Using Discrete Event Simulation to Optimise Donor Lung Allocation

Samuel John Kennedy

Thesis submitted for the degree of
Doctor of Philosophy

Faculty of Medical Sciences
Translational and Clinical Research Institute
Newcastle University

February 2024

Supervisors
Doctor William E. Scott III
Doctor Leo Freitas
Professor Andrew J. Fisher

This thesis is dedicated to my family: my wife, Charlotte, and daughters, Eliza and Lara, for their continuing love and support; and to my parents, Pauline and Peter, for always encouraging my interests and pursuit of knowledge.

Acknowledgements

I would like to express my gratitude to the individuals and organisations that have supported this work. While I have the great privilege of presenting the findings of this research in this thesis, none of this would have been possible without their hard work and support over the course of completing my PhD.

Firstly, I would like to extend my sincerest gratitude and thanks to my supervisors, Dr. William E. Scott III, Dr. Leo Freitas, and Prof. Andrew Fisher. This thesis would not have been completed without their hard work in providing feedback, guidance and support throughout the course of this project. They have collectively and individually gone above and beyond to provide me with opportunities for growth and support me through all of the 'ups' and 'downs' that have occurred. Thank you.

I am very grateful to Rachel Johnson, Sally Rushton, Lisa Mumford and Rachel Hogg for their feedback and assistance with the statistical and simulation methods that were key to this research.

I would also like to thank Dr. Jasvir Parmar and the Cardiothoracic Advisory Group (CTAG) for their invaluable support and feedback on this work, and Robert Burns for his key role in patient outreach and circulating the patient survey, a crucial component of this research.

The development of the Clinical Prioritisation Assistance Tool (CPAT) was made possible through the collaborative efforts of a large number of individuals spanning multiple organisations. I would like to especially thank Prof. Paul Watson and Stephen Dowsland (National Innovation Centre for Data), Prof. Cliff Jones, Andrew Hardy and Mark Turner (Newcastle University - Computing), Prof. Fiona Matthews (Newcastle University - Evidence Synthesis), Ian Nesbitt, David Cressey, Kenneth Baker, Chris Duncan and Chris Plummer (Newcastle upon Tyne Hospitals).

My PhD was funded by the NIHR Blood and Transplant Research Unit (BTRU) and The European Society for Organ Transplantation (ESOT) by means of their basic science fellowship. I am extremely grateful for this funding that has given me this incredible opportunity to complete my PhD and continue this research. The financial support has not only facilitated my research but also enabled my academic growth and contributions to the field.

Abstract

Background: Optimally allocating deceased donor lungs to candidates requiring a life saving lung transplant while balancing efficiency and equity is a difficult challenge, which is compounded by the limited availability of donor lungs in the UK, with less than 15% of offered lungs being utilised for transplantation. This thesis argues that the sequential, centre-based approach used by the existing UK lung allocation policy does not make optimal use of scarce donor lungs and leads to inequitable access to transplantation. This research identifies the need for a transparent, auditable, and equitable system that maximises the additional years of life recipients gain from transplant ('net benefit') by considering both clinical urgency and post-transplant outcomes. This research uses the concepts of the Lung Allocation Score (LAS) as a springboard to bridge the gap between the current sequential, centre-based UK lung allocation policy and a score-based national allocation policy.

Methods: Lung transplant datasets were provided by NHS Blood and Transplant that included data on adult (aged 16+), first-time, lung-only lung transplant candidates (n = 4280) and recipients (n = 2131) listed or transplanted between 2002 and 2021. Custom Cox proportional hazards models were developed to simulate waiting list and post-transplant survival durations, and a novel lung allocation policy simulation engine was developed that used discrete event simulation to predict the impact of a number of potential national lung allocation policies. Five initial allocation policies were simulated, focusing on different priority-ratios between waiting list survival (WL) and post-transplant survival (PTX). Five additional policies were simulated that maximise the use of single-lung transplants (SLT) for recipients with interstitial lung disease (ILD). Additional scenarios were simulated to assess the impact of increased utilisation for each of the standard and SLT policies. The key performance metrics recorded for each policy were: annual waiting list deaths, mean net benefit per recipient, and post-transplant survival rates at 1 and 5 years. The analytic hierarchy process (AHP) was used to collect and evaluate stakeholder preferences (i.e., candidates, recipients, their family members (n = 100), and clinicians (n = 62)) to identify which simulated allocation policies aligned most closely with the goals and values of the lung transplant community.

Results: The Cox models used for this work demonstrated reasonably strong predictive power for waiting list survival (C-statistic: training dataset = 0.73, validation dataset = 0.66), and moderate predictive power for post-transplant survival (C-statistic: training dataset = 0.60, validation dataset = 0.55). This demonstrates a significant improvement over the existing UK lung allocation policy, which had a C-statistic for waiting list survival of 0.51 (training dataset) and 0.56 (validation dataset), and for post-transplant survival: 0.54 (training dataset) and 0.52 (validation dataset).

The five initial simulated policies revealed that a national score-based system would significantly decrease waiting list mortality relative to the existing policy (90 annual waiting list deaths) regardless of choice of priority-ratio: the WL policy (i.e., prioritising clinical urgency) resulted in 46 annual waiting list deaths (48.9% decrease), and the PTX policy (i.e., prioritising post-transplant outcomes) resulted in 77 deaths (14.4% decrease).

The PTX policy yielded the highest net benefit (6.9 years, compared to 5.0 years with the existing policy), and post-transplant survival rates: 83.5% at 1 year and 59.5% at 5 years, compared to the existing policy with 80.2% and 53.3% respectively. The SLT policies further decreased waiting list mortality: when combined with the 1:2 WL:PTX priority-ratio, the SLT-1:2 policy reduced annual waiting list deaths to 31, a 65.6% decrease compared to the existing policy.

Simulations showed a non-proportional relationship between increased utilisation rates and reduction in waiting list mortality: for the standard policies, a 5% increase in utilisation resulted in a 10.9% reduction in mortality, a 10% increase resulted in a 21.7% reduction, and a 25% increase resulted in a 45.7% reduction. This non-proportionality was also observed for the SLT policies.

Survey responses highlighted a preference for policies that prioritise post-transplant survival, with the PTX policy aligning most closely with 52% of candidates, recipients and their family members, and 48.4% of clinicians (50.6% overall).

Conclusion: This thesis demonstrates a comprehensive approach to evaluating and identifying improvements to the UK lung allocation policy, through a novel combination of methods from the fields of survival analysis, operations research, and computer science. Simulations demonstrated that a national score-based allocation policy could significantly decrease waiting list mortality, increase post-transplant survival, and ensure donor lungs are efficiently allocated to recipients that will benefit most from transplant. The results of this thesis calls into question the historical trend of decreasing use of SLT, and argues for a reversal of this trend by utilising SLT for candidates with ILD.

Importantly, the methods developed and described in this thesis extend beyond lung transplantation, offering a framework that can be applied to other donor organs and other healthcare allocation challenges more generally. Furthermore, the proposed lung allocation scoring system would be the first in the world to integrate candidate and donor characteristics to ensure optimal matching between donor and recipient.

Overall, this thesis contributes to the field of transplant data science by demonstrating the novel application of methods to balance benefit, urgency, and community values, and highlights the importance of data-driven, transparent, community-aligned research in allocation policy development. Future work should aim to refine predictive models, expand the target population, and explore the practical implications of implementing these recommendations, ensuring a careful, monitored transition to any new allocation system to mitigate unforeseen consequences.

Structure of Thesis

This thesis covers concepts from computer science, statistics, operations research and medicine. As a result of this, some sections focus on a specific theme. This thesis is designed to be read in the order it is presented, however if a section or subsection has a particular theme they are labelled as follows: the 🩺 icon indicates a clinical focus, the 💻 icon indicates a computing/mathematical focus and the 📊 icon indicates a statistical focus.

Contents

List of Figures

List of Tables

Glossary

ABO Refers to the blood type of an individual, being either 'A', 'B', or 'O'. 25–27, 29–33, 35, 39, 44, 77, 101, 125, 130

AFT accelerated failure time. 203

AHP analytic hierarchy process. i, 24, 46, 48–51, 53, 55–57, 176–181, 201, 202, 207

ANP Analytic Network Process. 54

AUC area under the curve. 114

BLT bilateral lung transplant. 44, 45

BMI body mass index. 81, 82, 206

CAS composite allocation score. 44

CF Cystic Fibrosis. 15, 40, 41

CFS Clinical Frailty Scale. 66, 73, 74

CI consistency index. 179

CIT cold ischaemia time. 7, 43

COPD Chronic obstructive pulmonary disease. 14–16, 37, 39, 40, 42, 44, 45, 77, 153, 187

CPAT Clinical Prioritisation Assistance Tool. 3, 59, 60, 63, 65–68, 72–75

CR consistency ratio. 179

CTAG Cardiothoracic Advisory Group. 4, 180

DES Discrete event simulation. 77, 107

DSA donation service area. 42, 44

DSL Domain-Specific Language. 60, 63, 65, 66

ECD extended criteria donors. 43

ECMO extracorporeal membrane oxygenation. 27, 28, 41

EPR electronic patient record. 68

EPTS Estimated post transplant survival. 52

ET-LAS Eurotransplant Lung Allocation Score. 41

EV expected value. 100

FDM fuzzy delphi method. 51

FEV1 forced expiratory volume over one second. 204

FVC forced vital capacity. 204

GAN generative adversarial network. 204

HLA human leukocyte antigen. 27, 35, 51

HR hazard ratio. 82

HU High Urgency. 28, 30

iLA interventional lung assist. 27

ILD Interstitial lung disease. 15

IPCW Inverse Probability of Censoring Weighting. 44, 96

IPF Idiopathic Pulmonary Fibrosis. 8, 15, 37, 40, 41, 44, 45, 77, 102, 147, 153, 155, 157, 180

ISHLT The International Society for Heart and Lung Transplantation. 15

ITU Intensive Therapy Unit. 59, 201

KM Kaplan-Meier. 79–81

kPa kilo-pascals. 8

LAS lung allocation score. 27–32, 39–41, 44, 45, 77, 201, 203

LP linear predictor. 81, 82, 114

LR likelihood ratio. 92

MABP multi-armed bandit problem. i, 18, 21, 22

MAE Mean absolute error. 85

MCDM multi-criteria decision making. i, 18, 23, 24, 55–57

MELD Model for End-Stage Liver Disease. 53

NHS-BT National Health Service - Blood and Transplant. 4, 16, 17, 26, 35, 92, 100, 101, 134, 208, 209

NM nautical miles. 31

NT Not Transplantable. 28, 30

NULAS non-urgent lung allocation scheme. 4

NuTH Newcastle upon Tyne Hospitals. 59

OPTN Organ Procurement and Transplantation Network. 14, 15, 26–30, 32, 35, 54–56, 92

$PaCO_2$ partial pressure of CO_2 in arterial blood. 28

PAH Pulmonary arterial hypertension. 14, 40, 41

PaO_2 partial pressure of O_2 in arterial blood. 8, 28

PGD primary graft dysfunction. 203

PH proportional hazards. 80, 203

PTSD post-traumatic stress disorder. 59

PTX post-transplant. 100, 101, 103, 104, 145, 153, 155, 157, 159, 161–169, 172, 184, 187, 190, 193, 196, 198, 199

RCS restricted cubic splines. 82, 83, 135

REST Representational State Transfer. 68

SLT single lung transplant. 45, 77, 102, 147–149, 152, 153, 155, 157, 159, 161–172, 183, 186, 188, 189, 191, 192, 194, 195, 198, 199, 208

SQL Structured Query Language. 66

SRTR Scientific Registry of Transplant Recipients. 57, 107

SULAS super-urgent lung allocation scheme. 4

T Transplantable. 28

TLC total lung capacity. 33

TSAM Thoracic Simulated Allocation Model. 42, 57, 107

TSANZ The Thoracic Society of Australia and New Zealand. 26–35

UK United Kingdom. 4, 7, 10, 11, 15–17, 27–34, 39, 43, 57, 77, 101, 106, 121, 128, 201, 206, 208, 209

ULAS urgent lung allocation scheme. 4

US United States. 14–16, 27, 30, 31, 37, 39, 41–43, 45, 56, 57, 77, 201, 203

WL waiting list. 100–104, 145, 147, 151, 155, 157, 159, 161, 163–166, 168–170, 184, 187, 196

Chapter 1

Introduction

1.1 Motivation and Context: The UK Lung Allocation Policy

In the United Kingdom (UK) there are between 250 and 350 people with life-threatening lung disease on the lung transplant active waiting list at any one time. Of all donor lungs that are offered for transplant, less than 15% are utilised for transplant.[1] As a result, the number of donor lungs utilised for transplant is relatively small compared to the number of candidates, with approximately 100 to 230 lung transplants being performed each year.[2] This scarcity of donors means transplant candidates on the waiting must be prioritised appropriately when donor lungs become available. The criteria that are used to prioritise candidates to receive a specific donor organ are defined in the lung allocation policy. The criteria include both patient characteristics (clinical urgency, blood group etc.) and also the suitability of matching between donor and recipient, such as height mismatch and blood group compatibility.

