed number of patient arrivals at HMH in the the evaluation window covered by the July 2019 PSRs. This method generated 22 patient classes, illustrated in Figure 4.

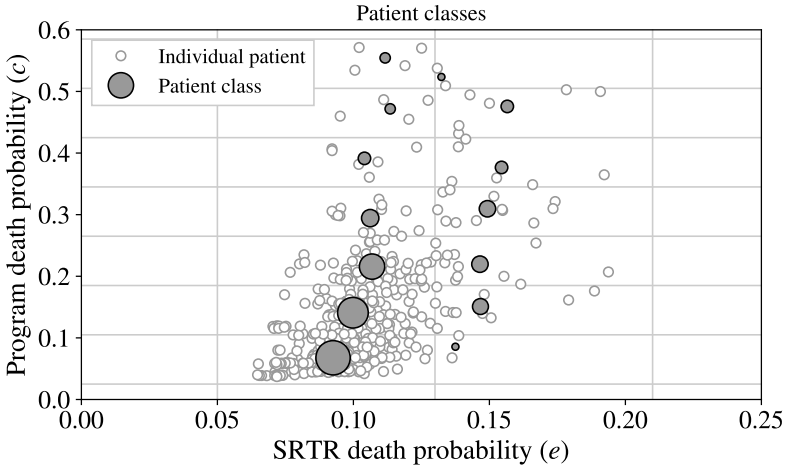

Figure 4: Patient classes created using the technique described in Section 6.1. The light gray grid partitions the unit square into square bins of width 0.08, and each patient pair (e, c) is indicated by an open circle. Filled circles indicate the resulting patient classes, with the size of the circle proportional to the (log of) the arrival rate λ_i. This figure crops outlying data points for visual clarity (complete plot shown in the electronic companion).

6.2 Program Response Prior to Flagging

We next explore how our model reacts to changes in the program's position relative to regulatory expectations. Our primary question is whether a program which is currently performing below regulatory expectations is incentivized to (temporarily) reduce its transplant volume in order to reach compliance.

To test this question, we solve the approximation model (MA) for various values of the starting parameters for each of the five active windows $w = 1, \ldots, 5$: E_w^{init} (the current expected number of deaths), $\widehat{\mu}_w + O_w^{\text{init}}$ (the current expected value of the observed number of deaths) and $\widehat{\sigma}_w^2$ (the current variance of the observed number of deaths). We assume that at the current week $t = 0$ there are three patients under consideration with (e, c) values given by $(10\%, 9\%)$, $(8\%, 12\%)$ and $(8\%, 15\%)$. We fix the parameters $\alpha_w = 3\%$, $\rho_w = 10^3$ and $L_w = 0$ for all windows w and choose T such that the model encompasses 15 complete evaluation windows.

The values E_w^{init}, O_w^{init}, $\widehat{\mu}_w$ and $\widehat{\sigma}_w^2$ were calculated from the actual situation of Houston Methodist during three weeks in 2016. The data for these three weeks are shown in Table 2. The values of E_w^{init} were computed using the SRTR survival model for the July 2019 PSRs. At the time listing decisions are made, the precise coefficients used by the SRTR are not known precisely, and thus the choice of the July 2019 PSR can be viewed as the program estimating the SRTR

Table 2: Starting position of Houston Methodist during three weeks in the second half of 2016. The "Exp." column reports the value E_w^{init}, and the "Obs." column reports the value $O_w^{\text{init}} + \widehat{\mu}_w \pm \widehat{\sigma}_w$.

	$w = 1$		$w = 2$		$w = 3$		$w = 4$		$w = 5$	
	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.
Jul. 18	17.0	20.9 ± 1.6	12.1	13.9 ± 1.6	8.1	10.9 ± 1.6	3.7	4.2 ± 1.4	0.1	0.1 ± 0.2
Sep. 26	18.1	20.9 ± 1.6	13.2	13.9 ± 1.6	9.3	10.9 ± 1.6	4.8	4.7 ± 1.5	1.2	1.2 ± 1.0
Dec. 19	19.1	20.5 ± 1.5	14.3	13.5 ± 1.5	10.3	10.5 ± 1.5	5.8	4.5 ± 1.5	2.3	1.6 ± 1.2

survival model in advance. The values O_w^{init} were computed using SRTR data, and the remaining values ($\widehat{\mu}_w$ and $\widehat{\sigma}_w^2$) were computed using the survival model of Chan et al. (2019), with a synthetic baseline hazard curve.

The results are illustrated in Figure 5. We note that the solutions satisfy the chance constraints for all windows. Moreover, the optimal solution to (MA) converges to the same steady state for all sets of initial conditions (cf. Theorem 1 and the ensuing discussion). These results demonstrate that if a program has experienced fewer deaths to date than the number expected by regulators, a self-interested program may be able to *increase* the number of transplants it performs, taking advantage of its favorable position. Conversely, if a program has a higher number of observed deaths than expected, a program may become more conservative, until its position has stabilized. This experiment shows that, under realistic circumstances, it may make sense for a self-interested transplant program to (temporarily) reduce its transplant volume in order to manage risk of penalization.

6.3 Program Response After Flagging

We now consider the response of a program after it receives a flag. Empirical studies in the medical literature indicate that flagged transplant programs see a corresponding decline in transplant volume (see, e.g., Schold et al. (2013)). However, the cause of this decline is not well understood. Because post-transplant outcomes and flagging information are publicly available, one hypothesis for the decline in transplant volume is that patients avoid low-performing programs, driving down transplant volume at such programs (Howard and Kaplan 2006).

Under the hypothesis that the decline in transplant volume after a flag is due to patient choice, we use our model to characterize the response of a transplant program to these changing conditions. To perform this test, we solve the model (MA) for a single window w satisfying $w \geq 6$, using the

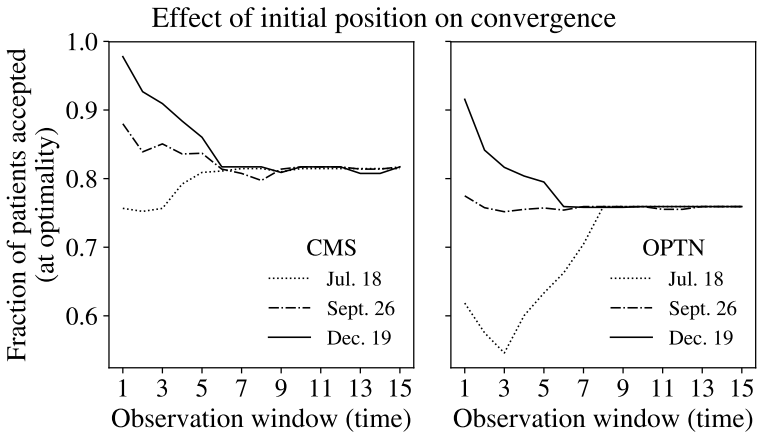

Figure 5: Results of the numerical experiment in Section 6.2. Regardless of the initial conditions, the program converges to the same steady-state solution (cf. Theorem 1). However, if a program finds itself in a favorable initial position, our model adjusts to perform more transplants before reaching steady-state than a program in an unfavorable position. The behavior is qualitatively similar for both CMS criteria (left panel) and OPTN criteria (right panel).

patient classes derived in Section 6.1. In order to model the decrease in volume due to patient choice, we uniformly scale all patient arrival rates λ_i by a scale factor $\beta \in (0, 1)$. That is, we assume that all patient types uniformly defect to other programs. We fix the parameters $\alpha_w = 3\%$, $\rho_w = 10^3$ and $L_w = 0$.

