SELECTION BIAS IN LUNG ALLOCATION: INFLUENCE ON LUNG ALLOCATION SCORE AND PHYSICIAN DECISION-MAKING

Erin M. Schnellinger

A DISSERTATION

in

Epidemiology and Biostatistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2022

Supervisor of Dissertation

Stephen E. Kimmel, MD, MSCE

Professor of Epidemiology

Co-Supervisor of Dissertation

Alisa J. Stephens-Shields, PhD

Assistant Professor of Biostatistics

Graduate Group Chairperson

Nandita Mitra, PhD, Professor of Biostatistics

Dissertation Committee

Julia E. Szymczak, PhD, Assistant Professor of Epidemiology

Edward Cantu III, MD, MSCE, Associate Professor of Surgery

Douglas E. Schaubel, PhD, Professor of Biostatistics

ACKNOWLEDGMENT

This work is the culmination of many years of conversations with colleagues, collaborators, and friends. I am incredibly grateful for their support, and would like to acknowledge them here.

To Dr. Stephen Kimmel, primary mentor: working with you these past few years has been the highlight of my time in graduate school. From the moment you welcomed me to your research team, you have taught me how to navigate the ups and downs of academic life while never losing sight of the ultimate goal: conducting high-quality research that has real-world impact. Through your insightful and thought-provoking questions, you challenged me to take my initial research ideas and develop them into a much stronger body of research. I am honored to have had your guidance throughout my dissertation, and look forward to continuing to work together in the future.

To Dr. Alisa Stephens-Shields, co-mentor: thank you for pushing me to think more deeply about statistical analyses and what they mean. I admire your ability to translate complex mathematical concepts and formulas into understandable language, and hope to someday attain half of the statistical savvy that you have. Your insight and feedback throughout this project undoubtedly made it stronger.

To Dr. Edward Cantu, clinical committee member: thank you for providing clinical insight throughout this entire project, from the initial research questions to the analysis and interpretation of results. Thank you also for connecting me with your colleagues and helping me to recruit transplant surgeons and pulmonologists for the qualitative aim of this dissertation. Without your help, this project would not have nearly the same clinical relevance as it does now.

To Dr. Douglas Schaubel, biostatistics committee member: thank you for sharing your expertise in statistics and organ transplantation research to make this dissertation

stronger. Thank you also for serving as the “voice of reason” when I started to get a little carried away with statistical analyses. Your practicality ensured that this project moved beyond the pages of this dissertation and out into the scientific community.

To Dr. Julia Szymczak, committee chair: you took me under your wing the moment I matriculated at Penn. Thank you for helping me progress through all the stages of graduate school, from bright-eyed and naïve first-year student, to overwhelmed and stressed second-year student, to “hey, I think I’m finding my groove” third-year student, to “just kidding, I actually have no idea what I’m doing” fourth-year student, and finally, to where I am today. Thank you also for teaching me why qualitative research is important and how to conduct it in a rigorous manner. You once advised me that “the best dissertation is a done dissertation”. I hope you find that this final product is not just “done”, but done well.

To Drs. John Jackson and Michael Harhay, mentors: thank you for helping me navigate the world of academia and find my footing as a statistician-turned-epidemiologist. Looking forward to continuing to work with you in the future.

To my cohort, Drs. Elle Saine, Amelia Van Pelt, and Emma Zheng: I am extremely grateful to have had the opportunity to work and learn alongside such brilliant and kind scholars. Through study sessions, happy hours, dinner dates, and game nights, you made me feel included and valued. Elle: thank you for teaching me – and reminding me, when needed – what it means to be a scholar, and how to do so with grace and humility. Thank you also for sharing my enthusiasm for charcuterie boards. Amy: thank you for listening to my endless worries about grad school and providing concrete advice on how to move forward. Thank you also for joining me on weekly walks/hikes/bike rides during the pandemic to stay sane. Emma: thank you for reviewing early versions of this dissertation (back when it was a proposal!) and providing invaluable feedback on my

ideas. Thank you also for the incredible kindness you have shown me these past few years. Looking forward to many more adventures with all of you.

To my friends, especially Taryn, Kelly, Abby, Leah, Barb, and Jessie: Thank you for reminding me that there is more to life than grad school. I will always treasure you.

To Mom, Dad, Rusty, and Victoria: You are my heroes. Thank you for your endless love and support. I hope I have made you proud.

To my qualitative interview respondents: thank you for being so generous with your time. Your thoughtful responses to my interview questions informed not only the qualitative analyses in this dissertation, but also helped me interpret my quantitative findings in the context of clinical practice. I hope that this work may prove useful to you as you continue to improve the field of organ transplant allocation.

I would also like to acknowledge the funding provided to me by the National Heart, Lung, and Blood Institute through an F31 Pre-doctoral Fellowship Award (NIH F31HL194338). This dissertation would not have been possible without such support.

ABSTRACT

SELECTION BIAS IN LUNG ALLOCATION: INFLUENCE ON LUNG ALLOCATION SCORE AND PHYSICIAN DECISION-MAKING

Erin M. Schnellinger

Stephen E. Kimmel

Alisa J. Stephens-Shields

In the U.S., donor lungs are allocated to recipients based on a lung allocation score (LAS). While the statistical models used to construct the LAS control for patients' demographic and clinical values, they do not account for selection bias, which arises because: (1) individuals are removed from the waitlist once they receive transplant (dependent censoring), and (2) in order to receive transplant, individuals must survive on the waitlist long enough for a suitable lung to become available (survivor bias). Failure to account for selection bias can lead to inaccurate predicted probabilities and suboptimal organ allocation. The goal of this dissertation is to improve the predictive accuracy of the LAS by mitigating selection bias so that lungs are allocated to the appropriate patients in the appropriate order. This goal was accomplished via three aims. First, we proposed a weighted estimation strategy to mitigate selection bias in the pre- and post-transplant LAS models, constructed a modified LAS score using these weights, and compared its performance to that of the existing LAS. Second, we examined the clinical impact of our modified LAS in both observed data and through simulations. Third, we conducted qualitative semi-structured interviews with lung transplant surgeons and pulmonologists throughout the U.S. to examine respondents' understanding of selection bias and how it may affect the LAS and organ distribution. We found that our modified LAS exhibited better discrimination and calibration than the existing LAS and led to changes in patient prioritization. Diagnosis group, six-minute walk distance, continuous mechanical ventilation, functional status, and age exhibited the largest impact on prioritization changes. Simulations suggest that one-year waitlist survival may improve

under the modified LAS, while one-year post-transplant and overall survival remain comparable to that under the existing LAS. Finally, our qualitative study demonstrates that selection bias can arise at several points along the transplantation pathway. To address such bias, transplant centers must consider both patient health and program health within constraints imposed by donor organ scarcity. We hope that this work can inform future revisions of the LAS and other prediction models in organ transplantation to ensure more equitable allocation of donor organs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1. INTRODUCTION

Dissertation Overview

In May 2005, the Organ Procurement and Transplantation Network (OPTN) modified their lung transplantation policy from one in which potential recipients were prioritized based on the amount of time spent on the waitlist, to one in which donor lungs were allocated to recipients based on a lung allocation score (LAS) (Egan et al., 2006; Gottlieb, 2017; U.S. Department of Health and Human Services, October 20, 1999; United Network for Organ Sharing, 2015). The LAS is calculated as the difference in predicted days of life between transplant benefit (post-transplant survival minus pre-transplant survival) and waitlist urgency (pre-transplant survival) and is normalized so that it ranges from 0 to 100, with higher values indicating greater priority for lung transplantation (Egan et al., 2006; United Network for Organ Sharing, 2015). This score was intended to balance equity, justice, beneficence, and utility in lung allocation (Gottlieb, 2017; Veatch & Ross, 2015). Since the adoption of the LAS, studies have shown that the mortality rate for patients on the lung transplant waitlist has decreased; a greater proportion of transplants have been performed for patients with pulmonary fibrosis compared to patients with other diagnoses; older and sicker patients are being transplanted at a higher rate; and one-year post-transplant survival has not changed significantly (Chen et al., 2009; Gries et al., 2007; Kotloff, 2013; Kozower et al., 2008; Lingaraju et al., 2006; McCue et al., 2008; Merlo et al., 2009). However, other studies have suggested that the LAS is "broken," (Halpern, 2015) as its implementation is associated with greater resource use (Maxwell et al., 2015) and geographic and gender disparities in lung allocation (Russo et al., 2013; Thabut et al., 2012; Wille et al., 2013).

A limitation of the LAS score is that the prediction of post-transplant survival that

is used in the score is limited to those who survive to transplant. Although the statistical models used to estimate transplant benefit and waitlist urgency control for patient characteristics, they do not account for the fact that in order to receive a lung transplant, an individual must survive on the waitlist long enough for a suitable donor lung to become available. Since individuals who survive one year or more on the waitlist might be inherently different from individuals who die, receive a transplant, or are censored prior to one year, selection bias can arise. Failure to incorporate such information in the models used to estimate transplant benefit and waitlist urgency can lead to inaccurate LAS predictions.

The goal of this dissertation is to improve the predictive accuracy of the LAS by mitigating selection bias so that lungs are allocated to the appropriate patients in the appropriate order. We undertook both quantitative and qualitative aims to achieve this goal. Quantitatively, we aim to mitigate selection bias in the LAS and improve its predictive accuracy. To do so, we build upon recent methodological research reported in both the transplant and causal inference literature that has begun to incorporate the problem of dependent censoring into the calculation of transplant benefit and waitlist urgency (Tayob & Murray, 2017; Vock et al., 2017; Vock et al., 2013; Xiang & Murray, 2012; Xiang et al., 2014). Qualitatively, we aim to understand how clinicians perceive and use information from the LAS in their decision-making processes. To do so, we conduct qualitative interviews with lung transplant surgeons and pulmonologists throughout the United States to determine how these clinicians use the LAS in practice, how they think about selection and survivor bias in lung allocation, and whether they feel the LAS should be modified to account for such biases.

Background and Significance

Lung Allocation in the United States

The use of lung transplantation as a treatment for end-stage lung diseases has become increasingly common in the United States, with over 2,000 lung transplants being performed each year (U.S. Department of Health & Human Services, 2018). However, waitlist mortality is still quite high (by age, it varied from 14.4 to 28.4 deaths per 100 waitlist years in 2018, the most recent year for which data were available), and waiting times vary depending on patients' diagnoses (U.S. Department of Health & Human Services, 2018). In fact, concerns about inequities in waitlist mortality among different types of patients led the Organ Procurement and Transplantation Network (OPTN) to adopt the Department of Health and Human Service's "Final Rule" in 2005. This Rule mandated the development of a new lung allocation system based on "medical urgency" rather than wait time (U.S. Department of Health and Human Services October 20, 1999).

Today, donor lungs are allocated to potential recipients via the LAS, which prioritizes patients based on a combination of estimated waitlist urgency and transplant benefit so that equity and utility can be maximized (Gottlieb, 2017; Veatch & Ross, 2015). More specifically, the LAS relies on clinical factors – such as demographic and laboratory values – to predict waitlist and post-transplant survival. These predictions are then used to estimate the number of days of life a person would gain over the next year if they receive transplant compared to if they do not receive transplant. Patients for whom this comparison is more favorable receive higher priority for transplant.

Performance of the LAS

Studies of LAS performance raise concerns about the accuracy of the predicted pre- and post-transplant survival components of the score. In a study of the relationship between transplant candidates' LAS scores and their pre- and post-transplant outcomes, Russo et al. questioned why 77% of lower-priority patients (i.e., patients with LAS scores between 40 and 49) undergo transplantation, whereas only 45% of higher-priority patients (i.e., patients with LAS scores of 80 or more) receive transplants (Russo et al., 2011), even though lower-priority candidates typically experience less survival benefit from transplant compared to higher-priority candidates. Perhaps more concerningly, Gries et al.'s attempt to develop and validate a predictive model for 5-year survival after lung transplantation yielded extremely low measures of predictive ability (e.g., areas under the receiver operating characteristic curve between 0.553 and 0.591) (Gries et al., 2010). Each of these studies suggest that the LAS has a serious problem: it does not adequately predict survival. Consequently, individuals are prioritized inappropriately, which, in turn, leads to inequitable allocation of donor lungs (Halpern, 2015; Maxwell et al., 2015; Russo et al., 2013; Thabut et al., 2012; Wille et al., 2013).

Updates to LAS Coefficients and Selection Bias

Although the coefficients of the LAS were updated in 2010 and 2015 by refitting the model using a more recent cohort of patients, the problems discussed above have persisted, as is evident by the most recent OPTN report (U.S. Department of Health & Human Services, 2018). One reason why re-estimation of LAS coefficients via model refitting is insufficient to alleviate these disparities is that this approach does not fully account for selection bias. Selection bias arises due to 1) dependent censoring

(Egleston et al., 2007; Freiman & Small, 2014; Wang et al., 2017): individuals are removed from the waitlist upon receipt of transplant; and 2) survivor bias (Glymour & Greenland, 2008): individuals must survive long enough in order to receive transplant. Failing to account for selection bias can lead to inaccurate predicted probabilities and suboptimal organ allocation. As an example, naïvely considering post-transplant survival probability only among individuals who survive on the waitlist might lead to the conclusion that a patient with a lower risk of dying on the waitlist is a better candidate for transplant than a patient with a higher risk of dying on the waitlist, even though the latter patient might actually experience a greater benefit from transplant than the former.

Causal Inference and Potential Outcomes

The "potential outcomes" framework allows us to address the selection bias issue by conceptualizing it in terms of counterfactual outcomes. Utilized extensively in the causal inference field, the "potential outcomes" framework defines individual causal effects as the difference (or ratio) of potential outcomes that would have been observed for a given individual under different exposure levels (Little & Rubin, 2000). Since only one potential outcome will be observed for any particular individual – namely, the outcome that resulted under the level of exposure that the individual actually experienced (Little & Rubin, 2000) – we cannot estimate individual causal effects using observed data. However, under assumptions, we can estimate an average causal effect across a well-defined population. Thus, application of the potential outcome framework to our problem requires identification of the population of interest. Here, we are interested in making inference about all individuals registered on the UNOS waitlist. That is, patients become eligible for our analysis with waitlist registration. Within this population, the target quantity is defined as the estimated survival benefit patients would

accrue if they received a transplant, compared to what they would experience had they not received a transplant (but remained on the waitlist). Individuals who die, are removed from the waitlist, or are otherwise lost to follow-up prior to transplantation are censored from our analyses. In Chapters 2 and 3, we incorporate these causal inference principles into existing prediction model frameworks to develop and evaluate an improved LAS algorithm.

Clinician Decision-Making and the LAS

When deciding whether to pursue lung transplantation, clinicians must determine whether the "potential benefits of transplant outweigh the potential risks and harms" (Yusen, 2009). The LAS ostensibly performs this comparison by comparing patients' predicted survival with versus without transplant and allocating organs to those for whom this comparison is more favorable. However, additional factors beyond the LAS can impact which patients are selected for – and ultimately receive – transplant, including the timing of referral (Weill et al., 2015) and geographic, gender, or racial/ethnic disparities in waitlist registration (Mooney et al., 2018; Ross-Driscoll et al., 2020; Thabut et al., 2012; Wille et al., 2013), pre- and post-transplant survival (Egan & Edwards, 2016; Maxwell et al., 2014; Russo et al., 2011), and donor organ availability (Benvenuto et al., 2018; Drolen et al., 2020; Ross-Driscoll et al., 2020). These factors can exacerbate the selection bias problem detailed above. For example, individuals who are referred to a transplant center too early or too late (Mooney et al., 2019; Ramos et al., 2019; Stephenson et al., 2021), or who seem (from the perspective of transplant clinicians) to have poor social and/or financial support (Blumenthal et al., 2017; Ladin et al., 2019), may not be registered on the waitlist, despite the fact that they could, in fact, benefit from transplant.

Although much statistical research has been conducted surrounding the LAS and selection bias (Vock et al., 2017; Vock et al., 2013; Xiang & Murray, 2012; Xiang et al., 2014), it is unclear how clinicians employ the LAS in their decision-making processes, and how they interpret and respond to selection bias. To date, qualitative studies in this space tended to restrict their focus to one specific aspect of the transplantation process (e.g., screening/evaluation (Blumenthal et al., 2017; Volk et al., 2011) or donor organ acceptance (Loss et al., 2013)) at the exclusion of other steps; emphasized patient-level selection factors rather than program-level factors (Blumenthal et al., 2017); and ignored how patients' LAS scores (or, for liver transplant candidates, Model for End Stage Liver Disease [MELD] scores) might influence candidate selection (Blumenthal et al., 2017; Volk et al., 2011).

Given that OPTN is currently developing a new organ allocation framework – the continuous distribution model (Kasiske et al., 2020; Snyder et al., 2018; U.S. Department of Health & Human Services) – and that this new framework is planned to be implemented in lung transplant selection, followed by other organs (e.g., kidney, liver, heart), it is important to develop a thorough understanding of where selection bias can arise in the current lung allocation system and how such biases can impact the success and fairness of lung allocation. Toward that end, in Chapter 4, we conducted a qualitative study of lung transplant surgeons and pulmonologists to understand the role that the LAS plays in clinical decision making, how transplant clinicians think about selection and survivor bias in lung allocation, and whether they feel the LAS should be modified to account for such biases. Studying these questions qualitatively can help inform the findings of our quantitative aims in the context of clinical practice and can generate hypotheses for future research.

Dissertation Aims

This dissertation is comprised of both quantitative and qualitative aims. The quantitative aim is to modify the LAS algorithm to account for selection bias and compare the performance of the modified and existing LAS. The qualitative aim is to understand the role that the LAS plays in clinical decision making. These two aims are accomplished through three separate analyses, which are thematically linked by the notion of selection bias and how it is a concern in lung transplantation.

In Chapter 2, we propose a weighted estimation strategy to mitigate selection bias in the pre- and post-transplant LAS models. We then compare the predictive performance of the modified and existing LAS models by 1) evaluating discrimination and calibration, and 2) comparing patient rankings and predicted survival. In Chapter 3, we examine the clinical impact of our modified LAS developed in Chapter 2. More specifically, we use observed data to investigate the demographic and clinical characteristics of patients who receive higher or lower priority under the modified LAS relative to the existing one. We also conduct a statistical simulation study to evaluate how the modified LAS might impact waitlist and post-transplant survival if it were implemented in clinical practice. In Chapter 4, we conduct semi-structured interviews with lung transplant surgeons and pulmonologists throughout the United States to examine respondents' understanding of selection bias and how it is a concern in lung transplantation and the LAS. In Chapter 5, we provide concluding remarks and suggest potential avenues for future research.

CHAPTER 2. MITIGATING SELECTION BIAS IN ORGAN ALLOCATION MODELS

This chapter has been published in *BMC Medical Research Methodology*, and is reprinted here with permission. The citation for this publication is:

Schnellinger EM, Cantu E, Harhay MO, Schaubel DE, Kimmel SE, Stephens-Shields AJ. Mitigating Selection Bias in Organ Allocation Models. *BMC Medical Research Methodology* 2021; 21: 191. doi: https://doi.org/10.1186/s12874-021-01379-7.

Abstract

Background. The lung allocation system in the U.S. prioritizes lung transplant candidates based on estimated pre- and post-transplant survival via the Lung Allocation Scores (LAS). However, these models do not account for selection bias, which results from individuals being removed from the waitlist due to receipt of transplant, as well as transplanted individuals necessarily having survived long enough to receive a transplant. Such selection biases lead to inaccurate predictions.

Methods. We used a weighted estimation strategy to account for selection bias in the pre- and post-transplant models used to calculate the LAS. We then created a modified LAS using these weights, and compared its performance to that of the existing LAS via time-dependent receiver operating characteristic (ROC) curves, calibration curves, and Bland-Altman plots.

Results. The modified LAS exhibited better discrimination and calibration than the existing LAS, and led to changes in patient prioritization.

Conclusions. Our approach to addressing selection bias is intuitive and can be applied to any organ allocation system that prioritizes patients based on estimated pre- and post-transplant survival. This work is especially relevant to current efforts to ensure more equitable distribution of organs.

Background

The Organ Procurement and Transplantation Network (OPTN) is responsible for allocating deceased-donor organs in the United States. The OPTN utilizes separate policies to govern the allocation of livers, kidneys, hearts, and lungs (Organ Procurement and Transplantation Network, 2020). For example, liver transplant candidates receive a Model for End-Stage Liver Disease (MELD) score; those wait-listed for kidney transplantation receive an Estimated Post-Transplant Survival (EPTS) score, and lung transplant candidates receive a Lung Allocation Score (LAS) (Egan et al., 2006; Organ Procurement and Transplantation Network, 2020; Veatch & Ross, 2015). More specifically, the LAS is derived from models that predict both pre-transplant and post-transplant survival and aims to balance each patient's predicted transplant benefit (i.e., difference between survival with versus without a lung transplant) against their waitlist urgency (Egan et al., 2006; Organ Procurement and Transplantation Network, 2020; Veatch & Ross, 2015).

Such models are particularly susceptible to selection bias. Estimates of pre-transplant survival are subject to selection bias in the form of dependent censoring because patients can be removed from the waiting list prior to one year of follow-up due to receipt of transplant, loss to follow-up, or other clinical reasons (e.g., the inability to withstand the transplant surgery). In each of these cases, the patients' true one-year pre-transplant survival is unobserved. Estimates of post-transplant survival similarly are subject to a type of selection bias – "survivor bias" (Egleston et al., 2007; Freiman & Small, 2014; Glymour & Greenland, 2008; Wang et al., 2017) – because they have been derived using information only among patients who received a transplant. Thus, post-transplant survival models are applied to all wait-listed patients, but are fitted using only transplanted patients. In particular, while statistical models used to estimate transplant

benefit account for variation in patient characteristics, they do not account for the fact that in order to receive a transplant, individuals must: 1) survive on the waitlist long enough for a suitable donor organ to become available, and 2) have sufficient priority to actually receive the organ. Since individuals who survive one year or more on the waitlist might be inherently different from individuals who die, receive a transplant, or are censored (e.g., lost to follow-up or removed from the waitlist for other clinical reasons, such as being too sick to withstand the transplant surgery) prior to one year, failure to incorporate such information in the models used to estimate transplant benefit and waitlist urgency can lead to inaccurate predictions.

Lung transplantation represents an important example to study because it is a highly effective treatment, but organs are scarce. Waitlist mortality is high, waitlist times vary, and there are concerns about inequities in waitlist mortality and organ allocation (U.S. Department of Health & Human Services, 2018). In fact, concerns about these inequities led the Department of Health and Human Services to mandate the development of the LAS based on medical need rather than wait time (Egan et al., 2006; Gottlieb, 2017; U.S. Department of Health and Human Services, October 20, 1999; United Network for Organ Sharing, 2015). Donor lungs are now allocated to recipients based on the LAS (Egan et al., 2006; Gottlieb, 2017; U.S. Department of Health and Human Services, October 20, 1999; United Network for Organ Sharing, 2015) which is calculated using the predicted difference between transplant benefit and waitlist urgency, with transplant benefit defined as one-year post-transplant survival minus one-year waitlist survival, and waitlist urgency defined as one-year waitlist survival. Conceptually, the LAS aims to determine the number of days of life a person would gain over the next year if they receive transplant compared to if they do not receive transplant, and prioritizes patients for whom this comparison is more favorable.

We use use directed acyclic graphs (DAG) to illustrate how selection bias can lead to inaccurate predictions for both the pre-transplant and post-transplant prediction models. More specifically, Figure 2.1A shows the hypothesized DAG of the relationship between patients' covariates at the time of their hypothetical organ offer [L(0)], receipt of transplant, and post-transplant survival prior to weighting. Whether or not a particular individual will survive long enough to receive transplant likely depends on post-baseline characteristics [L(t)], such as acute exacerbations of their illness and the amount of time they have already spent on the waitlist. These variables are also likely to influence how long the patient survives post-transplant. Before weighting, these variables are unaccounted for. At first, failure to account for these variables might seem fine, as Transplant is a collider on the path: Covariates informing LAS → Transplant ← L(t) → Post-transplant survival. However, the patients who survive long enough to receive transplant may differ from those who do not. Moreover, the estimate of post-transplant survival used by the current LAS is inherently restricted to the former individuals. Thus, a spurious association is induced between patients' covariates at the time of their hypothetical organ offer and post-transplant survival. This spurious association can lead to biased estimates of post-transplant survival, which implies that the current prioritization of lung transplant recipients may be inaccurate. This bias can occur even with measured covariates, because: 1) not all measured covariates which are available in the UNOS database are included in the LAS models (e.g., geographic differences in transplant listing and outcomes (Thabut et al., 2012)); and 2) even if a measured covariate is included in the LAS, the association between such a covariate and post-transplant survival can be different in the post-transplant subset than it is in the full waitlist population.

Figure 2.1B shows the same relationship between patients' covariates at the time of their hypothetical organ offer [L(0)], receipt of transplant, and post-transplant survival after weighting. Essentially, the weighting approach we propose below capture the information contained in L (i.e., post-baseline covariates and time on the waitlist), thereby removing the arrows into transplant. This approach should mitigate the survivor bias problem.

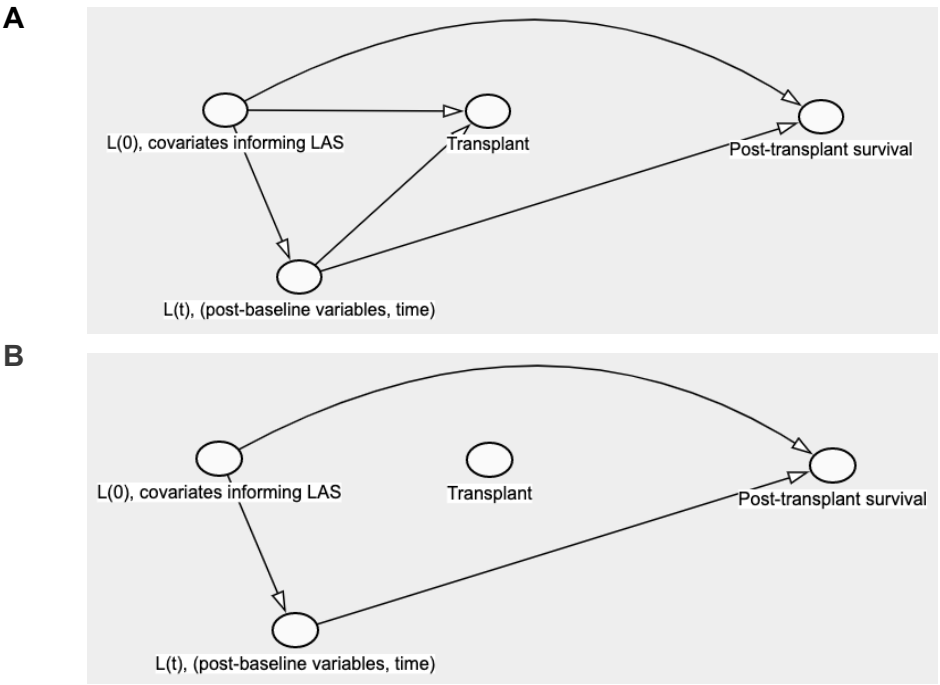


Figure 2.1. Hypothesized Directed Acyclic Graph

Hypothesized DAG depicting the relationship between patients' covariates, receipt of transplant, and post-transplant survival A) prior to weighting, and B) after weighting.

Although prior research has incorporated weights in the pre-transplant survival model (Xiang & Murray, 2012), the models did not capture important geographic differences in patient selection and survival, and no work, to our knowledge, has estimated weights to the post-transplant survival model.

In this study, we attempt to bring principles from causal inference into the existing prediction model framework employed by the LAS to improve organ allocation.

Specifically, we develop a modified LAS using inverse probability weighting to improve the accuracy of the LAS by accounting for selection bias in the pre- and post-transplant survival models. Our work incorporates additional factors in the pre-transplant weights to better address selection bias in the pre-transplant survival model. We also develop new weights to address selection bias in the post-transplant survival model.

Methods

Data Source & Setup

We used publicly available pre- and post-lung transplant data from the United Network for Organ Sharing (UNOS). Our development cohort consisted of all patients 18 years or older who were listed for single or bilateral lung transplantation in the United States between January 1, 2010, and December 31, 2013. This date range was chosen because it ensures that no patient experiences any person-time prior to the implementation of the LAS (i.e., prior to May 2005) and is consistent with the development cohort used to fit the current LAS models. Our testing cohort consisted of patients meeting these same criteria who were listed between January 1, 2016, and December 31, 2017 (the last complete year of available data). Patients listed during 2014 and 2015 were excluded from our analyses to ensure that 1) our development cohort is consistent with the development cohort used to fit the current LAS; and 2) our testing cohort does not include any patients for whom prior versions of the LAS were used in clinical practice. To avoid concerns about positivity violations associated with the likelihood of receiving a transplant, we removed individuals who had clinical contraindications to receiving transplant (e.g., those with panel reactive antibodies greater than 90%), and individuals with both restrictive lung disease (diagnosis group D) and height less than five feet who require such small donor organs that they rarely find a

match. In both cohorts, patients were followed from their initial listing date to their time of transplant, death, loss to follow-up, or removal from the waitlist due to other clinical reasons (e.g., inability to withstand the transplant surgery), whichever occurred first.

Our analysis uses daily time intervals. If individuals presented to clinic multiple times in one day, their last record in that day was retained; conversely, if individuals did not present to clinic in a given day, covariate information from their most recent visit was used (i.e., last observation carried forward, LOCF). The use of LOCF is consistent with UNOS guidelines (Organ Procurement and Transplantation Network, 2020). Any values which remained missing (i.e., due to the absence of recent covariate values for a particular patient) were then filled in following the UNOS value substitution policy (Organ Procurement and Transplantation Network, 2020). This policy stipulates that if any covariates remain missing after LOCF, then those values should be replaced with a default value, which represents a value either in the normal range for that covariate or which would yield the lowest LAS score for that patient. These default values are publicly available (Organ Procurement and Transplantation Network, 2020). After data cleaning, 8379 patients provided 1,751,912 total records. These data were divided into pre- and post-transplant subsets, with the pre-transplant subset containing daily time intervals, and the post-transplant subset consisting of a single record per patient.

Constructing the Weights

To estimate waitlist urgency and transplant benefit, the LAS relies on two separate outcome models. The first estimates one-year waitlist (pre-transplant) survival, while the second estimates one-year post-transplant survival. Circumventing survivor bias requires us to “map” the survival probabilities obtained among the post-transplant group back to the full waitlist population. To do so, we constructed inverse probability of

treatment weights (IPTW) and inverse probability of censoring weights (IPCW) to create a pseudo-population which reflects the characteristics of the full waitlist population, rather than just the subset of patients who were selected to receive transplant. This approach accounts for differences between the post-transplant subset and the full waitlist population that arise due to 1) measured covariates which are available in the UNOS database, but not included in the existing LAS models; and 2) measured covariates that are already included in the LAS, but which have a different effect on post-transplant survival in the post-transplant subset than they do in the full waitlist population; it does not account for differences due to unmeasured covariates.

We define the following notation: A indicates exposure (receipt of transplant); L represents covariates; the subscript i denotes observations pertaining to the same subject; and the subscript k denotes the day in which the exposure or covariates were observed. Thus, $\mathrm{A_{ki}}$ represents the observed exposure at day k for subject i, and can take on values $\mathrm{a_{ki}} = 1$ or $\mathrm{a_{ki}} = 0$ for transplanted or not, respectively. Similarly, $\mathrm{L_{ki}} = l_{\mathrm{ki}}$ represents the observed covariate values at day k for subject i. Overbars indicate exposure or covariate history. Thus, $\overline{\mathrm{A}}_{(\mathrm{k}-1)\mathrm{i}}$ represents the vector of exposure values for subject i up through day $\mathrm{k}-1$ (which, by design, will equal 0 up until the time subject i receives transplant); similarly, $\overline{\mathrm{L}}_{\mathrm{ki}}$ represents the vector of covariate values for subject i up through day k.

Weights were considered separately for the pre- and post-transplant outcome models. The IPTW for the pre-transplant model accounts for time-varying covariate values and variable time of transplant by patient. Stabilized IPTW (sw_i) were constructed as the cumulative product of the probability of receiving transplant at each day,

conditional on individuals not having received transplant yet (observed through the prior day ($\overline{\mathrm{A}}_{(\mathrm{k}-1)\mathrm{i}}$)) and covariate history observed through the current day ($\overline{\mathrm{L}}_{\mathrm{ki}}$):

$$sw_i = \frac{\prod_{k=0}^{K} \Pr\left[\mathrm{A_{ki}} = \mathrm{a_{ki}} | \overline{\mathrm{A}}_{(\mathrm{k}-1)\mathrm{i}} = \overline{0}\right]}{\prod_{k=0}^{K} \Pr\left[\mathrm{A_{ki}} = \mathrm{a_{ki}} | \overline{\mathrm{A}}_{(\mathrm{k}-1)\mathrm{i}} = \overline{0}, \overline{\mathrm{L}}_{\mathrm{ki}} = \overline{\mathrm{l}}_{\mathrm{ki}}\right]}$$

The denominator of the pre-transplant IPTW weights was estimated using pooled logistic regression, where predictors included the time-varying covariates of the published waitlist LAS model (i.e., age, bilirubin, body mass index, cardiac index, central venous pressure, continuous mechanical ventilation, serum creatinine, diabetes, diagnosis group, forced vital capacity, functional status, oxygen need at rest, partial pressure of carbon dioxide, pulmonary artery systolic pressure, six-minute walk distance), and also blood type, gender, race, and height (consistent with (Xiang & Murray, 2012)). Additional predictors included geography (to account for differences in survival across UNOS regions), time (days) since waitlist registration (modeled using restricted cubic splines with five knots; knots were chosen using Harrell's recommended percentiles (i.e., 5, 27.5, 50, 72.5, and 95), as implemented by STATA's *mkspline* function (Harrell, 2001)), and one-month lagged versions of the covariates of the published waitlist LAS (to capture patients' waitlist history). The numerator of the pre-transplant IPTW weights was also estimated using pooled logistic regression, where predictors included covariate values at the time of waitlist registration (i.e., baseline covariates) to improve the stability of the weights.