Designing a lung allocation policy that is transparent and equitable to transplant candidates that also makes optimal use of limited donor lungs is difficult. This difficulty arises from the multiple, often conflicting goals that an allocation policy might be designed to achieve. For example, prioritising the most seriously ill candidates will reduce the number of candidates dying on the waiting list, but this could also lead to lower survival rates post-transplant. Likewise, prioritising candidates with the highest chances of survival post-transplant may result in less seriously ill candidates being transplanted, improving post-transplant survival rates but increasing waiting list mortality.

In this chapter the current adult lung allocation policy in the UK will be described, key issues with the current policy will be identified and potential improvements will be proposed.

1.1.1 POL230/14 - Donor Lung Distribution and Allocation

In the United Kingdom, donor lungs for transplant are allocated according to the 'Donor Lung Distribution and Allocation (POL230/15) Policy'.[3] This policy was created by the Cardiothoracic Advisory Group (CTAG), a group within National Health Service - Blood and Transplant (NHS-BT) that focuses on heart and lung transplantation.

Lung transplant candidates can belong to one of three categories: adult (aged 16 or older, and height above 155cm), small adult (aged 16 or older, and height 155cm or less) or paediatric (aged under 16). There are also three different levels of urgency schemes that a candidate can be assigned to: the non-urgent lung allocation scheme (NULAS), the urgent lung allocation scheme (ULAS) and the super-urgent lung allocation scheme (SULAS).

The allocation policy prioritises candidates by taking into account their category, urgency, compatibility with the donor, waiting time and location. This is accomplished by

using multiple tiers, where each tier selects a subset of candidates on the waiting list. Offers are first extended to candidates in the higher tiers on a named basis for ULAS and SULAS listed candidates. If the offer is not accepted for the named candidate or there are no candidates in that tier then the offer is extended to transplant centres (i.e., *not* using named allocation) in the lower priority tiers.

There is also an adult lung transplant centre rotation which directs donor offers to transplant centres rather than individual candidates. When a centre accepts an offer outside of its zone on the non-urgent scheme, that centre is moved to the bottom of the rotation, allowing other centres which haven't recently accepted offers to have an opportunity to receive donor offers.

The allocation policy tiers are as follows (note: these tier labels have been added to aid referencing specific tiers within this thesis):

Tier 1: All **Adult, Paediatric and Small Adult** patients on the **super-urgent** scheme, ordered by waiting time

Tier 2: All **Paediatric and Small Adult** patients on the **urgent** scheme

- Tier 2.a: Patients with blood group identical to donor, ordered by waiting time
- Tier 2.b: Patients with blood group compatible with donor, ordered by waiting time

Tier 3: All **Adult** patients on the **urgent** scheme

- Tier 3.a: Patients with blood group identical to donor, ordered by waiting time
- Tier 3.b: Patients with blood group compatible with donor, ordered by waiting time

Tier 4: All **Adult, Small Adult and Paediatric** patients on the **non-urgent** scheme requiring a **lung-liver transplant**

Tier 5: All **Adult** patients on the **non-urgent** scheme

- Tier 5.a: All patients at the same centre as the donor (free centre choice)
- Tier 5.b: All patients at Great Ormond Street Hospital (free centre choice)
- Tier 5.c: All patients at remaining centres (prioritised according to the adult lung centre rota)

Tier 6: Patients in Republic of Ireland

Tier 7: Patients at Organ Exchange Organisation in EU countries

Tier 8: Group 2 patients

Candidates are split into 'Group 1' and 'Group 2'. The definitions for these groups are outlined in the 2005 NHS Blood and Transplant Directions document.[4] In brief, Group 1 includes:

- Residents of the UK
- Members of the Armed Forces, Crown servants and employees of the British council employed abroad (including family members under the age of 19)
- People who are entitled under the relevant regulations to medical treatment in the UK
- People entitled to medical treatment by means of a reciprocal health agreement

Group 2 includes all individuals not meeting the criteria for Group 1.

1.1.2 Key Issues

The policy is designed so that specific candidates on the super-urgent and urgent waiting lists are prioritised (i.e. "named allocation") in tiers 1 to 3, with the remaining tiers using centre-based allocation, where allocation is decided not by the policy, but by a "free centre choice". This brings us to the issues identified with the current allocation policy.

The named allocation system in tiers 1 to 3 prioritises urgent and super-urgent candidates using waiting time and blood group compatibility with the donor. It is important that the donor and recipient blood groups are compatible, otherwise an immune response would be triggered leading to hyperacute rejection of the transplanted lungs,[5] an ABO compatibility table is shown in table 1.1. Prioritising blood group identical candidates over blood group compatible candidates does not result in any difference in post-transplant outcomes.[6,7] In terms of post-transplant outcomes these rules do not appear justified. However these rules may be in place to allow equal access to transplant across blood groups. Group 'O' candidates can only be matched with group 'O' donors, making them the least likely group to access transplant. On the other hand, group 'O' donors can be matched with any blood group. Without these rules being in place, group 'O' candidates would have to contend with all other blood groups for access to transplant, further decreasing access.

There are also a number of issues with using waiting time for allocation that will be explained in more detail in chapter 2.4, the key issue is that this mechanism of allocation selects for candidates that can afford to wait longer on the waiting list. Within the higher urgency tiers, waiting time does not adequately take each candidate's individual clinical status into account. Using a mechanism for candidates with high clinical urgency that is better for candidates with low urgency implies this it not the most suitable method of allocation for urgent and super-urgent candidates.

Table 1.1: Blood group compatibility table: Leftmost column specifies donor blood group, topmost row specifies recipient blood type. The compatibility is determined by the intersection of the donor row and recipient column: 'Identical' indicates the donor and recipient have the same blood type. 'Compatible' indicates different blood types between the donor and recipient but the recipient does not have antibodies that will bind to the donor blood cells. 'Not Compatible' indicates different blood types between the donor and recipient, **and** the recipient has antibodies that will bind to the donor blood cells.

Donor \ Recipient	A	B	AB	O
A	Identical	Not Compatible	Compatible	Not Compatible
B	Not Compatible	Identical	Compatible	Not Compatible
AB	Not Compatible	Not Compatible	Identical	Not Compatible
O	Compatible	Compatible	Compatible	Identical

Tiers 4 to 8 make use of a "free centre choice" where the donor lungs are allocated to a centre and the recipient is selected based on the judgement of the transplant team receiving the donor offer. The time of day, who is on call at the time of offer, other activity already happening and experience levels could all impact the allocation decision. This process is highly subjective, not consistent and not transparent to candidates on the waiting list.

Candidates listed at a centre in the same allocation zone as the donor have priority over candidates located at other centres, resulting in access to transplant differing by geographic region. This decision may be an attempt to minimise the cold ischaemia time (CIT) of the donor lungs, defined as the time from cross-clamping the aorta in the donor to reperfusion in the recipient. However, one article observed no significant difference in post-transplant outcomes in the United States (US) lung transplant population with CIT under 8 hours.[8] With the UK being geographically much smaller than the US, prolonged ischaemia times significantly higher than 8 hours are infrequent. Donation after brainstem death (DBD) donors accounted for approximately 63% of transplants in the UK in 2022-2023 and had a median CIT of 6.8 hours (inter-quartile range: 5.5 to 8.8 hours).[2] The remaining 37% of donation after circulatory death (DCD) donors had a median CIT of 8.3 hours (inter-quartile range: 7.2 to 10.1 hours).[2] The overall median CIT was 7.2 hours (inter-quartile range: 6.2 to 9.2 hours).[2] As a result of these considerations, this mechanism may not be necessary for allocation in the UK.

The adult lung centre rota may appear to ensure some degree of fairness, but this is a matter of perspective: is it fair that a candidate does not receive a transplant because another candidate that is unrelated to them was transplanted at some point in the past, and happened to be listed at the same centre? Another issue is that the rota mechanism is very likely to lead to sub-optimal allocation, which will be explained in detail in the following mathematical interlude.

Centre-Based Allocation Leads to Sub-Optimal Allocation

The allocation centre rotation mechanism being sub-optimal can be explained using the following thought experiment. When donor lungs are allocated to a candidate based on their transplant centre, either because the centre is in the same zone as the donor centre, or the centre is at the top of the rota, what is the probability that there was a more suitable candidate who would have benefited more at a different centre? The exact value of this probability isn't important, however, if the probability of sub-optimal allocation $P(S)$ is greater than zero, then the following holds:

1. The probability of optimal allocation is: $P(O) = 1 - P(S)$
2. $P(S) > 0$, therefore: $P(O) < 1$
3. Over n iterations of this mechanism, the probability of making an optimal decision all n times is $P(O)^n$
4. As n approaches ∞, $P(O)^n$ approaches 0

Using this chain of logic, as this mechanism of allocation is repeated over time, the probability of making optimal allocation decisions approaches zero.

The final issue is that the tier-based policy can result in situations where without intervention, clinically similar candidates would have very different access to transplant. This is a result of using **hard boundaries** within allocation policy decisions.

For example, it is possible for two clinically similar adult Interstitial lung disease (ILD) candidates to be on the waiting list with slightly different measured partial pressure of O_2 in arterial blood (PaO_2) values (PO_2 is the partial pressure of oxygen measured in the blood, with the units being in kilo-pascals, or kPa - lower values indicate lower blood oxygen levels). If one candidate has a measured PaO_2 of 7.8 kilo-pascals (kPa) and the other 8.2 kPa, they will be listed at very different priority levels. The listing criteria for urgent ILD candidates in POL231/5[9] are: "Persisting hypoxia (PO_2 <8 kPa) despite continuous O_2 at 10 L/min"

If the policy was followed as it is written, then one candidate would be listed as urgent and the other as non-urgent. In the most extreme case, one candidate could be in allocation tier 3.*a.* and the other at a centre on the bottom of the adult lung centre rota in tier 5.*c.*

This problem of hard boundaries is not unique to the UK lung allocation policy, but is also seen in other lung allocation policies worldwide, more examples of hard boundaries and the issues they cause are explained in section 2.3.1.

In practice, if a candidate almost meets the criteria for being listed as urgent, then their case could go to an adjudication panel. However, the requirement for manual intervention in the allocation process indicates the potential for improving the allocation policy.