The results, shown in Figure 6, offer two insights. The first is that the decline in volume following a program flag may be driven by two factors. When a program's volume of incoming patients declines—nominally due to patients choosing to receive care at higher-performing programs—the fraction of patients the program can accept while maintaining a constant level of risk declines as well. Using the example in Figure 6, consider a program where, initially, an average of 2.0 patients arrive per week seeking transplant, with an average of 85.0% being accepted under CMS regulations. This yields an average of 1.70 transplants performed per week. Suppose that the program then receives a flag, and sees a corresponding decline of 25% in the number of patients arriving, uniformly across all patient classes $i \in \mathcal{I}$—i.e., the program now has an average of 1.50 patients arrive each week seeking transplant. According to our model of program behavior, the program will now accept only 81.6% of the arriving patients, leading to only 1.22 transplants being performed per week. This amounts to an additional 4% decline in transplant volume on top of the loss of patients to other programs. These results suggest that the observed decline in transplant volume after penalization could be caused simultaneously by patient choice and the rational reaction of programs to the

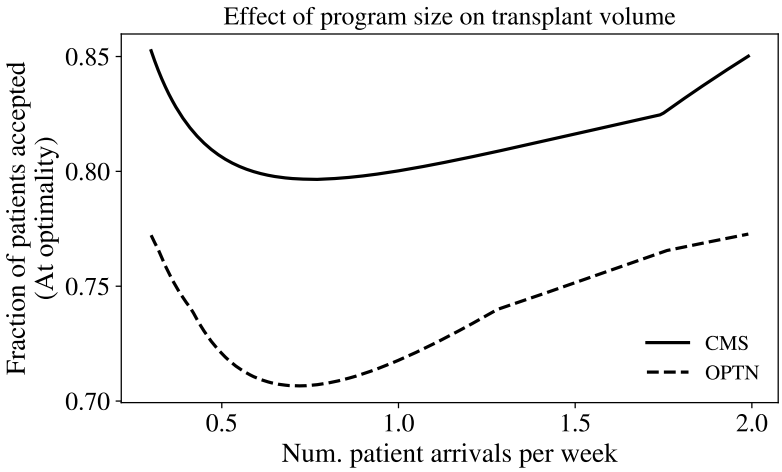

Figure 6: Results of the numerical experiment in Section 6.3. The non-monotonicity of the curves is due to the different regulatory criteria for smaller programs compared to medium and large programs. Note that medium programs in particular must curtail their waitlist acceptance fraction compared to other programs in order to maintain the same risk of flagging ($\alpha = 3\%$ for this figure).

corresponding decline in volume.

The second takeaway from this experiment is an unintended consequence of CMS/OPTN style regulation which, to our knowledge, has not been discussed in the literature. Namely, Figure 6 demonstrates that programs of different sizes but otherwise identical patient mixes are not penalized equally under these regulations. Specifically, medium-sized programs may be forced to accept relatively fewer patients than small- or large-sized programs in order to maintain the same level of penalization risk. This finding can be explained as follows: large programs are protected by the law of large numbers—that is, their variance in outcomes is lower due to larger sample sizes. On the other hand, small programs also have small variance in outcomes (e.g. variance in the number of deaths goes to zero as the number of transplants performed goes to zero). Medium programs, however, have the largest variance in outcomes, and thus may have a stronger incentive to engage in adverse selection in order to control penalization risk. This finding highlights an important area for improvement in future regulation, in order to ensure that programs with similar performance but varying sizes are regulated evenly.

6.4 What if the Risk Adjustment is Accurate?

To conclude our experiments, we consider the behavior of programs when the risk adjustment used to compute expected outcomes is accurate. That is, if a program's estimated death probabilities

(c_i) match those of regulators (e_i), is it always optimal for that program to accept all arriving patients to the waitlist? Ideally such programs, which are meeting regulatory expectations, should be able to accept a large majority—if not all—arriving patients.

To test this hypothesis, we generated three synthetic but reasonable transplant programs (small, medium and large) which satisfy $e_i = c_i$ for all patient classes $i \in \mathcal{I}$ (Table 3). We then solved the approximation model (MA) for a single window $w \geq 6$ for varying values of the risk parameter α, and recorded the resulting optimal fraction of patients accepted to the waitlist.

Figure 7 shows that there exist programs of realistic scale satisfying $c_i = e_i$ for all patients, for which 100% acceptance is *not* optimal at reasonable risk thresholds α (Figure 7). For all three programs we constructed, we see that 100% acceptance is optimal for α larger than approximately 4%. For any smaller risk values α, the optimal transplant volume drops almost immediately by approximately 20%, before decreasing steadily.

The CMS criteria exhibit similar behavior, with a threshold of approximately 2%. In fact, while the CMS criteria were in effect, there was significant debate in the literature over changing the value of the 1.5 parameter used in the CMS criteria (Hamilton 2016, Kasiske et al. 2016, Wright 2016). To test the effect of such changes, we repeat the previous experiment for the synthetic medium program using the CMS criteria, while varying the value of the 1.5 parameter. Results in Figure 8 indicate that raising the slope parameter to approximately 2 may alleviate the burden on transplant programs.

These results indicate that even when the risk adjustment used to compute expected outcomes is accurate, there may still exist rational programs who reject patients in order to keep their flagging risk within pre-specified bounds. Moreover, we also find that adjustments to the parameters used in the regulations (such as those discussed by Kasiske et al. (2016)) could lead to significant gains in transplant volume for some programs. We note that our model is able to quantify the precise and relatively severe effect that this regulatory risk aversion may have on waitlist (and thus transplant) volume.

6.5 Discussion

Collectively, our results demonstrate that it may make sense, under realistic circumstances, for a self-interested program to reject patients in order to avoid penalization under CMS/OPTN-like outcome-based regulations. This finding answers the question of Schold et al. (2013), who ask whether it can ever be rational for a program to decrease its transplant volume in response to

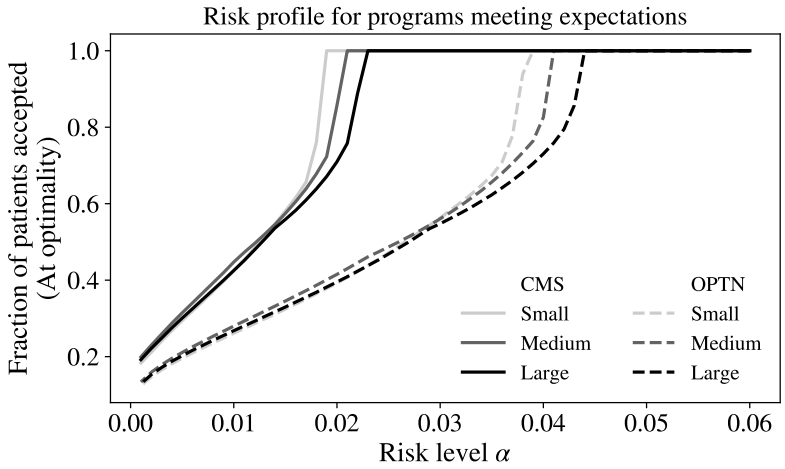

Figure 7: Effect of risk level α on the fraction of patients accepted to the waitlist, for three programs which exactly meet regulatory expectations (i.e. satisfy $c_i = e_i$ for all patient classes $i \in \mathcal{I}$). Note that even when programs meet regulatory expectations, they may be forced to significantly curtail their acceptance fraction in order to remain in regulatory compliance with reasonably high probability.

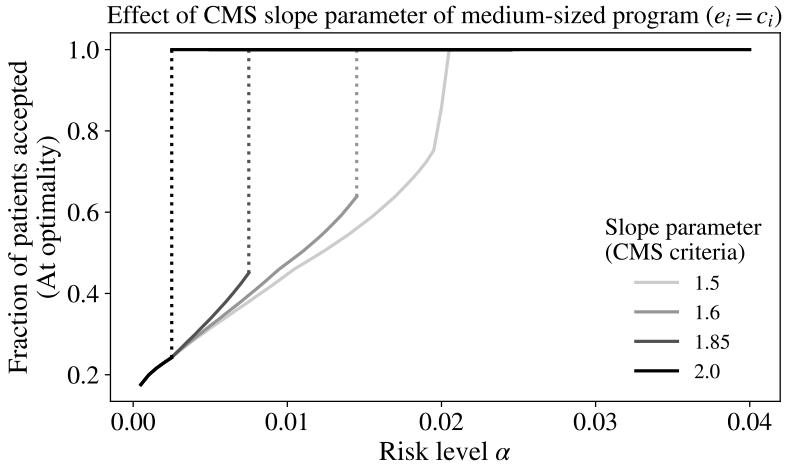

Figure 8: Effect of changing the "slope parameter" 1.5 used in the CMS criteria. We see that increasing this parameter to approximately 2 could greatly reduce the burden of false flagging on transplant programs.

Table 3: Arrival rates for three synthetic programs used in Section 6.4. Blank entries are assumed to be zero. The small program has 8 patient classes, the medium program has 10 patient classes, and the large program has 12 patient classes. The "Total" column is given in units of arrivals per week.