Stabilized IPCW (scw_i) were similarly constructed as the cumulative product of the probability of being censored at each day, conditional on individuals not having been censored or transplanted yet (observed through the prior day; $\overline{\mathrm{C}}_{(\mathrm{k}-1)\mathrm{i}}$ and $\overline{\mathrm{A}}_{(\mathrm{k}-1)\mathrm{i}}$, respectively) and covariate history observed through the current day ($\overline{\mathrm{L}}_{\mathrm{ki}}$):

$$sw\_c_i = \frac{\prod_{k=0}^{K} \Pr\left[\mathrm{C_{ki}} = 0 | \bar{\mathrm{C}}_{(\mathrm{k}-1)\mathrm{i}} = \bar{0}, \bar{\mathrm{A}}_{(\mathrm{k}-1)\mathrm{i}} = \bar{0}\right]}{\prod_{k=0}^{K} \Pr\left[\mathrm{C_{ki}} = 0 | \bar{\mathrm{C}}_{(\mathrm{k}-1)\mathrm{i}} = \bar{0}, \bar{\mathrm{A}}_{(\mathrm{k}-1)\mathrm{i}} = \bar{0}, \bar{\mathrm{L}}_{\mathrm{ki}} = \bar{\mathrm{l}}_{\mathrm{ki}}\right]}$$

These IPCW weights were obtained by fitting pooled logistic regression models with pre-transplant censoring (e.g., loss to follow-up or removal from the waitlist for other clinical reasons, such as being too sick to withstand the transplant surgery) as the outcome, and predictors including variables in the pre-transplant LAS, these same covariates lagged by one month, time spent on the waitlist, geography, gender, race, and height. The final, time-varying pre-transplant weight was then calculated as the product of the stabilized IPTW and stabilized IPCW at each time point: $swscw_i = sw_i * scw_i$.

Among patients who received transplant, the denominator of the IPTW is the same as the denominator of the pre-transplant IPTW described above. The numerator, however, includes the transplant indicator only. Covariate values at waitlist registration were excluded from the numerator of the post-transplant IPTW because these covariates are not present in the post-transplant outcome model (Robins et al., 2000; Yang & Joffe, 2012). The post-transplant IPTW was then taken to be the cumulative product of this marginally stabilized IPTW associated with the last record for each transplanted patient in the pre-transplant data set.

The post-transplant stabilized IPCW was estimated by fitting a logistic regression model with the post-transplant censoring variable (i.e., loss to follow-up after transplant) as the outcome and the covariates of the published post-transplant LAS as predictors (i.e., age, cardiac index, continuous mechanical ventilation, serum creatinine, diagnosis group, functional status, oxygen need at rest, six-minute walk distance), along with geography, gender, race, and time (days) spent on the waitlist (modeled using restricted cubic splines with five knots; knots were chosen using Harrell's recommended percentiles (i.e., 5, 27.5, 50, 72.5, and 95), as implemented by STATA's *mkspline*

function (Harrell, 2001)). The final post-transplant weight was calculated as the product of the stabilized IPTW and stabilized IPCW. To minimize the impact of extreme weights on our outcome models, we progressively truncated the final pre- and post-transplant weights following the procedure outlined in (Cole & Hernán, 2008), with weights finally truncated at the 0.25% (99.75%) percentile. Distributions of these weights appear in Appendix 2.1.

Fitting the outcome models

We fit a weighted Cox proportional hazards model to each patient's baseline record in the pre-transplant data, weighted by each patient's daily time-varying weight to estimate one-year pre-transplant survival accounting for dependent censoring (Buchanan et al., 2014). This approach allows us to incorporate information from time-varying covariates and time on the waitlist captured by the IPTW and IPCW models, while still retaining the same form of the outcome model as the current pre-transplant LAS. It also provides predicted probabilities of one-year survival on the waitlist, which is consistent with the definition of waitlist urgency used by the current LAS. Our weighted Cox proportional hazards model contains all covariates included in the existing pre-transplant LAS (but not follow-up time, as UNOS policy prohibits follow-up time from being included in the outcome model). Thus, the variables in the outcome model are the same in the modified LAS as in the existing LAS, but the coefficient estimates vary.

Similarly, we estimated one-year post-transplant survival by fitting a weighted Cox proportional hazards model to the post-transplant subset which included the covariates in the existing post-transplant LAS, and was weighted by each patient's post-transplant weight (which is fixed at the time of transplant). Consistent with UNOS policy, the post-transplant outcome model did not include time since waitlist registration. The

estimate of one-year post-transplant survival obtained from this weighted outcome model differs from that included in the current LAS, as the weighted outcome model provides an estimate of survival that reflects the *entire* waitlist population, whereas the current LAS estimates this quantity only among the subset of individuals who did, in fact, receive transplant. Taken together, the weighted pre- and post-transplant outcome models produce survival estimates that align more closely with the LAS's conceptual goal of comparing the number of days of life a person would gain over the next year if they receive transplant versus if they do not receive transplant.

Assessing model performance

To assess the discrimination of the pre- and post-transplant outcome models, we constructed time-dependent receiver operating characteristic (ROC) curves and evaluated the area under these curves (AUC) via nearest-neighbor smoothing at one year post-waitlist registration and one year post-transplantation, respectively (Cattaneo et al., 2017; Heagerty et al., 2000). This approach accommodates censoring by viewing survival time as a "time-varying binary outcome" at each possible time point, and estimating the sensitivity and specificity of the model among all patients who are still alive and at risk of the outcome at those time points (Cattaneo et al., 2017; Heagerty et al., 2000). Separate statistics were computed for the development and testing cohorts.

Calibration was evaluated graphically by defining low-, medium-, and high-risk categories based on tertiles of the linear predictor of the pre- and post-transplant outcome models, averaging the survival functions within each of these risk categories, and then overlaying the observed (Kaplan-Meier) and predicted survival curves for each risk category (Royston, 2015). This approach allows us to evaluate the calibration of the pre- and post-transplant outcome models at any time point after waiting list registration

among all patients who are alive and at risk of the outcome at those time points (Royston, 2015). Separate calibration plots were constructed for the development and testing cohorts.

Comparing the modified LAS to the existing LAS

The current LAS is composed of a pre-transplant outcome model and a post-transplant outcome model, where only the pre-transplant outcome model is weighted using a select number of covariates (Xiang & Murray, 2012). These outcome models are used to predict one-year waitlist and one-year post-transplant survival, which are combined into a raw score by computing one-year post-transplant survival minus two times one-year waitlist survival (Egan et al., 2006; Organ Procurement and Transplantation Network, 2020; United Network for Organ Sharing, 2015). This raw score is normalized so that it ranges from 0 to 100, with higher values indicating greater priority for transplantation (Egan et al., 2006; Organ Procurement and Transplantation Network, 2020; United Network for Organ Sharing, 2015). We construct modified pre- and post-transplant outcome models by applying weights to both models, as described above. We applied the weighted pre- and post-transplant outcome models to the testing cohort to estimate a modified LAS score for each patient considering all possible offer dates in 2016 and 2017. At each offer date, we subset the data to include only patients who were alive, registered on the waitlist, and not yet transplanted at that date. Then, we computed modified waitlist urgency and modified post-transplant survival measures following UNOS guidelines (Organ Procurement and Transplantation Network, 2020; United Network for Organ Sharing, 2015). We computed daily, person-specific survival estimates for the first year spent on the waitlist and the first year post-transplant using the baseline hazard, weighted model coefficients, and each individual's covariate values.

Each patients' resulting waitlist and post-transplant survival probabilities were summed to obtain the modified waitlist urgency (mWL_i) and modified post-transplant survival (mPT_i) measures:

$$mWait_i = \sum_{k=1}^{365} S_{Wait,(k-1)i} * 1\ day$$

$$mPT_i = \sum_{k=1}^{365} S_{PT,(k-1)i} * 1\ day$$

where $S_{Wait,(k-1)i}$ and $S_{TX,(k-1)i}$ respectively represent the pre-transplant (waitlist) and post-transplant survival probabilities for subject i at day $k-1$. The modified raw score was then computed as:

$$Modified\ Raw\ Score = mPT_i - 2 * mWait_i$$

Taking into account the maximum and minimum pre- and post-transplant survival, the modified raw score was normalized via the following equation, consistent with the existing LAS:

$$Modified\ LAS = \frac{100 * [(Modified\ Raw\ Score_i) + 730]}{1095}$$

Existing LAS scores were estimated for each patient by applying the published pre- and post-transplant LAS models to the testing cohort following the same procedure as above (Organ Procurement and Transplantation Network, 2020; United Network for Organ Sharing, 2015) with coefficients from the existing LAS. Last, we constructed two sets of rankings for eligible patients at each offer date: rankings based on their modified LAS scores and rankings for the same patients based on their existing LAS scores.

To assess the difference between the modified and existing LAS models, we constructed Bland-Altman plots of 1) the modified LAS score versus the existing LAS score; and 2) the modified patient rank versus the existing patient rank (Bland & Altman, 1986). We also created a scatter plot of the difference in estimated post-transplant survival versus the difference in pre-transplant survival obtained under the modified and existing LAS models to examine which of these factors drive changes in patient prioritization.

Analyses were conducted using Stata (StataCorp LLC, College Station, TX) and R (R Foundation for Statistical Computing, Vienna, Austria).

Results

Table 2.1 displays demographic and clinical characteristics of the complete waiting list population at the time of waitlist registration in the development and testing cohorts. Table 2.2 displays similar information at the time of transplantation for the subset of patients in each cohort who received transplant.

Table 2.1. Demographic and clinical characteristics of the complete waiting list population in the development and testing cohorts

Covariates are measured at the time of waitlist registration in both cohorts.

	Development Cohort	Testing Cohort
Total number of patients (N)	8379	5354
Waiting time (days), median (IQR)	73.0 (19.0, 237.0)	57.0 (16.0, 154.0)
Death on waitlist within 1 year of waitlist registration	56 (0.7%)	25 (0.5%)
All deaths on waitlist	751 (9.5%)	288 (5.7%)
Removal from waitlist due to transplant	7157 (85.4%)	4154 (77.6%)
REGION		
1	251 (3.0%)	185 (3.5%)
2	1264 (15.1%)	953 (17.8%)
3	951 (11.3%)	522 (9.7%)
4	1216 (14.5%)	636 (11.9%)
5	1203 (14.4%)	878 (16.4%)
6	208 (2.5%)	110 (2.1%)
7	607 (7.2%)	438 (8.2%)
8	502 (6.0%)	312 (5.8%)
9	285 (3.4%)	180 (3.4%)
10	999 (11.9%)	684 (12.8%)
11	893 (10.7%)	456 (8.5%)

GENDER		
F	3520 (42.0%)	2263 (42.3%)
M	4859 (58.0%)	3091 (57.7%)
RACE/ETHNICITY		
White	6966 (83.1%)	4224 (78.9%)
Black	738 (8.8%)	552 (10.3%)
Hispanic	494 (5.9%)	411 (7.7%)
Asian	131 (1.6%)	124 (2.3%)
Other	50 (0.6%)	43 (0.8%)
DIAGNOSIS		
A (Obstructive Disease)	2423 (28.9%)	1444 (27.0%)
B (Pulmonary Hypertension)	398 (4.7%)	299 (5.6%)
C (Cystic Fibrosis)	975 (11.6%)	538 (10.0%)
D (Pulmonary Fibrosis)	4583 (54.7%)	3073 (57.4%)
BLOOD TYPE		
A	3297 (39.3%)	2100 (39.2%)
AB	307 (3.7%)	195 (3.6%)
B	916 (10.9%)	607 (11.3%)
O	3859 (46.1%)	2452 (45.8%)
Age (years), median (IQR)	59.0 (49.0, 65.0)	60.0 (52.0, 66.0)
Bilirubin (mg/dL), median (IQR)	0.7 (0.7, 0.7)	0.7 (0.7, 0.7)
BMI (kg/m^3), median (IQR)	25.5 (21.6, 28.9)	26.0 (22.0, 29.2)
Height (feet), median (IQR)	5.6 (5.3, 5.8)	5.6 (5.3, 5.8)
Cardiac index <2 $L/min/m^2$	2096 (25.0%)	896 (16.7%)
Central venous pressure (mmHg), median (IQR)	5.0 (5.0, 8.0)	5.0 (5.0, 8.0)
Continuous mechanical ventilation	301 (3.6%)	217 (4.1%)
Creatinine (serum) (mg/dL), median (IQR)	0.8 (0.7, 1.0)	0.8 (0.7, 1.0)
Diabetes	1892 (22.6%)	1074 (20.1%)
Forced vital capacity % predicted, median (IQR)	48.0 (38.0, 61.0)	48.0 (38.0, 62.0)
Functional status (none)	1027 (12.3%)	604 (11.3%)
Oxygen need at rest (L/min), median (IQR)	3.0 (2.0, 6.0)	4.0 (2.0, 6.0)
pCO_2, median (IQR)	42.0 (40.0, 49.0)	43.0 (40.0, 50.0)
Pulmonary artery systolic pressure (mmHg), median (IQR)	37.0 (30.0, 47.0)	38.0 (31.0, 48.0)
Six-minute walk distance (feet), median (IQR)	870.0 (550.0, 1176.0)	880.5 (580.0, 1142.0)

Table 2.2. Demographic and clinical characteristics of the subset of individuals who received transplant in the development and testing cohorts

Covariates are measured at the time of transplantation in both cohorts.

	Development Cohort	**Testing Cohort**
Total number of patients (N)	7074	4154
Follow-up time post-transplant (days), median (IQR)	1439.0 (721.0, 1834.0)	199.0 (68.0, 365.0)
Death within 1 year of transplant	935 (13.2%)	346 (8.3%)
All deaths post-transplant	3063 (43.7%)	420 (10.9%)
REGION		
1	201 (2.8%)	160 (3.9%)
2	1068 (15.1%)	680 (16.4%)
3	784 (11.1%)	405 (9.7%)
4	1030 (14.6%)	534 (12.9%)
5	1031 (14.6%)	675 (16.2%)
6	174 (2.5%)	67 (1.6%)
7	503 (7.1%)	334 (8.0%)
8	432 (6.1%)	251 (6.0%)
9	241 (3.4%)	126 (3.0%)
10	794 (11.2%)	528 (12.7%)
11	816 (11.5%)	394 (9.5%)
GENDER		
F	2780 (39.3%)	1594 (38.4%)
M	4294 (60.7%)	2560 (61.6%)

RACE/ETHNICITY		
White	5938 (83.9%)	3339 (80.4%)
Black	581 (8.2%)	397 (9.6%)
Hispanic	405 (5.7%)	299 (7.2%)
Asian	107 (1.5%)	93 (2.2%)
Other	43 (0.6%)	26 (0.6%)
DIAGNOSIS		
A (Obstructive Disease)	2792 (39.5%)	1038 (25.0%)
B (Pulmonary Hypertension)	265 (3.7%)	182 (4.4%)
C (Cystic Fibrosis)	764 (10.8%)	436 (10.5%)
D (Pulmonary Fibrosis)	3253 (46.0%)	2498 (60.1%)
BLOOD TYPE		
A	2821 (39.6%)	1702 (41.0%)
AB	266 (3.7%)	157 (3.8%)
B	772 (10.8%)	474 (11.4%)
O	3272 (45.9%)	1821 (43.8%)
Age (years), median (IQR)	60.0 (50.0, 65.0)	61.0 (52.0, 66.0)
Bilirubin (mg/dL), median (IQR)	0.7 (0.7, 0.7)	0.7 (0.7, 0.7)
BMI (kg/m^3), median (IQR)	25.4 (21.6, 28.8)	25.9 (22.1, 29.0)
Height (feet), median (IQR)	5.6 (5.3, 5.8)	5.6 (5.3, 5.8)
Cardiac index <2 $L/min/m^2$	1406 (19.9%)	596 (14.3%)
Central venous pressure (mmHg), median (IQR)	5.0 (5.0, 8.0)	5.0 (5.0, 8.0)
Continuous mechanical ventilation	457 (6.5%)	320 (7.7%)
Creatinine (serum) (mg/dL), median (IQR)	0.8 (0.7, 1.0)	0.8 (0.7, 1.0)
Diabetes	1725 (24.4%)	971 (23.4%)
Forced vital capacity % predicted, median (IQR)	45.0 (35.0, 57.0)	45.0 (36.0, 58.0)
Functional status (none)	414 (5.9%)	209 (5.0%)
Oxygen need at rest (L/min), median (IQR)	4.0 (3.0, 6.0)	4.0 (2.5, 8.0)
pCO_2, median (IQR)	45.0 (40.0, 53.0)	45.0 (40.0, 53.0)
Pulmonary artery systolic pressure (mmHg), median (IQR)	38.0 (32.0, 48.0)	38.0 (31.0, 49.0)
Six-minute walk distance (feet), median (IQR)	774.0 (403.0, 1068.0)	800.0 (445.0, 1059.0)

Overall, the full waiting list population is comparable in the development and testing cohorts. The subset of transplanted individuals are also comparable across cohorts. However, the amount of pre- and post-transplant follow-up time and the number of deaths (outcome events) are noticeably smaller in the testing cohort compared to the development cohort. This discrepancy is due to the fact that the UNOS data to which we have access was administratively censored in 2018. Thus, any pre- or post-transplant deaths that occurred beyond this date are not counted. As more follow-up time accrues, the number of pre- and post-transplant deaths in the testing cohort should increase, which will lead to increased precision in the observed Kaplan-Meier survival estimates discussed below.

Table 2.3 displays the parameter estimates obtained for the modified (weighted) pre- and post-transplant outcome models.

Table 2.3. Parameter estimates obtained from the modified (weighted) pre- and post-transplant outcome models

Table 2.3A) Parameter estimates obtained for the modified pre-transplant outcome model

Covariate	Coefficient Estimate	95% Confidence Interval
Age at offer (years)	-0.0024318	-0.0113106, 0.0064469
Bilirubin (mg/dL)	0	(omitted)
Bilirubin increase ≥50%	0	(omitted)
Body mass index (BMI) (kg/m^3)	0.0104932	-0.0058752, 0.0268615
Cardiac index ($L/min/m^2$)	0.4723494	0.2710321, 0.6736667
Central venous pressure (CVP) (mmHg)	-0.0503087	-0.1253766, 0.0247593
Continuous mechanical ventilation	1.22271	0.6709733, 1.774447
Creatinine (serum) (mg/dL)	0.0864787	0.0193925, 0.1535648
Diabetes	0.3591762	0.137798, 0.5805544
Diagnosis Group A	REF	
Diagnosis Group B	0.665212	0.0004605, 1.329963
Diagnosis Group C	0.2546667	-0.2330827, 0.7424161
Diagnosis Group D	0.8609124	0.3824339, 1.339391
Forced vital capacity (FVC) % predicted	0.0682059	-0.0151386, 0.1515504
Functional status (none)	0.0379326	-0.32105, 0.3969151
Oxygen need at rest (L/min)	0.0520477	0.024774, 0.0793213
Oxygen-by-diagnosis interaction	0.0301022	-0.0152716, 0.0754759
pCO_2	0.1595511	0.0794078, 0.2396944
pCO_2 increase ≥15%	0	(omitted)
Pulmonary artery (PA) systolic pressure (mmHg)	0.1009163	0.0482246, 0.153608
PA-by-diagnosis interaction	0.0040516	-0.0967816, 0.1048847
Six-minute walk distance (feet)	-0.0130285	-0.0300434, 0.0039865

Table 2.3B) Parameter estimates obtained for the modified post-transplant outcome model

Covariate	Coefficient Estimate	95% Confidence Interval
Age at transplant (years)	0.0310272	0.0212575, 0.0407969
Cardiac index ($L/min/m^2$)	0.1035634	-0.0367848, 0.2439117
Continuous mechanical ventilation	0.5672613	0.1808594, 0.9536632
Creatinine (serum) (mg/dL)	0.2231819	0.0890585, 0.3573053
Creatinine increase ≥150%	0	(omitted)
Diagnosis Group A	REF	
Diagnosis Group B	0.3578133	0.0197239, 0.6959027
Diagnosis Group C	0.16691	-0.123965, 0.4577851
Diagnosis Group D	0.0524997	-0.1283522, 0.2333516
Functional status (none)	-0.2999765	-0.5940731, -.00588
Oxygen need at rest (L/min)	0.0166301	0.0014614, 0.0317989
Oxygen-by-diagnosis interaction	-0.0094877	-0.039502, 0.0205267
Six-minute walk distance (feet)	-0.0000709	-0.0002148, 0.0000729

These parameter estimates were used in conjunction with the baseline survival probabilities shown in Appendix 2.2 to construct modified LAS scores for patients in our testing cohort.

Table 2.4 displays the time-dependent AUC (Cattaneo et al., 2017; Heagerty et al., 2000) evaluated at one year post-waitlist registration and one year post-transplant for the modified and existing LAS models in the development and testing cohorts. In all cases, the AUC of the modified model is higher than that of the existing LAS, indicating that the modified model has better discrimination. However, the extent of improvement is larger in the pre-transplant population than in the post-transplant population.

Table 2.4. Time-dependent AUC for the modified and existing LAS

Time-dependent AUC estimated via nearest-neighbor smoothing at one year post-waitlist registration and one year post-transplant for the modified and existing LAS models. 95% confidence intervals were computed via the bootstrap percentile method; p-values were estimated using the normal approximation and standard deviation obtained from the bootstrap replicates. 1000 bootstrap replicates were used.

Cohort*		Modified LAS	Existing LAS	Difference	p-value of difference
Development	Pre-tx	0.732 (0.690, 0.760)	0.660 (0.619, 0.697)	0.071 (0.030, 0.106)	<0.001
	Post-tx	0.605 (0.580, 0.629)	0.560 (0.531, 0.585)	0.045 (0.026, 0.065)	<0.001
Testing	Pre-tx	0.750 (0.686, 0.792)	0.693 (0.631, 0.738)	0.057 (-0.004, 0.122)	0.083
	Post-tx	0.570 (0.536, 0.606)	0.540 (0.507, 0.576)	0.030 (0.003, 0.058)	0.030

* tx = transplant

Figure 2.2 depicts the time-dependent calibration of the modified pre- and post-transplant outcome models and existing pre- and post-transplant LAS models in the development cohort for the first two years post-listing and post-transplant. The predicted survival curves from the modified pre-transplant outcome model agree more closely with the observed survival curves for all three risk categories when compared with the existing LAS model. Conversely, predictions from the existing pre-transplant LAS model are noticeably different from the observed survival curves (Figure 2.2B).

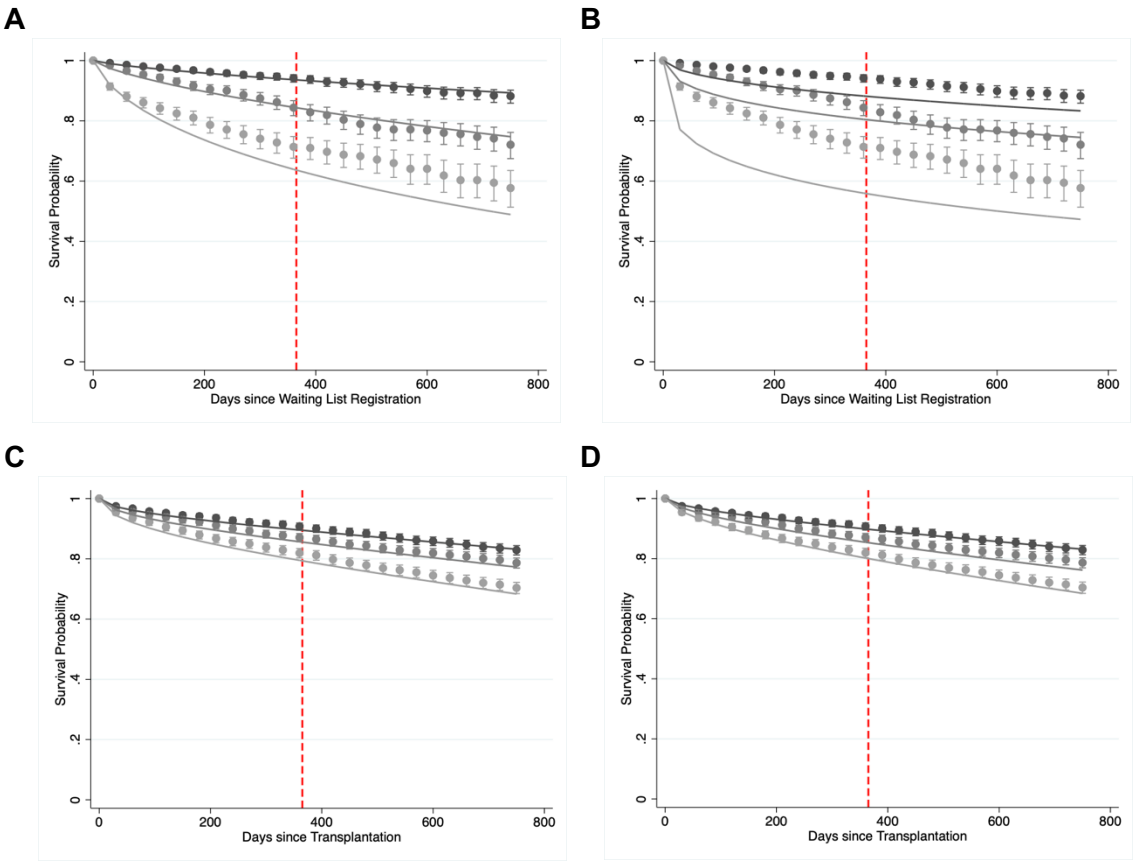


Figure 2.2. Observed vs. Predicted Survival in Development Cohort

Time-dependent calibration of A) the modified pre-transplant outcome model, B) the existing pre-transplant LAS model, C) the modified post-transplant outcome model, and D) the existing post-transplant LAS model, in the development cohort. Smooth, solid lines represent predicted survival probabilities; points with vertical error bars represent observed Kaplan-Meier estimates with their corresponding 95% confidence intervals. Estimates were plotted every 30 days to ease plot readability. Three risk groups are shown: low-risk/best survival (darkest lines), medium-risk/intermediate survival (medium-shaded lines), and high-risk/worst survival (lightest lines). A vertical, dashed, red line is placed at one year post-waitlist registration for reference.

This discrepancy is most prominent during the first year post-waitlist registration, but continues beyond this time point for all three risk groups (Figure 2.3). In contrast, predicted survival estimates from the modified post-transplant outcome model (Figure 2.2C) and the existing post-transplant LAS model (Figure 2.2D) closely match the observed survival curves. This observation is consistent with the AUC results in Table 2.4, and suggests that the extent of improvement in calibration is more noticeable for the pre-transplant model than the post-transplant model.

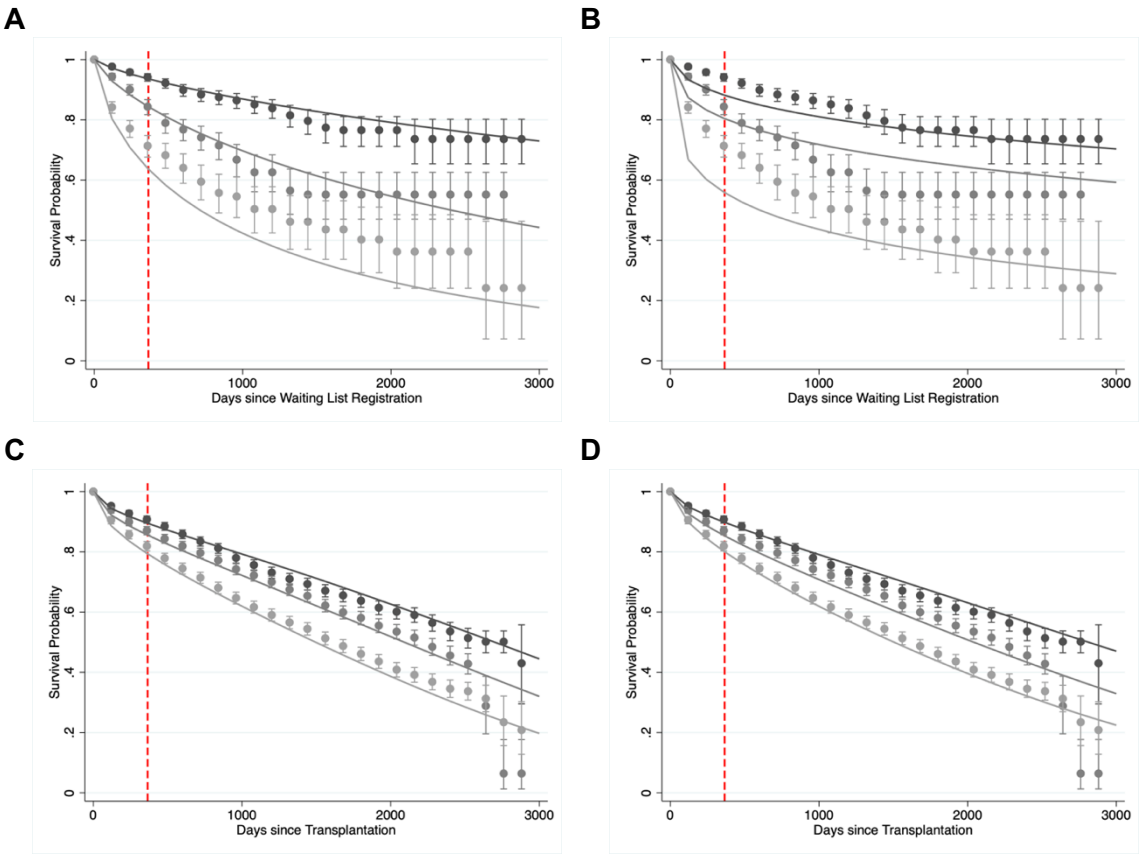


Figure 2.3. Long Term Observed vs. Predicted Survival in Development Cohort

Time-dependent calibration of A) the modified pre-transplant outcome model, B) the existing pre-transplant LAS model, C) the modified post-transplant outcome model, and D) the existing post-transplant LAS model, in the development cohort. Smooth, solid lines represent predicted survival probabilities; points with vertical error bars represent observed Kaplan-Meier estimates with their corresponding 95% confidence intervals. Estimates were plotted every 30 days to ease plot readability. Three risk groups are shown: low-risk/best survival (darkest lines), medium-risk/intermediate survival (medium-shaded lines), and high-risk/worst survival (lightest lines). A vertical, dashed, red line is placed at one year post-waitlist registration for reference.

Similar results were obtained in the testing cohort (Figure 2.4). In the testing cohort, the predicted survival estimates from the modified pre-transplant outcome model were consistent with the observed survival curves over time, regardless of risk group (Figure 2.4A). Predictions obtained from the existing pre-transplant LAS model, however, did not align as well with observed survival (Figure 2.4B). The modified and existing post-transplant models exhibit similar calibration in the testing cohort (Figures 2.4C and 2.4D, respectively). The calibration of both models is quite good in the first

year post-transplant but deteriorates considerably beyond one year for all three risk groups.

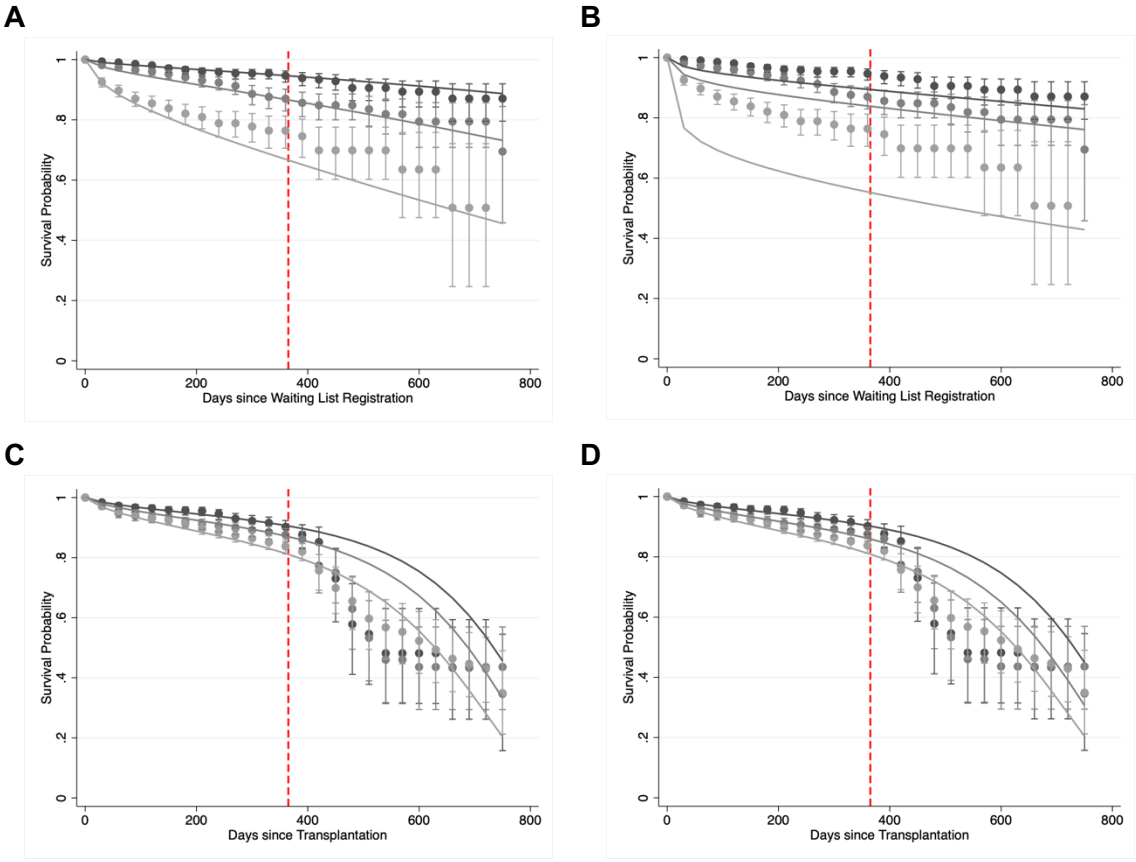


Figure 2.4. Observed vs. Predicted Survival in Testing Cohort

Time-dependent calibration of A) the modified pre-transplant outcome model, B) the existing pre-transplant LAS model, C) the modified post-transplant outcome model, and D) the existing post-transplant LAS model, in the testing cohort. Smooth, solid lines represent predicted survival probabilities; points with vertical error bars represent observed Kaplan-Meier estimates with their corresponding 95% confidence intervals. Estimates were plotted every 30 days to ease plot readability. Three risk groups are shown: low-risk/best survival (darkest lines), medium-risk/intermediate survival (medium-shaded lines), and high-risk/worst survival (lightest lines). A vertical, dashed, red line is placed at one year post-waitlist registration for reference.