1.2 Hypothesis, Goals and Contribution

1.2.1 Potential Improvements

In order to define a hypothesis and set goals for this work, it is first necessary to identify areas for potential improvements to the current UK allocation policy.

The named allocation system for urgent and super-urgent patients could be extended to include all patients at a national level. A national named allocation system would remove the impact of geographical restrictions to accessing transplant and would be more transparent and equitable to candidates on the waiting list.

The current UK lung allocation policy makes no use of predictive survival models for prioritising patients. If a national named system were implemented then candidates could be prioritised by their estimated survival on the waiting list, estimated survival after transplant or some ratio of the two. Since the same survival models would be used for all candidates nationally, the allocation decisions would be transparent, repeatable and more easily auditable.

Another potential benefit of using predictive models is that they can be used to develop an allocation score. The allocation score would rank candidates on a continuous scale, therefore removing the necessity for specifying hard numeric boundaries or allocation tiers. Ideally, a small change in candidate characteristics would result in a minor change in allocation priority, avoiding scenarios where clinically similar candidates would have significantly different access to transplant.

1.2.2 Hypothesis

Hypothesis: Improvements to the current UK lung allocation policy can be made by use of survival analysis and simulation techniques. If the current UK lung allocation policy can be simulated and performance metrics measured, then improvements can be identified by using statistical techniques to compare the current and potential alternative policies. In addition to this, if the current allocation policy is sub-optimal with respect to the performance metrics of interest, then there should exist at least one alternative policy that performs better according to the metrics of interest.

Null hypothesis: There is no statistically significant difference in performance metrics between the current UK lung allocation policy and any alternative simulated policy.

1.2.3 Goals

1. Design and implement a simulation engine to predict the impact of different lung allocation policies according to specific performance metrics
2. Quantify the relative weight (i.e., percentage priority, with all priorities summing to 100%) of goals and values that the lung transplant community (i.e., patients, clinicians and other stakeholders) believe should be part of an ideal allocation policy
3. Using the results generated from goal (2), identify which potential policy most closely aligns with the goals and values of the lung transplant community
4. Using goal (1) and the results from goal (3), compare the current and proposed policies using performance metrics of interest
5. Ensure policies are equitable: all candidates should be prioritised based solely on the same clinical criteria, it should not be possible to unfairly influence candidate rankings, and rankings should not be skewed to benefit or disadvantage specific groups of candidates
6. Ensure policies are auditable: it should be possible to justify allocation decisions and understand the exact reasoning that was undertaken at the time of allocation
7. Ensure policies are transparent to candidates: it should be clear to candidates how their position on the allocation rankings was determined

1.2.4 Contribution to Clinical Transplantation

If the goals of this work are successfully achieved, the following contributions to lung allocation in the UK will be made:

1. Development of predictive models for waiting list and post-transplant survival for the UK lung transplant population (chapter 4.3.1)
2. Development of a novel simulation engine to assess the impact of lung allocation design decisions on different groups of lung transplant candidates (chapter 4.2.7)
3. Identification and quantification of the goals and values of the UK transplant community (chapter 5.2)
4. Development of a framework that combines contributions (1), (2) and (3) to identify the lung allocation policy that most closely aligns with the goals and values of the UK transplant community (chapter 5.2)

While the contributions listed above are specific to UK lung transplantation, this work can generalise to other clinical decision making problems. Some more general contributions are:

1. A general-purpose simulation engine that can simulate the allocation of any limited resource (assuming sufficient data is available)
2. A general-purpose framework for comparing allocation policies and selecting the most desirable policy

In addition to the above, as a result of the pandemic, a rapidly deployable prioritisation system was developed for emergency use when limited or no data is available. If this system is successfully validated then another contribution would be a novel general-purpose emergency prioritisation system.

In the final chapter of this thesis, the hypothesis, goals and contributions will be revisited in the context of the results that were generated in chapters 3 - 5. The degree of success and contribution for each goal will be explored in more detail in section 6.2.

Chapter 2

Background

This section outlines the necessary background before tackling the main goals of this thesis. First, a general overview of lung transplantation will be given, along with a summary of the main indications for lung transplantation. Next, historical and current lung allocation policies are reviewed in order to examine the successes and failures of different policies. Finally, the gap in the literature will be identified and the goals of this thesis will be reviewed in the context of this research gap.

2.1 An Overview of Lung Transplantation

The number of lung transplants taking place each year has consistently increased since the early 90's. Worldwide, there were 33,891 adult lung transplants performed between January 2010 and June 2018, giving an average of just under 4000 per year.[10] This is over a three-fold increase in worldwide lung transplantation compared to the 1300 annual transplants that were performed between 1992 and 2000. The majority of lung transplants worldwide take place in North America (54.9%), followed by Europe (36.6%), with the rest of the world accounting for the remaining 8.5% of adult lung transplants.

2.1.1 Indications for Lung Transplant

Lung transplantation is used for patients with end-stage lung diseases who have failed to respond to maximal medical treatments (i.e., patients where lung function is severely compromised and the lungs are unable to adequately oxygenate their blood and their prognosis is limited). There are four major categories of lung disease, the Organ Procurement and Transplantation Network (OPTN) in the US uses four diagnosis groups labelled A, B, C and D, and these groups will be used throughout this thesis. The four diagnosis groups are: obstructive lung disease (group A), pulmonary vascular disease (group B), cystic fibrosis (group C) and interstitial lung disease (group D).

Group A - Obstructive Lung Disease

Chronic obstructive pulmonary disease (COPD) refers to a group of lung diseases including emphysema and chronic bronchitis. The main cause of COPD is smoking, however it is also possible for COPD to develop in adults who have never smoked. Long-term exposure to harmful fumes or dust can also lead to COPD.[11] Emphysema is a condition where the air sacs in the lungs are damaged and chronic bronchitis is the long-term inflammation of the airways. The main symptoms of COPD are breathlessness, a persistent cough, frequent chest infections and persistent wheezing.

Group B - Pulmonary Vascular Disease

Pulmonary arterial hypertension (PAH) is a condition where there is high blood pressure in the smaller branches of the pulmonary arteries. This is caused by the pulmonary artery walls thickening and becoming stiff, making them unable to expand to allow sufficient blood flow.[12] This puts strain on the right side of the heart which can lead to heart failure. PAH may not present with any symptoms until the condition has advanced. Some of the main symptoms are shortness of breath, tiredness, dizziness, chest pain, palpitations and swelling in the legs, ankles, feet or abdomen. Within the context of transplantation, PAH often refers to idiopathic pulmonary arterial hypertension, where the underlying

cause is unknown,[13] although other causes of pulmonary vascular disease are included in this diagnosis group.

Group C - Cystic Fibrosis

Cystic Fibrosis (CF) is a genetic condition that causes the mucus produced by the body to be thick and sticky, causing problems with the respiratory and digestive systems along with other complications.[14] Within the lungs, this can lead to an increased number of infections and also interferes with gas exchange into the bloodstream. The symptoms of cystic fibrosis are: recurring chest infections, wheezing, coughing, shortness of breath, damaged airways (bronchiectasis), difficulty gaining weight, jaundice and digestive issues. New treatments for CF have transformed the severity of the disease for many people with this condition, and it is now a much less common indication for lung transplantation than it was historically.[15,16]

Group D - Interstitial Lung Disease

The main diagnosis in group D is Idiopathic Pulmonary Fibrosis (IPF), which is part of a group of diseases referred to as interstitial lung disease (ILD). This category of diseases also includes sarcoidosis, hypersensitivity pneumonitis, connective tissue disease related ILD as well as a number of rarer diseases.[9] IPF causes the alveoli in the lungs to become scarred, causing the walls of the alveoli to become rigid, making it more difficult for oxygen to enter the bloodstream.[17] The main symptoms are shortness of breath, persistent cough, tiredness, loss of appetite and clubbed fingers.

2.1.2 Proportion of Waiting List Candidates and Lung Transplant Recipients by Diagnosis Group

Patients within different diagnosis groups experience differing rates of waiting list and post-transplant mortality. Changes to allocation policies result in different transplant rates between diagnosis groups, which in turn directly impacts the probability of a candidate dying on the waiting list and their chances of receiving a transplant. These probabilities indirectly affect the percentage of each diagnosis group on the waiting list. The next section will look at worldwide listing and transplant rates by diagnosis group, then look at rates in the US and UK.

Worldwide

The distribution of candidates within each diagnosis groups receiving a lung transplant has changed worldwide over time. The International Society for Heart and Lung Transplantation (ISHLT) gives an overview of transplant rates stratified by diagnosis group[10] summarised in table 2.1 (note the table also includes patients with alpha-1 antitrypsin deficiency, which results in a predisposition to develop COPD[18]). The two main global trends are:

1. The percentage of group A (COPD) recipients decreased over time from 37.3% in the period 1992-2000 to 26.5% in 2010-2018
2. the percentage group D (ILD) recipients increased from 15.0% in 1992-2000 to 29.0% in 2010-2018

Table 2.1: Worldwide change in transplant rates by diagnosis, 1992 - 2018. COPD = Chronic Obstructive Pulmonary Disease, PAH = Pulmonary Arterial Hypertension, CF = Cystic Fibrosis, ILD = Interstitial Lung Disease, A1ATD = Alpha-1 Antitrypsin Deficiency (results in a predisposition to developing COPD). Source: ISHLT Adult Lung Transplantation Focus Theme[10]

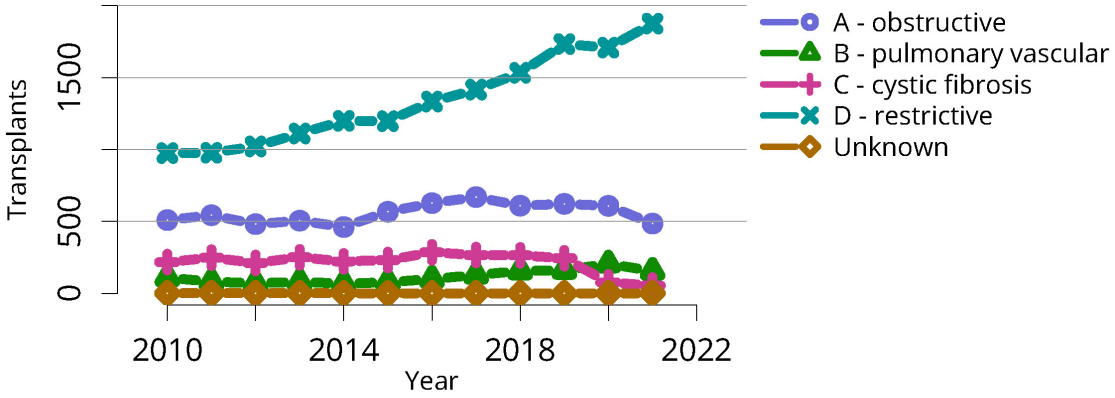
Diagnosis Group	Jan 1992 - Dec 2000	Jan 2001 - Dec 2009	Jan 2010 - Jun 2018
A - COPD	4,162 (37.3%)	7,102 (33.1%)	8,917 (26.5%)
B - PAH	578 (5.2%)	563 (2.6%)	945 (2.8%)
C - CF	1,717 (15.4%)	3,470 (16.2%)	4,771 (14.2%)
D - ILD	1,677 (15.0%)	4,899 (22.8%)	9,755 (29.0%)
A1ATD	1,169 (10.5%)	1,131 (5.3%)	1,054 (3.1%)
Retransplant	394 (3.5%)	921 (4.3%)	1,364 (4.1%)
Other	1,471 (13.2%)	3,392 (15.8%)	6,822 (20.3%)
Total Transplants	11,796	21,806	33,891

United States

The US 2021 annual report released by the OPTN[19] also reveals the same trends in diagnoses receiving a transplant as the global trends, these are displayed in figure 2.1. From 2010 to 2021 the number of group A (COPD) transplant recipients has remained stable at approximately 500 per year, whereas the number of group D (ILD) recipients increased from approximately 1000 per year to almost 1700 per year. As a proportion, group A (COPD) has decreased and group D (ILD) has increased dramatically.