	$e_i = c_i$ (%)												
Program	2	4	6	8	10	12	14	16	18	20	22	24	Total
Small					.229	.030	.030	.030	.030	.030	.030	.030	0.439
Medium			.298	.188	.020	.020	.020	.020	.020	.020	.020	.020	0.646
Large	.765	.700	.010	.010	.010	.010	.010	.010	.010	.010	.010	.010	1.565

flagging. It also supports the findings of White et al. (2015), who observe that the decline in the volume at flagged programs could be at least partially attributable to increased risk aversion.

The implications of this result are significant to policymakers. Most notably, we have demonstrated that the regulations create incentives for programs to engage in adverse selection of patients in order to control penalization risk. Although we do not explicitly consider program selection of candidate organs, our findings suggest that programs likely have incentives to adversely select organs as well as candidates—a result further supported by works cited above, which indicate that programs engage in adverse selection of both organs and candidates. Collectively, these findings imply that the regulations could exacerabte the already severe supply-demand mismatch of organs by incentivizing programs to reject medically-suitable organs and recipients for transplant.

We now discuss the relationship between our results and the risk adjustment used by the SRTR to compute the expected number of graft failures. It is understood in the clinical literature that the methods used by SRTR for risk adjustment do not adequately capture all factors relevant to patient and graft survival. There are three potential lenses to interpret our results:

1. The first case, most prominently discussed in the clinical literature, is that the SRTR risk adjustment is not adequate. Under this hypothesis, our findings suggest that rational programs may reject patients in order to overcome inaccuracies in the risk adjustment models.

2. The second case is that the SRTR risk adjustment is adequate, and $c_i = e_i$ for all patient classes i. In this situation, the results of Section 6.4 demonstrate that a rational program may still be incentivized to engage in adverse patient selection, if their risk tolerance is not sufficiently large. This finding suggests that even perfect risk adjustment may not eliminate the unintended consequences of the regulations.

3. Finally, there is the possibility that the risk adjustment is adequate, but $c_i \neq e_i$ for some

patient classes i. That is, it is possible that the risk adjustment e_i truly reflects national average outcomes for patients of class i, but that a particular program performs better or worse than the national average for some patient classes. In particular, $c_i > e_i$ implies that a program is truly underperforming for class i patients. Our results then indicate that a program may decline to transplant patients when those patients may receive better care (i.e. higher one-year post transplant survival) at other programs. This interpretation is plausibly in keeping with the spirit of the CMS regulations, which aimed to route patients towards higher-performing transplant programs. We note that this argument makes the large assumption that a patient will have the resources to select and travel to another program.

7 Conclusion

We have presented the first modeling framework to consider the problem of regulatory-induced risk aversion from the perspective of a transplant program. Using data provided by the SRTR and Houston Methodist Hospital, we highlighted several important clinical insights for transplant programs, as well as several shortcomings of the regulatory system currently in place in the United States. Notably, we demonstrated realistic conditions under which it may be rational for a program to accept fewer patients in order to remain in compliance with both OPTN and CMS regulations. Moreover, we illustrated that medium-sized transplant programs may need to decrease their volume more than larger or smaller programs with otherwise identical patient mixes.

Due to the significant implications of our work to patient care, we discuss some limitations of this study. We also suggest several extensions of our model as future work.

First, this paper does not explore the impact of outcome-based regulations on waitlist dynamics. Given the critical role of the transplant waitlist in the United States—particularly for organs other than lungs—any regulatory decisions should consider the impact of organ availability and waiting times. Furthermore, our results assume that programs are able to accurately predict post-transplant graft failure probabilities. While some predictive models have been developed for this task, no model is perfect, and it may be important to consider the effects of error and stochasticity in the model parameters we use.

Another limitation of our study is that the data used in our numerical experiments only included patients who were accepted for listing at Houston Methodist, not those who were declined. The effect that this lack of data may have on the final results is difficult to forecast. However, discussions

with surgeons and other decision makers indicate that patients who were not chosen for listing were generally those with the highest risk (roughly, the highest c_i values). Consequently, the results presented in this paper may be interpreted loosely as a "best case scenario," with relatively low-risk patients, but future study with (proprietary) data is warranted. Finally, we also note that the structure of our results depends on the method used to encode patient data (i.e. discrete patient classes characterized by (λ_i, e_i, c_i) triples). Future work could explore other possible encodings.

An additional extension beyond the scope of the present work is to empirically explore the methods used by programs when making waitlisting decisions. In particular, it is unclear whether program behavior can truly be classified as "rational," and which factors play the largest role in waitlisting decisions. A better understanding of this process among policymakers would help augment the results of models such as ours when designing regulations.

Other future work could extend the proposed model to include other organs, such as livers and kidneys. Because these organs have higher variance in graft quality as well as dramatically different waitlist dynamics, any model that incorporates them would likely need to relax some of our modeling assumptions. In addition to requiring the modeling of waitlist dynamics, these extensions would require additional modeling of the organ allocation process, which has been widely studied in the operations research literature. These extensions represent a promising direction for future work.

Acknowledgment

This research was supported by National Science Foundation grants CMMI-1826144, CMMI-1826297 and CMMI-1826323. Additional support was provided by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program. The authors would like to thank Temitayo Ajayi, Dr. Edward Graviss and Dr. Duc Nguyen for helpful comments and conversations.

References

Abecassis MM, Burke R, Cosimi AB, Matas AJ, Merion RM, Millman D, Roberts JP, Klintmalm GB (2008) Transplant center regulations–a mixed blessing? An ASTS council viewpoint. *American Journal of Transplantation* 8:2496–2502.

Abecassis MM, Burke R, Klintmalm GB, Matas AJ, Merion RM, Millman D, Olthoff K, Roberts JP (2009)

American Society of Transplant Surgeons transplant center outcomes requirements–a threat to innovation. *American Journal of Transplantation* 9:1279–1286.

Akan M, Alagoz O, Ata B, Erenay FS, Said A (2012) A broader view of designing the liver allocation system. *Operations Research* 60(4):757–770.

Alagoz O, Maillart LM, Schaefer AJ, Roberts MS (2007a) Choosing among living-donor and cadaveric livers. *Management Science* 53(11):1702–1715.

Alagoz O, Maillart LM, Schaefer AJ, Roberts MS (2007b) Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Operations Research* 55(1):24–26.

Andreoni KA (2020) Now is the time for the Organ Procurement and Transplantation Network (OPTN) to change regulatory policy to effectively increase transplantation in the United States; carpe diem. *American Journal of Transplantation* 20(8):2026–2029.

Beliën J, De Boeck L, Colpaert J, Devesse S, Van den Bossche F (2013) Optimizing the facility location design of organ transplant centers. *Decision Support Systems* 54:1568–1579.

Buccini LD, Segev DL, Fung J, Miller C, Kelly D, Quintini C, Schold JD (2014) Association between liver transplant center performance evaluations and transplant volume. *American Journal of Transplantation* 14(9):2097–2105.

Chan EY, Nguyen DT, Kaleekal TS, Goodarzi A, Graviss EA, Gaber AO (2019) The Houston Methodist lung transplant risk model–a validated tool for pre-transplant risk assessment. *The Annals of Thoracic Surgery* 108(4):1094–1100.

Charnes A, Cooper WW, Symonds GH (1958) Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Science* 4(3):235–263.

CMS (2020) Value-based programs. URL https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/Value-Based-Programs, retrieved March 10, 2021.

Cox DR (1972) Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* 34:197–220.

Davis A, Mehrotra S, Friedewald J, Ladner D (2013) Characteristics of a simulation model of the national kidney transplantation system. *2013 Winter Simulations Conference (WSC)*, 2320–2329 (IEEE).

Davis AE, Mehrotra S, Friedewald JJ, Daskin MS, Skaro AI, Abecassis MM, Ladner DP (2015) Improving geographic equity in kidney transplantation using alternative kidney sharing and optimization modeling. *Medical Decision Making* 35(6):797–907.

Diamond JM, Lee JC, Kawut SM, Shah RJ, Localio AR, Bellamy SL, Lederer DJ, Cantu E, Kohl BA, Lama VN, Bhorade SM, Crespo M, Demissie E, Sonett J, Wille K, Orens J, Shah AS, Weinacker A, Arcasoy S, Shah PD, Wilkes DS, Ware LB, Palmer SM, Christie JD (2013) Clinical risk factors for primary

graft dysfunction after lung transplantation. *American journal of respiratory and critical care medicine* 187(5):527–534, URL http://dx.doi.org/10.1164/rccm.201210-18650C.

Dickinson DM, Arrington CJ, Fant G, Levine GN, Schaubel DE, Pruett TL, Roberts MS, Wolfe RA (2008) SRTR program-specific reports on outcomes: A guide for the new reader. *American Journal of Transplantation* 8(2):1012–1026.