Figure 2.5 depicts Bland-Altman plots of A) the modified versus existing LAS score, and B) the modified versus existing patient rank, for 712 different organ offer dates in the testing cohort. Patients at the extremes tend to receive similar scores under the two models, while patients with intermediate scores tend to experience more changes under the modified LAS. Specifically, a distinct cluster of patients appears more

than 2 standard deviations below the mean; for these patients, the modified LAS predicts a lower score than the existing LAS does. The Bland-Altman plot of changes in rank (Figure 2.5B) exhibits a somewhat different pattern than that in Figure 2.5A due to constraints of ranks (i.e., individuals who receive a more favorable rank are necessarily balanced by those who receive a less favorable one).

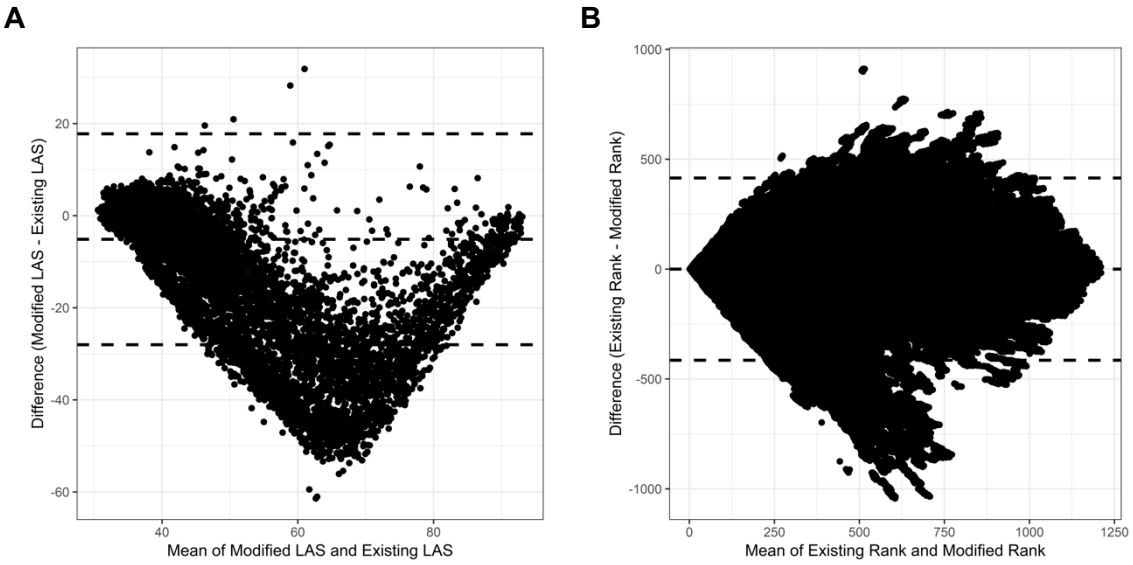


Figure 2.5. Bland-Altman plots

A) Difference between the modified and existing LAS scores versus the mean of the two scores, and B) difference between the modified and existing patient ranks versus the mean of the two ranks, from 712 different organ offer dates in the testing cohort. Horizontal, dashed lines are placed at ± 2 standard deviations of the mean.

Figure 2.6 illustrates how differences in estimated pre- and post-transplant survival under the modified and existing LAS models influence patients' LAS scores (Figure 2.6A) and prioritization (Figure 2.6B) for 712 organ offer dates in the testing cohort. Consistent with Figure 2.5A, the majority of patients in Figure 2.6A are shaded white or blue, indicating that these patients would receive the same or lower score under the modified model compared with the existing LAS. This lower score does not always translate into a lower (worse) priority, because the rank of a particular patient on a

particular organ offer date depends on the ranks of all other eligible patients on the waitlist at that date. Consequently, the distribution of shading in Figure 2.6B differs from that in Figure 2.6A. In both figures, however, differences in pre-transplant survival explain a greater proportion of the variability in the outcome (i.e., change in LAS score or change in priority) compared to differences in post-transplant survival.

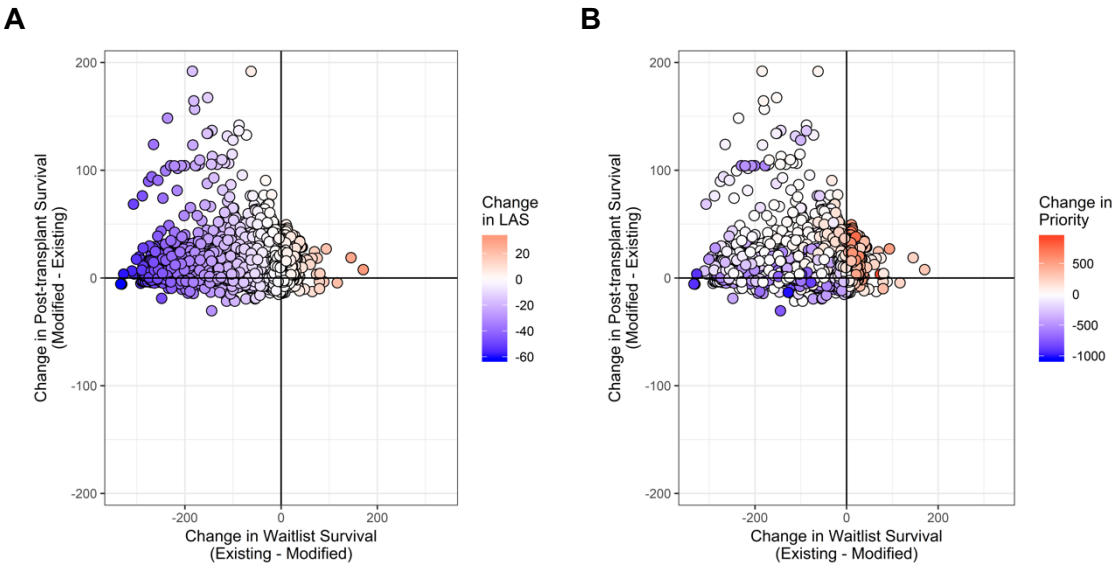


Figure 2.6. Scatter plots of the difference in post-transplant survival versus the difference in waitlist survival obtained under the modified and existing LAS in the testing cohort

Differences are computed for 712 organ offer dates in the testing cohort. Points represent patients, and are shaded based on the magnitude of change in A) score, with red representing increases in score and blue representing decreases in score; or B) rank, with red representing higher (better) priority and blue representing lower (worse) priority.

Discussion

We developed a weighted estimation strategy to account for selection bias in the pre- and post-transplant models used to calculate LAS scores in prioritizing patients for lung transplant. To our knowledge, we are the first to incorporate weighting into the fitting of post-transplant models to account for survivor bias and other forms of selection into the post-transplant population. We also improve upon weighted fitting of pre-transplant models by incorporating additional variables in our weight models to better

account for dependent censoring – most notably, geography. Since these variables are only included in the weight models (not the outcome models), they would only influence the performance of the outcome model if they are associated both with survival and with patients' selection for transplantation. The fact that discrimination and calibration improve under the modified pre- and post-transplant outcome models compared to the existing LAS models suggests that regional differences in patient selection may be important to consider when estimating pre- and post-transplant survival. That said, the extent of improvement is larger in the pre-transplant population (i.e., the full waiting list population) than in the post-transplant subset. There are two potential explanations for this observation: first, we have considerably more follow-up time in the pre-transplant population than in the post-transplant population; and second, selection bias appeared to have a larger impact on the estimate of waiting list urgency than on the estimate of post-transplant survival.

We postulated that the current LAS underestimates predicted transplant benefit due to survivor bias because it only predicts this quantity among people who were indeed selected to receive transplant, who tend to be older and sicker. Our results (e.g., Figures 2.6A and 2.6B) are consistent with this idea because they suggest that patients' estimated post-transplant survival under the modified LAS would be the same or greater than their estimated post-transplant survival under the existing LAS. Because the estimate of pre-transplant survival also tends to be longer under the modified LAS compared to the existing LAS, a sizable number of patients with intermediate scores under the current LAS would receive lower scores (lower priority) under the modified LAS.

Our study has several strengths. We demonstrate how inverse probability weighting can account for survivor bias in real-world lung transplant data. These weights

account for the probability of receiving transplant as well as the probability of censoring. By including additional variables in our weighting, such as geography and prior (i.e., lagged) clinical covariates, we account for regional variability in pre- and post-transplant survival and are able to capture patients' waitlist history and clinical trajectory more fully, while maintaining consistency between our modified outcome model and the existing LAS. This approach ensures a fair comparison between the modified and existing models in both the development and testing cohorts. Although our primary endpoints were one-year pre- and post-transplant survival, the results using greater extent of follow-up time demonstrate how these models perform over longer time frames. Such an evaluation can help inform future revisions of the LAS, and is especially relevant as the lung transplant community considers what role longer-term survival should play in lung allocation (Kasiske et al., 2020; Maxwell et al., 2014; U.S. Department of Health & Human Services). Finally, while our analysis focuses on lung allocation specifically, our approach can be applied to any organ allocation system that relies on estimates of post-transplant survival to prioritize patients (e.g., the United States' EPTS score for kidney allocation (Egan et al., 2006; Organ Procurement and Transplantation Network, 2020; Veatch & Ross, 2015), Germany's Lung Allocation Score (Gottlieb, 2017; Gottlieb et al., 2014), and the United Kingdom's Liver Transplant Benefit Score (NHS Blood and Transplant)).

Our study is not without limitations. First, insufficient information is available to distinguish between patients who were newly listed and those who were re-activated after temporary waitlist removal. Thus, the first record associated with each identification number was taken to be the initial registration date, follow-up time was counted from that date forward, and individuals who were subsequently lost to follow-up were censored at that time. Second, we exclude individuals on the waitlist who are highly unlikely to

receive transplant, due to certain patient characteristics that prevent them from finding a suitable donor organ match (e.g., high sensitization or small stature). Although this analytic decision ensures that the remaining individuals on the waitlist have at least some probability of receiving transplant, it also implies that we can only generalize our findings to individuals on the waitlist who do not have these clinical contraindications. Third, we cannot account for ascertainment bias/informed presence bias (i.e., we cannot account for the fact that presence in the UNOS database is not random, but rather indicates that the patient was ill enough to visit the hospital, undergo evaluation for transplant, and be registered on the waitlist). Fourth, transplant organ allocation is a highly selective process, and selection bias can occur at various stages throughout this process (e.g., decision to register a patient on the waitlist, decision to remove a waitlisted patient, decision to transplant). In this particular paper, we restrict our focus to selection bias that arises due to the fact that some individuals die or are otherwise censored prior to receiving transplant, and present a quantitative approach to mitigating this bias in the LAS. Although beyond the scope of this study, additional research – including qualitative work – is necessary to understand how to balance all the factors that go into making transplant decisions.

Our approach to addressing selection bias is intuitive and straightforward to implement, and demonstrates how principles from causal inference can be incorporated into existing prediction model frameworks to improve organ allocation. Additionally, it can be applied to any organ allocation system that relies on estimates of pre- and post-transplant survival to prioritize patients, including those used for different organs and in other countries. We anticipate that this work can inform future revisions of the LAS and other prediction models in organ transplantation to improve prediction and ensure fair and equitable organ allocation.

CHAPTER 3. CLINICAL IMPACT OF A MODIFIED LUNG ALLOCATION SCORE THAT MITIGATES SELECTION BIAS

This chapter is under review. The suggested citation is:

Schnellinger EM, Cantu E, Schaubel DE, Kimmel SE, Stephens-Shields AJ. Clinical Impact of a Modified Lung Allocation Score that Mitigates Selection Bias (under review).

Abstract

Background. The Lung Allocation Score (LAS) is used in the U.S. to prioritize lung transplant candidates. Selection bias, induced by dependent censoring of waitlisted candidates and prediction of post-transplant survival among surviving, transplanted patients only, is only partially addressed by the LAS. Recently, a modified LAS (mLAS) was designed to mitigate such selection biases. Here, we estimate the clinical impact of replacing the LAS with the mLAS.

Methods. We considered lung transplant candidates waitlisted during 2016-2017. Modified and existing LAS scores were computed for each registrant at each observed organ offer date; individuals were ranked accordingly. Patient characteristics associated with better priority under the mLAS were investigated via logistic regression and generalized linear mixed models. We also determined whether differences in rank were explained more by changes in predicted pre- or post-transplant survival. Simulations examined how one-year waitlist, post-transplant, and overall survival might change under the mLAS.

Results. Diagnosis group, six-minute walk distance, continuous mechanical ventilation, functional status, and age demonstrated the highest impact on differential allocation. Changes in predicted pre-transplant survival explained a greater proportion of variability in differences in rank than changes in predicted post-transplant survival, suggesting that selection bias has a larger impact on the estimate of waitlist urgency than on the estimate of post-transplant survival. Simulations suggest that for every 1000 waitlisted individuals, 12.8 (interquartile range, IQR: 5.2-24.3) fewer waitlist deaths per year would occur under the mLAS, without compromising post-transplant and overall survival.

Conclusions. Implementing a modified LAS that mitigates selection bias into clinical practice can lead to important differences in allocation and possibly modest improvement in waitlist survival.

Background

The U.S. lung allocation system prioritizes lung transplant candidates based on the Lung Allocation Score (LAS) (Organ Procurement and Transplantation Network, 2020). The LAS aims to predict how long a patient would survive with versus without a transplant, and is composed of two prediction models: one to predict pre-transplant survival and one to predict post-transplant survival (Egan et al., 2006). Patients' transplant benefit is estimated as the difference between predicted post- and pre-transplant survival, while their waitlist urgency is taken to be predicted pre-transplant survival (Egan et al., 2006; United Network for Organ Sharing, 2015). The LAS is computed as the difference between transplant benefit and waitlist urgency (equivalently, post-transplant survival minus two times pre-transplant survival), and is normalized so that it ranges between 0 and 100, with higher scores indicating greater priority for transplant (United Network for Organ Sharing, 2015).

Although LAS scores are calculated using each patient's most recent demographic and clinical variables, the models used to predict pre- and post-transplant survival underlying the LAS do not fully account for selection bias. Such bias arises due to dependent censoring (individuals are removed from the waitlist once they receive transplant) and survivor bias (individuals must survive long enough for a suitable donor organ to become available) (Schnellinger et al., 2021). In Chapter 2, we developed a modified LAS using inverse probability of treatment weighting (IPTW) and inverse probability of censoring weighting (IPCW) in order to mitigate selection bias. We then demonstrated improved discrimination and calibration for the modified LAS compared to the existing LAS. We also showed that the modified LAS would affect patient rankings considerably.

In this chapter, we evaluate the impact of implementing this modified LAS in clinical practice. More specifically, we use observed data to investigate the demographic

and clinical characteristics of individuals who would have received better priority under the modified LAS compared to the existing one. We also conduct a simulation study to estimate how waitlist, post-transplant, and overall survival might change under the modified LAS. Understanding the clinical impact of the modified LAS is especially relevant now, as the transplant community is currently developing and implementing a new organ allocation framework, the continuous distribution model, which aims to increase the flexibility of the scoring algorithm so that organs are allocated to patients more equitably. Note that the sources of bias we outline above would still be in play under this continuous distribution model unless they are recognized and addressed (Kasiske et al., 2020; Snyder et al., 2018; U.S. Department of Health & Human Services).

Methods

This study utilizes pre- and post-lung transplant data from the United Network for Organ Sharing (UNOS). Our cohort consisted of all patients 18 years or older who were listed for single or bi-lateral lung transplantation in the United States between January 1, 2016, and December 31, 2017. This cohort is consistent with the testing cohort used in Chapter 2 and our earlier paper (Schnellinger et al., 2021).

We applied the modified and existing LAS models to our cohort to estimate both a modified LAS score and an existing LAS score for each patient at each possible offer date, as in Chapter 2 and our previous paper (Schnellinger et al., 2021). At each offer date, eligible patients were ranked twice: first using their modified LAS scores, and second using their existing LAS scores. Patients were then grouped into two categories (better versus same or worse priority) based on how their rank would change under the modified LAS compared to the existing LAS. We were unable to analyze the better priority, same priority, and worse priority categories separately because the sample size

among individuals who received the same rank under the two LAS models was too small. Moreover, dropping individuals who received the same rank from the analyses entirely would alter the interpretation of the resulting probabilities (i.e., they would be interpreted as conditional rather than marginal probabilities). Thus, we combined individuals who received the same rank together with individuals who received worse priority. Although we could have instead grouped individuals who received the same rank with those who received better priority, we felt that keeping the better priority individuals separate would be more meaningful to patients and clinicians, as this category then represents individuals whose chances of receiving transplant would increase under the modified LAS.

To examine the demographic and clinical characteristics of patients who receive better priority under the modified LAS relative to the existing LAS, univariable and multivariable logistic regression models were fit. The outcome variable for these models was categorical change in rank (better priority versus no change or worse priority), and predictors included patients' diagnosis group, age, gender, race/ethnicity, primary payment source, education level, employment status, UNOS geographic region, body mass index (BMI), blood type, HLA mismatch, prior transplant, prior cardiac surgery, smoking status, diabetes, functional status, cardiac index, mechanical ventilation, six-minute walk distance, creatinine, and oxygen need at rest. Covariates were only included in the multivariable model if they were statistically significant at the $\alpha = 0.05$ level in the univariable models. Based on the multivariable model, bar charts of the predicted probability of receiving better priority under the modified LAS compared to the existing LAS were constructed for each covariate individually.

To evaluate the extent to which changes in prioritization are driven by changes in predicted pre- or post-transplant survival between the modified and existing LAS, we subset the data into two groups: one with only worse priority individuals, and another

with only better priority individuals. Generalized linear mixed models (GLMM) were fit to each subset using the continuous differences in ranks for each patient under the modified and existing LAS, and the difference in predicted pre-transplant survival and difference in predicted post-transplant survival as predictors. Generalized r^2 values were obtained for each GLMM model via the *r2glmm* package (Jaeger, 2017). We also conducted these analyses stratified by the demographic and clinical characteristics that were deemed to have a significant impact on prioritization changes based on the logistic regression models.

Finally, we undertook a statistical simulation – adapted from (Vock et al., 2013) – to investigate how the modified LAS would impact observed waitlist and post-transplant survival if it were implemented in clinical practice. Specifically, we:

1. Selected a random set of 364 offer dates from the observed offer dates in the testing cohort.
2. Assigned the number of organ offers per day by drawing from a Poisson distribution with a mean of 6, which equals the observed mean number of offers per day in the testing cohort.
3. Simulate characteristics of the hypothetical donor organs – e.g., donor region and donor blood type – by randomly sampling from the observed distribution of donor organ characteristics in the testing cohort.
4. Simulate waitlist survival (i.e., transplant-free survival) for all individuals by drawing from an exponential distribution with rate parameter depending on their predicted probability of waitlist death in the next month associated with their baseline (i.e., waitlist registration) record. Further details on how we obtained these predicted probabilities appear below.
5. At each offer date, compute a modified LAS score and an existing LAS score for each patient alive and eligible to receive transplant on that date. If an

individual's simulated death date exceeds their last observed visit date, linearly interpolate their LAS such that it increases from their final observed value to halfway between that value and 100 (as per (Vock et al., 2013)). This approach simulates how patients tend to become sicker the longer they remain on the waitlist.

6. At each offer date, conduct two separate allocations, one based on patients' existing LAS scores, and a second based on their modified LAS scores. In each allocation, we subset the set of eligible patients to include only those who match the characteristics of the hypothetical donor organ. Then, rank the patients in this subset by their modified and existing LAS scores, and allocate the hypothetical organ to the highest ranking person under each scoring system. Offers were first extended to recipients in the same region and same blood type as the donor lung. If no one met those criteria, we successively relaxed the matching criteria (same region, compatible blood type; nearby region, same blood type; nearby region, compatible blood type; any region, same blood type; any region, compatible blood type) until a match could be made.
7. If a patient was selected to receive the hypothetical organ offer in step (6), simulate their post-transplant survival by drawing from an exponential distribution with rate parameter depending on their predicted probability of one-year post-transplant death associated with the time at which they receive the hypothetical transplant. Further details on how we obtained these predicted probabilities appear below.
8. Repeat steps (6) and (7) until all hypothetical donor organs are allocated.
9. Generate patients' final pre- or post-transplant outcome status, as appropriate. Specifically:

a. if patients do not receive a hypothetical transplant in the simulation, generate their final waitlist outcome status by drawing from a Bernoulli distribution with a probability equal to their predicted probability of waitlist death in the next month given that they survived through the time point associated with their final waitlist record; set their post-transplant outcome status to missing.
b. if patients receive a hypothetical transplant in the simulation, censor their final waitlist outcome at the time they receive the hypothetical transplant; generate their final post-transplant outcome status by drawing from a Bernoulli distribution with a probability equal to their predicted probability of one-year post-transplant death that corresponds with the time they received the hypothetical transplant.

10. At the end of each iteration, compute Kaplan-Meier estimates of one-year waitlist, one-year post-transplant, and one-year overall survival among the entire simulated population. Also compute the proportion of patients who receive hypothetical transplant under both the existing LAS and the modified LAS (concordant set).
11. Among individuals who receive hypothetical transplant under one LAS but not the other (discordant set), compute the same Kaplan-Meier estimates as in (10). Also calculate means and proportions (as appropriate) of time-invariant demographic and clinical characteristics.

The predicted probabilities used to generate transplant-free survival in step (4) were obtained by fitting a discrete-time pooled logistic regression model with waitlist mortality as the outcome and the covariates of the existing waitlist LAS model (i.e., age, bilirubin, body mass index, cardiac index, central venous pressure, continuous mechanical

ventilation, serum creatinine, diabetes, diagnosis group, forced vital capacity, functional status, oxygen need at rest, partial pressure of carbon dioxide, pulmonary artery systolic pressure, six-minute walk distance), lagged versions of these covariates, time since waitlist registration, geography, race/ethnicity, gender, blood type, and height as predictors. This model was fit among individuals who did not receive transplant in the observed data (i.e., whose waitlist outcome was observed) and applied to all individuals to obtain (counterfactual) predicted probabilities of pre-transplant survival. These predicted probabilities were then used to generate hypothetical transplant-free survival and the corresponding counterfactual waitlist outcomes used in the simulation. Similarly, the predicted probabilities used to generate post-transplant survival in step (7) were obtained by fitting a discrete-time pooled logistic regression model with post-transplant mortality as the outcome and the same covariates listed above as predictors. This model was fit among individuals who received transplant in the observed data (i.e., whose post-transplant outcome was observed) and applied to all individuals to obtain (counterfactual) predicted probabilities of post-transplant survival. These predicted probabilities were then used to generate hypothetical post-transplant survival and the corresponding counterfactual post-transplant outcomes used in the simulation. This approach allows individuals' predicted probabilities of pre- and post-transplant survival to change over time based on their observed covariate data.

The entire simulation (i.e., steps 1-11, above) was repeated 100 times (due to computational constraints) to obtain distributions of the Kaplan-Meier estimates and demographic/clinical characteristics. We also conducted sensitivity analyses in which we varied the transplant offer rate (e.g., using Poisson means of 3 or 9 offers per day) to explore how lower or higher offer rates might impact results. Analyses were conducted in R (R Foundation for Statistical Computing, Vienna, Austria) and Stata (StataCorp LLC, College Station, TX).

Results

Observed Analyses

Table 3.1 summarizes the demographic and clinical characteristics of the complete waiting list population as well as the subset of individuals who received transplant. In the full waitlist population, covariates were measured at the time of waitlist registration; among the subset of transplanted individuals, covariates are shown both at the time of patients' waitlist registration and at the time they received transplant. In the full waitlist population, the median waiting time was 57 days (interquartile range, IQR: 16-154 days), 5.7% of these patients died on the waitlist during the two-year follow-up time, 0.5% died within one year of waitlist registration, and 77.6% received transplant. Among those who received transplant, the median waiting time was 39 days (IQR: 13-102), the median follow-up time after transplant was 199 days (IQR: 68-365), 10.9% died after transplantation during the two-year follow-up time, and 8.3% died within one year of transplantation. The distributions of some clinical characteristics differed between the transplanted group and the full waitlist population. For example, the proportion of patients in diagnosis group D was higher among transplanted individuals (60.1%) compared to the full waitlist population (57.4%). Conversely, the proportion of patients in diagnosis group A was lower among transplanted individuals (25.0%) compared to the full waitlist population (27.0%). At the time of waitlist registration, the distributions of continuous mechanical ventilation, functional status, and six-minute walk distance were comparable among the full waitlist population and among the subset of individuals who would eventually go on to receive transplant. Conversely, at the time of transplant, a larger proportion of individuals in the post-transplant group were on continuous mechanical ventilation (7.7%) compared to the full waitlist population (4.1%); a smaller proportion of individuals in the post-transplant group required no assistance with daily

living tasks (5.0%) compared to the full waitlist population (11.3%); and the median (IQR) six-minute walk distance was somewhat lower among transplanted individuals (800.0 [445.0, 1059.0] feet) compared to the full waitlist population (880.5 [580.0, 1142.0] feet).

Table 3.1. Demographic and clinical characteristics of the complete waiting list population and the subset of this population who received transplant

In the full waitlist population, covariates were measured at waitlist registration; among the subset of transplanted individuals, we display covariates measured at waitlist registration and at transplant.

	Waiting List	Post-transplant (at waitlist registration)	Post-transplant (at transplant)
Total number of patients (N)	5354	4154	4154
Waiting time (days), median (IQR)	57.0 (16.0, 154.0)	39.0 (13.0, 102.0)	---
Death on waitlist within 1 year of waitlist registration	25 (0.5%)	---	---
All deaths on waitlist	288 (5.7%)	---	---
Removal from waitlist due to transplant	4154 (77.6%)	4154 (100.0%)	---
Follow-up time post-transplant (days), median (IQR)	---	---	199.0 (68.0, 365.0)
Death within 1 year of transplant	---	---	346 (8.3%)
All deaths post-transplant	---	---	420 (10.9%)
REGION			
1	185 (3.5%)	160 (3.9%)	160 (3.9%)
2	953 (17.8%)	680 (16.4%)	680 (16.4%)
3	522 (9.7%)	405 (9.7%)	405 (9.7%)
4	636 (11.9%)	534 (12.9%)	534 (12.9%)
5	878 (16.4%)	675 (16.2%)	675 (16.2%)
6	110 (2.1%)	67 (1.6%)	67 (1.6%)
7	438 (8.2%)	334 (8.0%)	334 (8.0%)
8	312 (5.8%)	251 (6.0%)	251 (6.0%)
9	180 (3.4%)	126 (3.0%)	126 (3.0%)
10	684 (12.8%)	528 (12.7%)	528 (12.7%)
11	456 (8.5%)	394 (9.5%)	394 (9.5%)
GENDER			
F	2263 (42.3%)	1594 (38.4%)	1594 (38.4%)
M	3091 (57.7%)	2560 (61.6%)	2560 (61.6%)
RACE/ETHNICITY			
White	4224 (78.9%)	3339 (80.4%)	3339 (80.4%)
Black	552 (10.3%)	397 (9.6%)	397 (9.6%)
Hispanic	411 (7.7%)	299 (7.2%)	299 (7.2%)
Asian	124 (2.3%)	93 (2.2%)	93 (2.2%)
Other	43 (0.8%)	26 (0.6%)	26 (0.6%)
DIAGNOSIS			
A (Obstructive Disease)	1444 (27.0%)	1038 (25.0%)	1038 (25.0%)
B (Pulmonary Hypertension)	299 (5.6%)	182 (4.4%)	182 (4.4%)
C (Cystic Fibrosis)	538 (10.0%)	436 (10.5%)	436 (10.5%)
D (Pulmonary Fibrosis)	3073 (57.4%)	2498 (60.1%)	2498 (60.1%)
BLOOD TYPE			
A	2100 (39.2%)	1702 (41.0%)	1702 (41.0%)
AB	195 (3.6%)	157 (3.8%)	157 (3.8%)
B	607 (11.3%)	474 (11.4%)	474 (11.4%)
O	2452 (45.8%)	1821 (43.8%)	1821 (43.8%)
Age (years), median (IQR)	60.0 (52.0, 66.0)	61.0 (52.0, 66.0)	61.0 (52.0, 66.0)
Bilirubin (mg/dL), median (IQR)	0.7 (0.7, 0.7)	0.7 (0.7, 0.7)	0.7 (0.7, 0.7)
BMI (kg/m^3), median (IQR)	26.0 (22.0, 29.2)	26.0 (22.1, 29.1)	25.9 (22.1, 29.0)

Height (feet), median (IQR)	5.6 (5.3, 5.8)	5.6 (5.3, 5.8)	5.6 (5.3, 5.8)
Cardiac index <2 L/min/m^2	896 (16.7%)	717 (17.3%)	596 (14.3%)
Central venous pressure (mmHg), median (IQR)	5.0 (5.0, 8.0)	5.0 (5.0, 8.0)	5.0 (5.0, 8.0)
Continuous mechanical ventilation	217 (4.1%)	160 (3.9%)	320 (7.7%)
Creatinine (serum) (mg/dL), median (IQR)	0.8 (0.7, 1.0)	0.8 (0.7, 1.0)	0.8 (0.7, 1.0)
Diabetes	1074 (20.1%)	895 (21.5%)	971 (23.4%)
Forced vital capacity % predicted, median (IQR)	48.0 (38.0, 62.0)	48.0 (37.0, 61.0)	45.0 (36.0, 58.0)
Functional status (none)	604 (11.3%)	472 (11.4%)	209 (5.0%)
Oxygen need at rest (L/min), median (IQR)	4.0 (2.0, 6.0)	4.0 (2.5, 6.0)	4.0 (2.5, 8.0)
pCO_2, median (IQR)	43.0 (40.0, 50.0)	43.0 (40.0, 50.0)	45.0 (40.0, 53.0)
Pulmonary artery systolic pressure (mmHg), median (IQR)	38.0 (31.0, 48.0)	38.0 (30.0, 47.0)	38.0 (31.0, 49.0)
Six-minute walk distance (feet), median (IQR)	880.5 (580.0, 1142.0)	880.0 (562.0, 1146.0)	800.0 (445.0, 1059.0)

Table 3.2 (Figure 3.1) displays the odds ratio estimates (predicted probabilities) for better versus worse priority under the modified LAS by demographic and clinical characteristics at each possible offer date. Odds ratios greater than 1 for a given category indicate that a person in that category is more likely to receive better priority under the modified LAS relative to a person in the reference category.