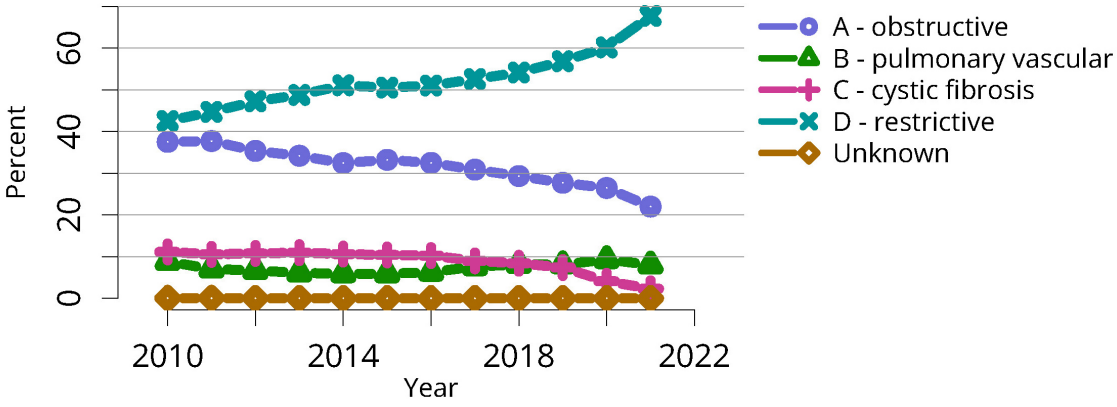
The report also shows the percentage of each diagnosis group on the waiting list.[19] In the US in 2010, group A (COPD) candidates accounted for just under 40% of the waiting list and group D (ILD) candidates just over 40%. The percentage of group A (COPD) candidates decreased to approximately 20% in 2021, and the percentage group D (ILD) candidates increased to slightly below 70%. These trends are displayed in figure 2.2.

Another more recent trend is the decrease in the percentage of group C (CF) candidates being listed and receiving a transplant in the years 2020 - 2021, the report states this was due to the approval of new CF treatments[20] which was discussed briefly on page 16.



OPTN/SRTR 2021 Annual Data Report

Figure 2.1: US transplants by year and stratified by diagnosis group. Note the large increase in the number of group D (interstitial lung disease) recipients and the recent decrease in group C (cystic fibrosis) recipients. Group A - chronic obstructive pulmonary disease, group B - pulmonary vascular disease, group C - cystic fibrosis, group D - interstitial lung disease. Source: OPTN/Scientific Registry of Transplant Recipients (SRTR)[20]



OPTN/SRTR 2021 Annual Data Report

Figure 2.2: US annual percentage of candidates by diagnosis group on the waiting list, from 2010 to 2021. Note the large increase in the number of group D (interstitial lung disease) recipients and the recent decrease in group C (cystic fibrosis) recipients. Group A - chronic obstructive pulmonary disease, group B - pulmonary vascular disease, group C - cystic fibrosis, group D - interstitial lung disease. Source: OPTN/SRTR[20]

United Kingdom

In the UK, NHS-BT publish annual organ-specific reports, the most recent pre-COVID report[21] shows the distribution of transplant rates by diagnosis group covering the period 1 April 2019 to 31 March 2020 (summarised in table 2.2). In that period a higher percentage of group A (COPD) patients received a transplant (36%) than group D (ILD) patients (26%), which is opposite of the distributions seen in the US and worldwide.

The published organ-specific reports only date back to 2013-2014,[22] and transplants by diagnosis have only been reported since 2016-2017,[23] so a comparison of long-term trends was not possible. However, one common trend between the UK and US is the decrease in group C (CF) recipients in 2020.

In the 2019-20 report,[21] CF candidates accounted for 20% of the waiting list, decreasing to 9% in 2020-21,[24] 5% in 2021-22,[25] and slightly increasing to 10% in 2022-23.[26]

This reduction in listing of candidates with CF is reflected in the reduced percentages of CF recipients: 24% in 2019-20,[21] 14% in 2020-21,[24] 8% in 2021-2022,[25] and 6% in 2022-2023.[26]

Summary

While the diseases in the four categories have various causes and symptoms, one thing they all have in common is that as the disease progresses lung transplantation is the only viable treatment option. The availability of donor lungs suitable for transplantation is relatively small compared to the number of patients on the waiting list. Therefore it is important to make optimal use of this precious and scarce resource. The next two sections look at policies that have been implemented globally as well as the history and evolution of lung allocation policies.

Table 2.2: UK transplant rates by diagnosis, 2019 - 2020 and 2022 - 2023. COPD = Chronic Obstructive Pulmonary Disease, PAH = Pulmonary Arterial Hypertension/Pulmonary Vascular Disease, CF = Cystic Fibrosis, ILD = Interstitial Lung Disease. Source: NHS-BT Annual Report on Cardiothoracic Organ Transplantation 2019-2020[21]

Diagnosis Group	Transplants in 2019-20 (%)	Transplants in 2022-23 (%)
A - COPD	56 (36%)	51 (25%)
B - PAH	6 (4%)	9 (5%)
C - CF	38 (24%)	19 (10%)
D - ILD	40 (26%)	101 (50%)
Other	16 (10%)	19 (10%)

2.2 Possible Frameworks 💻

There are many frameworks that can be used for looking at the problem of allocating donor lungs. The framing of the problem will inform the methods that are used, the types of results generated and the conclusions that are reached. As such, it is important to carefully consider the framework that will be used throughout this thesis. In this section three frameworks will be explored, and the justification for the chosen framework is given.

The first framing approaches allocation by looking at the potential benefit each patient gains from transplant along with the probability of a successful transplant. Each patient is then assigned a probability of being allocated the donor organs accordingly.

The next approach is the multi-armed bandit problem (MABP), where every possible donor-recipient combination is mapped to a range of probabilities of various outcomes. The goal is to maximise benefit to recipients in the presence of uncertainty of outcomes.

The final framework will look at lung allocation as a multi-criteria decision making (MCDM) problem. There are 'multiple criteria' that an allocation policy must consider, for example: waiting list mortality, post-transplant outcomes and equity of access to transplant. Designing an allocation policy to maximise one of these criteria often necessitates compromising on the other criteria. MCDM techniques can be used for optimal decision making in the presence of (sometimes conflicting) multiple criteria.

2.2.1 Social Welfare Functions

Donor lung allocation can be seen as attempting to maximise certain social welfare functions.[27] A social welfare function quantifies the benefit to a population resulting from a distribution of goods. In this case the 'goods' are the distribution of the *probability* of being transplanted given an allocation policy, and the 'benefit' is an abstract quantity that in practice could relate to minimising waiting list mortality, maximising survival post-transplant, or increasing the additional days of life gained from transplant.

This concept is described by Steven M. Goldman using three social welfare functions:[28]

1. *Utilitarianism* - the lungs would be allocated to the candidate with the highest *potential* benefit from transplant (i.e., the candidate has both a high probability of a successful transplant and a high amount of potential benefit.)

2. *Rawlsian Ethics* - the probabilities of any one candidate being selected would be distributed to maximise the expected utility for the *least well off individual*. (i.e., the candidate with the lowest quantity of potential benefit receives a higher probability of being allocated)

3. *Nash Social Welfare Function* - every candidate has an equal probability of being selected for transplant, regardless of their individual characteristics.

These social welfare functions each have a unique ethical philosophy guiding the approach to distributing donor lungs. There are four principles that should guide the underlying philosophy of any allocation policy:[29]

1. Principle of equity - in the context of organ allocation this refers specifically to equity of access to transplantation and can be considered analogous to the concept of *equality of opportunity*, and opposite to the concept of equality of outcome. This principle dictates that there should be no unfair bias in how candidates are prioritised for transplant, and that certain characteristics (such as ethnicity, wealth, religious or political affiliations) should not be considered in the allocation process.

2. Principle of justice - individuals receive that which is owed to them. In the context of transplantation, this could be giving priority to individuals that have previously donated an organ, or who have registered as organ donors. The inverse of this principle also applies: individuals **do not** receive that which is **not owed** to them, or in other words, no unjustified advantages should be given to one individual over another

3. Principle of beneficence - this is a multi-faceted principle:[30]

   (a) Do not actively harm others

   (b) Do not passively allow harm to happen to others, and actively prevent harm

   (c) Remove harm that has occurred to others

   (d) Any intervention or treatment should provide positive benefits

4. Principle of utility - any intervention or treatment should result in an equal or greater amount of net good compared to any other alternative action

It is important that any policy being evaluated or proposed for implementation should adhere to these principles. The degree to which each type of policy or allocation system adheres to each of the four principles will vary, but nonetheless these principles should be the standard all policies are held to.

There are pros and cons both practically and ethically for each of these social welfare functions. These functions are described mathematically by Goldman[28] and are summarised in the following mathematical interlude using the same notation given by Goldman.

Mathematical Interlude - Social Welfare Functions

Mathematically, Goldman defines the problem of organ allocation as a set of I candidates $M = \{1, ..., I\}$, where each candidate i receives some amount of *benefit* from receiving a transplant. The benefit for candidate i is defined by the utility function U^i, and the probability of a successful transplant is defined as p_i. The allocation policy is represented by a set of probabilities π, where π_i represents the probability of candidate i being allocated the donor organ, with the condition that every probability is zero or greater, and the sum of all probabilities equals one. A more formal re-statement of these conditions is:

$$(\forall i \in M.(\pi_i \geq 0)) \wedge \sum_{i=1}^{I} \pi_i = 1$$

Applying Goldman's definitions to the context of this thesis, 'benefit' could refer to additional life gained from transplant, increased probability of survival to a certain point in time, improvement in quality of life and so on. The probability of a successful transplant p_i depends on the individual candidate's attributes and the characteristics of the donor lungs.

A utilitarian policy would allocate lungs to the candidate that has the highest *potential* benefit from transplant, or mathematically the candidate for which $p_i U^i$ is the largest in the set of candidates M. This will tend to select candidates with both good chances of a successful transplant and a high expected benefit from transplant. The candidate that meets these criteria has a probability of being selected for transplant $\pi_i = 1$ and for all other candidates $\pi_i = 0$. This type of policy maximises the utility from donor lungs but also has the largest inequality in the distribution of probabilities for candidates being selected.

A Rawlsian policy assigns a probability π_i to each candidate so that the well-being of the *least well off individual* is maximised. Mathematically, π_i is distributed between candidates $\{1, ..., I\}$ so that *expected* benefit $(\pi_i p_i U^i)$ is equal across the population. This assumes that every candidate has a non-zero probability of a successful transplant and a non-zero benefit. The result of this is that candidates with high expected benefit are assigned low probabilities of receiving an organ and candidates with low expected benefit are assigned high probabilities of receiving an organ.

The probability assigned to every candidate i when using the Rawlsian policy can be calculated as follows:

$$\text{invsum} = \frac{1}{\sum_{i=1}^{I} p_i U^i}$$

$$\pi_i = \frac{\frac{1}{p_i U^i}}{\text{invsum}}$$

A policy of this type does not maximise utility but does minimise the inequality of expected benefit by ensuring $\pi_i p_i U^i$ is equal for all candidates. On the other hand, the probability of being selected for transplant (π_i) is still distributed unevenly.