Dickinson DM, Shearon TH, O'Keefe J, Wong HH, Berg CL, Rosendale JD, Delmonico FL, Webb RL, Wolfe RA (2006) SRTR center-specific reporting tools: Posttransplant outcomes. *American Journal of Transplantation* 6(2):1198–1211.

Dobrow RP (2016) *Stochastic processes in R and Mathematica* (Wiley and Sons).

Dolgin NH, Movahedi B, Martins PN, Goldberg R, Lapane KL, Anderson FA, Bozorgzadeh A (2016) Decade-long trends in liver transplant waitlist removal due to illness severity: The impact of Centers for Medicare and Medicaid Services policy. *Journal of the American College of Surgeons* 222(6):1054–1065.

Gurobi Optimization, LLC (2019) Gurobi optimizer reference manual. URL http://www.gurobi.com.

Hamilton TE (2013) Regulatory oversight in transplantation: Are the patients really better off? *Current Opinion in Organ Transplantation* 18(2):203–209.

Hamilton TE (2016) Solid transplant programs–outcome thresholds–revised guidelines. S&C: 16-24-Hospitals, URL http://go.cms.gov/2FuN7Lh.

Hillestad RJ, Jacobsen SE (1980) Linear programs with an additional reverse convex constraint. *Applied Mathematics and Optimization* 6(1):257–269.

Ho B, Skaro IA, Abecassis MM (2015) Should both UNOS and CMS provide regulatory oversight in kidney transplantation? *Current Transplantation Reports* 2(2):127–134.

Howard DH, Kaplan B (2006) Do report cards influence hospital choice? The case of kidney transplantation. *Inquiry* 43:150–159.

Howard RJ, Cornell DL, Schold JD (2009) CMS oversight, OPOs and transplant centers and the law of unintended consequences. *Clinical Transplantation* 23:778–783.

Joyce DL, Li Z, Edwards LB, Kobashigawa JA, Daly RC (2018) Predicting 1-year cardiac transplantation survival using a donor-recipient risk-assessment tool. *The Journal of Thoracic and Cardiovascular Surgery* 155(4):1580–1590.

Kasiske BL, McBride MA, Cornell DL, Gaston RS, Henry ML, Irwin FD, Israni AK, Metzler NW, Murphy KW, Reed AI, Roberts JP, Salkowski N, Snyder JJ, Sweet SC (2012) Report of a consensus conference on transplant program quality and surveillance. *American Journal of Transplantation* 12:1988–1996.

Kasiske BL, Salkowski N, Wey A, Israni AK, Snyder JJ (2016) Potential implications of recent and proposed changes in the regulatory oversight of solid organ transplantation in the United States. *American Journal of Transplantation* 16(12):3371–3377.

Kilambi V, Mehrotra S (2017) Improving liver allocation using optimized neighborhoods. *Transplantation* 101(2):350–359.

Kong N, Schaefer AJ, Hunsaker BK, Roberts MS (2010) Maximizing the efficiency of the U. S. liver allocation system through region redesign. *Management Science* 56(12):2111–2122.

Kremers WM (2014) Impact of the scientific registry of transplant recipients' new bayesian method on estimating center effects and flagging of centers as worse than expected. *American Journal of Transplantation* 14:1703–1704.

Lehr CJ, Skeans M, Valapour M (2019) Validating thoracic simulated allocation model predictions for impact of broader geographic sharing of donor lungs on transplant waitlist outcomes. *The Journal of Heart and Lung Transplantation* To appear.

Massie AB, Segev DL (2013) Rates of false flagging due to statistical artifact in CMS evaluations of transplant programs: Results of a stochastic simulation. *American Journal of Transplantation* 13:2044–2051.

McDiarmid SV, Pruett TL, Graham WK (2008) The oversight of solid organ transplantation in the United States. *American Journal of Transplantation* 8:739–744.

Nassir AB, Dean CE, Li S, Salkowski N, Solid CA, Schnitzler MA, Snyder JJ, Kim J, Kasiske BL, Linzer M, Israni AK (2015) Variation in cost and quality in kidney transplantation. *Transplantation* 99(10):2150–2157.

OPTN (2019) OPTN bylaws. URL https://optn.transplant.hrsa.gov/media/1201/optn-bylaws-effective-dec-3-2019.pdf.

OPTN Database (2021) The Organ Procurement and Transplantation Network data. URL https://optn.transplant.hrsa.gov/data.

Pelletier RP, Phillips G, Rajab A, Pesavento TE, Henry M (2014) Effects of cardiovascular comorbidity adjustment on SRTR risk-adjusted Cox proportional hazard models of graft survival. *Transplantation* 97(6):686–693.

Pieloch D, Dombrovskiy V, Osband AJ, DebRoy M, Mann RA, Fernandez S, Mondal Z, Laskow DA (2015) The Kidney Transplant Morbidity Index (KTMI): A simple prognostic tool to help determine outcome risk in kidney transplant candidates. *Progress in Transplantation* 25(1):70–76.

Russo MJ, Davies RR, Hong KN, Iribarne A, Kawut S, Bacchetta M, D'Ovidio F, Arcasoy S, Sonett JR (2009) Who is the high-risk recipient? Predicting mortality after lung transplantation using pretransplant risk factors. *The Journal of Thoracic and Cardiovascular Surgery* 138(5):1234–1238.e1.

Salkowski N, Snyder JJ, Zaun DA, Leighton T, Edwards EB, Israni AK, Kasiske BL (2014a) A Scientific Registry of Transplant Recipients Bayesian method for identifying underperforming transplant programs. *American Journal of Transplantation* 14:1310–1317.

Salkowski N, Snyder JJ, Zaun DA, Leighton T, Israni AK, Kasiske BL (2014b) Bayesian methods for assessing transplant program performance. *American Journal of Transplantation* 14:1271–1276.

Sandıkçı B, Maillart LM, Schaefer AJ, Alagoz O, Roberts MS (2008) Estimating the patient's price of privacy in liver transplantation. *Operations Research* 56(6):1393–1410.

Sandıkçı B, Maillart LM, Schaefer AJ, Roberts MS (2013) Alleviating the patient's price of privacy through a partially observable waiting list. *Management Science* 59(8):1836–1854.

Schold JA, Axelrod DA (2014) Changing our prior assumptions: Adapting to new Bayesian transplant center report cards. *American Journal of Transplantation* 14:1231–1233.

Schold JD (2020) The evolving role of regulatory reporting on patient and donor selection in organ transplantation. *Current Opinion in Organ Transplantation* 25(2):158–162.

Schold JD, Arrington CJ, Levine G (2010) Significant alterations in reported clinical practice associated with increased oversight of organ transplant center performance. *Progress in Transplantation* 20(3):279–287.

Schold JD, Buccini LD, Srinivas TR, Srinivas RT, Poggio ED, Flechner SM, Soria C, Segev DL, Fung J, Goldfarb DA (2013) The association of center performance evaluations and kidney transplant volume in the United States. *American Journal of Transplantation* 13:67–75.

Segev DL, Gentry SE, Melancon JK, Montgomery RA (2005a) Characterization of waiting times in a simulation of kidney paired donation. *American Journal of Transplantation* 5(10):2448–2455.

Segev DL, Gentry SE, Warren DS, Reeb B, Montgomery RA (2005b) Kidney paired donation and optimizing the use of live donor organs. *Journal of the American Medical Association* 293(15):1883–1890.

Shechter SM, Bryce CM, Alagoz O, Kreke JE, Stahl JE, Schaefer AJ, Angus DC, Roberts MS (2005) A clinically based discrete event simulation of end-stage liver disease and the organ allocation process. *Medical Decision Making* 25(2):199–209.

SRTR (2019a) Program-specific reports. URL https://www.srtr.org/reports-tools/program-specific-reports/.

SRTR (2019b) Technical methods for the program-specific reports. URL https://www.srtr.org/about-the-data/technical-methods-for-the-program-specific-reports/.

Stahl JE, Kong N, Shechter SM, Schaefer AJ, Roberts MS (2005) A methodological framework for optimally reorganizing liver transplant regions. *Medical Decision Making* 25(1):35–46.

Su X, Zenios SA (2006) Recipient choice can address the efficiency-equity trade-off in kidney transplantation: A mechanism design model. *Management Science* 52(11):1647–1660.

Thompson D, Waisanen L, Wolfe R, Merion RM, McCullough K, Rodgers A (2004) Simulating the allocation of organs for transplantation. *Health Care Management Science* 7(4):331–338.