Table 3.2. Estimated odds ratios for better versus worse priority under the modified LAS by demographic and clinical characteristics at each possible offer date

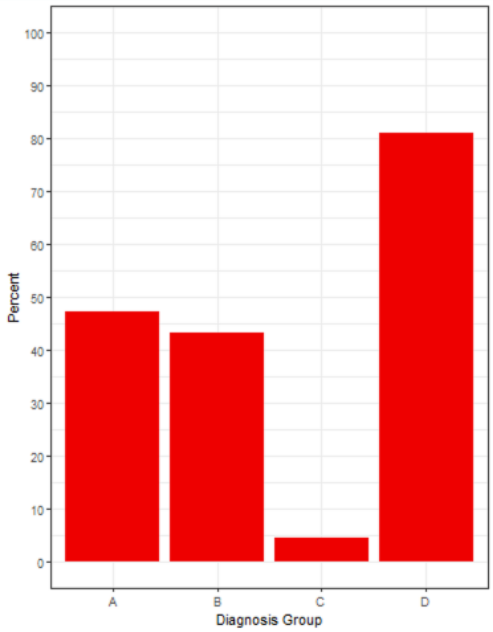
Covariate	Univariable (Unadjusted) Model			Multivariable (Adjusted) Model		
	Odds Ratio	95% Confidence Interval		Odds Ratio	95% Confidence Interval	
Diagnosis A	REF			REF		
Diagnosis B	1.17*	(1.15,	1.20)	0.78*	(0.75,	0.81)
Diagnosis C	0.14*	(0.13,	0.14)	0.01*	(0.01,	0.01)
Diagnosis D	7.24*	(7.17,	7.32)	13.6*	(13.3,	13.8)
Age (years) [18,50)	REF	N/A	N/A	REF	N/A	N/A
Age (years) [50,59)	2.16*	(2.13,	2.19)	0.59*	(0.58,	0.61)
Age (years) [59,65)	1.83*	(1.81,	1.85)	0.38*	(0.37,	0.38)
Age (years) [65,80]	1.39*	(1.38,	1.41)	0.15*	(0.14,	0.15)
Male	REF	N/A	N/A	REF	N/A	N/A
Female	0.81*	(0.81,	0.82)	0.83*	(0.82,	0.84)
BMI Underweight	0.05*	(0.05,	0.05)	0.01*	(0.01,	0.01)
BMI Normal	REF	N/A	N/A	REF	N/A	N/A
BMI Overweight	2.60*	(2.58,	2.63)	1.98*	(1.94,	2.01)
BMI Obese	3.71*	(3.64,	3.77)	2.07*	(2.02,	2.12)
Height (feet) [4.33,5.25)	REF	N/A	N/A			
Height (feet) [5.25,5.50)	1.17*	(1.16,	1.19)	Excluded due to collinearity with BMI		
Height (feet) [5.50,5.75)	1.27*	(1.26,	1.29)			
Height (feet) [5.75,6.92]	1.37*	(1.36,	1.38)			
Blood Type A	REF	N/A	N/A	REF	N/A	N/A
Blood Type AB	1.31*	(1.28,	1.33)	2.02*	(1.94,	2.11)
Blood Type B	1.07*	(1.05,	1.08)	0.98	(0.96,	1.00)
Blood Type O	1.11*	(1.10,	1.11)	1.07*	(1.06,	1.09)
No Previous Transplant	REF	N/A	N/A	REF	N/A	N/A
Previous Transplant	1.46*	(1.40,	1.53)	0.64*	(0.61,	0.67)

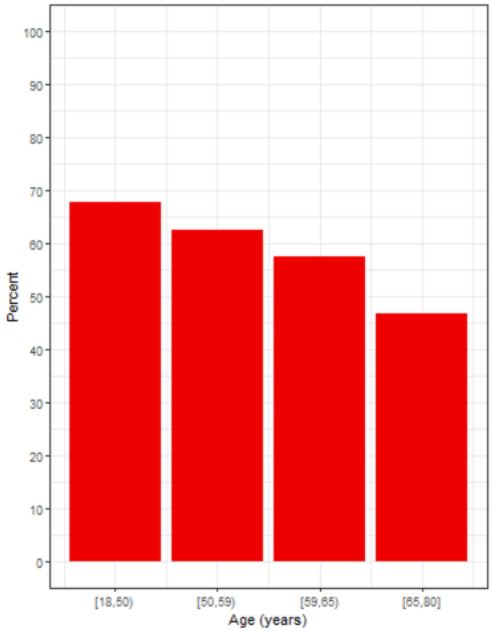
UNOS Region 1	REF	N/A	N/A	REF	N/A	N/A
UNOS Region 2	0.34*	(0.32,	0.35)	0.49*	(0.46,	0.51)
UNOS Region 3	0.45*	(0.44,	0.47)	0.56*	(0.53,	0.58)
UNOS Region 4	0.41*	(0.39,	0.42)	0.52*	(0.49,	0.55)
UNOS Region 5	0.43*	(0.41,	0.44)	0.51*	(0.48,	0.54)
UNOS Region 6	0.42*	(0.40,	0.44)	0.38*	(0.35,	0.42)
UNOS Region 7	0.35*	(0.34,	0.36)	0.88*	(0.84,	0.93)
UNOS Region 8	0.21*	(0.20,	0.22)	0.40*	(0.37,	0.42)
UNOS Region 9	0.38*	(0.37,	0.40)	0.39*	(0.36,	0.42)
UNOS Region 10	0.53*	(0.51,	0.55)	0.74*	(0.70,	0.79)
UNOS Region 11	0.40*	(0.38,	0.41)	0.72*	(0.68,	0.77)
No Cigarette Use	REF	N/A	N/A	REF	N/A	N/A
Cigarette Use	1.00	(0.99,	1.01)	0.91*	(0.90,	0.93)
HLA Mismatch (55% missing)						
0	REF	N/A	N/A			
1	0.24*	(0.16,	0.34)			
2	0.40*	(0.28,	0.55)	Excluded due to missingness		
3	0.45*	(0.32,	0.63)			
4	0.44*	(0.32,	0.62)			
5	0.54*	(0.39,	0.76)			
6	0.46*	(0.33,	0.64)			
Primary Payment Source/Insurance						
Private Insurance	REF	N/A	N/A			
Medicaid	0.69*	(0.68,	0.66)			
Medicare Fee for Service	0.88*	(0.86,	1.01)	Excluded due to ambiguity in how to combine groups to ensure sufficient sample sizes for stable parameter estimates		
Medicare & Choice	0.67*	(0.66,	0.79)			
Children's Health Insurance Program	0.01*	(0.00,	0.03)			
Department of VA	0.58*	(0.56,	0.60)			
Other government insurance	0.97*	(0.94,	1.00)			
Self	3.30*	(2.81,	3.87)			
Donation	drop (perfect prediction)					
Free Care	0.66*	(0.50,	0.87)			
Pending	19.9*	(10.8,	36.8)			
Foreign Government	0.84*	(0.80,	0.88)			
Education Level (3% missing)						
None	REF	N/A	N/A	REF	N/A	N/A
Grade School	1.01	(0.65,	1.59)	0.69*	(0.48,	0.99)
High School or GED	0.96	(0.61,	1.51)	0.43*	(0.30,	0.61)
Some College	1.09	(0.69,	1.71)	0.49*	(0.35,	0.71)
Associate or Bachelor's Degree	1.14	(0.73,	1.79)	0.43*	(0.30,	0.61)
Post-college or Graduate Degree	1.67*	(1.07,	2.63)	0.55*	(0.39,	0.78)
Employment Status (53% missing)						
Not Employed	REF	N/A	N/A	Excluded due to missingness and because variable is not clinically important		
Employed	1.47*	(1.44,	1.50)			
Prior Cardiac Surgery (2% missing)						
No	REF	N/A	N/A	REF	N/A	N/A
Yes	1.54*	(1.51,	1.57)	1.26*	(1.20,	1.32)
Prior lung surgery (52% missing)						
No	REF	N/A	N/A	Excluded due to missingness; use prior cardiac surgery instead		
Yes	1.86*	(1.81,	1.92)			
Diabetes						
No	REF	N/A	N/A	REF	N/A	N/A
Yes	0.85*	(0.84,	0.86)	2.00*	(1.97,	2.04)
Functional Status						
At least some assistance needed	REF	N/A	N/A	REF	N/A	N/A
No assistance needed	3.44*	(3.37,	3.52)	21.3*	(20.0,	22.7)
Cardiac index						
≥2 L/min/m^2	REF	N/A	N/A	REF	N/A	N/A
<2 L/min/m^2	1.50*	(1.49,	1.52)	3.31*	(3.24,	3.38)
Mechanical ventilation						
Not continuous	REF	N/A	N/A	REF	N/A	N/A
Continuous	1.61*	(1.55,	1.67)	18.7*	(16.5,	21.1)
Six-minute walk distance						
Q1: [0, 656) feet	REF	N/A	N/A	REF	N/A	N/A
Q2: [656, 928) feet	2.12*	(2.10,	2.15)	5.10*	(5.01,	5.18)
Q3: [928, 1160) feet	5.98*	(5.91,	6.05)	28.6*	(27.6,	29.6)
Q4: [1160, 4000) feet	3.00*	(2.97,	3.04)	7.44*	(7.28,	7.61)

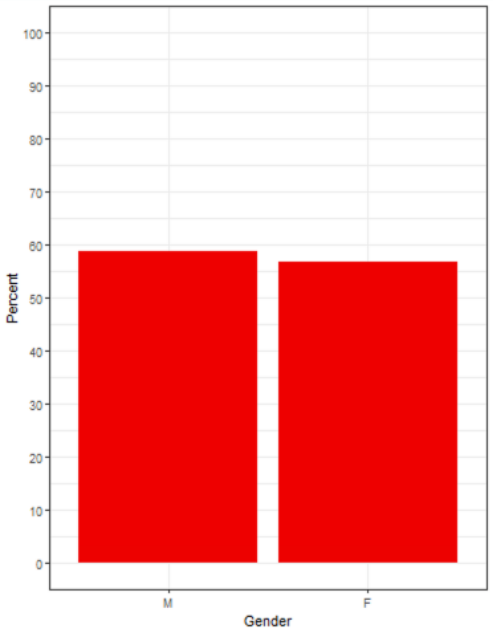
Creatinine (serum)						
Q1: [0.20,0.67) mg/dL	REF	N/A	N/A	REF	N/A	N/A
Q2: [0.67,0.80) mg/dL	1.32*	(1.31,	1.34)	0.47*	(0.46,	0.48)
Q3: [0.80,0.93) mg/dL	1.58*	(1.56,	1.60)	0.40*	(0.39,	0.41)
Q4: [0.93,6.83] mg/dL	1.29*	(1.27,	1.31)	0.19*	(0.19,	0.20)
Oxygen need at rest						
Q1: [0, 2) L/min	REF	N/A	N/A	REF	N/A	N/A
Q2: [2, 3) L/min	0.88*	(0.86,	0.89)	0.98*	(0.96,	1.00)
Q3: [3, 4) L/min	0.91*	(0.90,	0.91)	1.15*	(1.14,	1.17)
Q4: [4,35] L/min	1.19*	(1.18,	1.21)	1.17*	(1.15,	1.20)
Race/Ethnic: White	REF	N/A	N/A	REF	N/A	N/A
Race/Ethnic: Black	1.84*	(1.82,	1.87)	1.34*	(1.30,	1.37)
Race/Ethnic: Hispanic	2.73*	(2.67,	2.79)	1.17*	(0.13,	1.20)
Race/Ethnic: Asian	1.67*	(1.62,	1.72)	1.25*	(1.19,	1.32)
Race/Ethnic: Other	1.05*	(1.02,	1.08)	0.66*	(0.60,	0.71)
Baseline Odds (exp(Intercept))	N/A (varies by model)			4.87*	(1.74,	13.6)

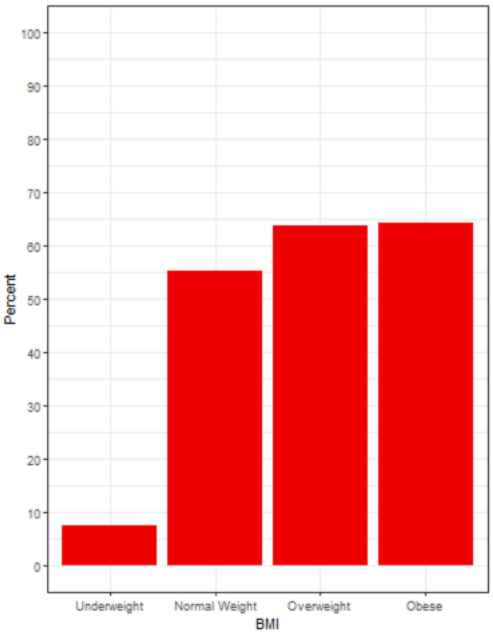
* Statistically significant at the $\alpha = 0.05$ level

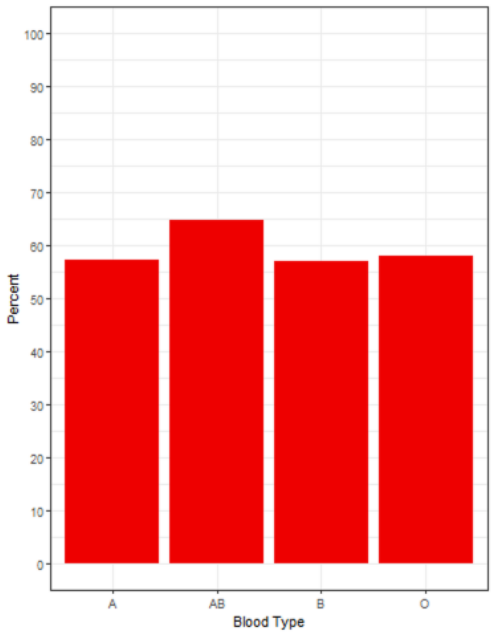
Figure 3.1 shows that diagnosis group, six-minute walk distance, continuous mechanical ventilation, functional status, and age exhibit the largest impact on prioritization changes. Individuals in diagnosis group D (pulmonary fibrosis) are 13.6 (95% confidence interval, CI: 13.3-13.8) times more likely to receive better priority under the modified LAS relative to individuals in diagnosis group A (obstructive diseases); individuals whose six-minute walk distance falls between 928-1160 feet are 28.6 (95% CI: 27.6-29.6) times more likely to receive better priority under the modified LAS relative to individuals whose six-minute walk distance falls between 0-656 feet; individuals receiving continuous mechanical ventilation are 18.7 (95% CI: 16.5-21.1) times more likely to receive better priority under the modified LAS relative to individuals who do not require continuous mechanical ventilation; individuals who require no assistance in daily living tasks are 21.3 (95% CI: 20.0-22.7) times more likely to receive better priority under the modified LAS relative to individuals who require at least some assistance with these tasks; and the odds of receiving better priority under the modified LAS decrease as age increases.

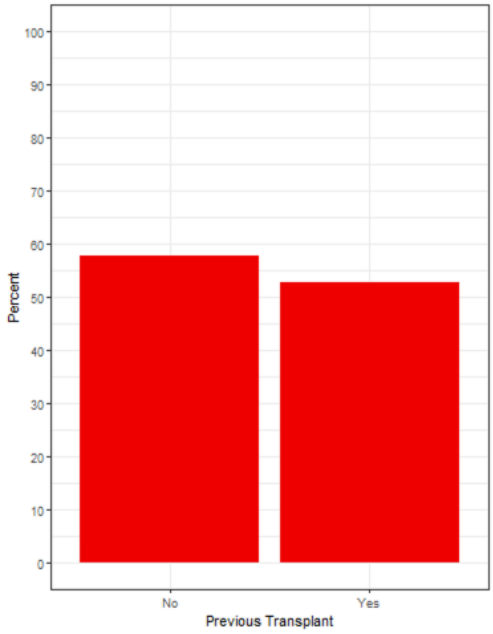

Percent
100
90
80
70
60
50
40
30
20
10
0
A
B
C
D
Diagnosis Group

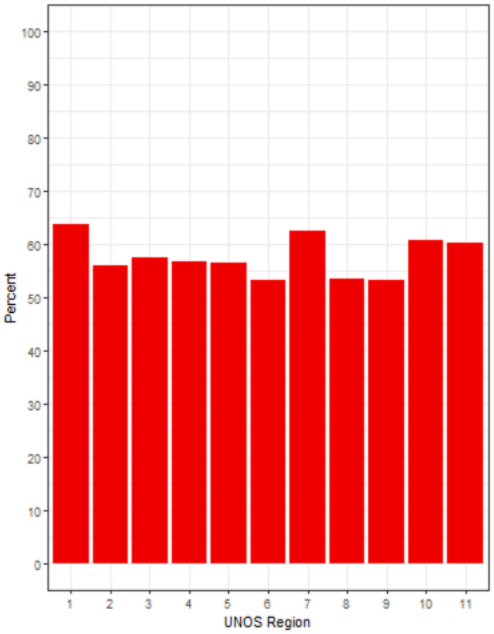

Percent
100
90
80
70
60
50
40
30
20
10
0
[18,50)
[50,59)
[59,65)
[65,80]
Age (years)

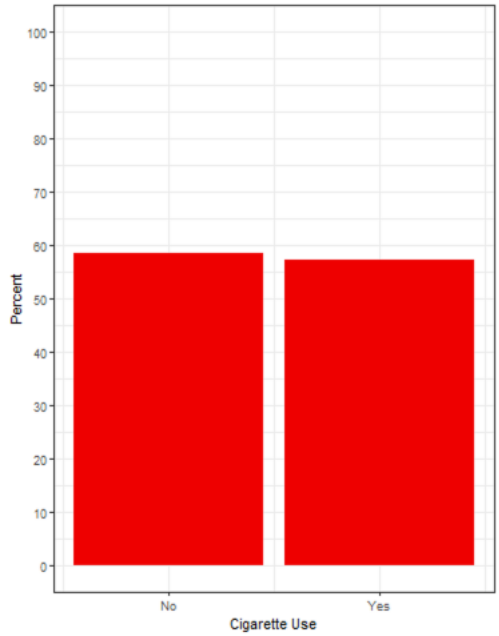

Percent
100
90
80
70
60
50
40
30
20
10
0
M
F
Gender

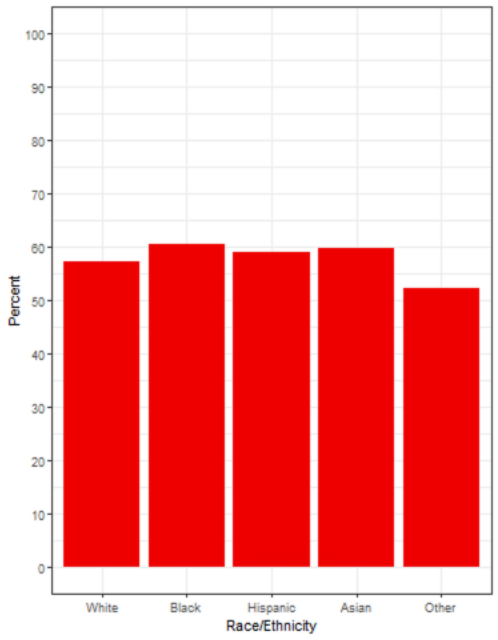

Percent
100
90
80
70
60
50
40
30
20
10
0
Underweight
Normal Weight
Overweight
Obese
BMI

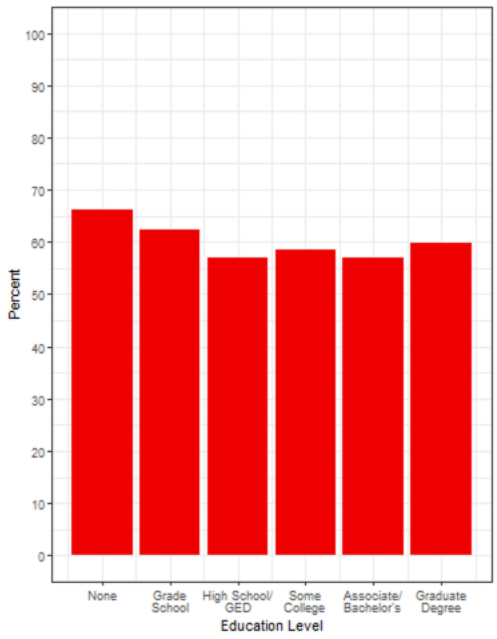

Percent
100
90
80
70
60
50
40
30
20
10
0
A
AB
B
O
Blood Type

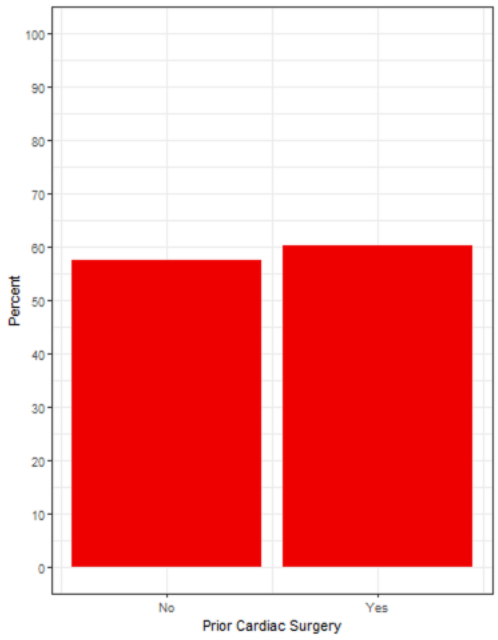

Percent
100
90
80
70
60
50
40
30
20
10
0
No
Yes
Previous Transplant

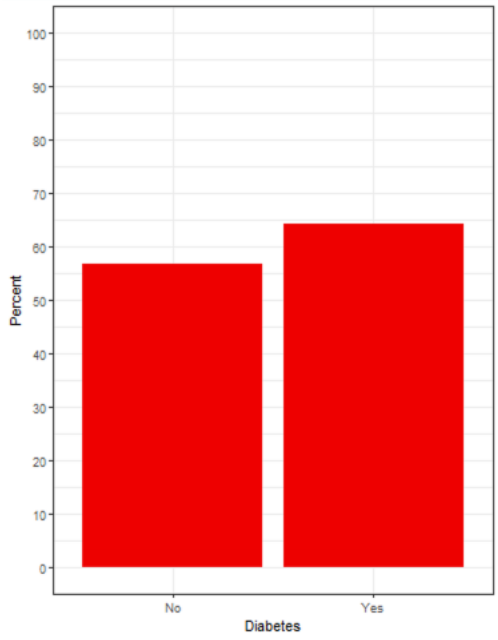

Percent
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
UNOS Region

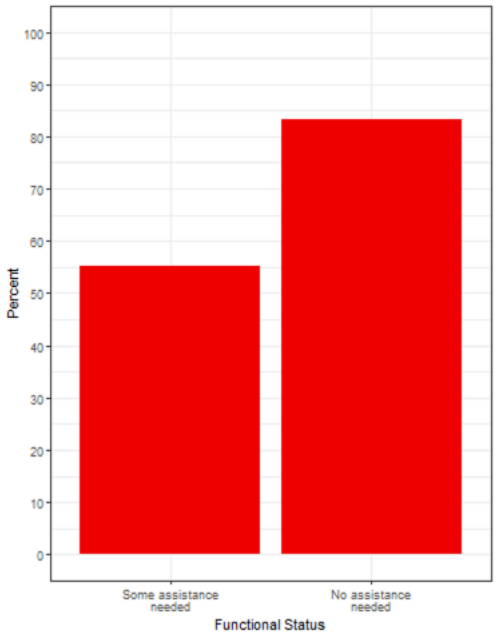

Percent
100
90
80
70
60
50
40
30
20
10
0
No
Yes
Cigarette Use

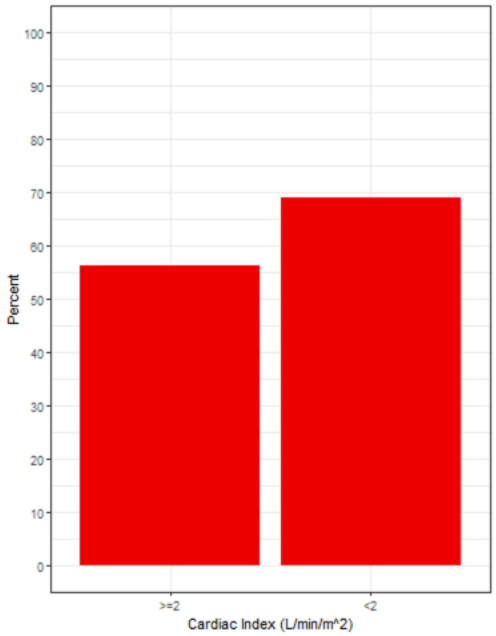

Percent
100
90
80
70
60
50
40
30
20
10
0
White
Black
Hispanic
Asian
Other
Race/Ethnicity

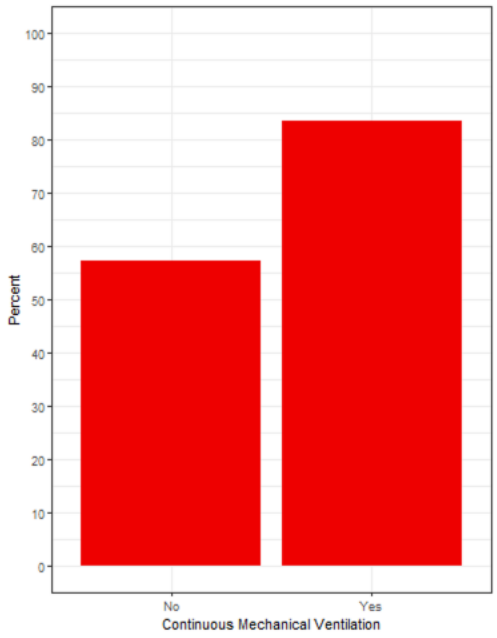

Percent
100
90
80
70
60
50
40
30
20
10
0
None
Grade School
High School/ GED
Some College
Associate/ Bachelor's
Graduate Degree
Education Level

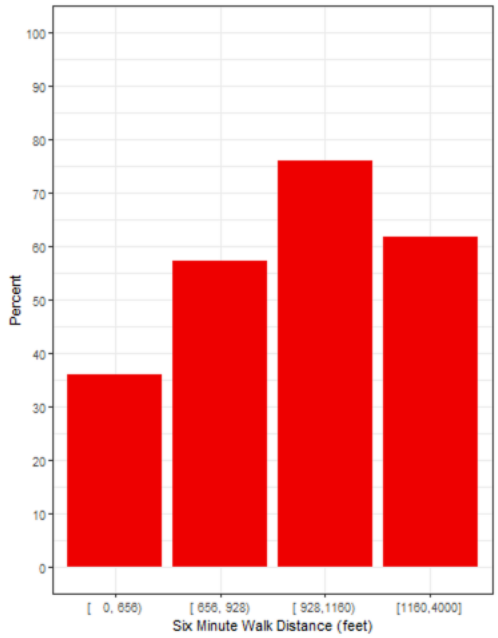

Percent
100
90
80
70
60
50
40
30
20
10
0
No
Yes
Prior Cardiac Surgery


Percent
100
90
80
70
60
50
40
30
20
10
0
No
Yes
Diabetes


Percent
100
90
80
70
60
50
40
30
20
10
0
Some assistance needed
No assistance needed
Functional Status


Percent
100
90
80
70
60
50
40
30
20
10
0
>=2
<2
Cardiac Index (L/min/m^2)


Percent
100
90
80
70
60
50
40
30
20
10
0
No
Yes
Continuous Mechanical Ventilation


Percent
100
90
80
70
60
50
40
30
20
10
0
[0, 656)
[656, 928)
[928,1160)
[1160,4000]
Six Minute Walk Distance (feet)

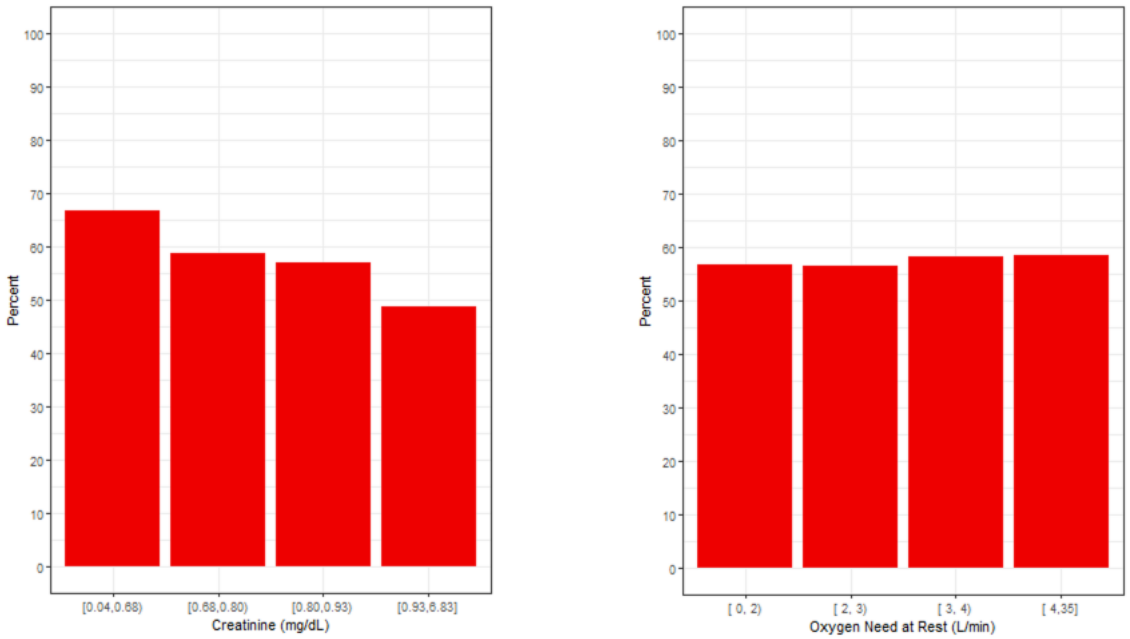


Figure 3.1. Bar charts of the predicted probability of receiving better priority under the modified LAS compared to the existing one for each covariate individually, after adjusting for all other covariates

In probability terms, individuals in diagnosis group D had an 81.1% probability of receiving better priority under the modified LAS. Conversely, individuals in diagnosis group A were almost equally likely to receive better priority (47.2%) versus worse priority or no change (52.8%) under the modified LAS (Figure 2). Individuals whose six-minute walk distance fell between 928-1160 feet had the highest probability of receiving better priority under the modified LAS (76.1%) compared to individuals whose walk distance fell below or above that range. Individuals receiving continuous mechanical ventilation or who required no assistance with daily living tasks also had higher probabilities of receiving better priority under the modified LAS. Finally, the probability of receiving better priority under the modified LAS also decreased with age.

To quantify the extent to which changes in prioritization are driven by changes in predicted pre- or post-transplant survival between the modified and existing LAS, separate generalized r^2 statistics were obtained for each panel in these plots, as well as

overall (Table 3.3). Histograms of the difference between patients' rank under the mLAS versus LAS are shown in Appendix 3.1 for reference.

Table 3.3. Generalized r^2 values quantifying the extent to which changes in prioritization under the modified versus existing LAS are driven by changes in predicted pre- or post-transplant survival

Values were obtained from generalized linear mixed models with continuous differences in ranks for each patient under the modified and existing LAS as the outcome and differences in predicted pre- and post-transplant survival as predictors.

Model	Covariate	Difference in Predicted Survival	Lower Priority	Higher Priority
Overall	---	Pre-transplant	0.144 (0.142, 0.147)	0.215 (0.212, 0.217)
		Post-transplant	0.028 (0.027, 0.030)	0.100 (0.098, 0.101)
Stratified	Diagnosis Group A	Pre-transplant	0.498 (0.494, 0.501)	0.249 (0.245, 0.254)
		Post-transplant	0.067 (0.064, 0.069)	0.248 (0.244, 0.253)
	Diagnosis Group B	Pre-transplant	0.049 (0.043, 0.055)	0.268 (0.257, 0.279)
		Post-transplant	0.033 (0.028, 0.038)	0.020 (0.016, 0.024)
	Diagnosis Group C	Pre-transplant	0.239 (0.233, 0.245)	0.184 (0.166, 0.201)
		Post-transplant	0.064 (0.060, 0.068)	0.088 (0.075, 0.103)
	Diagnosis Group D	Pre-transplant	0.018 (0.016, 0.021)	0.294 (0.291, 0.297)
		Post-transplant	0.006 (0.005, 0.008)	0.085 (0.083, 0.087)
Stratified	Six-minute walk Q1: [0, 656) feet	Pre-transplant	0.039 (0.037, 0.042)	0.244 (0.238, 0.250)
		Post-transplant	0.024 (0.022, 0.026)	0.025 (0.022, 0.028)
	Six-minute walk Q2: [656, 928) feet	Pre-transplant	0.227 (0.222, 0.232)	0.198 (0.193, 0.202)
		Post-transplant	0.045 (0.049, 0.042)	0.127 (0.123, 0.131)
	Six-minute walk Q3: [928, 1160) feet	Pre-transplant	0.091 (0.086, 0.097)	0.280 (0.276, 0.285)
		Post-transplant	0.073 (0.068, 0.079)	0.203 (0.199, 0.207)
	Six-minute walk Q4: [1160, 4000) feet	Pre-transplant	0.349 (0.343, 0.355)	0.207 (0.202, 0.211)
		Post-transplant	0.029 (0.026, 0.032)	0.159 (0.154, 0.163)

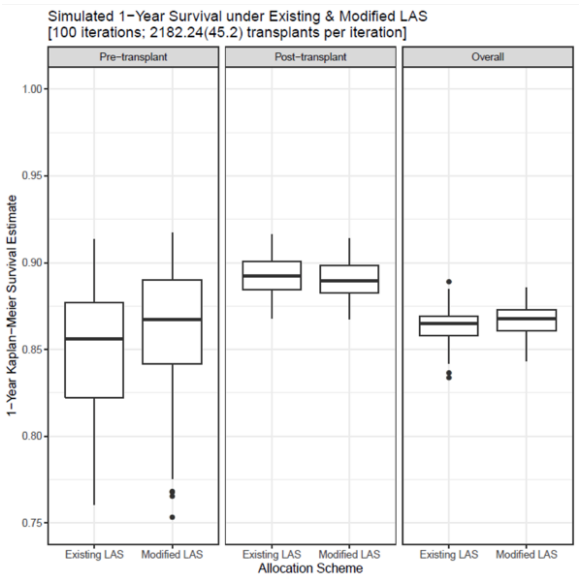
In both the worse and better priority subsets, changes in predicted pre-transplant survival explained a greater proportion of the variability in differences in rank than changes in predicted post-transplant survival. Among individuals who received worse priority under the modified LAS, changes in predicted pre-transplant survival accounted for 14.4% of the variability in differences in rank, while changes in predicted post-

transplant survival accounted for 2.8% of this variability. Similarly, among individuals who received better priority under the modified LAS, changes in predicted pre-transplant survival accounted for 21.5% of the variability in differences in rank, while changes in predicted post-transplant survival accounted for 10.0% of this variability. This pattern was observed for most of the analyses that stratified by diagnosis group and six-minute walk distance, with one notable exception: for individuals in diagnosis group A who received better priority, changes in predicted post-transplant survival accounted for nearly the same proportion of variability in differences in rank (i.e., 24.8%) compared to changes in predicted pre-transplant survival (i.e., 24.9%).