A policy based on the Nash social welfare function simply assigns a probability of $\pi_i = 1/I$ to each candidate. This minimises the inequality of the probability for each candidate being selected for transplant, but results in unequal expected benefit. A policy of this type also does not maximise the utility from the donor organs.

In summary, the utilitarian policy aims to maximise the total utility (i.e. benefit) across the population by allocating to candidates with a high probability of experiencing a large transplant benefit. The Rawlsian policy aims to equalise outcomes by assigning low probabilities of allocation to candidates with high expected benefit and vice versa. The Nash policy gives all candidates an equal probability of being transplanted, regardless of potential benefit or the probability of a good outcome.

In practice, a utilitarian policy is the only justifiable type of policy because it does not depend on random processes, making it the only approach that is transparent and auditable (i.e. repeatable). Donor lungs are a scarce resource and should be allocated to maximise the benefit to the recipient while also avoiding futile transplants with a low probability of success, which would lead to sub-optimal use of the donor lungs. Furthermore, the Rawlsian and Nash policies require a candidate to be selected at random. The goal for any policies proposed from this work is that they should be fair (i.e., equitable), auditable and transparent to candidates (see page 12). A random process that cannot be predicted or repeated would not meet the standard to achieve these goals.

As a result of these considerations, approaching the problem of organ allocation from the point of view of social welfare functions is not appropriate for evaluating and developing an allocation policy. However, what can be gained from this framing is the importance of clearly defining what is meant by 'benefit' and the ability to accurately predict outcomes and benefit given a candidate-donor pairing.

2.2.2 The Multi-Armed Bandit Problem MABP

In the previous section (2.2.1), it was assumed that each candidate had a fixed, known probability of a successful transplant and a fixed, known quantity of transplant benefit. When an individual is transplanted with lungs from a donor the probability of a successful transplant can be influenced by a number of factors relating to both the recipient, the donor, and the reaction of the recipient's immune system to the donor lungs. This probability of a successful transplant and the amount of benefit gained can be thought of as being drawn randomly from a probability distribution.

The multi-armed bandit problem (MABP) is a classic problem in decision theory,[31] operations research[32] and reinforcement learning,[33] a branch of probability theory and machine learning. The problem can be understood by the following analogy:

Imagine you have been offered several hundred free attempts at a row of slot machines at a casino. You want to maximise your potential payout, however each slot machine pays out with a different probability distribution. How many free attempts should be allocated to each slot machine in order to maximise your payout?

At the beginning of the scenario you have no knowledge of the probability distributions for the slot machines, however as the number of attempts increases more information is revealed about the payout probabilities. The next question arises: should you continue to pull the lever for the slot machine that has the best payout so far, or experiment with other machines that may have a better probability of paying out? This is the exploration versus exploitation problem - if you spend too much time exploring different slot machines you don't maximise your payout, conversely staying at the same slot machine ('exploitation') will prevent the discovery of a potentially better paying machine.

The MABP approach can be applied in the context of lung transplantation: the combination of recipient and donor characteristics constitute a single 'arm' of the multi-armed bandit. The waiting list and post-transplant survival durations of these combinations of characteristics can be thought of as being drawn from a probability distribution. The exploration versus exploitation problem also applies in the context of organ allocation: should organs be allocated to patients with characteristics that are known to have good outcomes? Or should they occasionally be allocated to patients with rare combinations of characteristics to discover more information on survival probabilities for these rare cases?

As the algorithms for optimising the payout from the MABP are probabilistic rather than deterministic, this framing of lung allocation does not appear to be appropriate for a fair, auditable and transparent allocation system, for the same reasons as outlined on page 23. However the concept of mapping patient and donor characteristics to probability distributions will be revisited later in this thesis, as this will be necessary to simulate the potential outcomes after lung transplant.

The details of the probabilistic nature of the commonly used MABP optimisation algorithms are given in the following mathematical interlude.

MABP Optimisation Algorithm Examples 💻

Two commonly used algorithms are the *epsilon greedy*[34] and *epsilon first*[35] algorithms. For both of these algorithms a value for epsilon (ϵ) needs to be specified, this determines the percentage of exploration in the strategy. For example, with $\epsilon = 0.1$, 10% of the strategy will be spent on exploration, with the remaining 90% exploiting the lever with the highest payout.

The epsilon first algorithm has a distinct exploration phase followed by an exploitation phase. If N lever pulls are going to be taken in total, then ϵN lever pulls will be performed at random. The lever with the highest payout is identified and pulled repeatedly for the remaining $(1 - \epsilon)N$ pulls.

The epsilon greedy strategy randomly switches between exploration and exploitation with a probability ϵ for exploration and $1 - \epsilon$ for exploitation. In the case of exploration, one of the levers is pulled at random. In the case of exploitation the lever with the best payout so far is pulled.

2.2.3 Multi-Criteria Decision Making MCDM

Allocation can be seen as making an optimal decision between two or more competing choices. What is meant by 'optimal' can be subjective and depends on the goals of the decision maker(s). Multi-criteria decision making (MCDM) is the discipline that studies decision making in situations where there are multiple conflicting criteria. This is the framing that will be used for this research and throughout this thesis.

The first concept of MCDM that will be explored is the 'Pareto set'. The relevance of the Pareto set can be explained via an analogy of choosing a house to purchase. For simplicity only two criteria will be used in this example: house price and house size, with the competing goals being to minimise the amount of money spent, and maximise the size of the purchased house. The prices and sizes of the houses used in this example are shown in table 2.3.

To find the Pareto set, simply select any house and remove all other houses from the list that are both *more expensive* and *smaller*. This makes logical sense, as there is no point paying more money for a smaller house. This process is repeated for each house until there is a set of houses remaining that have not been removed.

Table 2.3: Example house prices and sizes to choose from for a hypothetical multi-criteria decision making (MCDM) problem to illustrate the concept of the Pareto set. In this example, there are two conflicting goals ('criteria'): (1) purchase the largest house, and (2) spent as little money as possible. House 1 is the cheapest (but smallest) and house 7 is the largest (but most expensive). This set of houses can be reduced to the Pareto set by removing any house where there is a larger option that is also cheaper. Houses 4 and 5 can be removed, since house 3 is both larger and cheaper. *Italics* indicate the options that do not belong to the Pareto set.

Option	House Price (Thousands)	House Size (Sq Ft)
1	85	1000
2	115	1300
3	150	1700
4	*170*	*1400*
5	*200*	*1500*
6	375	2100
7	500	3000

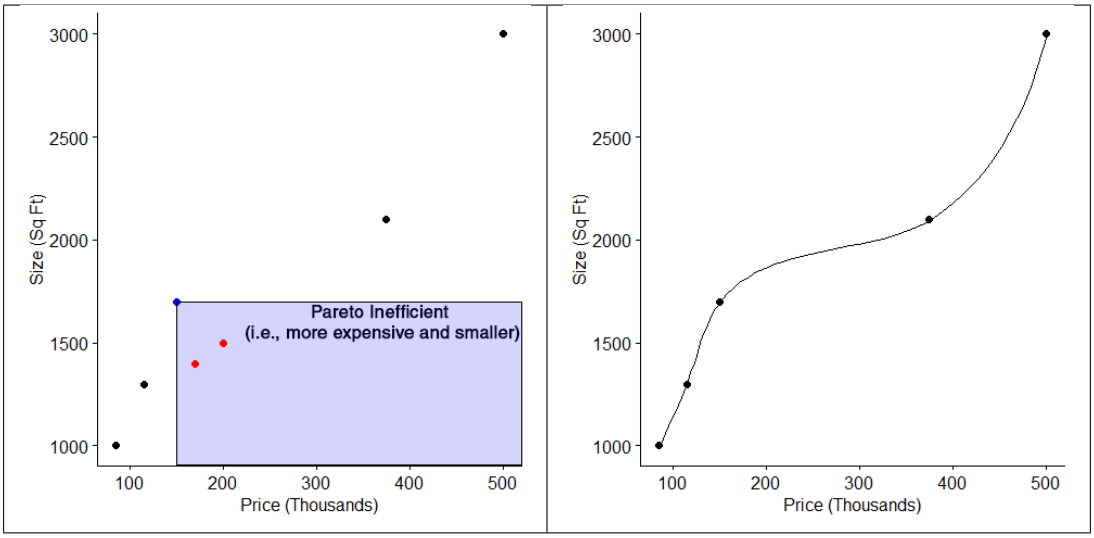


Figure 2.3: Left: Scatter plot of house price vs house size. The house points in red are said to be 'Pareto inefficient' due to them being more expensive and smaller than the house shown with a blue point. Right: The Pareto set / Pareto front of houses that have prices that are Pareto efficient with respect to their size and other houses on the market.

In this example, house 4 costs 170 thousand with a size of 1400 sq ft, and house 5 costs 200 thousand with a size of 1500 sq ft. Both of these options can be removed as they do not belong to the Pareto set (indicated by *italics* in table 2.3) and red in the left side of figure 2.3, since house 3 is cheaper (150 thousand) and also larger (1700 sq ft) than houses 4 and 5 (shown in blue in the left of figure 2.3). In the MCDM vernacular, houses 4 and 5 are said to be 'dominated' by house 3.

The remaining houses are referred to as being 'in the Pareto set', 'on the Pareto front', or being 'Pareto optimal'. The houses remaining in the Pareto set have the following property: when comparing any two houses, one will have a lower price and the other will have a larger size, but not both. The houses that have been inefficiently priced in relation

to their size and other available houses have been removed.

House 1 in table 2.3 represents the cheapest option and house 7 represents the largest option. Houses 2, 3 and 6 represent a trade-off between cost and house size. The house that represents the subjective "best" option depends on the decision maker's opinion of which house represents the optimal trade-off between price and size.

If for example, 100% of the decision was determined by the size of the house, then price is no longer a consideration and house 7 would be chosen as it is the largest. However, if 60% of the decision is determined by price and 40% by size, then the decision becomes more nuanced.

Weighted Criteria

The percentage of preference for each of the criteria, such as the 60% / 40% assigned to price and size in the previous example are referred to as criteria *weights*. How can the percentage values for these weights be determined?

Within the discipline of MCDM, a process can be used for determining the weights of multiple criteria for a specific decision maker or group of decision makers called the analytic hierarchy process (AHP).[36] The AHP will be explored later in chapter 5, along with the principles of Pareto optimality and weighted criteria.

These same principles can be applied to a waiting list for transplant candidates. Criteria like clinical urgency, anticipated cold ischaemic time, blood group compatibility, predicted survival benefit and sensitisation status can be used. The "best" decision depends on the candidate with the optimal balance between all of these criteria. Depending on which criteria are deemed to be more important, it may be decided to select a recipient that is more distant from the donor because they have higher clinical urgency, or someone that is ABO compatible instead of identical because they have a higher predicted survival benefit.

These principles can be applied again at the policy level. Rather than focusing on *how* to prioritise patients on the waiting list, the focus can be shifted to *what* an allocation policy should achieve. Several policies could be compared using criteria such as number of waiting list deaths, survival duration after transplant and average waiting time for a transplant. Depending on the relative importance of the criteria, one of the policies will be identified as most ideal.

Justification

Framing lung allocation as a MCDM problem appears to be the most appropriate option for this research. The process of reducing a waiting list to a Pareto set is deterministic and can be repeated, thus making this technique auditable and repeatable (goal 6). Candidates on the waiting list could be assigned a number of points for each of the (clinically relevant) criteria used for allocation. This would allow any candidate to see exactly why they were

ranked at a certain position on the waiting list for a given lung offer, resulting in a system that is transparent (goal 7). The use of MCDM methods also avoids the shortcomings of the MABP, and the Rawlsian and Nash social welfare policies, which required random processes that were not transparent (see section 2.2.1 on page 20).