UNOS (2018) Visit from the MPSC? "Don't freak out". URL https://unos.org/news/visit-from-the-mpsc-dont-freak-out/, retrieved March 10, 2021.

Vielma JP, Nemhauser GL (2011) Modeling disjunctive constraints with a logarithmic number of binary variables and constraints. *Mathematical Programming* 128(1–2):49–72.

Wald A (1944) On cumulative sums of random variables. *The Annals of Mathematical Statistics* 15(3):283–296.

Weinhandl ED, Snyder JJ, Israni AK, Kasiske BL (2009) Effect of comborbidity adjustment on CMS criteria for kidney transplant center performance. *American Journal of Transplantation* 9:506–516.

White SL, Zinsser DM, Paul M, Levine GN, Shearon T, Ashby VB, Magee JC, Li Y, Leichtman AB (2015) Patient selection and volume in the era surrounding implementation of Medicare conditions of participation for transplant programs. *Health Services Research* 50(2):330–350.

Wright DR (2016) Transplant centers: Citation for outcome requirements. S&C: 17-13-Transplant, URL http://go.cms.gov/2G0a3Q9.

Zahiri B, Tavakkoli-Moghaddam R, Mohammadi M, Jula P (2014) Multi-objective design of an organ transplant network under uncertainty. *Transportation Research Part E* 72:101–124.

Zenios SA, Chertow GM, Wein LM (2000) Dynamic allocation of kidneys to candidates on the transplant waiting list. *Operations Research* 48(4):549–569.

Appendix A Summary of Notation

Notation is summarized in Table 4 through Table 7.

Table 4: Summary of notation—sets.

Sets	
$\mathcal{I}$	Set of patient classes, with $I = \lvert\mathcal{I}\rvert$
$\mathcal{T}$	Set of decision epochs, with $T = \lvert\mathcal{T}\rvert - 1$
$\mathcal{W}$	Set of evaluation windows, with $W = \lvert\mathcal{W}\rvert$
$\mathcal{T}_w$	Set of decision epochs in window $w \in \mathcal{W}$
$\mathcal{P}$	Set of patients arriving and seeking transplant in week 0

Appendix B Proofs of Formal Results

B.1 Proof of Proposition 1

Proof. Proof of Proposition 1. Let $R^\ell(u) := O - m_\ell E - b_\ell$. Let $u^* \in [0,1]^{I\times D}$ be optimal for (M1). For each ℓ, Assumption 4 allows us to compute the characteristic function $\varphi_\ell(\,\cdot\,; u^*)$ of $R^\ell(u^*)$, given by

$$\varphi_\ell(s; u^*) = \exp\left[-\iota\beta s + \sum_{i\in\mathcal{I}} \xi_{i\ell}(s) \sum_{t=1}^{D} u^*_{it} \right]$$

Table 5: Summary of notation—parameters.

Parameters	
c_i, e_i	Program- and SRTR-expected one-year post-transplant death probability for patients of class $i \in \mathcal{I}$
λ_i	Arrival rate of patient of type $i \in \mathcal{I}$
c_0^j, e_0^j	Program- and SRTR-expected death probability for patient $j \in \mathcal{P}$
α_w	Risk threshold for window $w \in \mathcal{W}$
D	Number of decision epochs per evaluation window
L_w, U_w	Lower and upper bound on the expected number of transplants in window $w \in \mathcal{W}$
ρ_w	Penalty for violating the chance constraint in window $w \in \mathcal{W}$
M_w	Big-M value for the chance constraint in window $w \in \mathcal{W}$
N_w^{init}	Constant number of patients added to the waitlist in window $w \in \mathcal{W}$ before $t = 0$
O_w^{init}	Constant number of patient transplanted in window $w \in \mathcal{W}$ who did not reach the one-year post transplant survival mark, and died before $t = 0$
E_w^{init}	SRTR-expected number of deaths from patients transplanted in window $w \in \mathcal{W}$ before $t = 0$
$f(\cdot)$	CMS p-value criteria function
$g(\cdot)$	Nonlinear function generically representing both CMS and OPTN regulations
L	Number of piecewise linear segments used to approximate the function g
K	Number of piecewise linear segments used to approximate the square root function in (MA)
m_ℓ, b_ℓ	Slope and y-intercept of segment $\ell \in \{1, \ldots, L\}$ used to approximate g
β_k, γ_k	Slope and y-intercept of segment $k \in \{1, \ldots, K\}$ used to approximate the square root function
$\widehat{\mu}_w$, $\widehat{\sigma}_w^2$	Expectation and variance of the random variable $\widehat{O}_w$ (see Table 7) for window $w \in \mathcal{W}$

Table 6: Summary of notation—decision variables.

Decision Variables	
z_j	Binary variable indicating whether or not patient $j \in \mathcal{P}$ is added to the waitlist
u_{it}	Probability that a patient of type $i \in \mathcal{I}$ is added to the waitlist in week $t \in \mathcal{T}$
y_w	Binary variable indicating whether program performs transplants in window $w \in \mathcal{W}$
x_w	Binary variable indicating whether the chance constraint is satisfied in window $w \in \mathcal{W}$

Table 7: Summary of notation—random variables.

Random Variables	
A_{it}^j	Bernoulli random variable indicating whether the j^{th} patient of type $i \in \mathcal{I}$ arriving in week $t \in \mathcal{T}$ is added to the waitlist
X_{it}^j	Bernoulli random variable indicating whether the j^{th} patient of type $i \in \mathcal{I}$ added to the waitlist in week $t \in \mathcal{T}$ fails to survive one year after transplantation
Y_{it}	Number of patients of type $i \in \mathcal{I}$ added to the waitlist in week $t \in \mathcal{T}$
Z_{it}	Number of patients of type $i \in \mathcal{I}$ arriving in week $t \in \mathcal{T}$ added to the waitlist
N_w	Number of patients added to the waitlist in window $w \in \mathcal{W}$
O_w	Observed number of deaths for patients transplanted in window $w \in \mathcal{W}$
E_w	SRTR-expected number of deaths for patients transplanted in window $w \in \mathcal{W}$
$\widehat{O}_w$	Number of patients transplanted in window $w \in \mathcal{W}$ before $t = 0$ who are still alive at $t = 0$, but die before reaching the one-year post-transplant mark
X_0^j	Bernoulli random variable indicating whether patient $j \in \mathcal{P}$ fails to survive for one year post-transplant

where

$$\xi_{i\ell}(s) = \lambda_i \left[(1 - c_i) \exp\big(- \iota m_\ell e_i s \big) + c_i \exp\big(\iota (1 - m_\ell e_i) s \big) - 1 \right]$$

for $i \in \mathcal{I}$, and ι is the imaginary unit. From the form of $\varphi_\ell(s; u^*)$, we can immediately verify that $\varphi_\ell(s; u^*) = \varphi_\ell(s; \overline{u})$, where $\overline{u} \in [0,1]^{H \times D}$ is given by $\overline{u}_{ik} = \frac{1}{D} \sum_{t=1}^{D} u_{it}^*$ for all $i \in \mathcal{I}$, $k \in \{1, \ldots, D\}$. This implies that $R^\ell(u^*)$ and $R^\ell(\overline{u})$ follow the same distribution for $\ell \in \{1, \ldots, L\}$. Hence

$$\mathbb{P}\left[\min_{\ell \in \{1,\ldots,L\}} \left\{ R^\ell(\overline{u}) \right\} \leq 0 \right] = \mathbb{P}\left[\min_{\ell \in \{1,\ldots,L\}} \left\{ R^\ell(u^*) \right\} \leq 0 \right] \geq 1 - \alpha,$$

so $\overline{u}$ is feasible (checking the bounds $0 \leq \overline{u}_{it} \leq 1$ is trivial). Because $\mathbb{E}[Z_{it}]$ is assumed time homogeneous, direct calculation shows that both u^* and $\overline{u}$ produce the same objective value, and thus $\overline{u}$ is optimal. □

B.2 Proof of Theorem 1

Proof. Proof of Theorem 1. We consider the optimal solutions of three different models. Let $u^* \in [0,1]^{I \times T}$ denote the optimal solution to (M1) for full window set $\mathcal{W} = \{1, \ldots, W\}$. Let $\widehat{u} \in [0,1]^{I \times T}$ denote the optimal solution to (M1) for only a single window $\widehat{w} \geq 6$ and risk parameter