Simulation Study

Across all scenarios, 31.1% of records, on average, had their LAS and mLAS scores interpolated due to simulated death dates exceeding patients' last observed visit dates. Our reference simulation involved 364 offer dates and a transplant rate of approximately 6 offers per day, as this rate mimics current donor capacity. Kaplan-Meier estimates of one-year pre-transplant (waitlist) survival, one-year post-transplant survival, and one-year overall survival among all individuals appear in Figure 3.2A. Waitlist survival improved under the modified LAS compared to the existing one, with a median difference of 1.28% (interquartile range, IQR: 0.52%-2.43%). In a waitlist population of 1000 individuals, this result translates into 12.8 (IQR: 5.2-24.3) fewer waitlist deaths per year under the modified LAS compared to the existing one. Post-transplant and overall survival remained comparable across models.

A



B

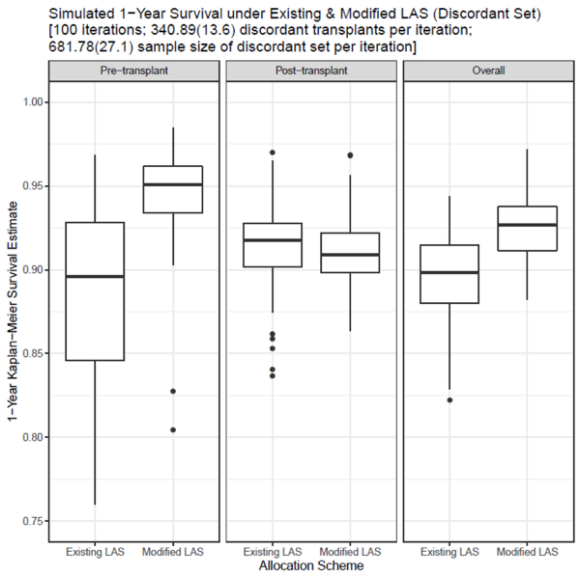
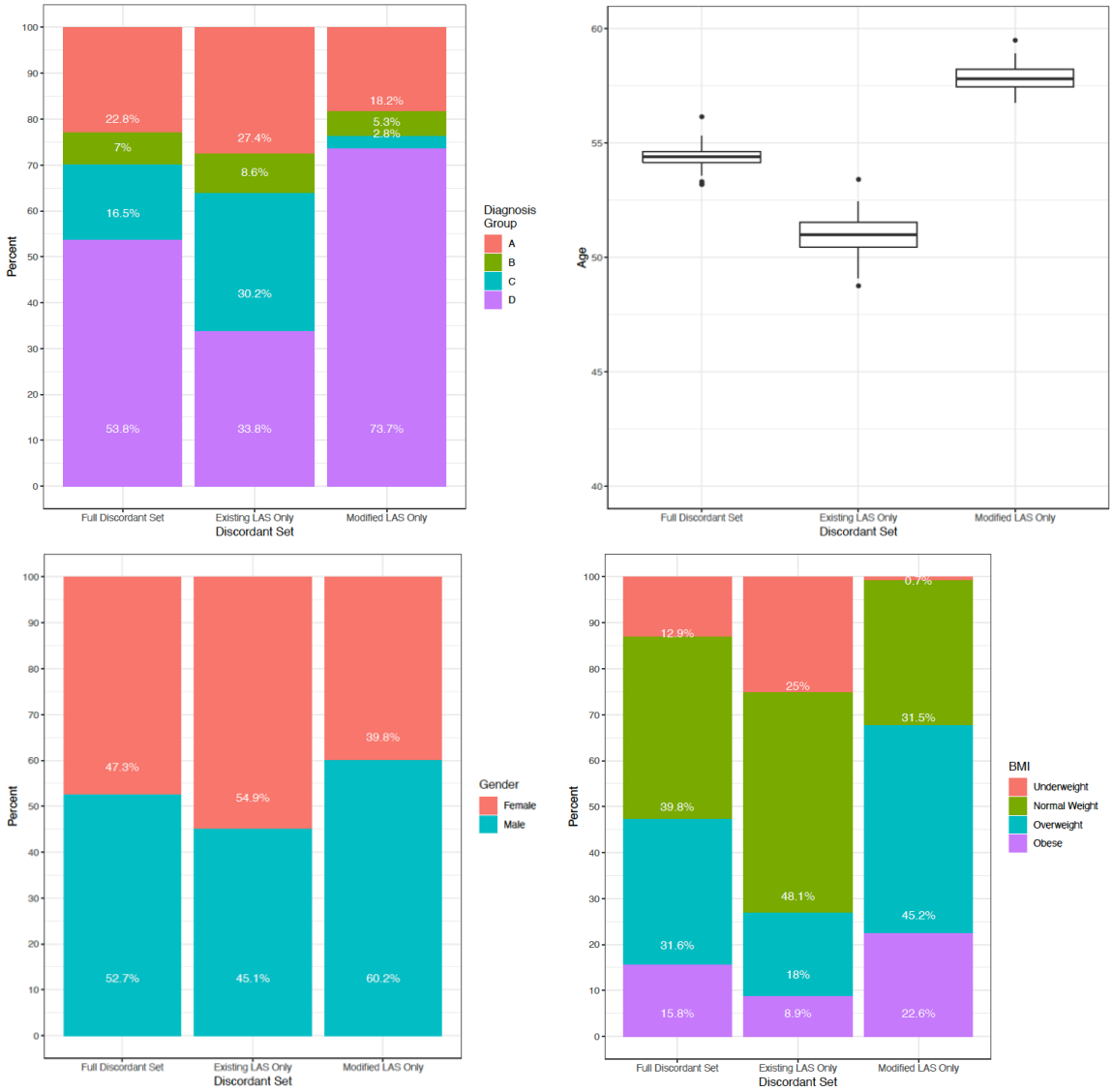


Figure 3.2. Survival estimates under the reference simulation (6 offers per day)

Kapan-Meier estimates of one-year pre-transplant (waitlist) survival, one-year post-transplant survival, and one-year overall survival from the reference simulation (i.e., 6 offers per day) for the overall simulated population (A) and the discordant set (B).

The majority (84.4% [IQR: 83.9%-84.9%]) of individuals received hypothetical transplant under both LAS models. Among the subset of individuals (15.6% [IQR: 15.1%-16.1%]) to whom organs were differentially allocated under the modified LAS (discordant set), the median difference in waitlist survival was 6.89% (IQR: 2.46%-11.5%) (Figure 3.2B). Overall survival in the discordant set appears slightly more

favorable under the modified LAS as well, with a median difference of 2.52% (IQR: 0.90%-4.16%). In a population of 1000 transplanted individuals, these results imply that 844 (IQR: 839-849) of these individuals would be the same under the two LAS models, but the remaining 156 (IQR: 151-161) individuals would differ. That is, in 156 (IQR: 151-161) instances, the modified LAS selects a different person for transplant than the existing LAS. Among these differentially transplanted individuals, 68.9 (IQR: 24.6-115.0) fewer waitlist deaths and 25.2 (IQR: 9.0-41.6) fewer overall deaths occur per year under the modified LAS compared to the existing one. Demographic and clinical characteristics of individuals in the discordant set appear in Figure 3.3.




Percent
Discordant Set
Full Discordant Set
Existing LAS Only
Modified LAS Only
Blood Type
A
AB
B
O
39%
1.7%
9.9%
49.5%
39%
1.7%
9.9%
49.4%
39%
1.6%
9.9%
49.5%
Previous Transplant
No
Yes
97%
3%
96.7%
3.3%
97.3%
2.7%
Cigarette Use
No
Yes
48.9%
51%
56.2%
43.6%
41.5%
58.5%
Prior Cardiac Surgery
No
Yes
93.2%
5%
93.8%
4.2%
92.5%
5.7%
Diabetes
No
Yes
78.7%
21.3%
76.1%
23.9%
81.2%
18.8%
Race/Ethnicity
White
Black
Hispanic
Asian
Other
78.1%
10.5%
8%
81.4%
7.6%
7.2%
74.9%
13.4%
8.8%

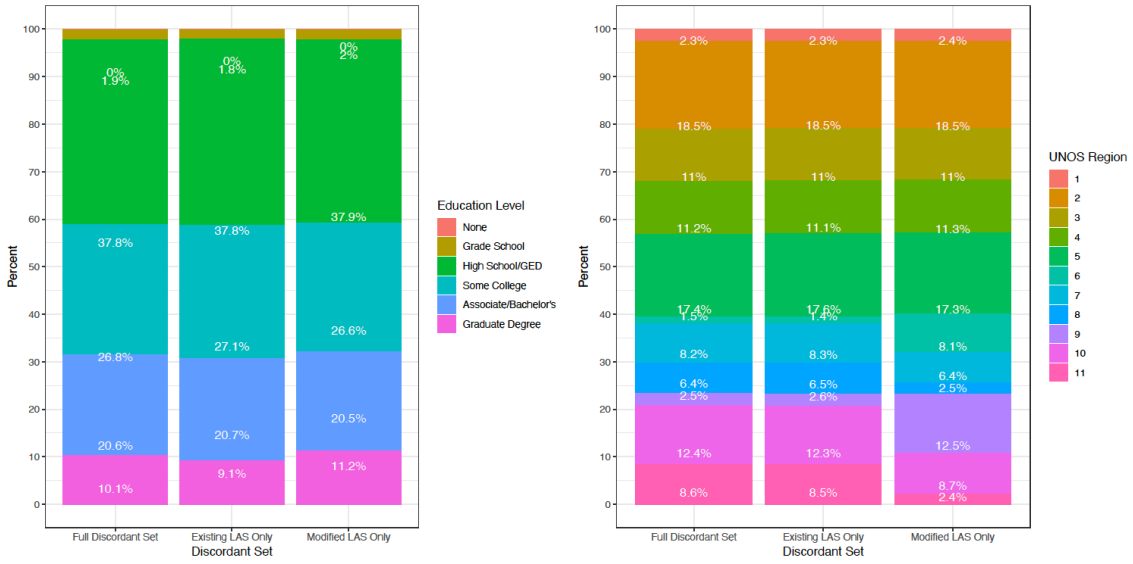
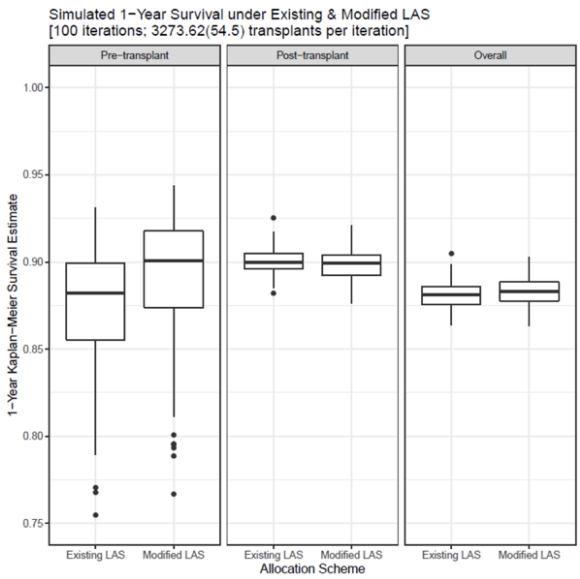


Figure 3.3. Distributions of demographic and clinical characteristics of the discordant set from the reference simulation (6 offers per day)

Increasing or decreasing the transplant rate (to 9 or 3 offers per day, respectively) did not substantially change our results (Figures 3.4 and 3.5). However, the concordance index and Kaplan-Meier estimates of waitlist survival are both higher under the "high" transplant rate scenario compared to the "low" transplant rate scenario, consistent with the fact that a much larger proportion of our simulated population is transplanted in the "high" scenario. For example, in the "high" transplant rate scenario, 91.4% (IQR: 91.2%-91.7%) of individuals received hypothetical transplant under both LAS models, whereas in the "low" transplant rate scenario, the median concordance was only 73.2% (IQR: 72.5%-23.8%). Similarly, in the "high" transplant rate scenario, the median (IQR) Kaplan-Meier estimates for waitlist survival under the existing and modified LAS were 87.3% (80.4%-89.8%) and 89.7% (85.9%-91.4%), respectively. Conversely, in the "low" transplant rate scenario, the median (IQR) Kaplan-Meier estimates for waitlist survival under the existing and modified LAS were 83.0% (78.9%-85.2%) and 84.2% (80.1%-86.3%), respectively.

A



B

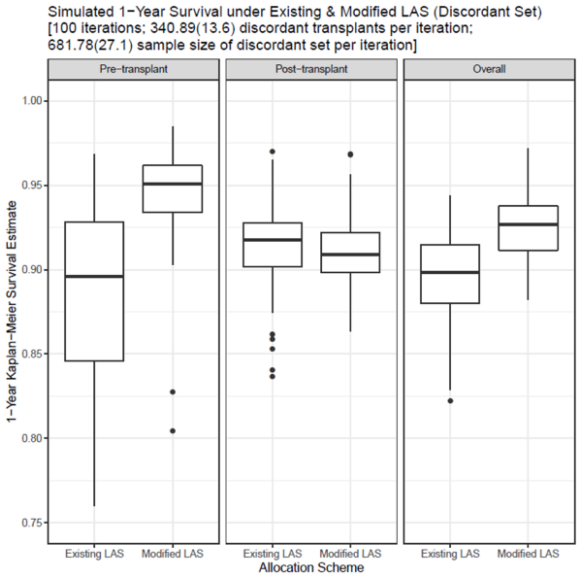
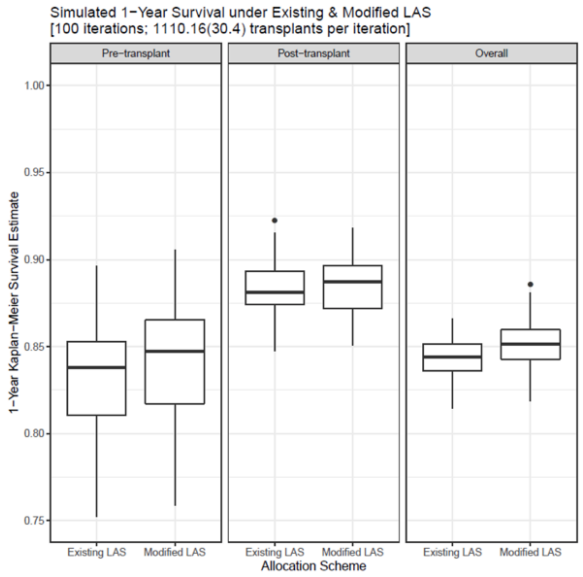


Figure 3.4. Survival estimates under the high transplant rate simulation (9 offers per day)

Kapan-Meier estimates of one-year pre-transplant (waitlist) survival, one-year post-transplant survival, and one-year overall survival from the high transplant rate scenario (i.e., 9 offers per day) for the overall simulated population (A) and the discordant set (B).

A



B

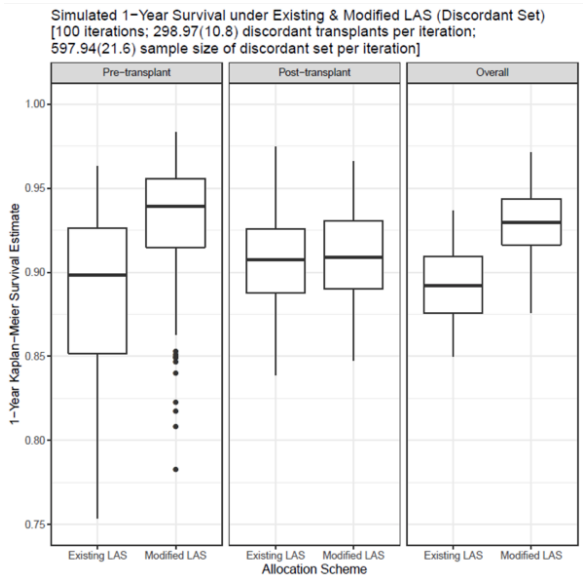


Figure 3.5. Survival estimates under the low transplant rate simulation (3 offers per day)

Kapan-Meier estimates of one-year pre-transplant (waitlist) survival, one-year post-transplant survival, and one-year overall survival from the low transplant rate scenario (i.e., 3 offers per day) for the overall simulated population (A) and the discordant set (B).

Discussion

In this study, we used observed and simulated data to examine the clinical impact of a modified LAS score designed to mitigate selection bias. We found that changes in prioritization were more pronounced for patients with certain demographic and clinical characteristics, such as diagnosis group (individuals in group D were more likely to receive better priority), six-minute walk distance (individuals in the third quartile of walk distance had the highest probability of receiving better priority), continuous mechanical ventilation (individuals receiving continuous mechanical ventilation were more likely to receive better priority), and functional status (individuals who required no assistance in daily living tasks had a higher probability of receiving better priority). Changes in predicted pre-transplant survival tended to explain a greater proportion of the variability in differences in rank than changes in predicted post-transplant survival. Simulations suggest that one-year waitlist survival may improve slightly under the modified LAS relative to the existing one, whereas one-year post-transplant and overall survival remain comparable under the two models. Among individuals who received hypothetical transplant under one LAS but not the other, waitlist and overall survival appears more favorable under the modified LAS; post-transplant survival remains similar to that under the existing LAS.

The fact that changes in prioritization under the modified LAS may be more pronounced for some patients but not others is clinically relevant, as it implies that implementation of the modified LAS could lead to important differences in allocation. For example, individuals in diagnosis group D (pulmonary fibrosis) – who tend to be sicker and unable to wait long for transplant – had an 81.1% probability of receiving better priority under the modified LAS (Figure 3.1). Conversely, individuals in diagnosis group A (obstructive diseases) – who tend to survive longer without transplant, albeit with lower quality of life – were almost equally likely to receive better priority (47.2%) versus worse

priority or no change (52.8%) under the modified LAS. Moreover, for individuals in diagnosis group A who received better priority, changes in predicted post-transplant survival accounted for nearly the same proportion of variability in differences in rank (i.e., 24.8%) compared to changes in predicted pre-transplant survival (i.e., 24.9%). Such observations suggest that while improvements in the prediction of waitlist urgency tend to have the largest impact on patient prioritization, changes in the prediction of post-transplant survival can impact prioritization for some subgroups of patients. This result is important to consider, especially as the transplant community continues to debate the role that post-transplant survival should play in organ allocation (Kasiske et al., 2020; Maxwell et al., 2014).

With regards to six-minute walk distance, continuous mechanical ventilation, functional status, and age, our analyses indicate that after adjusting for other covariates, younger, healthier individuals may have a higher probability of receiving better priority under the modified LAS compared to older, frailer individuals. These results are consistent with recent research advocating for more transparency around the impact of age and frailty on pre- and post-transplant survival (Blumenthal et al., 2017; Schaenman et al., 2020). Such findings could also help inform discussions on what role these factors should play in organ allocation (Persad et al., 2009; Veatch & Ross, 2015).

Our simulation study demonstrated that the modified LAS may yield modest improvements in waitlist survival. The extent of improvement in waitlist survival was larger among the subset of individuals to whom organs were differentially allocated under the modified LAS. These findings suggest that alternative organ allocation schemes can confer as much or more survival benefit to patients as the existing LAS. However, they also imply that evaluating the performance of allocation systems solely based on survival might not capture the full extent of transplant benefit. This observation is consistent with literature calling for greater emphasis on quality of life when evaluating

the risks and benefits of transplantation (Finlen Copeland et al., 2013; Singer et al., 2015; Yusen, 2009).

Our study has some limitations. First, the GLMM approach we took to estimate the proportion of variability (r^2) in the differences in rank explained by changes in predicted pre- versus post-transplant survival does not account for the correlation among differences in rank. We mitigated this issue by splitting the data into worse-priority and better-priority subsets, and fitting separate GLMMs to each subset. However, future methodologic research is needed to extend generalized r^2 statistics to the situation where the outcome of interest is differences in ranks (e.g., perhaps by building off of transformation approaches commonly used in genetics studies, such as (Beasley et al., 2009)). Second, while we used observed data to compare the order in which patients were prioritized under the two LAS models and examined the demographic and clinical characteristics of patients who received better priority under the modified LAS relative to the existing LAS, we cannot know whether patients who died on the waitlist under the existing LAS would, in fact, have received transplant had the modified LAS been used instead. This limitation motivated us to conduct the simulation study where we examined how patients' waitlist and post-transplant survival might change if the modified LAS were implemented in clinical practice. Limitations of this simulation include a simplified allocation policy based on blood type and region only (i.e., ignoring other salient, but difficult to measure, factors to organ matching, such as donor/recipient size and sensitization, donor organ quality and acceptance criteria, and travel constraints that arise when delivering donor organs to transplant candidates). Future work should aim to incorporate these additional factors into the simulation to see how the modified LAS might perform in more realistic settings.

Overall, this study demonstrates that implementing a modified LAS that mitigates selection bias into clinical practice can lead to important differences in allocation and

possibly modest improvement in one-year waitlist survival, without adversely affecting one-year post-transplant or overall survival. These findings can inform ongoing efforts to modify lung allocation policy, such as the continuous distribution model.

CHAPTER 4. A CONCEPTUAL MODEL FOR SOURCES OF SELECTION BIAS IN LUNG TRANSPLANT ALLOCATION

This chapter is under review. The suggested citation is:

Schnellinger EM, Cantu E, Kimmel SE, Szymczak JE. A Conceptual Model for Sources of Differential Selection in Lung Transplant Allocation (under review).

Abstract

Background. In the U.S., donor lungs are allocated to transplant candidates based on the lung allocation score (LAS). However, additional factors beyond the LAS can impact who is transplanted, including referral, listing, and donor organ acceptance practices. These factors can result in selection bias, undermining the fairness of lung allocation. Yet they have not been systematically studied using formal qualitative methods.

Methods. We conducted a semi-structured qualitative interview study of lung transplant surgeons and pulmonologists in the U.S. between June 2019 and June 2020 to understand the role that the LAS plays in clinical decision making, how transplant clinicians think about selection bias in lung allocation, and whether they feel the LAS should be modified to account for such bias.

Results. Our 51 respondents (30 transplant surgeons and 21 pulmonologists) identified many sources of selection bias arising at several points along the pathway from referral to transplantation. We developed a conceptual model synthesizing these sources of selection bias into five factors: 1) transplant center's level of risk tolerance and accountability, 2) successfulness and fairness of the LAS in mitigating selection bias, 3) donor organ availability and regional competition, 4) patient health versus program health, and 5) access to care versus responsible stewardship of organs.

Conclusions. Our conceptual model demonstrates how selection bias can arise throughout lung transplantation and can facilitate further study of such bias. As the transplant community continues to develop and implement new allocation models for lungs and other organs, the sources of selection bias described here should be considered carefully to ensure that the resulting allocation scheme is more equitable.

Background

In May 2005, the Organ Procurement and Transplantation Network (OPTN) implemented a new allocation policy that allocates donor lungs to candidates based on a lung allocation score (LAS) (Egan et al., 2006; Gottlieb, 2017; U.S. Department of Health and Human Services, October 20, 1999; United Network for Organ Sharing, 2015). The LAS is calculated as the difference in days of life between transplant benefit and waiting list urgency, and is normalized so that it ranges from 0 to 100, with higher values indicating greater priority for lung transplantation (Egan et al., 2006; United Network for Organ Sharing, 2015). This score was intended to balance equity, justice, beneficence, and utility in lung allocation (Gottlieb, 2017; Veatch & Ross, 2015). However, additional factors beyond the LAS can impact which patients are selected for – and ultimately receive – transplant, including the timing of referral (Weill et al., 2015) and geographic, gender, or racial/ethnic disparities in waitlist registration (Mooney et al., 2018; Ross-Driscoll et al., 2020; Thabut et al., 2012; Wille et al., 2013), pre- and post-transplant survival (Egan & Edwards, 2016; Maxwell et al., 2014; Russo et al., 2011), and donor organ availability (Benvenuto et al., 2018; Drolen et al., 2020; Ross-Driscoll et al., 2020). Collectively, such factors can give rise to selection bias, which occurs because individuals are removed from the waitlist upon receipt of transplant, and because in order to receive transplant, individuals must survive long enough for a suitable donor organ to become available (Glymour & Greenland, 2008; Schnellinger et al., 2021). Such bias can undermine the fairness of organ allocation.

While much statistical research has been conducted surrounding the LAS and selection bias (Schnellinger et al., 2021; Vock et al., 2017; Vock et al., 2013; Xiang & Murray, 2012), less is known about how lung transplant surgeons and pulmonologists employ the LAS in practice, or how these clinicians interpret and respond to selection bias. A few qualitative studies have examined candidate (Blumenthal et al., 2017; Volk et

al., 2011) and organ selection (Loss et al., 2013); however, these studies focused almost exclusively on the patient screening/evaluation phase of transplantation, rather than the complete pathway from referral through transplantation.

Given that OPTN is currently developing a new organ allocation framework – the continuous distribution model (Kasiske et al., 2020; Snyder et al., 2018; U.S. Department of Health & Human Services) – and that this new framework is planned to be implemented in lung transplantation followed by other organs (e.g., kidney, liver, heart), it is important to develop a thorough understanding of where selection bias can arise in the current lung allocation system and how such bias can impact the successfulness and fairness of lung allocation. Synthesizing such information into a conceptual model can facilitate further study of the patient- and program-level effects of selection bias in lung transplantation. Toward that end, we conducted a qualitative study of lung transplant surgeons and pulmonologists between June 2019 and June 2020 to understand the role that the LAS plays in clinical decision making, how transplant clinicians think about selection bias in lung allocation, and whether they feel the LAS should be modified to account for such bias.

Methods

Study Design and Target Population

We conducted a qualitative study between June 2019 and June 2020 to understand the role that the Lung Allocation Score (LAS) plays in clinical decision making. Our target population consisted of lung transplant surgeons and pulmonologists practicing in the United States. The study protocol was reviewed by the University of Pennsylvania Institutional Review Board (IRB Protocol #833089) and determined to meet exemption criteria authorized by 45 CFR 46.104, category #4,2. We also followed

the Consolidated criteria for Reporting Qualitative Research (COREQ) checklist (Tong et al., 2007) (see Appendix 4.1).

Participants and Recruitment

Participants were recruited through the "Pulmonology and Cardiothoracic/vascular surgery" listserv of the International Society for Heart and Lung Transplant (ISHLT). Individuals were included in our study if they were lung transplant surgeons or pulmonologists practicing in the United States; individuals who did not practice in the United States, were still in training (e.g., medical student or resident), or who focused on other organ transplant (e.g., heart) were excluded. Eligible individuals were invited to participate via email by the lead author who has training in qualitative methods (EMS). We purposively sampled individuals to introduce variation by discipline (lung transplant surgeon or pulmonologist), and also sought to include respondents with diverse characteristics relevant to our study question (e.g., geography, years in practice). The lead and senior author (JES, a sociologist with expertise in qualitative methods) monitored for thematic saturation during the data collection period to determine sample size adequacy. We maintained a data collection memo where key concepts were recorded following each interview. A saturation grid based on the interview guide was created to track the emergence of novel themes across respondents (Brod et al., 2009). Recruitment continued until thematic saturation across the main domains of our interview guide was reached (Miles et al., 1994; Miles et al., 2014; Weiss, 1994).

Data Collection

Interviews were conducted either in-person or by phone by the lead author using a semi-structured guide (see Appendix 4.2). The interview guide was piloted and

questions refined to ensure complete capture of relevant concepts, comprehension of questions, and length. The final guide included questions in the following domains: 1) how lung transplant surgeons and pulmonologists use the LAS in clinical practice; 2) factors which clinicians deem important for prioritizing lung transplant patients, but which might not be captured by the LAS; 3) how clinicians think about selection bias in lung allocation; and 4) whether clinicians think the LAS should be modified to account for this bias, and if so, how. Participants consented verbally and gave permission for their interview to be recorded.

Analyses

Interviews were professionally transcribed, de-identified, and uploaded to NVivo qualitative data analysis software (NVivo Release 1.0, QSR International 2020, Doncaster, Australia) for analyses. Data were analyzed by two authors (EMS, JES) in consultation with the research team using a modified Framework Method (Gale et al., 2013). Analysis proceeded in four phases. First, interview transcripts and data collection memos were reviewed to create an index codebook (Deterding & Waters, 2021). Second, the index codebook was applied to the transcripts line-by-line. The index codebook was open to revision as patterns emerged in the data, and existing codes were further developed into subcodes. Revisions to the codebook were performed in consultation with the research team and discrepancies were resolved by consensus. Third, once the transcripts were coded we charted the data into a framework matrix, exploring overlapping codes and comparing themes within and across respondents. At this stage we confirmed that thematic saturation was reached both overall and within subgroups defined by discipline (transplant surgeon, pulmonologist) (Gale et al., 2013). Fourth, we developed a conceptual model for sources of selection bias based on the

framework matrix. The model was refined through discussions amongst the research team and negative case analysis.

Self-reported respondent characteristics, including gender, race/ethnicity, years of experience, and UNOS region, were ascertained at the end of each interview and summarized using counts (proportions) or medians (interquartile ranges), as appropriate (STATA 15, StataCorp LLC, College Station, TX, USA).

Results

Respondents

Email invitations were sent to 218 potential respondents, with 76 (35%) responding. Of the 76 total individuals who responded, 19 declined to participate, 4 did not meet our study's inclusion/exclusion criteria (e.g., they were still in training or focused on heart transplant), and 2 did not respond to follow-up emails to schedule an interview. We interviewed the remaining 51 participants (30 [59%] pulmonologists and 21 [41%] lung transplant surgeons). The majority of participants were male (84%), white (61%), and had practiced at more than one institution (55%), with a median duration of experience of 15 years (Table 4.1).

Table 4.1: Demographic characteristics of interview respondents

Factor	Value
N	51
Gender	
Female	8 (16%)
Male	43 (84%)
Race/ethnicity	
African American	4 (8%)
Asian/Indian	6 (12%)
Hispanic/Latino	7 (14%)
Other	3 (6%)
White	31 (61%)
Discipline	
Pulmonologist	30 (59%)
Transplant Surgeon	21 (41%)
UNOS Region	
1	1 (2%)
2	18 (35%)

3	5 (10%)
4	1 (2%)
5	3 (6%)
6	1 (2%)
7	5 (10%)
8	2 (4%)
9	2 (4%)
10	5 (10%)
11	8 (16%)
Prior institution	28 (55%)
Years of experience [median (IQR)]	15.0 (9.0, 20.0)
Began practice before LAS was implemented	24 (47%)

Sources of Selection Bias Along the Lung Transplantation Pathway

In each of the sections below, we describe a step along the transplantation pathway where respondents identified that selection bias might arise. These sources of selection bias are supported by exemplar quotes and mapped to five overarching concepts (as explained further in the Conceptual Model section).

Referral

Respondents identified several time points at which selection bias influences lung transplantation. First, they identified referral to a transplant center. However, because lung transplant surgeons and pulmonologists are not typically involved at this stage of the process, respondents had less to say about the way selection bias shapes referral. That said, some respondents suggested that referring patterns depend on how knowledgeable patients and their primary pulmonary providers are about transplantation:

> *"At some point, a physician, sometimes at [hospital] but sometimes in the community, decides that a patient's lung disease is advanced enough to be referred to [transplant center] for transplant evaluation. […] So there's this pre-screening step that happens way before we even meet some of these patients" [pulmonologist].*

Other respondents advocated for more data on patients who are referred to a transplant center, but ultimately not listed:

> *"I think something that doesn't get captured very well by research though is how many people present for evaluation for a transplant and then pass away. Because there's many people that may present for evaluation that we turn down or get deactivated and that's not necessarily captured very well. It would be interesting if [the UNOS] registry could be something that captured that as well, like captured all people that were considered for lung transplant. Whereas right now it only is capturing people that actually go on to undergo a transplant. So you know, there's no national registry that captures all patients who present for evaluation. And so while it may capture people that die on the waitlist, it doesn't capture any of the people that were deemed not to be good candidates for transplant so that were never listed." [pulmonologist]*

Such data could help clarify the role selection plays in referral processes.

Screening and Waitlist Registration

Respondents identified the transition between transplant screening/evaluation and waitlist registration as the second major source of selection bias after referral. The screening/evaluation process involves a series of evaluations conducted by the transplant center – including assessments of physical and mental health and consideration of social and financial support – to determine whether a particular patient is sick enough to be registered on the waitlist yet healthy enough to withstand transplant, and whether the patient has sufficient social and financial support to undergo such a major procedure. Several respondents referred to this timeframe as the patient's *"transplant window"*, and described how patients listed too early or too late with respect to this window may have a difficult time finding a suitable donor organ match and receiving transplant. While professional societies and consensus documents provide guidance to the lung transplant community as to the appropriate timing of – and criteria for – listing patients, respondents suggested that listing decisions ultimately come down to clinician judgment and subjective impression:

> *"I think you're interpreting somebody's capabilities [of surviving transplant] based on how they look and how robust are they and do they look frail and those types of things which I think are one reflector of survival that kind of goes into the gestalt when you see someone. Part of that assessment is do they look like they're about to keel over or not? And I think there's some patients where that*

probably is accurate and there are some patients where it's really not." [pulmonologist]

Bias can also come into play more subtly when evaluating patients' extent of social and financial support, their level of adherence with protocols, and how that might impact their transplantation outcomes:

"The atrial pressure of your heart, that's hard data; your PFTs [pulmonary function tests] are hard data. So [selection committees] shouldn't be biased about how sick your heart is, or how sick your lungs are, but there is a lot of bias regarding how good of a...how compliant you're gonna be with your medications; what kind of family support you have; what is the level of outside, your stress, on how you're gonna manage to take care of yourself after transplant? And sometimes you can say, 'Well, you know what? This patient might survive a transplant, but they will [be on] chronic pain medications, or taking painkillers, or at some point they smoke marijuana, or at some point...' So, you know, all those factors come into play, particularly from the social aspects, which are the most difficult to define. And how do you define those in a way that you don't stigmatize your patient, and you eventually don't provide them with an opportunity." [surgeon]

Respondents suggested that listing decisions are also influenced by a transplant center's level of risk tolerance within the context of accountability-promoting regulatory pressures. Decisions about individual patients are made with the center's "numbers" (e.g., average one-year post-transplant survival calculated by the Scientific Registry of Transplant Recipients [SRTR] and acceptable performance thresholds determined by UNOS and/or Centers for Medicare & Medicaid Services [CMS] to monitor program performance) in mind.