At the policy level, stakeholders in relation to lung transplantation (i.e., medical professionals, lung transplant candidates and recipients) could be involved in the process of determining the weights of allocation policy goals. The justification for selecting one specific policy out of a set of possible policies can be given using the weights of allocation policy goals and the fact that those goals were determined by both medical professionals and individuals who have accessed, or require access to lung transplant.

Whether it's a candidate that is being selected for transplant, or a policy that is being selected for implementation, if the selection is from the Pareto set then you are guaranteed that the choice will meet the *minimum* conditions for being considered optimal: the choice is not sub-optimal (i.e., there are no options that are comparatively better on all criteria of interest) but not guaranteed to be the most optimal. The most optimal choice would be determined by the weights of the criteria relevant to the decision being made.

2.3 Global Lung Allocation Policies

The allocation of donor lungs to potential recipients must be carefully considered in order to ensure the best possible outcomes for patients, as motivated in section 1.1. There are multiple criteria that need to be considered when deciding who the optimal recipient should be. The specific criteria and relative importance of these criteria are a matter of policy. This section will give an overview of how lung allocation policies differ worldwide.

Lung allocation policy documents were sourced from the relevant organisations in each country. Each policy document was reviewed in order to identify the main features of how the allocation policies function in that country. Each feature of the policy was classified as 'common' between multiple policies or 'unique'.

Features that are common between policies were identified and are described in section 2.3.1. Features that differed between policies are discussed in section 2.3.2. The organisations and documents for each country/region are shown in table 2.4.

While this section is structured into common and unique features, to aid in navigating this section, the country/region(s) being discussed are shown in **bold** at the beginning of each relevant paragraph. The UK/NHS-BT policy is abbreviated **UK**, the Eurotransplant policy is abbreviated **ET**, the US/OPTN policy is abbreviated **US**, the Scandiatransplant policy is abbreviated **SC**, and the The Thoracic Society of Australia and New Zealand (TSANZ) policy is abbreviated **AU**.

2.3.1 Common Features

This section compares the general features that are common between most or all of the reviewed allocation policies. Although the general features are shared between policies, each policy has its own variation of the feature. The features discussed in this section are: medical urgency, paediatric status, ABO matching, waiting time, geographic boundaries and sequential allocation.

Common Feature #1 - Allocation Philosophy

Each policy has a set of guiding principles for the allocation of organs, referred to here as the 'allocation philosophy'. The allocation philosophy of a policy can be determined by looking at which social welfare function it maximises. The two main social welfare functions observed were utilitarian and Rawlsian (For more information on social welfare functions, see section 2.2.)

The utilitarian approach allocates the donor lungs to the individual who would benefit most from receiving a transplant. Overall, the utilitarian approach seeks to maximise the benefit of the group as a whole, regardless of how that benefit is distributed.

The Rawlsian approach on the other hand would allocate the donor lungs to the individual who is perceived to be 'worst off'. Note that the definition of 'worst off' is

Table 2.4: Lung transplant policies, the organisation responsible for lung allocation, and the corresponding countries/regions of implementation that were reviewed.

Abbr.	**Organisation**	**Countries Served**	**Document Title**
UK	NHS-BT	England Scotland Wales Northern Ireland	POL 230/11[37]
ET	Eurotransplant	Austria Belgium Croatia Germany Hungary Luxembourg The Netherlands Slovenia	ET Thoracic Allocation System (EThAS)[38]
US	OPTN	The United States of America	OPTN Policies[39]
SC	Scandiatransplant	Denmark Finland Iceland Norway Sweden Estonia	Guidelines for Organ Exchange in the Scandiatransplant Area[40]
AU	The Thoracic Society of Australia and New Zealand (TSANZ)	Australia New Zealand	Clinical Guidelines for Organ Transplantation from Deceased Donors[41]

subjective and can take many criteria into account - it should not be interpreted as meaning the individual with the worst prognosis. Overall, the Rawlsian approach aims to maximise the benefit of the least-benefited individual.

US The lung allocation score (LAS) defines benefit as the net difference between the candidate's expected survival on the waiting list (in days) and expected survival after transplant, with twice as much weight being assigned to waiting list survival as post transplant survival.[42] A purely utilitarian approach would allocate the lungs to the individual with the highest LAS, however the Eurotransplant and OPTN policies take additional criteria into account.

ET The Eurotransplant policy tracks the total number of organs offered and accepted from each country. If a country receives more organs than it offers to other countries, it is said to have a positive balance of exchange, conversely, countries that have received fewer organs than they have offered have a negative balance of exchange. The Eurotransplant policy prioritises candidates in countries with a negative total balance of exchange (see section 2.3.2 for more on this). The LAS is used in two countries within the Eurotransplant region: Germany and the Netherlands.

US The OPTN policy initially prioritises candidates who are within 250 nautical miles of the donor. This is a pragmatic approach to the allocation of donor lungs in a country the size of the US where prolonged CIT are more likely, however, the OPTN final rule[43] does state allocation policies "shall not be based on the candidate's place of residence or place of listing, except to the extent required [...]". Additional sub-criteria are also used in both policies, however the overall philosophy of the policies can be categorised as utilitarian.

UK, SC In the UK and Scandinavia, allocation is primarily determined by medical urgency, as urgency (or 'priority') is one of the first criteria considered in the allocation policy. Prioritising medical urgency reduces waiting list mortality and is an attempt to reduce the probability that a candidate experiences the worst possible outcome (i.e., dying without a transplant). This leans more in the direction of the Rawlsian philosophy, however there are still elements of a utilitarian approach in these countries. Candidates are not listed for transplant if they would not benefit from it,[9] therefore benefit is a factor. Nevertheless, avoiding waiting list mortality is more heavily weighted in these countries than in the rest of the Eurotransplant region and in the US.

AU The TSANZ policy prioritises compatibility (in terms of size, ABO and human leukocyte antigen (HLA) matching) between the donor and recipient. This can also be classed as a utilitarian approach if 'benefit' is thought of in terms of: (1) - the benefit that a recipient would receive from transplant and (2) - the probability of the candidate receiving that benefit. While the Eurotransplant and OPTN policies seek to maximise (1), the TSANZ policy maximises the probability that the patient does not experience post-transplant complications, hence receives the benefit of transplant (2).

Common Feature #2 - Medical Urgency

UK In the UK medical urgency is defined as three tiers: super-urgent, urgent and non-urgent. There are clearly defined criteria for listing within a specific tier.[9] For example, the super-urgent lung allocation scheme is for patients supported via extracorporeal membrane oxygenation (ECMO) or interventional lung assist (iLA) as a bridge to transplant. The

prioritisation follows a logical order: super-urgent candidates take priority over urgent candidates, who take priority over non-urgent candidates.

ET The Eurotransplant policy defines urgency as a combination of High Urgency (HU), Transplantable (T), Not Transplantable (NT) and the region to which it applies: international, national, regional or local. There is one exception: there is no HU international status due to the use of LAS.

US The OPTN policy has two priority tiers for paediatric candidates. Priority 1 is for candidates less than 12 years old who have evidence of either respiratory failure or pulmonary hypertension. Priority 2 is assigned to candidates who are not priority 1. There is no specific indication of medical urgency for adult candidates due to the use of LAS.

SC The Scandiatransplant scheme has four priority tiers: Priority 0, 1, 2 and 3. Priority 0 is similar to the UK's Super-urgent classification and is for patients on extra-corporeal support. Priority 1 is for patients with rapidly progressing organ failure and a poor prognosis. Priority 2 is for all other transplantable candidates and priority 3 is for candidates who are currently not transplantable. There are limitations to the number of candidates who can be listed as priority 0/1; each centre is limited to three candidates per year.

AU TSANZ define clinical urgency using the level of support required as well as the rapidity of deterioration of the candidate. The level of support includes: ECMO, invasive mechanical ventilation, non-invasive ventilation, requiring high/low flow O_2, prolonged/recurrent hospitalisation and use of other support devices. The rapidity of deterioration is determined by significant changes in measurements of lung function such as partial pressure of oxygen and carbon dioxide (PaO_2 and partial pressure of CO_2 in arterial blood ($PaCO_2$)) or 6-minute walk test distance, development of certain conditions or an escalation in level of support.

Common Feature #3 - Paediatric Status

UK In the UK, a donor under the age of 16 at the time of death is classified as a paediatric donor, likewise a candidate under the age of 16 at the time of listing can be classed as a paediatric candidate. Paediatric candidates keep their paediatric status even if they reach their 16^{th} birthday while on the waiting list. There are two lung transplantation centres in the UK which perform paediatric transplants: Newcastle (Freeman Hospital) and London (Great Ormond Street Hospital). The combination of age and centre of listing determine the paediatric status, for example, a 15 year old listed at a non-paediatric centre will be listed as requiring adult-sized organs. However, a 15 year old listed at one of the paediatric centres would be listed as a paediatric candidate. It is not possible to be listed

simultaneously as both an adult and a paediatric candidate. The category a candidate is listed as determines their chances of accessing transplant, as paediatric donor lungs are first offered to paediatric candidates in each of the urgency tiers before being offered to other patient categories.

ET, US Both the Eurotransplant scheme and OPTN define a paediatric candidate as aged under 12 years old. For candidates aged 12 years or older the LAS is used for allocation. Paediatric candidates registered on the Eurotransplant scheme automatically receive a LAS of 100. The OPTN policy allows exceptions to be made (an 'approved adolescent classification exception'), where a candidate aged less than 12 years old will receive offers based on their calculated LAS.

The allocation of paediatric donor lungs differs between the Eurotransplant and OPTN policies, as shown in tables 2.5 and 2.6.

Table 2.5: Eurotransplant prioritisation of candidates based on paediatric donor age.

Rank	Donor aged < 12 years	Donor aged $12 - 17$ years
1	Candidates aged < 12 years	Candidates aged $12 - 17$ years
2	Candidates aged $12 - 17$ years	Candidates aged < 12 years
3	Candidates aged 18+ years	Candidates aged 18+ years

Table 2.6: OPTN prioritisation of candidates based on donor paediatric status. *While LAS is used for candidates aged 12 or over, for candidates aged less than 12 waiting time is used to sort candidates.

Rank	Donor aged < 18 years	Donor aged 18+ years
1	Candidates aged < 12 years*	ABO compatible/identical candidates aged 12+ years
2	Candidates aged < 1 year*	Candidates aged < 12 years*
3	-	Candidates aged < 1 year*

SC The Scandiatransplant policy does not explicitly mention age or paediatric status. This was the only policy where this was the case.

AU The TSANZ policy mentions the nationally funded Alfred Hospital which is the sole paediatric lung transplant centre in Australia. The centre recommends an age range between 6 and 16 years old for referral. There is no specific allocation algorithm outlined for paediatric donors or recipients.

Common Feature #4 - ABO Blood Group Matching

All countries have ABO matching rules, however the rules for matching the lung donor to candidates based on ABO compatibility differ. The relative importance of ABO compatibility is also different depending on the policy in place in each country.

UK, US In the UK and US, there is a strict approach of blood group identical candidates always being prioritised above blood group compatible candidates. This is not supported by the literature in terms of being predictive of outcomes post-transplant,[44] and in fact may be causing increased waiting list mortality.[45]

ET Eurotransplant defines two different rule groups for donor-recipient ABO matching: ABO modified and ABO compatible. The ABO compatible rules are the standard compatibility rules for blood type matching. The ABO modified rules are similar to ABO identical matching rules but there are two exceptions: a blood type 'A' donor can be allocated to a blood type 'AB' recipient and a blood type 'O' donor can be allocated to a blood type 'B' recipient. The ABO modified rules are used for allocation before the ABO compatible rules.