α_0. Finally, let $\widetilde{u} \in [0,1]^I$ denote the optimal solution to (M2) with risk parameter α_0. Then

$$\sum_{t\in\mathcal{T}_{\widehat{w}}}\sum_{i\in\mathcal{I}} \lambda_i u^*_{it} \leq \sum_{t\in\mathcal{T}_{\widehat{w}}}\sum_{i\in\mathcal{I}} \lambda_i \widehat{u}_{it} = D\sum_{i\in\mathcal{I}} \lambda_i \widetilde{u}_i, \tag{12}$$

where the inequality follows because $\{u^*_{it}\}_{i\in\mathcal{I},t\in\mathcal{T}_{\widehat{w}}}$ is feasible for the same model for which $\widehat{u}$ is optimal, and the equality follows from Proposition 1. Next, because T is evenly divisible by D and window $w = 6$ begins in week $t = 1$, there exists a set of windows $\widehat{\mathcal{W}} \subset \mathcal{W}$ such that $w \geq 6$ for all $w \in \widehat{\mathcal{W}}$, $|\widehat{\mathcal{W}}| = T/D$, and $\{\mathcal{T}_w\}_{w\in\widehat{\mathcal{W}}}$ partitions $\mathcal{T}$ (i.e., $\bigcup_{w\in\widehat{\mathcal{W}}} \mathcal{T}_w = \mathcal{T}$ and $\mathcal{T}_{w_1} \cap \mathcal{T}_{w_2} = \varnothing$ for all $w_1 \neq w_2 \in \widehat{\mathcal{W}}$). Consequently, we have that

$$z^*_T(\alpha) = \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}} \lambda_i u^*_{it} = \sum_{w\in\widehat{\mathcal{W}}}\sum_{t\in\mathcal{T}_w}\sum_{i\in\mathcal{I}} \lambda_i u^*_{it} \leq D\sum_{w\in\widehat{\mathcal{W}}}\sum_{i\in\mathcal{I}} \lambda_i \widetilde{u}_i = T\sum_{i\in\mathcal{I}} \lambda_i \widetilde{u}_i,$$

where the inequality follows from (12). The result follows from dividing through by T.

To see that this bound is tight, take $\alpha_w = 1$ for $w = 1,\ldots,5$. Then the solution $u_{it} = \widetilde{u}_i$ for all $i \in \mathcal{I}$, $t \in \mathcal{T}$ is feasible for (M1), and satisfies (12) with equality.

□

B.3 Proof of Theorem 2

The proof of Theorem 2 follows directly from the following lemma:

Lemma 1 (Hillestad and Jacobsen (1980)). *Given $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and continuous, strictly concave $g : \mathbb{R}^n \to \mathbb{R}$, if the optimization problem $\max\{c^{\mathrm{T}}x \mid Ax \leq b, g(x) \leq 0\}$ has an optimal solution, then it has an optimal solution which lies on an edge of $F_A = \{x \in \mathbb{R}^n \mid Ax \leq b\}$.*

Proof. Proof of Theorem 2. Because $u \equiv 0$ is feasible for (M3), there exists an optimal solution for the same. The result then follows directly from Lemma 1 where the function g given by (M3-b) is strictly concave as the minimum of strictly concave functions. □

B.4 Proof of Theorem 3

To prove Theorem 3, we derive a convex relaxation of the model (M4), then use the KKT conditions to obtain our final result. By replacing the values u_i in (M4-d) with u_i^2, the concave term $\Phi^{-1}(1-\alpha)\sqrt{\sigma^2}$ becomes convex, yielding the model:

$$\max\ D\lambda^{\mathrm{T}}u \tag{13a}$$

$$\text{s.t. } \mathcal{F}(u) = \eta^{\mathrm{T}} u + \varphi \|N^{1/2} u\|_2 \leq 0 \tag{13b}$$

$$0 \leq u_i \leq 1 \text{ for } i \in \mathcal{I} \tag{13c}$$

where the vector $\eta \in \mathbb{R}^I$ is given by $\eta_i = D\lambda_i(c_i - m_L e_i)$, the matrix $N \in \mathbb{R}^{I \times I}$ is given by $N = \text{diag}(\nu)$ for $\nu_i = D\lambda_i\big[(c_i - 1.5e_i)^2 + c_i(1 - c_i)\big] > 0$, and $\varphi \equiv \Phi^{-1}(1 - \alpha)$.

Proposition 3. *Suppose $\varphi > 0$ (that is, $\alpha < 1/2$). Then the model (13) is convex. Furthermore, the feasible region of (13) is a relaxation of the feasible region of (M4).*

Proof. Proof. To see that (13) is convex, note that if $\varphi > 0$, then model (13) is a second-order conic program, and hence convex. To see that (13) is a relaxation of (M4), note simply that $\|N^{1/2}u\|_2 = (\sum_{i \in \mathcal{I}} \nu_i u_i^2)^{1/2} \leq (\sum_{i \in \mathcal{I}} \nu_i u_i)^{1/2}$ for $u_i \in [0, 1]$. The result follows immediately. □

Beacuse the model (13) is convex, the KKT conditions are sufficient for optimality. That is, if there exists a solution u to (13), multipliers $\mu \in \mathbb{R}$, $\sigma^+, \sigma^- \in \mathbb{R}^I$ and a vector $z \in \partial\mathcal{F}(u)$ (the subdifferential of $\mathcal{F}$ at u) such that

$$D\lambda_i = \mu z_i + \sigma_i^+ - \sigma_i^- \text{ for all } i \in \mathcal{I} \qquad \text{(Stationarity)} \tag{14a}$$

$$\mu \mathcal{F}(u) = 0 \qquad \text{(Complementary Slackness \#1)} \tag{14b}$$

$$\sigma_i^+(u_i - 1) = 0 \text{ for all } i \in \mathcal{I} \qquad \text{(Complementary Slackness \#2)} \tag{14c}$$

$$\sigma_i^- u_i = 0 \text{ for all } i \in \mathcal{I} \qquad \text{(Complementary Slackness \#3)} \tag{14d}$$

$$u \in [0, 1]^I \text{ and } \mathcal{F}(u) \leq 0 \qquad \text{(Primal Feasibility)} \tag{14e}$$

$$\mu \geq 0, \sigma_i^+ \geq 0, \sigma_i^- \geq 0 \text{ for all } i \in \mathcal{I} \qquad \text{(Dual Feasibility)} \tag{14f}$$

then u is optimal for (13). We note that for $u \neq 0$, $\partial\mathcal{F}(u)$ is simply the gradient of $\mathcal{F}$, and $\partial\mathcal{F}(0) = \{\eta + \varphi z \mid \|N^{-1/2}z\| \leq 1\}$. The following lemma is a statement of the KKT conditions (14) for (13) at $u \equiv 0$:

Lemma 2. *Suppose there exists a scalar $\mu \geq 0$, vector $\sigma \in \mathbb{R}^I_+$ and a vector $z \in \partial\mathcal{F}(0)$ (the subdifferential of $\mathcal{F}$ at 0) such that $D\lambda_i = \mu z_i - \sigma_i$ for all $i \in \mathcal{I}$. Then $u \equiv 0$ is optimal for (M4).*

Proof. Proof. Suppose such a μ, σ and z exist. Then $u \equiv 0, \mu, \sigma$ satisfy the KKT conditions (14) for (13).Because the model (13) is convex, the KKT conditions are sufficient, and thus $u \equiv 0$ is optimal for (13). By Proposition 3, $u \equiv 0$ must also be optimal for (M4). □

Proof. Proof (of Theorem 3). We prove the contrapositive. Suppose that

$$\sqrt{D \sum_{i \in \mathcal{I}^-} \lambda_i \frac{(c_i - 1.5e_i)^2}{(c_i - 1.5e_i)^2 + c_i(1-c_i)}} < \Phi^{-1}(1-\alpha).$$