"I guess the fact that we do 80, 90 transplants per year allows us to reason… well, given the number, we can be a little more aggressive and try to help these [high-risk] patients, which in a smaller center, probably that would give them a bigger hit on their overall numbers. I guess our center, we're large enough, and also experienced enough not to make a strong decision [for or against listing a particular patient]." [surgeon]

Programs with larger transplant volumes may have more experienced surgical teams and increased access to medical technology (e.g., extracorporeal membrane oxygenation) which may enable them to manage high-risk patients more easily. This

perception of program capacity and expertise was understood by our respondents as one potential influence on decision-making:

> *"Everything has some bias involved, bias of the treatment team, bias of the center treating, how big your program is, how experienced are your surgeons, what is your workforce, and what can you take care of, because the same patient can be looked after in a different center [and] based on how they can do things there and the complexity of the patient, if they're used to dealing with better, maybe they can pull it off as opposed to a center which is not used to doing that many transplants or it doesn't have the facility to do that. So I think the bias does exist in one way or the other based on available resources and experience of that center." [surgeon]*

Additionally, their large transplant volume effectively dilutes the impact that poor patient outcomes might have on their program's pre- or post-transplant survival metrics:

> *"The other half of that is in when it comes to candidates, I think there is an awareness of what our current kind of mortality and waitlist statistics are as to whether we're really capable of absorbing a really high-risk candidate. So we're a really aggressive center I think compared to other centers too. But even at a certain point, if you've had a string of bad outcomes or if you've had some waitlist deaths and somebody comes up to the [listing] committee who is a marginal candidate who has the potential to have a bad outcome, depending on where the program is, I think we have to take a harder look at that case and say like, can we absorb another bad outcome in those case? So I think there's a perception that it affects both candidates and recipients." [pulmonologist]*

Thus, it was explained that large-volume centers are perceived to have greater flexibility when choosing which patients to register on their waiting list, whereas small-volume centers may be forced to be more selective in order to promote the continued well-being of their program.

Waitlist Registration and Waiting Period

The third major source of selection bias identified by respondents occurs between the waitlist registration and waiting period phases in the patient pathway. This part of the pathway is ostensibly determined by the LAS score, which was designed to facilitate lung allocation by prioritizing patients for whom transplant benefit (i.e., post-transplant survival minus pre-transplant survival) exceeds waitlist urgency (i.e., pre-

transplant survival); equivalently, the LAS prioritizes individuals for whom post-transplant survival minus two times pre-transplant survival is favorable. In practice, the LAS determines the order in which patients receive donor organ offers as well as the number and quality of offers they receive, with higher-LAS patients having *"first bid" [surgeon]* on donor organ offers. Since the score itself is calculated based on patients' demographic and clinical characteristics, some respondents viewed the LAS as a *"byproduct" [pulmonologist]* that reflects patients' medical urgency, but does not on its own influence providers' decisions to register patients on the waitlist or transplant them:

> "*The score is something that I think we're all cognizant of in the back of our minds, but it's not necessarily a factor to say someone should or shouldn't be a candidate. Because the LAS is almost a byproduct, right? Someone has to be a transplant candidate, then they get the LAS score for whatever it is […] The LAS will be just whatever it is based on the specific tests of that patient." [pulmonologist]*

Some respondents lauded the LAS's ability to update as patients become sicker, thereby ensuring that patients receive more donor organ offers as their waitlist urgency increases, and hence, mitigating survivor bias (especially relative to the previous first-come, first-serve system of lung allocation):

> *"if someone's disease advances and they end up on mechanical ventilation or something like that, their LAS goes through the roof and they're going to get an organ. So in many ways, the LAS mitigates survivor bias" [pulmonologist]*.

Other respondents, however, suggested that the LAS is imperfect, mitigating some – but not all – sources of bias. Specific concerns include: 1) the LAS score's reliance on predicted one-year pre- and post-transplant survival, which may not be as relevant to patients or clinicians as longer term (e.g., three- or five-year) survival or quality of life; 2) the LAS score's failure to include variables important for predicting survival among some diagnosis groups but not others (e.g., FEV1 is clinically relevant for some diagnoses, but not statistically significant across all diagnoses); and 3) discrepancies in LAS among patients who are dual-listed at multiple centers.

These perceived *"gaps" [pulmonologist]* in the LAS led many respondents to question the fairness of lung allocation for particular patients. One respondent described being more concerned about survivor bias for patients with low LAS than for patients with high LAS scores, as the former tend to remain on the waitlist behind the latter, eventually becoming too old or sick to receive transplant:

> *"Yeah, it [survivor bias] is a concern. And it actually is a concern that I primarily have in regards to patients that have a relatively low score at the time of listing, especially patients with emphysema. You know, since the Lung Allocation Score favors patients with interstitial lung disease, patients with emphysema usually get a lower score. And we certainly have a number of patients on our waitlist where I kind of wonder are they ever going to get lungs? And, you know, are they eventually going to be too old, or are they going to develop comorbidities that would prevent them from being suitable for transplant anymore?" [pulmonologist]*

The challenge of advocating for patients within the existing LAS framework can lead to differences in patient management practices – including decisions to de-list patients – across centers.

These differences led some respondents to voice concerns about the potential for gaming, or the use of tactics that exploit the ambiguity and flexibility of the LAS to achieve desired outcomes while ostensibly honoring the framework. Overall, two broad types of gaming were identified by our respondents: 1) patient-level: manipulating a specific patient's LAS to increase their chances of receiving transplant by capitalizing on *"system inefficiencies" [pulmonologist]* (e.g., variables that are open to clinical interpretation, and thus can be changed while maintaining honesty); and 2) program-level: avoiding transplanting patients with the highest LAS or being more conservative with which donor organs one accepts to ensure that program-level accountability metrics are maintained.

The first type of gaming was described as patient advocacy, where clinicians worked to ensure that patients receive the score which truly reflects their waitlist urgency:

> *"Everybody is trying to get an angle for their own patients, you know, everybody feels responsible to find a strategy for their patient to get the transplant. So, they're gonna think of things that they can do that are, in general, are almost always I think are honest and have integrity. I don't think people cheat in any substantial degree, I guess, but they will use every honest angle that is available. We see that in all of our lists. But that's what you want your doctor to do for you, right? We assume that we'll be audited, so anything we do we want to be able to justify and explain why or that we were completely honest with how we did it." [surgeon]*

Having the flexibility to update patients' LAS as their clinical condition changes was seen as a positive attribute of the LAS. All respondents said that changes in patients' LAS must be supported by documentation because centers are subject to site visits and/or audits from UNOS, CMS, and insurance companies. However, respondents mentioned that the guidelines around some variables are vague (e.g., how much oxygen to use for six-minute walk test) and may even have conflicting goals (e.g., the walk test is used to assess both the severity of the patient's disease and the patient's ability to withstand the transplant procedure). If centers have enough resources, they are able to conduct such tests multiple times to fulfill all goals, but if not, some sort of compromise must be made:

> *"At [previous transplant center] we had this requirement that you had to be able to walk 1000 feet in 6 minutes to qualify for listing. So we would give them as much oxygen as they needed it and we would push them to walk really far. And then when they changed the LAS algorithm and we saw, like, all...that, like, these patients, these people who are really, really sick, with really advanced lung disease, their scores went down significantly and we weren't getting offers for them anymore, then we had to change how we did it. And we had the resources there to say, 'Okay, we're going to do an LAS six-minute walk test,' which is where we did a six-minute walk test based on their resting oxygen requirements. But then we would continue to do a six-minute walk test to assess their functional capacity. And so we would have both pieces of information, which was helpful, and then we'd use the LAS six-minute walk to put into a unit. Here, we don't have those kinds of resources to be doing it twice. And so we just accept the fact that, well, we do it on something sort of in the middle and try to interpret the data as best we can." [pulmonologist]*

Respondents also acknowledged that it can sometimes be difficult to distinguish between acceptable and unacceptable patient management strategies:

> *"it's hard to define where gaming the system versus differences in management practices, where that line is drawn" [surgeon]*.

In general, respondents supported the idea of exception requests, which allow physicians to petition UNOS to increase a particular patient's score when the LAS does not seem to accurately capture the severity/urgency of the patient's clinical condition. That said, respondents expressed frustration at the inconsistency with which certain appeals are evaluated (appeals are reviewed by a rotating board of clinicians, and acceptance criteria vary). Some respondents also expressed concern that dual-listed patients did not always have the exact same LAS in each transplant center (this observation was attributed to lack of communication/data sharing among transplant centers, rather than intentional dishonesty). Inconsistencies in LAS, while not necessarily indicative of gaming, were still viewed as having considerable impact on donor organ sharing, as even small differences in LAS can pull donor organs from one center to a different one:

> *"I think the [LAS] score is reasonable, but how people populate their lists is very variable, and my concern is that people lean on the score as being a vetted, objective, consistent measure of priority, and it's not. People will use different variables to their advantage, and listing practices are so variable that we can't assume that an LAS of 40 means the same thing at different centers. In fact, we've seen patients who go to different centers have very different lung allocation scores. And when you have variability in interpretation of how to score someone, it makes then the concept of broader regional sharing grossly unfair and vulnerable to gaming. This big push for broader regional sharing has to be predicated upon making listing behaviors entirely consistent across the country, or there will be gross iniquities manifest." [surgeon]*

The second type of gaming was conceptualized by our respondents as advocacy on behalf of a program. Selectivity when making listing decisions ensures that the program does not take on patients who are too high-risk for the transplant team to take care of (e.g., due to the technical skills of their transplant team or the amount of technological resources they have to support patients). While selectivity in listing could be viewed as "*cherry picking" [surgeon]*, most respondents felt that selectivity is sometimes necessary to preserve the integrity of the program and ensure that the

program's ratings do not fall below UNOS/CMS's acceptability thresholds. However, three respondents – all surgeons – suggested that larger programs may have enough resources to operate a *"substitute"* or back-up list in which individuals who are not currently sick enough for official listing are monitored until they become sick enough. The problem with this approach arises when the substitute list is not shared with UNOS, as then programs can adjust the size of their waitlist to manipulate their program-level transplant rate:

> *"Well, the [program-level] metrics are grossly imperfect. And the reason is that there are centers that will only list one patient who is size and blood type available in a given range, a given sort of size and blood type parameters. We don't do that. So we...if someone meets criteria, and is listable, in a practical and medically appropriate sense, they get on the list. So we run a large list, and a pared down list, and we do that to maintain a sense of connectivity and consideration of everyone who's on our list. The problem is, the metric that you're talking about is called the transplant rate, and it's not only determined by how many transplants you do, it's determined by the size of your list. So if I do a hundred transplants, and my transplant list is 100 patients long, I'm going to look like I'm less busy than someone who maintains a list of 10 patients and does 20 transplants per year. So it's… it's… It's a thing that gets often manipulated, and it's not an indication of how busy or aggressive a center is doing, but it's related to more the size of that list. It's called a gameable statistic." [surgeon]*

Waiting Period and Receipt of Transplant

The fourth major source of selection bias identified by respondents occurs between the waiting period and receipt of transplant phases in the patient pathway. Here, receipt of transplant is influenced by donor organ availability and regional competition:

> *"I think it depends on how easy access to organs your center has and how much competition too for those organs you have. Because if you're in a 250-mile radius, and you have 10 centers in that radius, then you're competing for the same organs across 10 centers. If instead, you were three centers in that radius or two or one, then you have a lot of offers for your patients. And so, you can wait until he gets sicker or you just transplant him, or you can put them on an LAS very low, and you still will have first bid to those organs" [surgeon]*

This respondent suggests that transplant centers in less competitive regions may have more flexibility in determining which patients receive which organs, because they do not

have to compete against other centers whose patients have higher LAS. A related concern is that the current allocation system encourages transplant centers to list high-risk patients, but discourages centers from accepting high-risk donor organs. Consequently, programs with primarily lower-score (lower-risk) patients are forced to accept donor organs that were turned down by the centers with higher-score (higher-risk) patients:

> "*it is unfair that you can have a very high LAS score and you could decide whatever lungs you wanna take" [surgeon]*.

While respondents admitted that they had no proof of higher-volume centers gaming the system by being overly selective with which donor organs they accept, they also pointed out that when the best-quality donor organs continually go to the highest-volume centers, *"it makes you second guess the system" [surgeon]*.

The above concerns suggest that there is a tension between improving the health of individual patients versus maintaining the health of the overall transplant program. Such tension was often described as being driven by the UNOS/CMS accountability metrics to which programs are held*:*

> *"These regulatory metrics, 30-day, one year mortality, have a very heavy impact on how surgeons and pulmonologists number one, list patients, it becomes a selective bias. And number two, how they transplant them, i.e., when they get an offer, they may be very risk adverse in certain settings or with certain donors or with any additional confounders that are encountered at the time of the* offer." *[surgeon]*

While centers can adjust their patient listing criteria and donor-organ acceptance criteria to balance patient versus program health, the successfulness of these efforts is often constrained by donor organ availability:

> *"We're handcuffed because there's no donor organs, our cohort is getting sicker and sicker, and conditional survival based on your frailty or condition at the time of transplant is compromised. [With a] precious resource that's in short supply, your cohort is likely to get sicker, and the outcomes are likely to be worse, so that the longer they sit on the list, the worse the outcomes are." [surgeon]*

To many respondents, such donor organ scarcity implied an ethical duty to select patients who are more likely to survive transplantation in order to ensure the responsible stewardship of resources:

> *"I think we are intentionally biased to transplant patients that we think are gonna survive…because it's a scarce resource. If you're thinking about providing a treatment that's a scarce resource that might compromise access to other patients, then I think it is reasonable to consider giving it to the patients where it may be more likely to be efficacious" [surgeon]*.

Thus, the tension between patient and program health can also be framed as a conflict between increasing access to care and responsible stewardship of donor organs.

Conceptual Model

Respondents identified many sources of selection bias that can arise at various points along the pathway from referral to transplant (Figure 4.1, unshaded boxes). These sources of selection bias were synthesized into five main concepts (Figure 4.1, shaded boxes): 1) transplant center's level of risk tolerance and accountability, 2) successfulness and fairness of the LAS in mitigating selection/survivor bias, 3) patient health versus program health, 4) donor organ availability and regional competition, and 5) access to care versus responsible stewardship of organs.

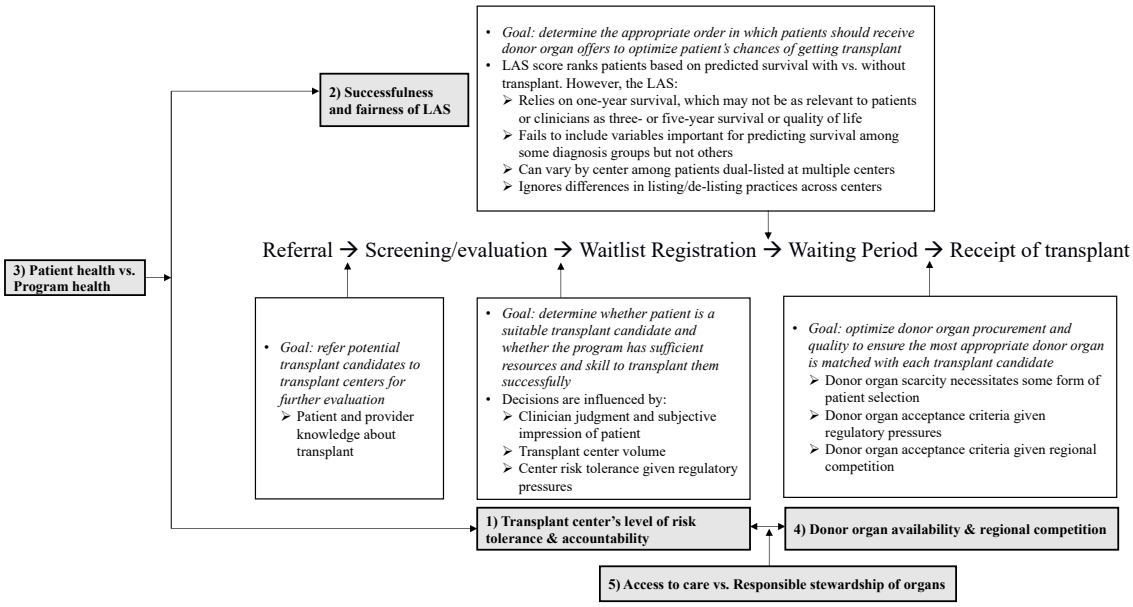


Figure 4.1. Conceptual model of sources of selection bias in lung allocation

Conceptual model depicts points at which selection bias can arise in lung transplantation and synthesizes these sources of selection bias into common themes and relationships between them.

The vast majority of these concepts were endorsed by both transplant surgeons and pulmonologists, with one exception: issues related to donor organ screening and procurement (including travel logistics and donor organ quality assessments) were primarily discussed by surgeons, not pulmonologists. That is, while the overall conceptual model holds for both pulmonologists and transplant surgeons, these two types of respondents generally approached the discussion of selection bias in lung transplantation from the perspective most aligned with their role in the transplant process. Specifically, pulmonologists – who provide longitudinal patient care to particular patients both before and after transplantation – tended to focus on *"optimizing the recipient"* so that the best patient is selected for each donor organ:

> *"I think the main thing that we're looking to do and I'm not sure it's the LAS number, the LAS number, or severity. I think we're certainly always looking to maximize...I mean, maximize isn't the right word. I think we want to be sure that the score as accurately as possible reflects the patient's severity of illness. So whether that's updating various testing parameters periodically. It's certainly an*

awareness of the patient's underlying condition and relatively aggressive assessment of changes to that condition in an effort to try to optimize the allocation score" [pulmonologist]

This perspective aligns most closely with concept (2) in the model.

Conversely, transplant surgeons – who are typically responsible for procuring donor organs in addition to performing the actual transplant procedure – tended to focus on *"optimizing the donor"* so that the best donor organ is selected for each patient:

"The lung allocation score only creates the potential offer. It doesn't control the quality of the offer. If you have a situation with a patient with a high score, who is likely to become a candidate for a number of offers, then a lung that is marginal may not be readily accepted. And instead, one would say, 'Hold out for a better organ so we have a better outcome proposition.' But what is a better organ is in the eye of the beholder. And there is some science, but no real class evidence or truly binding guidelines or regulations feasible because it remains a big black box. It's just experienced surgeons and physicians trying to extrapolate data that helps them decide what would or wouldn't be a functional organ. Because you want to avoid the high risk recipient with a marginal donor lung because the combination of makes for an extremely difficult postoperative course, in the vast majority of cases, and therefore, increases dramatically your chance of 30-day mortality or one year mortality" [surgeon]

This perspective aligns most closely with concept (3) in the model.

Discussion

Based on this semi-structured interview study, we developed a conceptual model (Figure 4.1) of selection bias in lung transplantation that contains five concepts. 1: Transplant center's level of risk tolerance and accountability. This concept captures selection bias arising between screening and waitlist registration, where clinicians' goal is to determine a) whether patients are suitable transplant candidates, and b) whether the program has sufficient resources and skill to transplant these patients successfully. 2: Successfulness/fairness of the LAS, which encompasses waitlist registration and the waiting period. Here, the goal is to determine the appropriate order in which patients should receive donor-organ offers to optimize their chances of transplantation. 3: Patient versus program health, captures the tension between clinicians' dual responsibilities of

advocating for individual patients while maintaining the health of their overall program. 4: Donor-organ availability and regional competition. This concept focuses on the transition between waiting period and transplantation, where the goal is to optimize donor-organ procurement and quality to ensure that the most appropriate donor organ is matched with each transplant candidate. 5: Access to care versus responsible stewardship of organs, which captures conflicts between “optimal” patient selection and “optimal” donor-organ selection.

While most respondents recognized that the LAS is partially successful in mitigating survivor bias at the point of organ allocation, many also suggested that the LAS ignores upstream sources of selection bias (e.g., listing/de-listing) that influence which patients are registered on the waitlist and remain active candidates. Existing literature on upstream sources of selection bias tends to frame the problem as “access to care” and not also an issue of “selection bias”. For example, disparities in waitlist registration are often attributed to inequities in referrals without consideration of inconsistencies in screening or listing decisions (which also creates inequities). Although referral and listing guidelines exist (Weill et al., 2015), recent research has advocated for earlier referral or multiple listing of specific patient populations (e.g., cystic fibrosis) (Mooney, Yang, et al., 2019; Ramos et al., 2019; Stephenson et al., 2021). Others have called for “standardized recognition and reporting of factors (e.g., frailty, social support, quality of life) which are relevant to lung transplant patient selection but which are not included in the formal evaluation process” (Blumenthal et al., 2017; Ladin et al., 2019; Schaenman et al., 2020). Our respondents suggest that the extent to which these criteria matter during listing decisions depends on each center’s transplant volume and level of risk tolerance.

Initial research surrounding the LAS examined its impact on pre- and post-transplant survival (Egan & Edwards, 2016; Maxwell et al., 2014; Russo et al., 2011).

Yet the LAS, by itself, simply prioritizes waitlisted patients; it does not fully address selection bias occurring at other points along the transplantation pathway. Consequently, transplant centers must manage their waitlist through other means (e.g., listing criteria), while considering potential tradeoffs between patient and program health. For example, relaxing listing criteria can provide higher-risk patients with an opportunity for transplant, but may also necessitate greater patient management and/or donor-organ selectivity to maximize patients' chances of surviving surgery and minimize their risk of post-transplant complications (e.g., graft rejection) (Hart & Engels, 2021). Similarly, while listing less-urgent patients lessens program-level risk, it may also hinder transplant access for higher-risk patients, who might not have enough social/financial resources to be listed at another center (Wagener, 2020).

Such findings are consistent with, and add significantly to, research on decision-making in organ transplant selection committees (Blumenthal et al., 2017; Volk et al., 2011). Prior studies included few transplant surgeons and pulmonologists; focused solely on candidate screening rather than the complete pathway from referral through transplantation; did not examine how patients' LAS might influence candidate selection; and only considered patient-level factors – not program-level factors – that might influence decision-making (Blumenthal et al., 2017; Volk et al., 2011). Our study addresses these limitations.

Although statistical methods have been developed to address censoring among waitlisted patients (e.g., due to death, transplantation, etc.) (Schnellinger et al., 2021; Vock et al., 2017; Vock et al., 2013; Xiang & Murray, 2012; Xiang et al., 2014), few studies have investigated disparities in de-listing, despite the fact that de-listing can influence patients' chances of transplantation and their long-term health (Rudasill et al., 2019). Our respondents suggest that clinicians employ careful waitlist management strategies – e.g., "substitute" waitlists –to maintain program health.

Research centered on receipt of transplant primarily focuses on geographic disparities in donor lung availability (Benvenuto et al., 2018; Drolen et al., 2020; Ross-Driscoll et al., 2020) and the role organ procurement organizations (OPOs) have in increasing donor-organ supply through broader organ sharing (Mooney, Bhattacharya, et al., 2019) and better procurement practices (Doby et al., 2020; Halpern et al., 2021). Yet transplant programs ultimately have the final say over which donor organs they accept, with some programs being more willing to accept marginal-quality/high-risk organs (Levitsky et al., 2017; Van Pilsum Rasmussen et al., 2020; Whitford et al., 2020) than others (Loss et al., 2013; Van Pilsum Rasmussen et al., 2019). Research on pancreas donors in Germany concluded that the evaluation of donor organs was “highly inconsistent” across transplant centers and that both “very cautious” and “very permissive” acceptance criteria “may render the allocation process less efficient” (Loss et al., 2013). For example, accepting marginal-quality donor organs can enable more patients to receive transplant, but may come at the cost of more post-transplant complications or reduced post-transplant survival (Schwarz et al., 2020). Such conflicts between access to care and responsible stewardship of organs are consistent with the findings of (Volk et al., 2011), and have both practical and ethical implications (Persad et al., 2009; Veatch & Ross, 2015). Thus, understanding how clinicians decide which donor organs to accept – and making this source of selection bias more transparent throughout the transplant community – is just as important as increasing donor-organ supply.

Our study has several strengths. Rigorous qualitative analyses allowed us to capture a diversity of perspectives on selection bias in lung transplantation and the LAS. Including transplant surgeons and pulmonologists from transplant centers throughout the U.S. strengthens the robustness of our conceptual model, and allowed us to understand how center volume and location may shape screening, waitlisting, and transplantation decisions. Finally, considering the entire pathway from referral through transplantation

enabled us to build upon existing literature and examine how patient- and program-level factors influence selection bias in lung transplantation.

Our study also has limitations. First, the coronavirus pandemic necessitated switching from in-person to phone interviews midway through data collection. This change did not impact interview quality, and actually increased surgeons' participation rate. Second, the qualitative design is limited in its ability to generalize to the concerns of all transplant providers. However, we purposively sampled by discipline (transplant surgeon, pulmonologist) and sought to include respondents with diverse characteristics relevant to our study question (e.g., geography, years in practice). Third, although our response rate was low we were able to obtain thematic saturation both overall and within subgroups defined by discipline. Finally, although 84% of respondents were male, this proportion is consistent with the distribution of men in the transplant workforce (Cooke et al., 2019; Erhunmwunsee et al., 2019; Valbuena et al., 2021).

Overall, our study demonstrates how selection bias can arise throughout lung transplantation, and the proposed conceptual model can facilitate further study of such bias. Although this conceptual model was developed among lung transplant surgeons and pulmonologists, it may be applicable to other organ allocation systems that aim to determine which patients to register for transplant and how to prioritize them. With respect to lung transplantation, OPTN is currently developing a new allocation framework, the "continuous distribution model" (Kasiske et al., 2020; Snyder et al., 2018; U.S. Department of Health & Human Services), that aims to prioritize patients via composite scores consisting of the LAS alongside other patient attributes, such as candidate biology, patient access, and placement efficiency (U.S. Department of Health & Human Services). This approach is a first step toward recognizing the complex nature of organ allocation, which requires "high performance, alignment, and accountability from donor hospitals, OPOs, and transplant programs" (Hauerwaas & Weisenfeld, 2020;

O'Connor & Glazier, 2021). However, our findings suggest that the other attributes included in the composite score may be susceptible to selection bias as well. Thus, the continuous distribution model should be monitored closely during development and implementation to ensure that the resulting allocation scheme is more equitable and does not inadvertently exacerbate selection bias in transplantation.

CHAPTER 5. CONCLUSIONS

Summary of Findings

In Chapter 2 we developed a modified LAS score using inverse probability weighting to mitigate selection bias in both the pre- and post-transplant survival models, and demonstrated improved predictive performance of the modified LAS relative to the existing one. Our findings suggest that the additional variables incorporated into our weights (e.g., geography) may be important to consider when estimating pre- and post-transplant survival. Additionally, selection bias appeared to have a larger impact on the estimate of waitlist urgency than on the estimate of post-transplant survival. Finally, although the estimate of post-transplant survival under the modified LAS was generally the same or greater than that under the existing LAS, the estimate of pre-transplant survival also tends to be longer under the modified LAS. Consequently, a sizable number of patients with intermediate scores under the existing LAS would receive lower scores under the modified LAS.

In Chapter 3, we explored the demographic and clinical characteristics of patients who receive better or worse priority under the modified LAS compared to the existing LAS and investigated how the modified LAS might impact pre- and post-transplant survival if it were implemented in clinical practice. Our findings suggest that changes in prioritization were more pronounced for certain demographic and clinical characteristics, such as diagnosis group, six-minute walk distance, continuous mechanical ventilation, functional status, and age. Moreover, changes in predicted pre-transplant survival tended to explain a greater proportion of variability in the differences in rank than did changes in post-transplant survival. Finally, our statistical simulation study suggests that one-year waitlist survival may improve under the modified LAS, while one-year post-transplant and overall survival remain comparable to that under the existing LAS.

In Chapter 4, we conducted a qualitative interview study of lung transplant surgeons and pulmonologists throughout the United States to understand how they perceive the LAS and selection bias. Our findings suggest that selection bias can arise at several points along the transplantation pathway, and that these sources of selection bias can be synthesized into five factors: 1) transplant center's level of risk tolerance and accountability, 2) successfulness and fairness of the LAS in mitigating selection bias, 3) donor organ availability and regional competition, 4) patient health versus program health, and 5) access to care versus responsible stewardship of organs.

Benefit of Incorporating Quantitative and Qualitative Objectives

While we presented our quantitative and qualitative objectives sequentially, they very much worked in parallel, informing each other. For example, findings from our qualitative study encouraged us to expand the focus of our quantitative study from survivor bias specifically to selection bias more generally. Similarly, our quantitative finding that selection bias had a greater impact on waitlist urgency than on post-transplant survival is consistent with some interview respondents' comments suggesting that upstream sources of selection bias (e.g., referral and listing) may be more clinically relevant than downstream ones (e.g., donor organ acceptance criteria). Thus, by incorporating both quantitative and qualitative research into this dissertation, we were able to develop a more comprehensive and nuanced understanding of selection bias and its impact in lung allocation.

Selection versus Selection-Bias

Throughout our interviews, some respondents raised concerns about the phrases "selection bias" and "survivor bias", pointing out that some form of patient selection is often necessary in the face of donor organ scarcity and accountability-promoting

regulatory pressures. Since a considerable component of patient selection is based on subjective clinical judgment, some respondents emphasized the importance of including a wide variety of perspectives in their selection committees to ensure that patient selection occurs in a fair manner. Other respondents acknowledged the enormity of the decisions being made in the presence of imperfect and incomplete information. Taken together, these comments highlight how the epidemiologic notion of selection bias (and its removal) does not always agree with the clinical notion of selection (and its necessity). These differences in perspectives on "good" versus "bad" selection bias will be important to consider when designing and developing new allocation frameworks, such as the continuous distribution model (Kasiske et al., 2020; Snyder et al., 2018; U.S. Department of Health & Human Services).

Limitations

The studies undertaken in this dissertation are not without limitations. First, when developing our proposed modified LAS (Chapter 2), insufficient information was available to consistently distinguish between patients who were newly listed and those who were re-activated after temporary waitlist removal. Thus, the first record associated with each identification number was taken to be the initial registration date, follow-up time was counted from that date forward, and individuals who were subsequently lost to follow-up were censored at that time. Given that only about 1.25% of registered individuals in our dataset were lost to follow-up prior to transplantation, this analytic decision is unlikely to influence our results. However, future studies could consider including variables indicating whether patients are currently active or inactive on the waitlist, how many times they transition to and from the active and inactive states, and the frequency of their visits to the transplant center to better capture their waitlist history and disease trajectory.

Another limitation of our modified LAS is that we developed it after excluding individuals on the waitlist who were highly unlikely to receive transplant due to certain patient characteristics (e.g., high sensitization or small stature) that prevent them from finding a suitable donor organ match. This exclusion was necessary to ensure that we fulfill the positivity assumption of IPTW & IPCW weighting – that is, to ensure that all individuals have at least some non-zero probability of receiving transplant. However, it also implies that we cannot generalize our findings to patients who have these clinical contraindications. Although the proportion of individuals excluded under this criteria was small (i.e., 1.79%) and unlikely to influence our results, future work should explore how we might mitigate selection bias among patients with these clinical contraindications. For example, we could build off of recent research on population decomposition by decomposing our study population into several groups: 1) those who survive and receive transplant, 2) those who die prior to transplant, and 3) those who survive indefinitely without receiving transplant (i.e., the individuals with clinical contraindications who were excluded from our study); then, we could fit a separate regression model among each subgroup and re-weight our outcome models based on the inverse of the joint probability of each subgroup-specific regression model (Haneuse & Daniels, 2016).

A third limitation of our quantitative study is that we cannot account for the fact that presence in the UNOS database is not random, but rather indicates that the patient was ill enough to visit the hospital, undergo evaluation for transplant, and be registered on the waitlist. Consequently, selection bias associated with waitlist registration can still affect predicted mortality. Future work should strive to collect referral data so that we can mitigate this additional source of selection bias.

Our assessment of the clinical impact of the modified LAS versus the existing LAS (Chapter 3) has some limitations as well. First, the generalized linear mixed model (GLMM) approach we took to estimate the proportion of variability (r^2) in the differences

in rank explained by changes in predicted pre- versus post-transplant survival does not account for the correlation among differences in rank. We mitigated this issue by splitting the data into worse-priority and better-priority subsets, and fitting separate GLMMs to each subset. However, future methodologic research is needed to extend generalized r^2 statistics to the situation where the outcome of interest is differences in ranks.

Second, while we used the observed testing cohort data to compare the order in which patients were prioritized under the two models and examine the clinical characteristics of patients who received better or worse priority under the modified LAS relative to the existing LAS, we cannot know whether patients who died on the waiting list under the existing LAS would, in fact, have received transplant had the modified LAS been used instead. This limitation motivated us to conduct the pilot simulation study described in Chapter 3, where we examined how patients' waitlist and post-transplant survival might change if the modified LAS were implemented in clinical practice. However, this simulation was limited in scope, assuming an artificially simple allocation policy based on blood type and region only (i.e., ignoring other salient factors to organ matching, such as donor/recipient size and sensitization, donor organ quality and acceptance criteria, and travel constraints that arise when delivering donor organs to transplant candidates). Future work should aim to incorporate these additional factors into the simulation to see how the modified LAS might perform in more realistic settings.