SC The Scandiatransplant policy states that ABO compatibility is required for organ exchange, however the donor-recipient match is the responsibility of the transplant centre. There is no explicit algorithm giving priority to either compatible or identical candidates.

AU The TSANZ policy is unique in that ABO compatibility is the first allocation factor, making it the top priority for allocation. It is prioritised above clinical urgency and logistic concerns such as the proximity of the donor to the recipient.

Common Feature #5 - Waiting Time

UK In the UK waiting time is used for allocation in the case of a paediatric or an adult donor. Waiting time is the final criterion that candidates listed on the super-urgent and urgent schemes are prioritised by.

ET The Eurotransplant policy uses waiting time as the fourth criterion on which patients are ordered by. Age group, ABO compatibility and LAS are used to prioritise candidates before waiting time. In some cases, time listed as HU is used instead of total waiting time. Candidates do no accumulate waiting time while being listed as NT.

US The OPTN policy uses waiting time differently depending on the age of the candidate. Candidates at least 12 years old are first prioritised by LAS and then prioritised by total active waiting time. For candidates less than 12 years old, paediatric priority waiting time is first used to prioritise candidates, followed by total waiting time.

SC, AU There is no mention of waiting time in the Scandiatransplant policy and the TSANZ policy only uses waiting time as a tie-breaker when all other factors are equal, making it the lowest priority of the allocation criteria.

Common Feature #6 - Geographic Boundaries

Each policy has a concept of geographic boundaries, however the exact rules on organ exchange across boundaries varies; as well as how boundaries are defined.

UK In the UK each transplantation centre has a defined geographic region. For non-urgent candidates the offer is first made to individuals at the same centre before being offered to candidates in other centres (based on rotas).

ET The Eurotransplant policy has different rules depending on the age of the donor, the country the donor is located and the LAS category (high/low) that candidates are categorised as. Geographic boundaries follow the national boundaries of each country. However, in some cases candidates from one country are listed alongside candidates in another country. LAS is used for international allocation, but the balance of organs shared between countries is considered

US In the US, boundaries are concentric circles with the radius measured in nautical miles (NM). This is a recent change that came about from a lawsuit, and will be discussed in more detail in section 2.4 on page 47. For adult donors the range starts at 250NM and increases to 500NM, 1000NM, 1500NM, 2500NM and then the entire nation. Paediatric lungs are first offered to paediatric candidates within a 1000NM radius, and then to various subsets of candidates on the waiting list starting again at 250NM from the donor.

SC The Scandiatransplant policy first offers lungs to local high priority (priority 0/1) candidates, then to high priority candidates in the same country, then to high priority candidates in other centres. If the offer is still not accepted at this point, the offer is then made to priority 2 candidates locally or nationally and then to priority 2 candidates located at other centres.

AU The TSANZ policy follows state boundaries and the offer is first made to the recognised lung transplant unit in the same state as the donor. If the home state does not accept the offer, then the offer is made to units outside the home state according to a rotation (see section 2.3.2 for more on rotation or rota-lists.) Units outside of the home state have 30 minutes to respond to the offer.

Common Feature #7 - Sequential / Tier-based Allocation

Although each policy takes a different approach to allocation, the sequential nature that the policies are applied is common between all policies. Within each policy a sequence of criteria are specified and an offer is made to all candidates matching these criteria. If the offer is not accepted or no candidates match the criteria, then the next group of criteria in

the sequence are applied and the process is repeated until the offer is accepted or rejected by all candidates.

UK, US For example, in the UK the offer is first made to all super-urgent candidates, then to blood group identical candidates on the urgent scheme, then to blood group compatible candidates on the urgent scheme and so on. In the US even though the LAS is used the policy is still sequential. The first group of candidates are those at least 12 years old, ABO identical to the donor, and within 250NM of the donor, they are then prioritised by LAS. The next group is the same as before but for candidates who are ABO compatible with the donor. This approach is repeated for different priority and age groups with increasing range from the donor.

To use a sequential approach, hard boundaries need to be specified. Two examples of hard boundaries in the OPTN policy are the 250 nautical mile radius and the paediatric age threshold of 12 years old. This leads to the problem of edge cases that can be illustrated with two nearly identical candidates. Candidate 1 is 12 years old, is registered within 250 nautical miles of the donor and has an identical blood type to the donor. Candidate 2 is 11 years and 364 days old, registered 251 nautical miles from the donor and has an identical blood type. The only difference is 1 day in age and 1 nautical mile, however the order of the offering is very different. Candidate 1 is in the first group to receive an offer, whereas candidate 2 is in the ninth group.

Another consequence of using a sequential approach is that allocation is heavily weighted towards the criteria in the first few steps of the allocation policy.

UK, SC In the UK the main priority is candidates on the super-urgent and urgent lists leading to medical urgency being the primary factor in allocation. The Scandiatransplant policy is similar, it first allocates to priority 0/1 candidates, also placing emphasis on medical urgency.

ET The Eurotransplant policy primarily emphasises international high LAS candidates in countries which have a negative total balance with the donor country. Since international high LAS candidates in countries with a zero or positive balance are not considered, the main factor in the Eurotransplant policy is balance of organ exchange.

US In the OPTN policy, any candidate not within 250NM of the donor is not considered at first. Next, priority is given to candidates who are blood type identical to the donor, they are then prioritised by LAS (highest to lowest). The main factor in this policy is the proximity of the candidate to the donor.

AU The TSANZ policy first allocates based on ABO compatibility, size compatibility and then cross-matching. The first three factors are all related to how compatible the

recipient is with the donor, making donor-recipient compatibility the main goal.

2.3.2 Unique Features

In this section the unique features of policies are outlined, they are: candidate size, country balance, rota lists and cross-matching.

Unique Feature #1 - Candidate Size

UK, ET, AU The UK, Eurotransplant and TSANZ policies take the size of the candidate into account for allocation.

UK The UK has the only policy that gives a definition for a small adult. A small adult is defined as a candidate with a height at or below 155cm. The small adult category is only used for super-urgent and urgent candidates. All candidates on the super-urgent and urgent schemes can specify gender-specific minimum and maximum donor height criteria, they will be screened if the donor parameters are not within the specified bounds.

ET The Eurotransplant policy uses total lung capacity (TLC) to assess compatibility between the donor and recipient. The TLC is calculated using a formula that includes the height and sex of the donor. The policy has a unique feature where if a donor is not accepted by a certain point in the offering sequence, then the acceptable threshold for minimum and maximum TLC expands to 10% lower and 20% higher in Germany, or 10% lower and 10% higher in countries other than Germany and the Netherlands.

AU The TSANZ policy determines size compatibility by using chest x-ray measurements and TLC. Size compatibility is the second criterion in the allocation algorithm, behind ABO compatibility.

Unique Feature #2 - Country Balance

ET The Eurotransplant policy has a unique feature that attempts to equalise the distribution of lungs between countries. This is accomplished by calculating the country balance of two countries 'X' and 'Y' using the following formula, where $T_{x,y}$ refers to the number of transplants performed in country x using lungs from country y:

$$\text{Country Balance} = T_{x,y} - T_{y,x}$$

The balance is calculated taking all lung transplants into account from 1st September 2004.

If country X has a negative balance relative to country Y then country X has donated more lungs to country Y than it has received from country Y. The allocation policy

prioritises candidates in countries with a negative total balance to the donor country over countries with a zero or positive total balance.

While on the surface this may appear to accomplish some type of fairness, it in fact is almost guaranteed to lead to sub-optimal allocation. The proof follows the same logic as described on page 10:

If there are N countries but a country does not prioritise those with a positive balance then there are $N-N^+$ possible allocations. The probability that the most optimal recipient was not excluded from consideration is $\frac{N-N^+}{N}$. If this process is repeated j times then the probability that the most optimal recipient was not excluded becomes $(\frac{N-N^+}{N})^j$. This value approaches zero as j increases.

In practice there may be multiple candidates which would all be considered close to an optimal choice, therefore the above formulae are not an exact model, however the principle is the same: if populations of candidates are excluded based on non-medical criteria then the probability increases of sub-optimal allocation occurring.

The unfairness of this type of allocation can be seen by looking at an individual level. A candidates who could receive a life-saving transplant can be overlooked because someone else completely unrelated to them - who just happened to live in the same country - already received imported donor lungs.

The final problem with this approach is that it enforces a vision of what the distribution of donor lungs *should* be rather than what would benefit the population as a whole. If one country has an excess of donor organs and another country has a greater need and allocation is based solely on medical criteria, then interfering with the supply and demand can only do harm by preventing organs being exported to countries with a greater than average need. This decision may be influenced by political considerations rather than utilitarian ones, however, political influences on organ allocation are beyond the scope of this thesis.

Unique Feature #3 - Centre Rotation-lists

UK, SC Both the UK and Scandinavia make use of 'rota-lists' which, like country balance, attempt to equalise the distribution of donor lungs between areas. Rota-lists create the same problems as country balance, just on a centre-level instead of a national level.

UK In the UK three separate rota-list are maintained: the Adult Lung Centre Rota, the Paediatric Lung Centre Rota and the Small Adult Rota. Each rota-list contains a ranked list of lung transplant centres. When a centre accepts an offer for an Adult/Small Adult/Paediatric recipient, that centre is moved to the last position on the relevant rota.

SC In the Scandiatransplant area, rota-lists are separated by the type of transplant (heart, lung or heart-lung) and divided into six regions: Copenhagen/Aarhus, Gothenburg,

Helsinki, Lund/Stockholm, Tartu and Oslo. Each region has a rank on each of the three rota-lists, and the accepting centre is moved to the last position on the rota-list.

UK In the UK, rota-lists are only used for non-urgent candidates whereas in Scandinavia rota-lists are only used for the exchange of organs across national boundaries, where there are no high priority local or national candidates accepting the offer.

AU The TSANZ policy also allocates according to a rotation that is kept by the donor coordination team in each state. The exact mechanism and rules on how the rotation operates is not outlined in the policy document.

Unique Feature #4 - Cross-matching

AU The TSANZ policy is the only policy which has 'absence of a T-cell cross match' and 'anti-HLA antibody profiles' as a factor in allocation. This is to reduce the probability of the recipient experiencing a hyperacute immune response to the donor lungs.[5]

UK The NHS-BT lung selection policy 231/4[9] mentions that HLA typing and antibody screening is performed as one of the immunology blood tests, however it is not mentioned explicitly in the allocation policy.

US The OPTN policy allows allocation exceptions to be made for highly sensitised candidates. If there are candidates ranked above the highly sensitised candidate at a different transplant centre, then a request can be made for the other candidates to turn down the offer.

Unique Feature #5 - Single Lung Transplantation

UK, AU The logistics for offering a single lung are only outlined in the NHS-BT and TSANZ policies. The NHS-BT policy states that if a centre is only accepting a single lung then it **must** specify which side (left or right) is being accepted. This allows the other centres to know exactly which side is being offered.

AU In the TSANZ policy, the type of transplantation is categorised under 'Logistics' in the fourth tier of the allocation algorithm. The final decision on the type of transplant (lobar, single, bilateral or heart/lung) is made by the accepting lung transplant unit.

2.3.3 Summary

In summary, each country has its own approach to prioritising candidates for the allocation of donor lungs. There are many features in common between policies, and all of the policies reviewed use a tier-based/sequential approach to allocation. All policies had different

priority levels according to medical urgency and paediatric status, generally prioritising paediatric candidates and candidates with higher clinical urgency. Geographic boundaries, ABO matching, and to a lesser extent, waiting time were all featured in the allocation policies reviewed, however the relative priority of these criteria differed between countries.