This may be expressed in more compact notation as $\|N^{-1/2}\hat{y}\|_2 < 1$, where $\hat{y} = -\min\{\eta, 0\}/\varphi$, and the $\min\{\cdot\}$ is taken component-wise. We have that $\eta + \varphi\hat{y} = \eta - \min\{\eta, 0\} = \max\{\eta, 0\} \geq 0$, and hence, there exists a sufficiently small perturbation vector $\varepsilon \in \mathbb{R}^I$ such that $\eta + \varphi\tilde{y} > 0$ and $\|N^{-1/2}\tilde{y}\|_2 \leq 1$ for $\tilde{y} = \hat{y} + \varepsilon$ (note in particular that $\eta + \varphi\tilde{y} \in \partial\mathcal{F}(0)$). Next, define

$$\mu \equiv \max_{i \in \mathcal{I}} \left\{ \frac{D\lambda_i}{\eta_i + \varphi\tilde{y}_i} \right\},$$

and, for all $i \in \mathcal{I}$, $\sigma_i \equiv \mu(\eta_i + \varphi\tilde{y}_i) - D\lambda_i$. By construction, $\sigma \in \mathbb{R}^I_+$, $\mu \geq 0$, $\eta + \varphi\tilde{y} \in \partial\mathcal{F}(0)$, and $D\lambda + \sigma = \mu(\eta + \varphi\tilde{y})$. By Lemma 2, $u^* \equiv 0$ is optimal for (13). By Proposition 3, it is also optimal for (M4), as was to be shown. □

B.5 Proof of Proposition 2

Proof. Proof of Proposition 2. Let (z^*, u^*, y^*, x^*) be an optimal solution to (MA), with $x^*_w = 0$ for all $w \in \mathcal{W}$. It suffices to show that (z^*, u^*) satisfy the constraints (M0-b) for all $w \in \mathcal{W}$. We have that, for any $w \in \mathcal{W}$,

$$\begin{aligned}
& \min_{\ell \in \{1,\ldots,L\}} \left\{ \mu_{w\ell} + \Phi^{-1}(1-\alpha)\widetilde{\sigma}_{w\ell} \right\} \leq 0 \\
\implies & \min_{\ell \in \{1,\ldots,L\}} \left\{ \mu_{w\ell} + \Phi^{-1}(1-\alpha)\sigma_{w\ell} \right\} \leq 0 && \text{(By condition 1)} \\
\implies & \max_{\ell \in \{1,\ldots,L\}} \mathbb{P}\big[O_w - m_\ell E_w - b_\ell \leq 0\big] \geq 1 - \alpha_w && \text{(By condition 3)} \\
\implies & \mathbb{P}\left[\min_{\ell \in \{1,\ldots,L\}} \big\{O_w - m_\ell E_w - b_\ell\big\} \leq 0 \right] \geq 1 - \alpha_w && \text{(By (6))} \\
\implies & \mathbb{P}\left[O_w \leq \max_{\ell \in \{1,\ldots,L\}} \big\{m_\ell E_w + b_\ell\big\} \right] \geq 1 - \alpha_w \\
\implies & \mathbb{P}\Big[O_w \leq g(E_w) \Big] \geq 1 - \alpha_w && \text{(By condition 2),}
\end{aligned}$$

completing the proof. □

Appendix C Additional Results

We now demonstrate that, in the limit $\alpha \to 0$, the optimal objective value of (M3) goes to zero. This result is illustrated in Figure 10.

Proposition 4. *Let $z^*(\alpha)$ denote the optimal objective value of (M3) as a function of $\alpha \in (0,1)$. Suppose that $b_\ell \geq 0$ for all $\ell = 1, \ldots, L$. Then $\lim_{\alpha \downarrow 0} z^*(\alpha) = 0$.*

Althought the assumption $b_\ell \geq 0$ for all ℓ in Proposition 4 can be relaxed slightly, we note that the choices of b_ℓ in Section 6 satisfy this condition as stated, and thus we leave it for simplicity. For clarity, we introduce the notation $r_{i\ell} \equiv c_i - m_\ell e_i$ and $s_{i\ell} \equiv r_{i\ell}^2 + c_i(1-c_i)$ for all $i \in \mathcal{I}$ and $\ell \in \{1, \ldots, L\}$.

Proof. Proof (of Proposition 4). Let $\alpha^* \in (0,1)$ satisfy

$$\Phi^{-1}(1-\alpha^*)\sqrt{\hat{s}} + \min\{\hat{r}, 0\}\sqrt{D\sum_{i\in\mathcal{I}} \lambda_i} \;=\; 0.$$

where $\hat{r} = \min_{i,\ell}\{r_{i\ell}\}$ and $\hat{s} = \min_{i,\ell}\{s_{i\ell}\}$. Then, because $\hat{s} > 0$, we have that

$$\Phi^{-1}(1-\alpha)\sqrt{\hat{s}} + \min\{\hat{r}, 0\}\sqrt{D\sum_{i\in\mathcal{I}} \lambda_i} \;>\; 0, \tag{15}$$

for all $\alpha \in (0, \alpha^*)$. Fix such an α, and let u^α be the corresponding optimal solution to (M3). If u^α is the zero vector, then $z^*(\alpha) = 0$. Otherwise, because u^α is feasible, we have that

$$\begin{aligned}
0 &\geq \min_\ell \left\{ -b_\ell + \sum_{i\in\mathcal{I}} r_{il}(D\lambda_i u_i^\alpha) + \Phi^{-1}(1-\alpha)\sqrt{\sum_{i\in\mathcal{I}} s_{i\ell}(D\lambda_i u_i^\alpha)} \right\} \\
&\geq \min_\ell\{-b_\ell\} + \min_\ell \left\{ \sum_{i\in\mathcal{I}} r_{i\ell}(D\lambda_i u_i^\alpha) \right\} + \min_\ell \left\{ \Phi^{-1}(1-\alpha)\sqrt{\sum_{i\in\mathcal{I}} s_{i\ell}(D\lambda_i u_i^\alpha)} \right\} \\
&= \min_\ell\{-b_\ell\} + \min_\ell \left\{ \sum_{i\in\mathcal{I}} r_{i\ell}(D\lambda_i u_i^\alpha) \right\} + \Phi^{-1}(1-\alpha)\sqrt{\min_\ell \left\{ \sum_{i\in\mathcal{I}} s_{i\ell}(D\lambda_i u_i^\alpha) \right\}} \\
&\geq \min_\ell\{-b_\ell\} + \min_\ell \left\{ \min_i\{r_{i\ell}\} \sum_{i\in\mathcal{I}} (D\lambda_i u_i^\alpha) \right\} + \Phi^{-1}(1-\alpha)\sqrt{\min_\ell \left\{ \min_i\{s_{i\ell}\} \sum_{i\in\mathcal{I}} (D\lambda_i u_i^\alpha) \right\}} \\
&= \min_\ell\{-b_\ell\} + \hat{r}\sum_{i\in\mathcal{I}} D\lambda_i u_i^\alpha + \Phi^{-1}(1-\alpha)\sqrt{\hat{s}\sum_{i\in\mathcal{I}} D\lambda_i u_i^\alpha} \\
&= \min_\ell\{-b_\ell\} + \hat{r} z^*(\alpha) + \Phi^{-1}(1-\alpha)\sqrt{\hat{s} z^*(\alpha)}.
\end{aligned}$$

Re-arranging this expression gives that

$$\Phi^{-1}(1-\alpha)\sqrt{\hat{s}} \leq -\frac{\min_\ell\{-b_\ell\}}{\sqrt{z^*(\alpha)}} - \hat{r}\sqrt{z^*(\alpha)} \leq -\frac{\min_\ell\{-b_\ell\}}{\sqrt{z^*(\alpha)}} - \min\{\hat{r},0\}\sqrt{\sum_{i\in\mathcal{I}} D\lambda_i}.$$

Re-arranging, using the condition (15) and the fact that $b_\ell \geq 0$ for all ℓ gives

$$z^*(\alpha) \leq \frac{-\min_\ell\{-b_\ell\}}{\Phi^{-1}(1-\alpha)\sqrt{\hat{s}} + \min\{\hat{r},0\}\sqrt{\sum_i D\lambda_i}}. \tag{16}$$

Because the relation (16) holds for all $\alpha \in (0,\alpha^*)$, and $z^*(\alpha) \geq 0$ for all $\alpha \in (0,1)$, we have that

$$0 \leq \lim_{\alpha\downarrow 0} z^*(\alpha) \leq \lim_{\alpha\downarrow 0} \frac{-\min_\ell\{-b_\ell\}}{\Phi^{-1}(1-\alpha)\sqrt{\hat{s}} + \min\{\hat{r},0\}\sqrt{\sum_i D\lambda_i}} = 0,$$

as was to be shown. □

Appendix D Convex PWL Approximation to OPTN and CMS Criteria

The models presented throughout this paper require a convex piecewise linear approximation of the criteria function g^{CMS} and g^{OPTN}. Recall that $g^{\text{CMS}}(x) = \max\{x+3, 1.5x, f^{-1}(x)\}$ where the function f is given by

$$f(O) \equiv O\left(1 - \frac{1}{9\cdot O} - \frac{1.96}{3\sqrt{O}}\right)^3.$$

To obtain a piecewise linear approximation, we replace the nonlinear segment f^{-1} with a linear interpolation between the two circled points in Figure 2 (i.e., between the intersection of the curve $f^{-1}(x)$ and the line $x+3$, and the intersection of the curve $f^{-1}(x)$ and the line $1.5x$). The resulting approximation to g^{CMS} is $\widehat{g}^{\text{CMS}}(x) = \max\{x+3, 1.364x+2.579, 1.5x\}$ (cf. condition 2 of Proposition 2).