With regards to our qualitative study (Chapter 4), the conceptual model we developed aims to describe how and when selection bias can arise in transplantation and the LAS. Although we sought to capture a diversity of perspectives on the LAS by sampling purposively by discipline (lung transplant surgeon, pulmonologist) and by including respondents from a variety of geographic regions with different amounts of experience (years in practice), we cannot generalize our findings to the concerns of all transplant providers. To the extent that transplant decisions are not made solely by lung

transplant surgeons and pulmonologists, future studies should consider interviewing patients or other providers (e.g., referring physicians, transplant nurses, and social workers) to understand their perspectives on selection bias in transplantation. Such an approach could inform the design of interventions aimed at improving referral rates and decreasing loss to follow-up.

Future Directions

The findings from our studies highlight the importance of recognizing and addressing selection bias throughout the transplantation process and are applicable to other organ allocation systems which rely on estimates of pre- and/or post-transplant survival to prioritize patients, such as the United States' Estimated Post-Transplant Survival (EPTS) score for kidney allocation (Egan et al., 2006; Organ Procurement and Transplantation Network, 2020; Veatch & Ross, 2015), Germany's lung allocation score (Gottlieb, 2017; Gottlieb et al., 2014), and the United Kingdom's Liver Transplant Benefit Score (NHS Blood and Transplant). Findings from our study could also inform the development of pre- and post-transplant survival models for organs which do not yet use such models to guide allocation decisions. This step will become important as other organs begin to adopt the continuous distribution allocation framework (Kasiske et al., 2020; Snyder et al., 2018; U.S. Department of Health & Human Services). Assessments of time-dependent discrimination and calibration for both the existing LAS and the modified LAS (Chapter 2) are relevant here as well, especially as the transplant community considers what role longer-term survival should play in organ allocation (Kasiske et al., 2020; Maxwell et al., 2014; U.S. Department of Health & Human Services).

Our pilot simulation (Chapter 3) provides preliminary evidence of how waitlist and post-transplant survival might change if the modified LAS were implemented in clinical

practice. However, as mentioned above, this simulation was limited in scope, ignoring many factors that are relevant to organ allocation and donor-recipient matching. In addition to incorporating these factors into the simulation, future work should also aim to provide more flexibility in terms of which simulation parameters can be varied (e.g., allowing the user to specify alternative scoring systems or geographic sharing rules). This approach would enable researchers to test a variety of allocation policy changes and evaluate their impact on a hypothetical (simulated) population before deciding whether to implement such policies in clinical practice.

Our conceptual model (Chapter 4) can also inform the development of new allocation systems, such as the continuous distribution model. Specific attention should be paid to sources of selection bias outside of the LAS, such as referral, listing/de-listing decisions, and donor organ acceptance criteria. These "external" sources are especially important to consider in the context of the continuous distribution model, which aims to combine the components of the LAS (i.e., pre- and post-transplant survival) with other attributes – such as candidate biology (e.g., blood type, sensitization, height), patient access (e.g., increase access for individuals under age 18 and for prior living donors), and placement efficiency (e.g., distance from donor hospital and travel efficiency) – in a weighted average composite score (Kasiske et al., 2020; Snyder et al., 2018; U.S. Department of Health & Human Services). While the goal of this composite score is to minimize situations where certain individuals are given absolute priority over other "clinically similar" individuals (U.S. Department of Health & Human Services) (e.g., because they happen to have a particular clinical characteristic which drives their LAS score), findings from our study suggest that the other attributes included in the new composite score can be susceptible to selection bias as well. Thus, the development, implementation, and performance of the continuous distribution model should be

monitored closely to ensure that it does not inadvertently exacerbate selection bias in transplantation.

Concluding Remarks

Collectively, the findings from this dissertation suggest that selection bias is important to address throughout the transplantation process. We hope that these findings, as well as our recommendations for future research, can inform future revisions of the LAS and other prediction models in organ transplantation to improve prediction of pre- and post-transplant survival, mitigate selection biases at other points in the transplantation pathway, and ensure more equitable allocation of donor organs.

APPENDIX

Appendix 2.1. Weight Truncation

To minimize the impact of extreme weights on our outcome models, we progressively truncated the final pre- and post-transplant weights following the procedure outlined in (Cole & Hernán, 2008). More specifically, we replaced weights that were below (above) a certain percentile with the value at that percentile. By progressively increasing (decreasing) the percentiles from 0% (100%) to 5% (95%), we could explore the trade-off between bias and variance. Appendix Table 2.1 displays the distribution of the pre-transplant weights, alternatively-stabilized pre-transplant weights (i.e., those used to construct the post-transplant weights; see main text for details), and post-transplant weights under various truncation percentiles. As the extent of truncation increases, the mean of the weights moves further away from one, indicating a greater degree of bias; conversely, the variance of the weights moves closer to zero, suggesting greater precision. Ultimately, the final weights were truncated at the 0.25% (99.75%) percentiles, as these were the ones which centered the weights around one and which reduced the 1/minimum and maximum weights by at least one order of magnitude (Cole & Hernán, 2008).

Appendix Table 2.1. Distribution of pre- and post-transplant weights under various truncation percentiles

The distribution of the final weights appears in bold.

Appendix Table 2.1A) Distribution of pre-transplant weights

Truncation Percentiles	Mean	Variance	Minimum	Maximum
0, 100	1.127	26.91	1.28E-09	551.0
0.25, 99.75	**1.034**	**0.817**	**0.001**	**10.29**
0.5, 99.5	1.024	0.666	0.003	7.386
1, 99	1.006	0.485	0.010	4.832
2.5, 97.5	0.976	0.316	0.042	2.902
3.5, 96.5	0.962	0.265	0.069	2.396
5, 95	0.948	0.223	0.113	2.025

Appendix Table 2.1B) Distribution of alternatively-stabilized pre-transplant weights

Truncation Percentiles	Mean	Variance	Minimum	Maximum
0, 100	1.148	530.7	9.81E-05	1.01E+04
0.25, 99.75	**0.959**	**1.492**	**0.001**	**14.88**
0.5, 99.5	0.941	1.080	0.001	9.656
1, 99	0.915	0.729	0.003	6.020
2.5, 97.5	0.871	0.427	0.011	3.070
3.5, 96.5	0.856	0.370	0.017	2.572
5, 95	0.839	0.316	0.028	2.147

Appendix Table 2.1C) Distribution of post-transplant weights

Truncation Percentiles	Mean	Variance	Minimum	Maximum
0, 100	1.175	75.33	1.50E-04	599.8
0.25, 99.75	**0.962**	**1.625**	**0.009**	**17.09**
0.5, 99.5	0.936	0.998	0.023	9.531
1, 99	0.905	0.621	0.033	4.948
2.5, 97.5	0.875	0.445	0.059	2.969
3.5, 96.5	0.862	0.394	0.077	2.503
5, 95	0.847	0.345	0.103	2.099

Appendix 2.2. Baseline Survival Probabilities for Modified Outcome Models

In this section, we display the baseline survival probabilities (Appendix Table 2.2) obtained from the modified (weighted) pre- and post-transplant outcome models. These survival probabilities were used in conjunction with the parameter estimates shown in Table 2.3 in Chapter 2 to construct modified LAS scores for patients in our testing cohort.

Appendix Table 2.2. Baseline survival probabilities for the modified pre- and post-transplant LAS

Appendix Table 2.2A) Baseline survival probabilities for the modified pre-transplant outcome model

Time (days)	Waitlist Survival
0	1.00000000
1	0.99949890
2	0.99861080
3	0.99847020
4	0.99824380
5	0.99809990
6	0.99787550
7	0.99771830
8	0.99762050
9	0.99752650
10	0.99738530
11	0.99721320
12	0.99715210
13	0.99700950
14	0.99690890
15	0.99680770
16	0.99673020
17	0.99669440
18	0.99663090
19	0.99651170
20	0.99637040
21	0.99633550
22	0.99630000
23	0.99622360
24	0.99618820
25	0.99607970
26	0.99606270
27	0.99595380
28	0.99588110
29	0.99580880
30	0.99580880
31	0.99575660
32	0.99575660
33	0.99555590
34	0.99543440
35	0.99535690
36	0.99529670
37	0.99504970
38	0.99498210
39	0.99482720
40	0.99478720
41	0.99470050
42	0.99467110
43	0.99464450
44	0.99457450
45	0.99455430
46	0.99449440
47	0.99449440
48	0.99439620
49	0.99436790
50	0.99429490
51	0.99427350
52	0.99413210
53	0.99409040
54	0.99397610
55	0.99387900
56	0.99387900
57	0.99361910
58	0.99361910
59	0.99353570
60	0.99351250
61	0.99343780
62	0.99334730
63	0.99327860
64	0.99303570
65	0.99298630
66	0.99294020
67	0.99294020
68	0.99278280
69	0.99270620
70	0.99225440
71	0.99217070
72	0.99211770
73	0.99205410
74	0.99195170
75	0.99192470
76	0.99188070
77	0.99188070
78	0.99183380
79	0.99180110
80	0.99174440
81	0.99174440
82	0.99171980
83	0.99169260
84	0.99169260
85	0.99165900
86	0.99161170
87	0.99157360
88	0.99144450
89	0.99127280
90	0.99127280
91	0.99127280
92	0.99123600
93	0.99105800
94	0.99097330
95	0.99069050
96	0.99065850
97	0.99057640
98	0.99057640
99	0.99057640
100	0.99057640
101	0.99057640
102	0.99050870
103	0.99045280
104	0.99045280
105	0.99042720
106	0.99032910
107	0.99032910
108	0.99032910
109	0.99032910
110	0.99028200
111	0.99028200
112	0.99025570
113	0.99022540
114	0.99015130
115	0.99009470
116	0.99009470
117	0.99009470
118	0.99009470
119	0.99004220
120	0.98999600
121	0.98999600
122	0.98990010
123	0.98990010
124	0.98987070
125	0.98987070
126	0.98979010
127	0.98975310
128	0.98970750
129	0.98957860
130	0.98945660
131	0.98945660
132	0.98945660
133	0.98938810
134	0.98927390
135	0.98927390
136	0.98924580
137	0.98924580
138	0.98924580
139	0.98901980
140	0.98898180
141	0.98898180
142	0.98891120
143	0.98881060
144	0.98881060
145	0.98881060
146	0.98862540
147	0.98862540
148	0.98859800
149	0.98859800
150	0.98855410
151	0.98855410
152	0.98850870
153	0.98846440
154	0.98846440
155	0.98838650
156	0.98833870
157	0.98833870
158	0.98833870
159	0.98828630
160	0.98828630
161	0.98810580
162	0.98810580
163	0.98807000
164	0.98807000
165	0.98804610
166	0.98799710
167	0.98799710
168	0.98799710
169	0.98799710
170	0.98792760
171	0.98789130
172	0.98784470
173	0.98784470
174	0.98778040
175	0.98778040
176	0.98762660
177	0.98752250
178	0.98752250
179	0.98752250
180	0.98752250

181	0.98747760
182	0.98686000
183	0.98643590
184	0.98606400
185	0.98558710
186	0.98555720
187	0.98547530
188	0.98542620
189	0.98535930
190	0.98530210
191	0.98530210
192	0.98530210
193	0.98517610
194	0.98517610
195	0.98517610
196	0.98517610
197	0.98498220
198	0.98489450
199	0.98480400
200	0.98465610
201	0.98465610
202	0.98465610
203	0.98457030
204	0.98457030
205	0.98457030
206	0.98452710
207	0.98444500
208	0.98444500
209	0.98440380
210	0.98440380
211	0.98440380
212	0.98440380
213	0.98440380
214	0.98422070
215	0.98422070
216	0.98406430
217	0.98406430
218	0.98406430
219	0.98398360
220	0.98398360
221	0.98398360
222	0.98398360
223	0.98398360
224	0.98392360
225	0.98384980
226	0.98378170
227	0.98378170
228	0.98378170
229	0.98353850
230	0.98353850
231	0.98329040
232	0.98317210
233	0.98310490
234	0.98304650
235	0.98304650
236	0.98304650
237	0.98304650
238	0.98300730
239	0.98295540
240	0.98242220
241	0.98236150
242	0.98229290
243	0.98229290
244	0.98229290
245	0.98229290
246	0.98229290
247	0.98229290
248	0.98217620
249	0.98217620
250	0.98217620
251	0.98202600
252	0.98202600
253	0.98202600
254	0.98202600
255	0.98202600
256	0.98195530
257	0.98185160
258	0.98185160
259	0.98176630
260	0.98176630
261	0.98139880
262	0.98131640
263	0.98125940
264	0.98125940
265	0.98125940
266	0.98104180
267	0.98090980
268	0.98080010
269	0.98080010
270	0.98039670
271	0.98032500
272	0.98032500
273	0.98032500
274	0.98032500
275	0.98032500
276	0.98021850
277	0.98001860
278	0.98001860
279	0.97993420
280	0.97993420
281	0.97985390
282	0.97985390
283	0.97985390
284	0.97985390
285	0.97985390
286	0.97985390
287	0.97985390
288	0.97976220
289	0.97976220
290	0.97976220
291	0.97976220
292	0.97975350
293	0.97975350
294	0.97963670
295	0.97954550
296	0.97954550
297	0.97954550
298	0.97939460
299	0.97930350
300	0.97930350
301	0.97930350
302	0.97920520
303	0.97912270
304	0.97912270
305	0.97912270
306	0.97908000
307	0.97908000
308	0.97908000
309	0.97908000
310	0.97908000
311	0.97896840
312	0.97896840
313	0.97896840
314	0.97896840
315	0.97896840
316	0.97896840
317	0.97894650
318	0.97894650
319	0.97881190
320	0.97881190
321	0.97874950
322	0.97874950
323	0.97874950
324	0.97874950
325	0.97874950
326	0.97874950
327	0.97874950
328	0.97874950
329	0.97874950
330	0.97874950
331	0.97874950
332	0.97874950
333	0.97868070
334	0.97868070
335	0.97868070
336	0.97868070
337	0.97854430
338	0.97848880
339	0.97848880
340	0.97807370
341	0.97807370
342	0.97807370
343	0.97801020
344	0.97766270
345	0.97766270
346	0.97766270
347	0.97766270
348	0.97761010
349	0.97761010
350	0.97761010
351	0.97761010
352	0.97758430
353	0.97737180
354	0.97737180
355	0.97737180
356	0.97721110
357	0.97721110
358	0.97721110
359	0.97721110
360	0.97705880
361	0.97705880
362	0.97705880
363	0.97705880
364	0.97705880

Appendix Table 2.2B) Baseline survival probabilities for the modified post-transplant outcome model

Time (days)	Post-transplant Survival
0	0.998379000
1	0.996209000
2	0.995379400
3	0.994953900
4	0.993997900
5	0.992298200
6	0.991685700
7	0.991396500
8	0.991039700
9	0.990595300
10	0.989299400
11	0.988702900
12	0.987841900
13	0.987356600
14	0.987100000
15	0.986487900
16	0.986323600
17	0.986016400
18	0.985647500
19	0.985276200
20	0.984772100
21	0.984616200
22	0.984040200
23	0.983644800
24	0.982689800
25	0.981989400
26	0.980997600
27	0.980587300
28	0.980077000
29	0.979670700
30	0.979599900
31	0.979448300
32	0.979286400
33	0.979130000
34	0.978923600
35	0.978855600
36	0.978692100
37	0.978400000
38	0.978172400
39	0.977904100
40	0.977776900
41	0.977734100
42	0.977639500
43	0.977488500
44	0.977392400
45	0.977168300
46	0.977003700
47	0.976958500
48	0.976732100
49	0.976600200
50	0.976471600
51	0.976109700
52	0.976018100
53	0.975722700
54	0.974344400
55	0.973900600
56	0.973895700
57	0.973762800
58	0.973460100
59	0.973216100
60	0.973103000
61	0.973052600
62	0.972896600
63	0.972679700
64	0.972528200
65	0.972342200
66	0.972271300
67	0.972008500
68	0.971873800
69	0.971872000
70	0.971683500
71	0.971650800
72	0.970162600
73	0.970044900
74	0.969789700
75	0.969757400
76	0.969443500
77	0.969000000
78	0.968453000
79	0.967538600
80	0.967524800
81	0.967168700
82	0.966833000
83	0.966816600
84	0.966757400
85	0.966510500
86	0.966219100
87	0.965916700
88	0.965883700
89	0.965768100
90	0.965748000
91	0.965563800
92	0.965503500
93	0.965338700
94	0.965007600
95	0.964726900
96	0.964699200
97	0.964074500
98	0.963914900
99	0.963844300
100	0.963640400
101	0.963522400
102	0.963408300
103	0.963045300
104	0.962898400
105	0.962898400
106	0.962100800
107	0.962046800
108	0.962046800
109	0.961838900
110	0.961772900
111	0.961755900
112	0.961686100
113	0.961529700
114	0.961294400
115	0.961167500
116	0.961068800
117	0.961046200
118	0.960706400
119	0.960585500
120	0.960303000
121	0.960256100
122	0.960174400
123	0.959945900
124	0.959896700
125	0.959729300
126	0.959527900
127	0.959465300
128	0.958906800
129	0.958779200
130	0.957531500
131	0.957405300
132	0.957405300
133	0.957353700
134	0.957239700
135	0.957239700
136	0.957233100
137	0.957141500
138	0.957141500
139	0.957070300
140	0.956980500
141	0.956749500
142	0.956740300
143	0.956274000
144	0.956183900
145	0.956071300
146	0.955805500
147	0.955735200
148	0.955626400
149	0.955425500
150	0.955280800
151	0.955147300
152	0.955049300
153	0.954756700
154	0.954693900
155	0.954568900
156	0.954542000
157	0.954471600
158	0.954321000
159	0.954193400
160	0.954054100
161	0.954054100
162	0.953888700
163	0.953675300
164	0.953573600
165	0.953506300
166	0.953472300
167	0.953452700
168	0.953325100
169	0.953237900
170	0.953237900
171	0.952996100
172	0.952936600
173	0.952710200
174	0.952282200
175	0.952176300
176	0.952116600
177	0.952017100
178	0.951952000
179	0.951952000
180	0.951952000
181	0.951950000
182	0.951911600
183	0.951601600
184	0.951466600
185	0.951463200
186	0.951400900
187	0.951214000
188	0.951092900
189	0.950993600

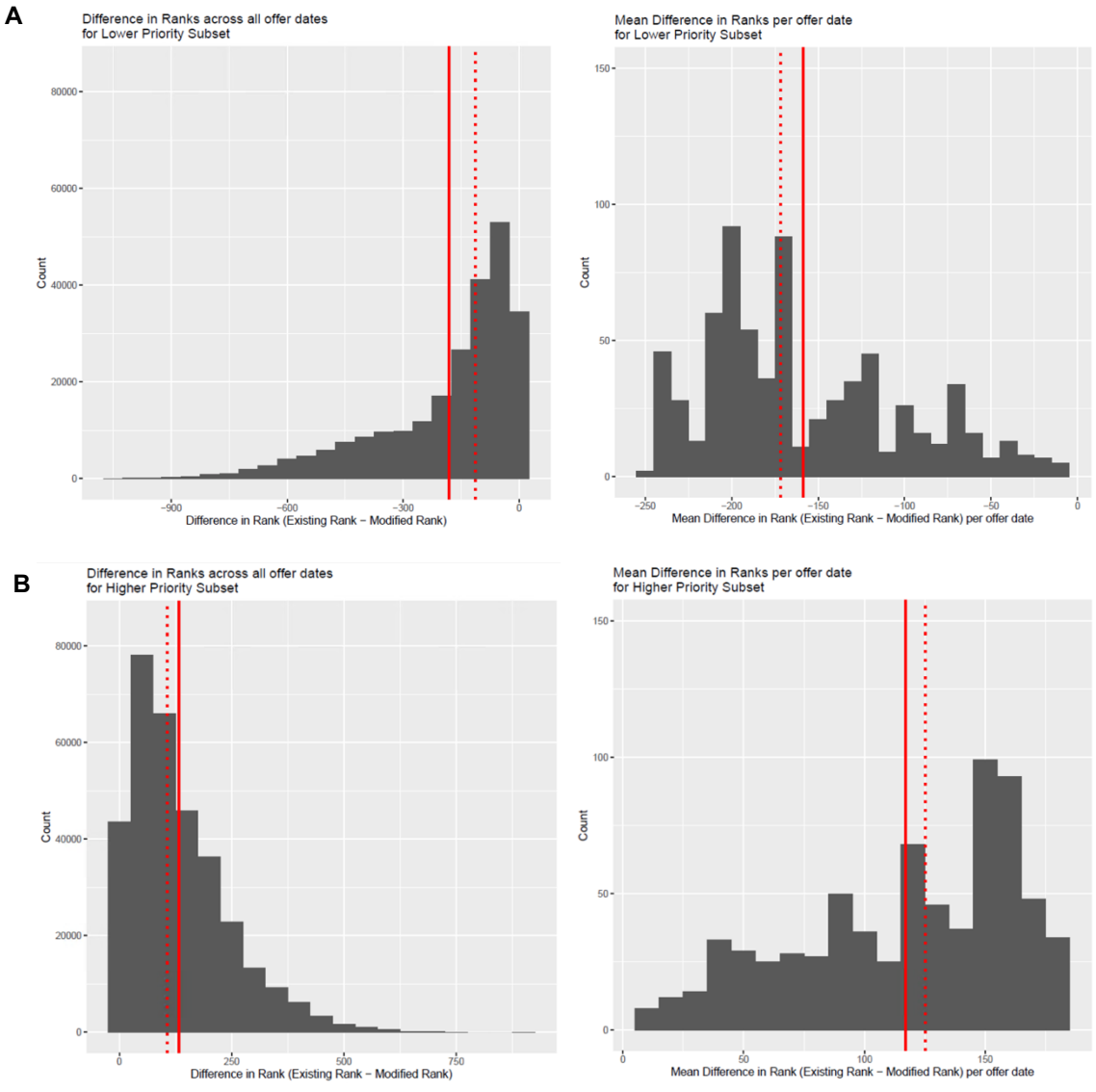
190	0.950847200
191	0.950584600
192	0.950504200
193	0.950463300
194	0.950147200
195	0.950147200
196	0.949776700
197	0.949665600
198	0.949574900
199	0.949529500
200	0.949529500
201	0.949463900
202	0.949422900
203	0.949243500
204	0.949027900
205	0.948980500
206	0.948826800
207	0.948815700
208	0.948689100
209	0.948545700
210	0.948230000
211	0.948183900
212	0.948036900
213	0.947931400
214	0.947726400
215	0.947523600
216	0.947122800
217	0.946956200
218	0.946857200
219	0.946711000
220	0.946613100
221	0.946086500
222	0.945987800
223	0.945961600
224	0.945961600
225	0.945942200
226	0.945703000
227	0.945703000
228	0.945670900
229	0.945435400
230	0.945435400
231	0.945277600
232	0.945208200
233	0.945014400
234	0.944705400
235	0.944600100
236	0.944469400
237	0.944377700
238	0.944296600
239	0.944296600
240	0.943992600
241	0.943724900
242	0.943714900
243	0.943588400
244	0.943588400
245	0.943468900
246	0.943468900
247	0.943394300
248	0.943310700
249	0.943023900
250	0.942614300
251	0.942614300
252	0.942598800
253	0.942598800
254	0.942598800
255	0.942313600
256	0.942255900
257	0.942201500
258	0.942201500
259	0.942201500
260	0.941973900
261	0.941789600
262	0.941781600
263	0.941551900
264	0.941510100
265	0.941169200
266	0.941169200
267	0.940625100
268	0.940375000
269	0.940290800
270	0.940290800
271	0.940124700
272	0.939783100
273	0.939713600
274	0.939536700
275	0.939360300
276	0.939360300
277	0.939224500
278	0.939219000
279	0.939219000
280	0.939210000
281	0.938749000
282	0.938478000
283	0.938264200
284	0.937949200
285	0.937949200
286	0.937933700
287	0.937672600
288	0.937491100
289	0.937464300
290	0.937374400
291	0.937180800
292	0.937130700
293	0.936952600
294	0.936940400
295	0.936704100
296	0.936677300
297	0.936576600
298	0.936576600
299	0.936544200
300	0.936416600
301	0.936385300
302	0.936285700
303	0.936217100
304	0.936217100
305	0.936097700
306	0.936017000
307	0.935961000
308	0.935924900
309	0.935584800
310	0.935510100
311	0.935276100
312	0.935252500
313	0.935242800
314	0.935242800
315	0.935234600
316	0.935234600
317	0.935234600
318	0.935081600
319	0.935081600
320	0.935081600
321	0.934945400
322	0.934927100
323	0.934898800
324	0.934809200
325	0.934624700
326	0.934624700
327	0.934587700
328	0.934535600
329	0.934449000
330	0.934449000
331	0.934414200
332	0.934267800
333	0.933950700
334	0.933911100
335	0.933706200
336	0.933689000
337	0.933689000
338	0.933604400
339	0.933426600
340	0.933280600
341	0.933167000
342	0.932964300
343	0.932928200
344	0.932928200
345	0.932928200
346	0.932729000
347	0.932729000
348	0.932472400
349	0.932472400
350	0.932273800
351	0.932266900
352	0.932266900
353	0.932230800
354	0.932166200
355	0.932071200
356	0.931930800
357	0.931914700
358	0.931643900
359	0.931368300
360	0.931017800
361	0.931009700
362	0.931009700
363	0.930992900
364	0.930986600

Appendix 3.1. Histograms of differences in rank

Here, we display histograms of the difference in patients' rank under the modified versus existing LAS for individuals receiving worse or better priority (Appendix Figure 3.1). Both cumulative change in rank and average change in rank per day are shown.

Appendix Figure 3.1. Histograms of differences in rank (overall)

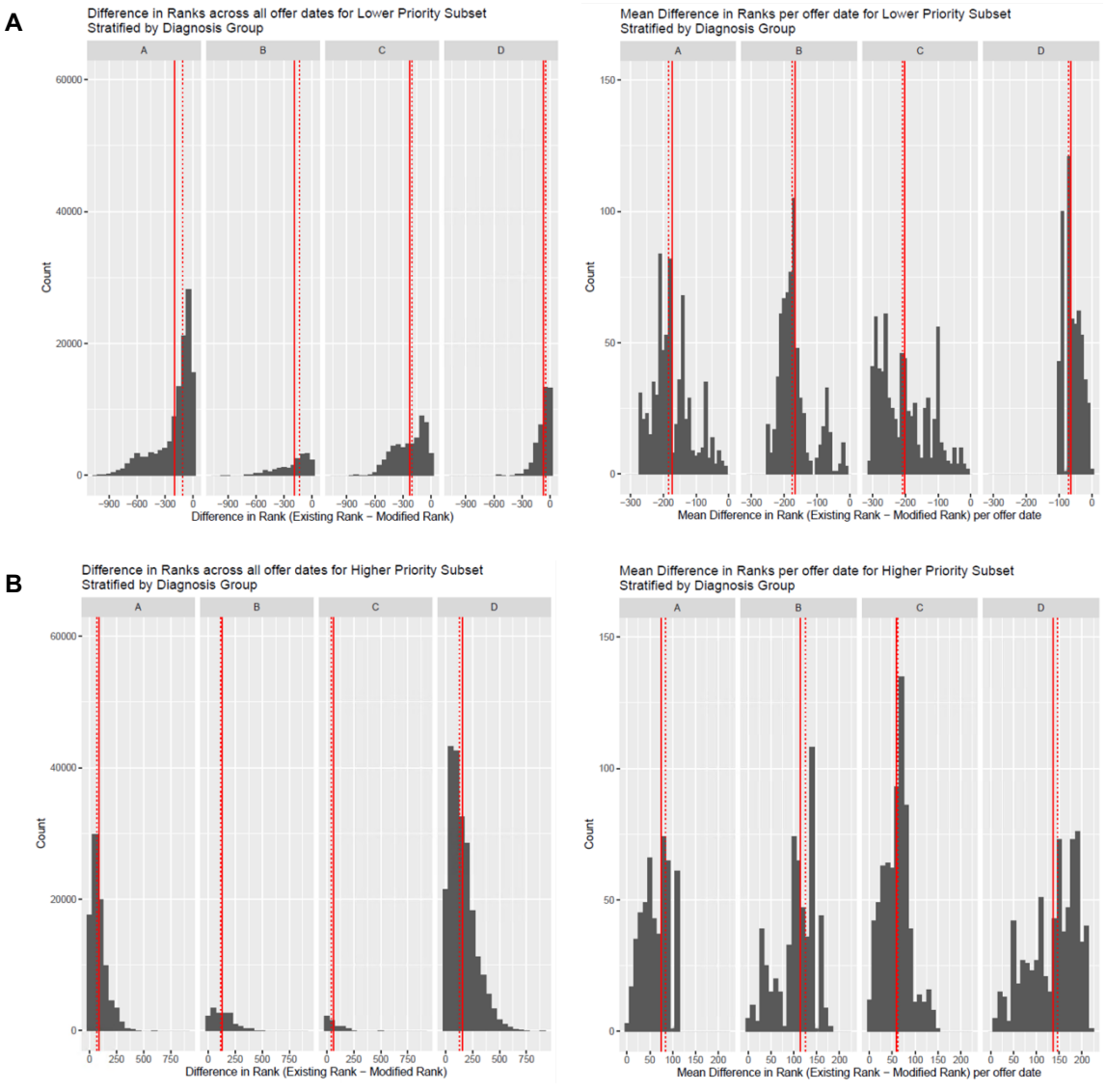
Histograms of the difference in patients' rank under the modified versus existing LAS for the subset of individuals receiving A) lower (worse) priority, or B) higher (better) priority. Solid red line indicates the mean shift in rank, while dotted red line represents the median shift in rank. Left column represents change in rank across all offer dates (cumulative change in rank); right column represents average change in rank per day.



We also plot differences in rank after stratifying by diagnosis group (Appendix Figure 3.2) and six-minute walk distance (Appendix Figure 3.3), two clinical characteristics that were shown in Chapter 3 to exhibit the largest impact on prioritization changes.

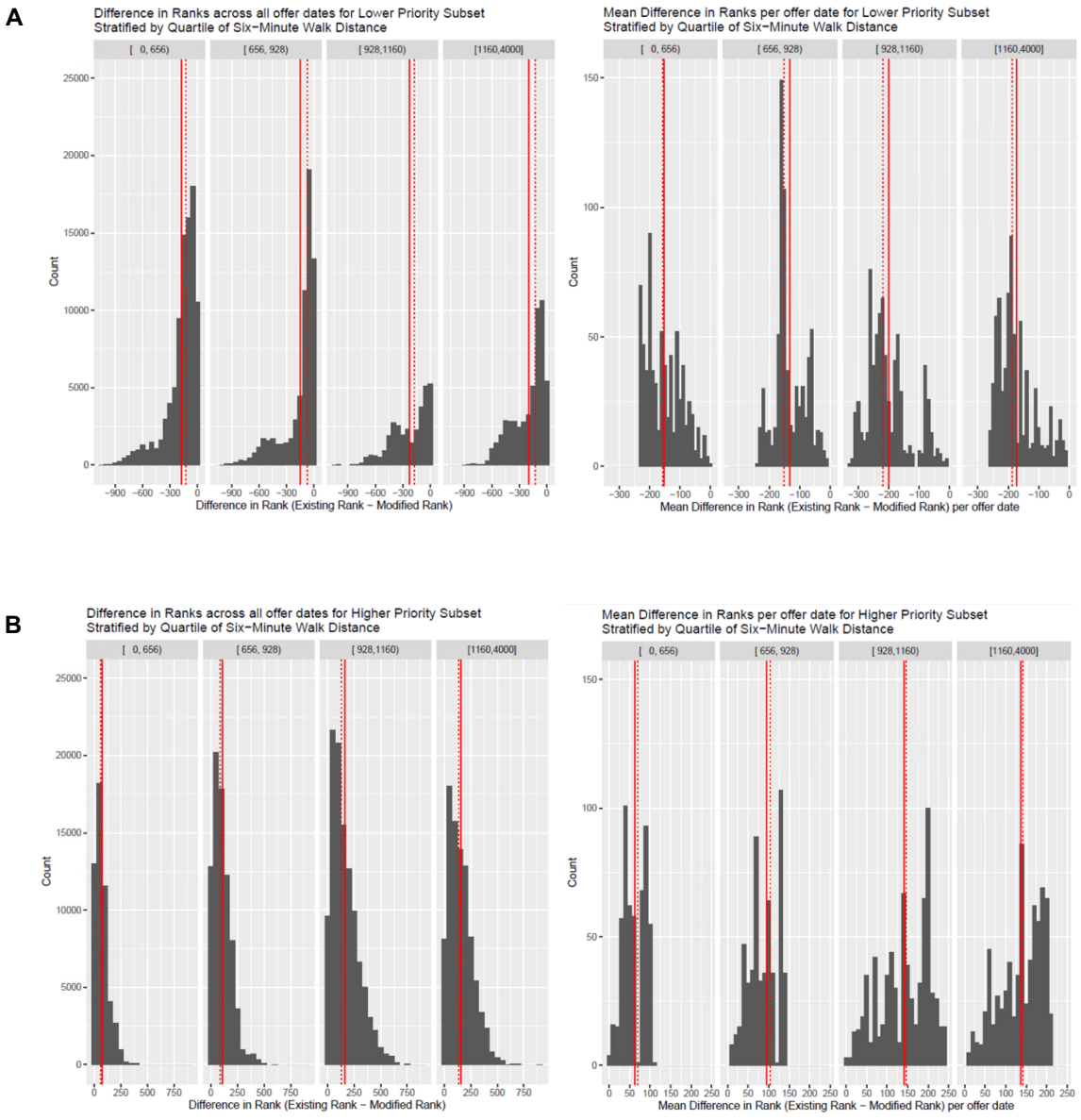
Appendix Figure 3.2. Histograms of differences in rank (stratified by diagnosis group)

Histograms of the difference in patients' rank under the modified versus existing LAS for the subset of individuals receiving A) lower (worse) priority or B) higher (better) priority, stratified by diagnosis group. Solid red line indicates the mean shift in rank, while dotted red line represents the median shift in rank. Left column represents change in rank across all offer dates (cumulative change in rank); right column represents average change in rank per day.