To approximate to the function g^{OPTN}, we use a line of best fit, resulting in the approximation $\widehat{g}^{\text{OPTN}}(x) = 1.298x + 2.265$. Note that this approximation is affine, and does not satisfy condition 2 of Proposition 2 for all $x \geq 0$. Nevertheless, we will see in Section E.2 that the resulting approximation is still conservative in practice.

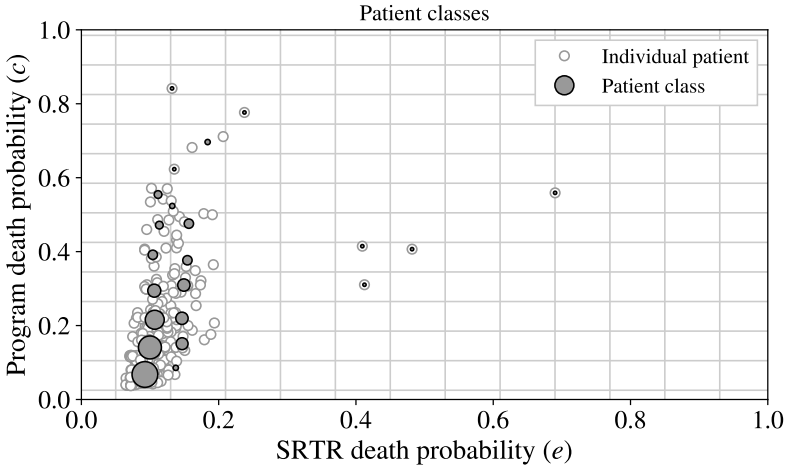

Figure 9: Patient classes created using the technique described in Section 6.1. The light gray grid partitions the unit square into square bins of width 0.08, and each patient pair (e, c) is indicated by an open circle. Filled circles indicate the resulting patient classes, with the size of the circle proportional to the (log of) the arrival rate λ_i.

Appendix E Sensitivity Analysis

E.1 Risk Parameter Selection

We explore the sensitivity of optimal program behavior to the choice of risk parameter α_w. Using the patient classes from Section 6.1, we solve the model (MA) for a single window $w \geq 6$, with $L_w = 0$, $\rho_w = 10^3$ and varying values of α_w. As proven in Proposition 4, the optimal fraction of accepted patients goes to zero as $\alpha_w \to 0$, consistent with intuition. We see in Figure 10 that this decline happens largely for $\alpha_w \leq 2\%$. This result suggests that a transplant program can only reasonably control its risk of false flagging up to approximately 2%, at which point it must cut transplant volume drastically. Potential modifications to OPTN/CMS-style regulation could focus on decreasing this number for transplant programs.

E.2 Poisson Arrivals

We assess the sensitivity of our model to the Poisson arrivals assumption (Assumption 4). As a byproduct of this analysis, we also demonstrate the effect of using piecewise linear approximations for the true penalization curves g (Figure 2). We begin by randomly perturbing a Poisson distribution (described in detail below). We next solve the model (MA) using the unperturbed Poisson distribution to obtain an optimal acceptance policy u^*, and use the perturbed distribution

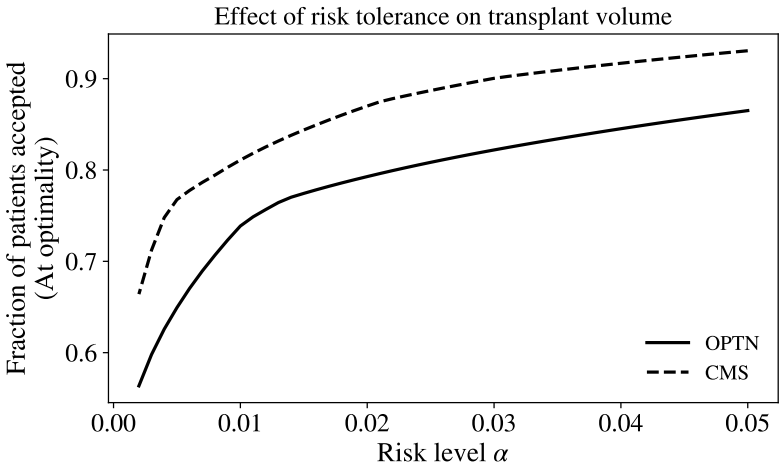

Figure 10: Sensitivity of optimal acceptance fraction to risk parameter selection. As proven in the appendix, the optimal fraction of accepted patients goes to zero as $\alpha_w \downarrow 0$. We observe that, for the patient classes in Figure 9, the fraction of patients accepted to the waitlist at optimality is relatively level until approximately $\alpha = 1\%$, at which point it begins to drop significantly.

and u^* to generate $n = 50,000$ pairs (E, O) of expected and observed deaths using the statistical model introduced in Section 4 (excluding Assumption 4). We then record the fraction of the $50,000$ sample windows which resulted in a flag under both the true criteria function g and the piecewise linear criteria function g^{PWL}. We repeat this process for 100 different random perturbations to the Poisson distribution, and report the mean, minimum and maximum flagging rates for these 100 trials. We repeat this entire process for both the CMS and OPTN criteria, and for risk values $\alpha \in \{0.97, 0.98, 0.99\}$.

We create perturbed distributions as follows. For each patient class $i \in \mathcal{I}$, we truncate a Poisson distribution with mean λ_i at the 10^{th} spot—i.e., we assume that no more than 10 patients from class i will arrive each week. Let $p^i \in \mathbb{R}^{11}$ denote the resulting vector of probabilities, so that p^i_j is the probability that j patients of class i arrive in a single week. The perturbed distribution $\tilde{p}^i$ is

Table 8: Sensitivity of program flagging rate to the Poisson arrivals assumption (Assumption 4). All values are percentages. Reported values are averaged over 100 different sampled distributions. For each sampled distribution, the flagging rate was estimated from $50,000$ trials.

		OPTN		CMS	
	Target $(1-\alpha)$	Mean	(Min, Max)	Mean	(Min, Max)
True	1	1.33	(1.17, 1.45)	1.17	(1.08, 1.34)
	2	2.40	(2.25, 2.56)	2.24	(2.08, 2.39)
	3	3.46	(3.26, 3.64)	3.25	(3.08, 3.42)
PWL	1	1.44	(1.27, 1.56)	1.40	(1.29, 1.55)
	2	2.55	(2.38, 2.70)	2.45	(2.29, 2.60)
	3	3.62	(3.40, 3.79)	3.45	(3.28, 3.68)

sampled uniformly from the set of vectors $p \in [0,1]^{11}$ satisfying

$$\sum_{j=0}^{10} p_j = 1, \tag{17a}$$

$$\sum_{j=0}^{10} j \cdot p_j = \lambda_i, \tag{17b}$$

$$p_0 \geq p_0^i, \tag{17c}$$

$$\|p - p^i\| \leq \varepsilon = 0.1. \tag{17d}$$

Constraint (17a) ensures that $\tilde{p}^i$ is a probability distribution, while (17b) ensures that $\tilde{p}^i$ has expectation λ_i (the same as p^i). Constraint (17c) is based on the empirical observation (Section 4) that, in practice, the Poisson distribution underestimates the probability of zero patient arrivals in a given week. Finally, (17d) ensures that the sampled distribution is not too different from the truncated Poisson distribution, consistent with empirical data.

The results in Table 8 show that, by removing the Poisson assumption, the observed flag rate stays within 0.4% of the target value $1-\alpha$ for the CMS criteria, and 0.6% for the OPTN criteria. These results indicate that the combined effect of the Poisson assumption and piecewise linearization used in (MA) is quite small—on the order of one half percent. We note also that the observed flag rates under the PWL criteria are higher than the true criteria because the PWL criteria were chosen conservatively, so that $g(E) \geq g^{\text{PWL}}(E)$ for a majority of the E values of interest.