Appendix Figure 3.3. Histograms of differences in rank (stratified by six-minute walk distance)

Histograms of the difference in patients' rank under the modified versus existing LAS for the subset of individuals receiving A) lower (worse) priority or B) higher (better) priority, stratified by six-minute walk distance. Solid red line indicates the mean shift in rank, while dotted red line represents the median shift in rank. Left column represents change in rank across all offer dates (cumulative change in rank); right column represents average change in rank per day.



Appendix 4.1. COREQ Checklist

In this section, we present additional supporting information that fulfills the requirements outlined in the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist.(Tong et al., 2007)

Characteristics of Research Team

EMS (female epidemiologist and PhD candidate trained in qualitative data collection and analyses) conducted the semi-structured in-person or telephone interviews between June 2019 and June 2020. JES (female sociologist, PhD, with mixed-methodology expertise) provided training and mentorship in data collection and analyses. EC (male lung transplant surgeon, MD, MSCE) provided clinical expertise to aid the construction of the interview guide and subsequent analyses. SEK (male epidemiologist, MD, MSCE) supervised the team.

Relationship with Participants

A relationship with participants was not established prior to study commencement. Individuals were recruited to participate in our study through the "Pulmonology and Cardiothoracic/vascular surgery" listserv of the International Society for Heart and Lung Transplant (ISHLT). Potential participants were informed by email of the researchers' and interviewer's occupation, credentials, and training, as well as the goals for conducting the research.

Interviews & Setting

Interviews were conducted in-person or via phone by the lead author using a semi-structured guide (supporting information). In-person interviews occurred in the

respondents' office. Phone interviews were conducted from the interviewer's office. In all cases, no one else was present besides the participants and researchers. Each respondent was interviewed once. Interviews typically lasted 30-50 minutes. Transcripts were not returned to participants for feedback.

Codebook

Appendix Table 4.1 below displays our index codebook. After applying this codebook to all transcripts in a line-by-line manner, we examined overlapping codes and charted data into a framework matrix to compare themes within and across respondents.

Appendix Table 4.1. Index codebook

Code Name	Description
01. motivation to work in lung transplant	Used to capture respondent's rationale for pursuing a career in lung transplantation.
02. region – compare/contrast	Includes descriptions where respondents explicitly compare and contrast practices at one institution/region vs. another
03. institution lung transplant decision process	High-level description of the lung transplant decision-making process at the respondent's institution, including information on other individuals involved in the decision
04. pre-consultation - approach	Details on how the respondent approaches surgical/pulmonology consultation conversations, including important factors the respondent considers when making treatment recommendations to the patient
05. consultation - conversation with patients	Details on what is discussed with patients during the surgical/pulmonology consultation conversation, including information on the LAS and how patients respond to it.
06. post-consultation - next steps	Details on how respondent proceeds after determining that a patient should pursue transplantation, including how the LAS influences these next steps
07. accountability	Information on how transplant centers are rated, how these ratings influence the respondent's decision to pursue transplant for a particular patient, and how/whether these rating metrics conflict with a patient's LAS
08. satisfaction with the LAS	How satisfied is the respondent with the way the current LAS prioritizes patients for lung transplantation?
09. fairness of allocation	How successful (or unsuccessful) does the respondent perceive the LAS to be in allocating donor lungs to patients in a fair manner?

Code Name	Description
10. factors to add to LAS	Additional factors which the respondent believes should be included in the LAS, and their rationale for inclusion.
11. gaming	Includes the respondent's thoughts on whether variables in the LAS can be "adjusted/modified" to ensure that patients receive a more favorable position on the waitlist, as well as if they have encountered this practice. Can also include information on whether the procedures used to measure modifiable variables are standardized across institutions.
12. appeals	Includes information on the UNOS appeals process, or other avenues through which clinicians can ensure their patients receive higher LAS scores.
13. audits	Includes information on the UNOS audit process, as well as whether this process curbs the practice of gaming.
14. characteristics of patients who die or survive	This code is used to capture whether the respondent has an intuitive sense of which patients might die prior to transplantation, and if so, which characteristics distinguish patients who die from those who survive long enough to receive transplant
15. survivor bias concern	Indicates whether the respondent thinks survivor bias is a concern in lung transplantation, and if so, how.
16. accounting for survivor bias	Indicates whether the respondent thinks survivor bias should be accounted for in lung transplantation, and if so, how.
17. complicating the LAS	Used to capture the respondents' thoughts on how a model which accounts for (survivor) bias might affect the decisions they make regarding the care of their patients. Can also include general thoughts on whether a more complicated - but more accurate - LAS would influence the way in which the respondent uses the model in practice.
18. if respondents were in charge of allocation	If respondents were in charge of allocating donor lungs to potential recipients, how would they go about doing it?
19. incredibly interesting, but not sure how to label it	Used to capture random comments which are incredibly interesting but don't necessarily answer any of our questions.

Appendix 4.2. Interview Guide

The semi-structured interview guide was developed by EMS based on literature review and discussion with the other authors. Participants were asked all questions, with additional probing when needed. While the domains of the guide are identical for lung transplant surgeons and pulmonologists, the specific wording of some questions vary slightly due to the fact that these two types of clinicians play different roles in the lung transplant decision-making process. These discrepancies are noted when applicable.

A. Individual and organizational characteristics

i. What motivated you to work in the field of lung transplantation?

ii. How many years of experience have you had in this field?

iii. In what region do you practice?

iv. Have you ever practiced in a different institution or different region?
   a) If YES, which institution/region?

v. How are decisions made about lung transplantation in your institution? Walk me through the process from the initial patient visit, through the decision to register them on the waiting list, through the decision to transplant. [Can you provide me with an example?]
   a) Are there other people involved in this decision? If so:
      1. Who are they?
      2. How would you describe your interactions with these people?

B. Attitudes regarding end-stage lung disease treatment

i. How do you approach [surgical/pulmonology] consultations for adults with end-stage lung disease? What factors are important to you when working up the patient and making recommendations for their treatment?

C. Decision to pursue lung transplantation

i. Can you explain to me the general components of a [surgical/pulmonology] consultation conversation with adult patients with end-stage lung disease?
   a) What things are discussed?
   b) Do you bring up the LAS in this discussion?
   c) If YES, how do you bring it up? How do patients react to it?
   d) Can you explain your rationale for discussing it/not discussing it?

ii. If you decide that a patient should pursue transplantation, how do you proceed? How does the LAS influence these next steps if it does at all? [Use below probes if necessary]:
    a) How might the patient's LAS influence whether you refer the patient to see a [surgeon/pulmonologist]?
    b) How might the patient's LAS influence the way in which you present the patient to the Lung Listing Conference?

iii. Transplant centers are often rated based on the number of transplants they perform and their patients' average 30-day and 1-year post-transplant mortality rates. How do these metrics influence your decision to pursue transplant for a particular patient?
    a) Have you encountered a scenario where the decision to pursue transplant is different depending on whether these metrics or the LAS is used? If so, how do you go about resolving the issue?

D. Satisfaction with current LAS

i. How satisfied are you with the LAS (e.g., the way in which the current LAS prioritizes patients for lung transplantation)?
    a) What factors make you feel this way?
    b) How do these feelings influence the way in which you use the LAS to make treatment decisions?

ii. The goal of the LAS is to allocate donor lungs – a scarce resource – to patients with chronic lung conditions in a fair manner. How successful (or unsuccessful) do you perceive the LAS to be in accomplishing this goal? Can you explain to me why you feel this way?

iii. Are there certain factors which you believe should be included in the LAS, but which are not in the current model? Can you explain why you think these factors should be included?

iv. We have heard that several variables in the LAS are modifiable, and can be adjusted to ensure that patients receive a more favorable position on the waiting list. Have you witnessed this? What are your thoughts on this?
    a) If YES, how does this influence the way in which you use the LAS to make treatment decisions?
    b) If NO, how does this influence the way in which you perceive the LAS?
    c) Are the procedures used to measure modifiable variables standardized across institutions?

v. We have also heard that transplant programs can submit an appeal to UNOS if they feel a particular patient should receive a higher LAS score. Have you participated in such an appeal before? Can you explain what the appeal process is like?

E. Understanding of survivor bias

i. In your experience, have you noticed that certain patients die while on the UNOS lung transplant waiting list, while others survive long enough to receive transplant?
   a) What characteristics distinguish these two groups of people?
   b) Do you have an intuitive sense of which patients might die prior to transplantation?
      1. If YES, how does that affect your decision to place patients on the waiting list?

ii. In epidemiology, "survivor bias" is the idea that if treatments are offered to patients conditional on their survival, the resulting estimate of the effectiveness of treatment might be inaccurate. My team believes that survivor bias might play a role when predicting outcomes of lung transplant, because in order to receive a transplant, patients must survive long enough for a suitable donor lung to become available.
   a) Do you think survivor bias might be a concern in lung transplantation? In what way?

F. Thoughts regarding modification of the LAS

i. Do you think survivor bias should be accounted for in the allocation of donor lungs?
   a) If YES, how?
   b) If NO, why do you feel that way?

ii. Adjusting for survivor bias in the LAS is possible, but it could lead to a more complicated allocation algorithm.
   a) How would having a more complicated – but more accurate – LAS model influence the way in which you use this model in clinical practice?
   b) How might it affect the decisions you make regarding the care of patients with end-stage lung disease?

iii. If you were in charge of allocating donor lungs to potential recipients, how would you go about doing it?

G. Demographic Questions

i. How would you report your gender?
ii. How would you report your race/ethnicity?
iii. Where did you complete your medical training?
   a) What year did you graduate?

BIBLIOGRAPHY

Beasley, T. M., Erickson, S., & Allison, D. B. (2009). Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav Genet*, *39*(5), 580-595. https://doi.org/10.1007/s10519-009-9281-0

Benvenuto, L. J., Anderson, D. R., Kim, H. P., Hook, J. L., Shah, L., Robbins, H. Y., . . . Program, F. t. C. U. L. T. (2018). Geographic disparities in donor lung supply and lung transplant waitlist outcomes: A cohort study. *Am J Transplant*, *18*(6), 1471-1480. https://doi.org/10.1111/ajt.14630

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*(8476), 307-310.

Blumenthal, N. P., Petty, M. G., & McCorkle, R. (2017). Missing Domains of Lung Transplant Patient Selection. *Prog Transplant*, *27*(1), 90-97. https://doi.org/10.1177/1526924816679840

Brod, M., Tesler, L. E., & Christensen, T. L. (2009). Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res*, *18*(9), 1263-1278. https://doi.org/10.1007/s11136-009-9540-9

Buchanan, A. L., Hudgens, M. G., Cole, S. R., Lau, B., Adimora, A. A., & Study, W. s. I. H. (2014). Worth the weight: using inverse probability weighted Cox models in AIDS research. *AIDS Res Hum Retroviruses*, *30*(12), 1170-1177. https://doi.org/10.1089/AID.2014.0037

Cattaneo, M., Malighetti, P., & Spinelli, D. (2017). Estimating receiver operating characteristic cures for time-dependent outcomes: The stroccurve package. *The Stata Journal*, *17*(4), 1015-1023.

Chen, H., Shiboski, S. C., Golden, J. A., Gould, M. K., Hays, S. R., Hoopes, C. W., & De Marco, T. (2009). Impact of the lung allocation score on lung transplantation for pulmonary arterial hypertension. *Am J Respir Crit Care Med*, *180*(5), 468-474. https://doi.org/10.1164/rccm.200810-1603OC

Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*, *168*(6), 656-664. https://doi.org/10.1093/aje/kwn164

Cooke, D. T., Olive, J., Godoy, L., Preventza, O., Mathisen, D. J., & Prager, R. L. (2019). The Importance of a Diverse Specialty: Introducing the STS Workforce on Diversity and Inclusion. *Ann Thorac Surg*, *108*(4), 1000-1005. https://doi.org/10.1016/j.athoracsur.2019.07.007

Deterding, N. M., & Waters, M. C. (2021). Flexible Coding of In-depth Interviews: A Twenty-first-century Approach. *Sociological Methods & Research*, *50*(2), 708-739. https://doi.org/10.1177/0049124118799377

Doby, B. L., Hanner, K., Johnson, S., Purnell, T. S., Shah, M. B., & Lynch, R. J. (2020). Results of a data-driven performance improvement initiative in organ donation. *Am J Transplant*. https://doi.org/10.1111/ajt.16442

Drolen, C., Cantu, E., Goldberg, H. J., Diamond, J. M., & Courtwright, A. (2020). Impact of the elimination of the donation service area on United States lung transplant practices and outcomes at high and low competition centers. *Am J Transplant*, *20*(12), 3631-3638. https://doi.org/10.1111/ajt.16098

Egan, T. M., & Edwards, L. B. (2016). Effect of the lung allocation score on lung transplantation in the United States. *J Heart Lung Transplant*, *35*(4), 433-439. https://doi.org/10.1016/j.healun.2016.01.010

Egan, T. M., Murray, S., Bustami, R. T., Shearon, T. H., McCullough, K. P., Edwards, L. B., . . . Grover, F. L. (2006). Development of the new lung allocation system in the United States. *Am J Transplant*, *6*(5 Pt 2), 1212-1227. https://doi.org/10.1111/j.1600-6143.2006.01276.x

Egleston, B. L., Scharfstein, D. O., Freeman, E. E., & West, S. K. (2007). Causal inference for non-mortality outcomes in the presence of death. *Biostatistics*, *8*(3), 526-545. https://doi.org/10.1093/biostatistics/kxl027

Erhunmwunsee, L., Backhus, L. M., Godoy, L., Edwards, M. A., & Cooke, D. T. (2019). Report from the Workforce on Diversity and Inclusion-The Society of Thoracic Surgeons Members' Bias Experiences. *Ann Thorac Surg*, *108*(5), 1287-1291. https://doi.org/10.1016/j.athoracsur.2019.08.015

Finlen Copeland, C. A., Vock, D. M., Pieper, K., Mark, D. B., & Palmer, S. M. (2013). Impact of lung transplantation on recipient quality of life: a serial, prospective, multicenter analysis through the first posttransplant year. *Chest*, *143*(3), 744-750. https://doi.org/10.1378/chest.12-0971

Freiman, M. H., & Small, D. S. (2014). Large sample bounds on the survivor average causal effect in the presence of a binary covariate with conditionally ignorable treatment assignment. *Int J Biostat*, *10*(2), 143-163. https://doi.org/10.1515/ijb-2013-0039

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol*, *13*, 117. https://doi.org/10.1186/1471-2288-13-117

Glymour, M. M., & Greenland, S. (2008). Survivor Bias. In K. J. Rothman, S. Greenland, & T. L. Lash (Eds.), *Modern Epidemiology* (3rd edition ed., pp. pp. 197-198). Lippincott Williams & Wilkins.

Gottlieb, J. (2017). Lung allocation. *J Thorac Dis*, *9*(8), 2670-2674. https://doi.org/10.21037/jtd.2017.07.83

Gottlieb, J., Greer, M., Sommerwerck, U., Deuse, T., Witt, C., Schramm, R., . . . Smits, J. M. (2014). Introduction of the lung allocation score in Germany. *Am J Transplant*, *14*(6), 1318-1327. https://doi.org/10.1111/ajt.12752

Gries, C. J., Mulligan, M. S., Edelman, J. D., Raghu, G., Curtis, J. R., & Goss, C. H. (2007). Lung allocation score for lung transplantation: impact on disease severity and survival. *Chest*, *132*(6), 1954-1961. https://doi.org/10.1378/chest.07-1160

Gries, C. J., Rue, T. C., Heagerty, P. J., Edelman, J. D., Mulligan, M. S., & Goss, C. H. (2010). Development of a predictive model for long-term survival after lung transplantation and implications for the lung allocation score. *J Heart Lung Transplant*, *29*(7), 731-738. https://doi.org/10.1016/j.healun.2010.02.007

Halpern, S. D. (2015). Beyond "the LAS is broken". Ways to improve lung allocation. *Am J Respir Crit Care Med*, *191*(3), 245-246. https://doi.org/10.1164/rccm.201412-2304ED

Halpern, S. E., McConnell, A., Peskoe, S. B., Raman, V., Jawitz, O. K., Choi, A. Y., . . . Hartwig, M. G. (2021). A three-tier system for evaluation of organ procurement organizations' willingness to pursue and utilize nonideal donor lungs. *Am J Transplant*, *21*(3), 1269-1277. https://doi.org/10.1111/ajt.16347

Haneuse, S., & Daniels, M. (2016). A General Framework for Considering Selection Bias in EHR-Based Studies: What Data Are Observed and Why? *EGEMS (Wash DC)*, *4*(1), 1203. https://doi.org/10.13063/2327-9214.1203

Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer.

Hart, A., & Engels, E. A. (2021). Balancing uncertain risks in candidates for solid organ transplantation with a history of malignancy: Who is safe to transplant? *Am J Transplant*, *21*(2), 447-448. https://doi.org/10.1111/ajt.16366

Hauerwaas, A., & Weisenfeld, U. (2020). The impact of systemic innovations for transforming transplant systems. Lessons learned from the German lung transplantation system. A qualitative study. *Health Syst (Basingstoke)*, *9*(1), 76-93. https://doi.org/10.1080/20476965.2019.1604086

Heagerty, P. J., Lumley, T., & Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, *56*(2), 337-344. https://doi.org/10.1111/j.0006-341x.2000.00337.x

Jaeger, B. (2017). *r2glmm: Computes R Squared for Mixed (Multilevel) Models*. In (Version 0.1.2) https://cran.r-project.org/web/packages/r2glmm/index.html

Kasiske, B. L., Pyke, J., & Snyder, J. J. (2020). Continuous distribution as an organ allocation framework. *Curr Opin Organ Transplant*, *25*(2), 115-121. https://doi.org/10.1097/MOT.0000000000000733

Kotloff, R. M. (2013). Risk stratification of lung transplant candidates: implications for organ allocation. *Ann Intern Med*, *158*(9), 699-700. https://doi.org/10.7326/0003-4819-158-9-201305070-00011

Kozower, B. D., Meyers, B. F., Smith, M. A., De Oliveira, N. C., Cassivi, S. D., Guthrie, T. J., . . . Jones, D. R. (2008). The impact of the lung allocation score on short-term transplantation outcomes: a multicenter study. *J Thorac Cardiovasc Surg*, *135*(1), 166-171. https://doi.org/10.1016/j.jtcvs.2007.08.044

Ladin, K., Emerson, J., Berry, K., Butt, Z., Gordon, E. J., Daniels, N., . . . Hanto, D. W. (2019). Excluding patients from transplant due to social support: Results from a national survey of transplant providers. *Am J Transplant*, *19*(1), 193-203. https://doi.org/10.1111/ajt.14962

Levitsky, J., Formica, R. N., Bloom, R. D., Charlton, M., Curry, M., Friedewald, J., . . . Volk, M. (2017). The American Society of Transplantation Consensus Conference on the Use of Hepatitis C Viremic Donors in Solid Organ Transplantation. *Am J Transplant*, *17*(11), 2790-2802. https://doi.org/10.1111/ajt.14381

Lingaraju, R., Blumenthal, N. P., Kotloff, R. M., Christie, J., Ahya, V. N., Sager, J. S., . . . Hadjiliadis, D. (2006). Effects of lung allocation score on waiting list rankings and transplant procedures. *J Heart Lung Transplant*, *25*(9), 1167-1170. https://doi.org/10.1016/j.healun.2006.05.007

Little, R. J., & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health*, *21*, 121-145. https://doi.org/10.1146/annurev.publhealth.21.1.121

Loss, J., Drewitz, K. P., Schlitt, H. J., & Loss, M. (2013). Accept or refuse? Factors influencing the decision-making of transplant surgeons who are offered a pancreas: results of a qualitative study. *BMC Surg*, *13*, 47. https://doi.org/10.1186/1471-2482-13-47

Maxwell, B. G., Levitt, J. E., Goldstein, B. A., Mooney, J. J., Nicolls, M. R., Zamora, M., . . . Dhillon, G. S. (2014). Impact of the lung allocation score on survival beyond 1 year. *Am J Transplant*, *14*(10), 2288-2294. https://doi.org/10.1111/ajt.12903

Maxwell, B. G., Mooney, J. J., Lee, P. H., Levitt, J. E., Chhatwani, L., Nicolls, M. R., . . . Dhillon, G. S. (2015). Increased resource use in lung transplant admissions in the lung allocation score era. *Am J Respir Crit Care Med*, *191*(3), 302-308. https://doi.org/10.1164/rccm.201408-1562OC

McCue, J. D., Mooney, J., Quail, J., Arrington, A., Herrington, C., & Dahlberg, P. S. (2008). Ninety-day mortality and major complications are not affected by use of lung allocation score. *J Heart Lung Transplant*, *27*(2), 192-196. https://doi.org/10.1016/j.healun.2007.11.001

Merlo, C. A., Weiss, E. S., Orens, J. B., Borja, M. C., Diener-West, M., Conte, J. V., & Shah, A. S. (2009). Impact of U.S. Lung Allocation Score on survival after lung transplantation. *J Heart Lung Transplant*, *28*(8), 769-775. https://doi.org/10.1016/j.healun.2009.04.024

Miles, M., Huberman, A., & Class of 1924 Book Fund. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Sage.

Miles, M., Huberman, A., & Saldana, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd Edition ed.). Sage.

Mooney, J. J., Bhattacharya, J., & Dhillon, G. S. (2019). Effect of broader geographic sharing of donor lungs on lung transplant waitlist outcomes. *J Heart Lung Transplant*, *38*(2), 136-144. https://doi.org/10.1016/j.healun.2018.09.007

Mooney, J. J., Hedlin, H., Mohabir, P., Bhattacharya, J., & Dhillon, G. S. (2018). Racial and ethnic disparities in lung transplant listing and waitlist outcomes. *J Heart Lung Transplant*, *37*(3), 394-400. https://doi.org/10.1016/j.healun.2017.09.017

Mooney, J. J., Yang, L., Hedlin, H., Mohabir, P., & Dhillon, G. S. (2019). Multiple listing in lung transplant candidates: A cohort study. *Am J Transplant*, *19*(4), 1098-1108. https://doi.org/10.1111/ajt.15124

NHS Blood and Transplant, N. *National Liver Offering Scheme*. Retrieved 22 June 2021 from https://www.odt.nhs.uk/odt-structures-and-standards/odt-hub-programme/national-liver-offering-scheme/

O'Connor, K., & Glazier, A. (2021). OPO performance improvement and increasing organ transplantation: Metrics are necessary but not sufficient. *Am J Transplant*. https://doi.org/10.1111/ajt.16545

Organ Procurement and Transplantation Network, O. (2020). *Organ Procurement and Transplantation Network (OPTN) Policies*. Retrieved 10 March 2020 from https://optn.transplant.hrsa.gov/media/1200/optn_policies.pdf

Persad, G., Wertheimer, A., & Emanuel, E. J. (2009). Principles for allocation of scarce medical interventions. *Lancet*, *373*(9661), 423-431. https://doi.org/10.1016/S0140-6736(09)60137-9

Ramos, K. J., Smith, P. J., McKone, E. F., Pilewski, J. M., Lucy, A., Hempstead, S. E., . . . Committee, C. L. T. R. G. (2019). Lung transplant referral for individuals with cystic fibrosis: Cystic Fibrosis Foundation consensus guidelines. *J Cyst Fibros*, *18*(3), 321-333. https://doi.org/10.1016/j.jcf.2019.03.002

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550-560. https://doi.org/10.1097/00001648-200009000-00011

Ross-Driscoll, K., Axelrod, D., Lynch, R., & Patzer, R. E. (2020). Using Geographic Catchment Areas to Measure Population-based Access to Kidney Transplant in the United States. *Transplantation*, *104*(12), e342-e350. https://doi.org/10.1097/TP.0000000000003369

Royston, P. (2015). Tools for checking calibration of a Cox model in external validation: Prediction of population-averaged survival curves based on risk groups. *The Stata Journal*, *15*(1), 275-291.

Rudasill, S. E., Sanaiha, Y., Kwon, M., Mardock, A. L., Khoury, H., Omari, B., . . . Benharash, P. (2019). Understanding lung transplant listing practices: Survival in lung transplant candidates who improve clinically to delisting. *Surgery*, *166*(6), 1142-1147. https://doi.org/10.1016/j.surg.2019.07.009

Russo, M. J., Meltzer, D., Merlo, A., Johnson, E., Shariati, N. M., Sonett, J. R., & Gibbons, R. (2013). Local allocation of lung donors results in transplanting lungs in lower priority transplant recipients. *Ann Thorac Surg*, *95*(4), 1231-1234; discussion 1234-1235. https://doi.org/10.1016/j.athoracsur.2012.11.070

Russo, M. J., Worku, B., Iribarne, A., Hong, K. N., Yang, J. A., Vigneswaran, W., & Sonett, J. R. (2011). Does lung allocation score maximize survival benefit from lung transplantation? *J Thorac Cardiovasc Surg*, *141*(5), 1270-1277. https://doi.org/10.1016/j.jtcvs.2010.12.028

Schaenman, J. M., Diamond, J. M., Greenland, J. R., Gries, C., Kennedy, C. C., Parulekar, A. D., . . . Bhorade, S. (2020). Frailty and aging-associated syndromes in lung transplant candidates and recipients. *Am J Transplant*. https://doi.org/10.1111/ajt.16439

Schnellinger, E. M., Cantu, E., Harhay, M. O., Schaubel, D. E., Kimmel, S. E., & Stephens-Shields, A. J. (2021). Mitigating selection bias in organ allocation models. *BMC Med Res Methodol*, *21*(1), 191. https://doi.org/10.1186/s12874-021-01379-7

Schwarz, S., Rahimi, N., Kifjak, D., Muckenhuber, M., Watzenböck, M., Benazzo, A., . . . Group, V. L. T. (2020). Comparison of donor scores in bilateral lung transplantation-A large single-center analysis. *Am J Transplant*. https://doi.org/10.1111/ajt.16402

Singer, J. P., Chen, J., Katz, P. P., Blanc, P. D., Kagawa-Singer, M., & Stewart, A. L. (2015). Defining novel health-related quality of life domains in lung transplantation: a qualitative analysis. *Qual Life Res*, *24*(6), 1521-1533. https://doi.org/10.1007/s11136-014-0875-5

Snyder, J. J., Salkowski, N., Wey, A., Pyke, J., Israni, A. K., & Kasiske, B. L. (2018). Organ distribution without geographic boundaries: A possible framework for organ allocation. *Am J Transplant*, *18*(11), 2635-2640. https://doi.org/10.1111/ajt.15115

Stephenson, A. L., Ramos, K. J., Sykes, J., Ma, X., Stanojevic, S., Quon, B. S., . . . Goss, C. H. (2021). Bridging the survival gap in cystic fibrosis: An investigation of lung transplant outcomes in Canada and the United States. *J Heart Lung Transplant*, *40*(3), 201-209. https://doi.org/10.1016/j.healun.2020.12.001

Tayob, N., & Murray, S. (2017). Statistical consequences of a successful lung allocation system - recovering information and reducing bias in models for urgency. *Stat Med*, *36*(15), 2435-2451. https://doi.org/10.1002/sim.7283

Thabut, G., Munson, J., Haynes, K., Harhay, M. O., Christie, J. D., & Halpern, S. D. (2012). Geographic disparities in access to lung transplantation before and after implementation of the lung allocation score. *Am J Transplant*, *12*(11), 3085-3093. https://doi.org/10.1111/j.1600-6143.2012.04202.x

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*, *19*(6), 349-357. https://doi.org/10.1093/intqhc/mzm042

U.S. Department of Health & Human Services, O. P. T. N. (2018). *OPTN/SRTR 2018 Annual Data Report: Lung*. Retrieved 5 June 2020 from https://srtr.transplant.hrsa.gov/annual_reports/2018/Lung.aspx

U.S. Department of Health & Human Services, O. P. T. N. *Continuous Distribution*. Retrieved 6 October 2020 from https://optn.transplant.hrsa.gov/governance/policy-initiatives/continuous-distribution/

Final Rule, 56649-56661 (October 20, 1999).

United Network for Organ Sharing, U. (2015). *A Guide to Calculating the Lung Allocation Score*. Retrieved 15 June 2018 from https://www.unos.org/wp-content/uploads/unos/lung_allocation_score.pdf

Valbuena, V. S. M., Obayemi, J. E., Purnell, T. S., Scantlebury, V. P., Olthoff, K. M., Martins, P. N., . . . Simpson, D. C. (2021). Gender and racial disparities in the transplant surgery workforce. *Curr Opin Organ Transplant*, *26*(5), 560-566. https://doi.org/10.1097/MOT.0000000000000915

Van Pilsum Rasmussen, S. E., Seaman, S., Brown, D., Desai, N., Sulkowski, M., Segev, D. L., . . . Sugarman, J. (2020). Patient's Perspectives of Experimental HCV-Positive to HCV-Negative Renal Transplantation: Report from a Single Site. *AJOB Empir Bioeth*, *11*(1), 40-52. https://doi.org/10.1080/23294515.2019.1670277

Van Pilsum Rasmussen, S. E., Zhou, S., Thomas, A. G., Segev, D. L., & Nicholas, L. H. (2019). Transplant community perceptions of the benefits and drawbacks of alternative quality metrics for regulation. *Clin Transplant*, *33*(4), e13500. https://doi.org/10.1111/ctr.13500

Veatch, R. M., & Ross, L. F. (2015). Part III: Allocating Organs. In *Transplantation Ethics* (2nd Edition ed.). Georgetown University Press.

Veatch, R. M., & Ross, L. F. (2015). The Role of Age in Allocation. In *Transplantation Ethics* (Second ed., pp. 334-354). Georgetown University Press.

Vock, D. M., Durheim, M. T., Tsuang, W. M., Finlen Copeland, C. A., Tsiatis, A. A., Davidian, M., . . . Palmer, S. M. (2017). Survival Benefit of Lung Transplantation in the Modern Era of Lung Allocation. *Ann Am Thorac Soc*, *14*(2), 172-181. https://doi.org/10.1513/AnnalsATS.201606-507OC

Vock, D. M., Tsiatis, A. A., Davidian, M., Laber, E. B., Tsuang, W. M., Finlen Copeland, C. A., & Palmer, S. M. (2013). Assessing the causal effect of organ transplantation on the distribution of residual lifetime. *Biometrics*, *69*(4), 820-829. https://doi.org/10.1111/biom.12084

Volk, M. L., Biggins, S. W., Huang, M. A., Argo, C. K., Fontana, R. J., & Anspach, R. R. (2011). Decision making in liver transplant selection committees: a multicenter study. *Ann Intern Med*, *155*(8), 503-508. https://doi.org/10.7326/0003-4819-155-8-201110180-00006

Wang, L., Zhou, X. H., & Richardson, T. S. (2017). Identification and estimation of causal effects with outcomes truncated by death. *Biometrika*, *104*(3), 597-612. https://doi.org/10.1093/biomet/asx034

Wagener, G. (2020). Multiple Listings: Good for a Few, but No Solution for the Organ Shortage. *Transplantation*, *104*(4), 671-672. https://doi.org/10.1097/TP.0000000000002966

Weill, D., Benden, C., Corris, P. A., Dark, J. H., Davis, R. D., Keshavjee, S., . . . Glanville, A. R. (2015). A consensus document for the selection of lung transplant candidates: 2014--an update from the Pulmonary Transplantation Council of the International Society for Heart and Lung Transplantation. *J Heart Lung Transplant*, *34*(1), 1-15. https://doi.org/10.1016/j.healun.2014.06.014

Weiss, R. (1994). *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. The Free Press.

Whitford, H., Kure, C. E., Henriksen, A., Hobson, J., Snell, G. I., Levvey, B. J., . . . McGiffin, D. C. (2020). A donor PaO. *J Heart Lung Transplant*, *39*(1), 53-61. https://doi.org/10.1016/j.healun.2019.08.021

Wille, K. M., Harrington, K. F., deAndrade, J. A., Vishin, S., Oster, R. A., & Kaslow, R. A. (2013). Disparities in lung transplantation before and after introduction of the lung allocation score. *J Heart Lung Transplant*, *32*(7), 684-692. https://doi.org/10.1016/j.healun.2013.03.005

Xiang, F., & Murray, S. (2012). Restricted mean models for transplant benefit and urgency. *Stat Med*, *31*(6), 561-576. https://doi.org/10.1002/sim.4450

Xiang, F., Murray, S., & Liu, X. (2014). Analysis of transplant urgency and benefit via multiple imputation. *Stat Med*, *33*(26), 4655-4670. https://doi.org/10.1002/sim.6250

Yang, W., & Joffe, M. M. (2012). Subtle issues in model specification and estimation of marginal structural models. *Pharmacoepidemiol Drug Saf*, *21*(3), 241-245. https://doi.org/10.1002/pds.2306

Yusen, R. D. (2009). Technology and outcomes assessment in lung transplantation. *Proc Am Thorac Soc*, *6*(1), 128-136. https://doi.org/10.1513/pats.200809-102GO