In general, the allocation systems can be categorised into one or more of the following:

Centre-based allocation Donor organs are allocated to transplant centres in the same geographic region as the donor hospital according to pre-defined geographic boundaries. These systems can minimise CIT in line with the principle of beneficence, preventing harm to the donor lungs and thus the recipient. However, geographic boundaries can and do lead to inequitable access to transplantation based on location (as discussed in section 1.1.2 and will be discussed in greater detail in section 1.1.2), violating the principle of equity, and by extension, the principle of justice. In section 1.1.2 on page 10 it was mathematically proven that centre-based allocation also violates the principle of utility.

National Urgency Candidates are prioritised at a national level according to their degree of clinical urgency. This is also in line with the principle of beneficence, by actively preventing harm (i.e., death on the waiting list) and removing harm that has occurred (i.e., the lung disease requiring listing). It may be considered to not be in line with the principle of equity, due to the fact that access to transplantation varies by clinical urgency, however, the principal of equity refers to *unfair* bias, and biasing allocation towards those with greatest need can be considered a fair bias.

Score-based Allocation Candidates are prioritised according to an allocation score. This strongly emphasises the principle of utility (at least in the case of the LAS) as this prioritises candidates according to expected benefit (i.e., additional lifespan gained from transplant). This approach also aligns with the principle of beneficence in that clinical urgency is also considered in the score. There are limitations to this approach however: if allocation scores use survival models as part of their calculation (as the LAS does), the scores are subject to the assumptions that guided the preparation of data and generation of the models, and survival models aren't always accurate. While they may make correct predictions the majority of the time, there will be some percentage of cases where they prioritise the wrong candidate (see 'Accuracy of Survival Times' in section 4.4.4 on page 161 for more on this). This fact weakens the alignment with the principles of beneficence and justice as there will always be cases where allocating according to the score isn't as optimal as what should have been done in reality.

The next section will review historical lung allocation policies, how they have changed over time, and how those changes have impacted different groups of patients.

2.4 History and Impact of Lung Allocation Policies ☿

This section focuses on policy-level decisions for allocating donor lungs to potential transplant recipients. The key historical changes in allocation policy are outlined along with the impacts and consequences of these policy changes on different candidate and recipient demographics. Proposals for alternative allocation algorithms are also evaluated for their potential to improve lung allocation.

While much of the existing literature focuses on the consequences of changes to policy on candidates and recipients with certain diagnoses, this section summarises many of the consequences of policy-level decisions in one place. These high-level decisions have significant impact on patient's lives, and influence factors such as waiting list mortality, post-transplant survival and quality of life. This section aims to provide a summary for policy makers to understand the impact of policy-level decisions, outline the pros and cons of historical policy decisions, and highlight where current policies can be improved.

Methods

Three databases were queried: Scopus,[46] Web of Science[47] and PubMed[48] using the following search term:

```
"lung allocation" AND ("policy" OR "algorithm")
```

The titles, abstracts and keywords were queried and the results were limited to journal papers in the English language. All articles from Jan 2000 until March 2021 were combined from the three databases to give 71 articles. The following exclusion criteria were used:

1. Articles not focused specifically on lung (2 excluded)
2. Articles relating to healthcare costs (3 excluded)
3. Articles unrelated to the policy-level decisions of selecting a transplant recipient for donor lungs (6 excluded)
4. Articles that were paywalled and which we did not have institutional access (12 excluded)
5. Other review articles (4 excluded)
6. Focuses exclusively on paediatric lung transplantation (5 excluded)

This led to 32 of the 71 articles being excluded, leaving 39 articles for the review. Within each article the affected population(s) were noted along with the positive and negative consequences of the allocation policy.

Results

Allocation Based on Waiting Time

In the US, lung allocation was originally based on waiting time and resulted in candidates who could survive waiting a longer duration for a transplant being favoured.[8] This is evidenced by a median waiting time of just over 3 years, and candidates aged 11 to 64 having reduced waiting list mortality despite older candidates having lower waiting times. Allocation was primarily to candidates with COPD (group A) despite the higher waiting list mortality of other lung diseases.[49–52]

Impact by Diagnosis Candidates with group A (COPD) diagnoses had the greatest chance of receiving a transplant due to their condition being stable and predictable. In the US in 2003, 40% of transplants were for recipients with COPD despite many COPD recipients not experiencing a survival benefit (i.e., additional days of life gained) from undergoing transplantation.[8] Therefore, prioritising based on waiting time selects for candidates who would experience the lowest benefit from transplant.[49] There is also mixed evidence for the survival benefit gained from lung transplantation for recipients with COPD (compare[50] and[8]). For a large number of recipients, this method of allocation did not result in survival benefit.[53]

Recipients with group B (PAH) diagnoses accounted for only 4% of transplant recipients in 2003 and group C (CF) recipients accounted for 16%.[8] Recipients with group D (ILD) diagnoses were at a major disadvantage under a waiting time based system due to the rapid progression of their lung disease.[8] In the US in March 1995, candidates with IPF (ILD - group D) were listed with an additional 90 days of waiting time, however it was common practice to list candidates early before they required transplantation so they could accumulate time on the waiting list.[49] The consequence of this practice was inflation in waiting times: 64% of candidates waited more than a year for transplant and 44% had waited more than 2 years for a transplant with a median waiting time of just over 3 years.[8] In 2003, group D recipients still only accounted for 22% of transplants in the US. Candidates in groups C and D had excessive waiting list deaths compared to candidates in group A (COPD).[49,54]

Impact by Sex Men and women experienced different waiting times and mortality rates. There were 10 to 20 percent more women on the waiting list in the US between 1995 and 2005, possibly leading to longer waiting times for women.[8] Men were more likely to have more severe lung diseases, leading to a higher annual death rate for men (141 deaths per 1000 patient years for men vs 121 for women).

Impact in Europe In the Eurotransplant region, similar observations were made to those seen in the US in a study following transplant candidates and recipients from 1990 - 1996.[51] Despite having the lowest waiting list mortality rate (15% two years after listing), COPD candidates (group A) had the highest transplant rate, with 62% of COPD candidates being transplanted two years after listing. Pulmonary fibrosis (ILD - group D) candidates had a comparable transplant rate to COPD candidates (59%) however they had just over twice the rate of waiting list mortality (31%).

The next highest transplant rate was for cystic fibrosis (group C) candidates (55% at two years), with a waiting list mortality of 34%. Candidates with PAH (group B) had the lowest transplant rates (49%) but also comparable waiting list mortality (27%). A summary of these figures are shown in table 2.7.

Two years post-transplant, cystic fibrosis recipients had the highest survival rate (72%), followed by COPD and pulmonary fibrosis recipients (56%) with the lowest survival rate occurring in recipients with PAH (group B - just under 50%). The post-transplant outcomes are summarised in figure 2.4.

Waiting list outcome, 2 years after listing

Table 2.7: Outcome on waiting list for candidates listed from 1st Jan 1990 to 1st Jan 1999 (n = 744) in the Eurotransplant region. Source: De Meester et al.[51] COPD = Chronic Obstructive Pulmonary Disease, PAH = Pulmonary Arterial Hypertension/Pulmonary Vascular Disease, CF = Cystic Fibrosis, ILD = Interstitial Lung Disease.

Diagnosis Group	% Transplanted	% Death	% Removal	% Waiting
A - COPD	62	15	7	16
B - PAH	49	27	4	20
C - CF	55	34	1	10
D - ILD	59	31	2	8

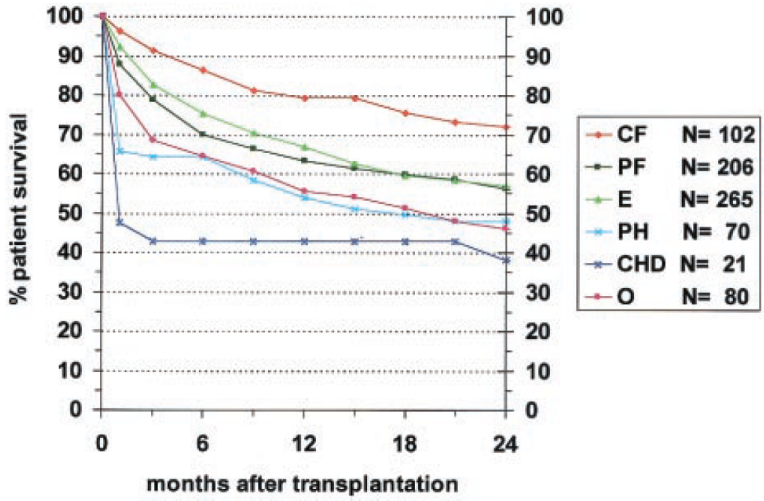


Figure 2.4: Post-transplant survival curve, stratified by diagnosis group. CF = Cystic Fibrosis, PF = Pulmonary Fibrosis, E = Emphysema, PH = Pulmonary Hypertension, CHD = Coronary Heart Disease, O = Other. Source: De Meester et al.[51]

Allocation Based on Free-Centre Choice

In the UK, lung allocation was primarily centre-based (and still is for non-urgent candidates - see page 6), with lungs being allocated according to a free centre choice.

A study in 2009 looked at the survival benefit gained by adult lung transplant recipients in the UK, stratified by disease.[50] At the time of the study, lungs were allocated by geographical proximity, with the centre allocating according to ABO and size compatibility. If the local centre did not accept the lungs then the offer would go out to other centres according to a rota (see page 38 for an explanation of allocation rota-lists). The national urgent and super-urgent schemes were not implemented at this time, however urgency may have been taken into account when lungs were allocated locally.

There are mixed results when looking at transplant benefit for COPD (group A) recipients. Previous studies conducted in the US show that for the majority of COPD recipients there was no survival benefit from transplant,[8] however the 2009 UK study[50] used a different methodology for calculating benefit and showed that COPD recipients had a clear survival benefit. It was shown that 83% of COPD recipients survived to the point where post-transplant risk of death was less than the risk of death on the waiting list, and that the survival benefit experienced was significant. Pulmonary fibrosis (ILD - group D), and cystic fibrosis (group C) recipients also experienced a clear survival benefit, with cystic fibrosis recipients most rapidly gaining survival benefit after transplant.

Introduction of the Lung Allocation Score

The LAS was introduced in the US in 2005 to replace the waiting time system.[53] The LAS is calculated for each suitable candidate on the waiting list, where higher scores indicate greater potential benefit from transplant. Survival models were developed to predict the duration of survival on the waiting list and post-transplant. The difference between these two predicted survival durations is the number of additional days of life gained from transplant ('net benefit'). Waiting list survival was given twice the weight of post-transplant survival, giving additional priority to candidates with high clinical urgency. The calculation is as follows, according to United Network for Organ Sharing (UNOS):[42]

Definitions:

PTX = Predicted days of life post-transplant over the next year

WL = Predicted days of life on the waiting list over the next year

Calculation:

$$\text{Raw Score} = \text{PTX} - (2 \times \text{WL})$$

$$\text{LAS} = 100 \times \frac{\text{Raw Score} + 730}{1095}$$

Introduction of LAS in US Introducing the lung allocation score still resulted in reasonable access to transplant across all diagnosis and age groups[53,55] however the distribution of recipients did change. The LAS was successful in reducing the size of the waiting list and the number of offers required to allocate the donor lungs. Within the first few weeks of implementation the active waiting list decreased from 1700 to less than 1500 candidates and the median number of offers to place the lungs dropped from 12 to 5. This helped address the the high turndown rates under the waiting time based system.

Shifting the emphasis away from waiting time led to waiting times decreasing and more medically urgent candidates being listed who previously would not have been able to accumulate enough waiting time to be transplanted.[56] Waiting time decreased from 680.9 ± 528.3 days to 445.6 ± 516.9 days (a 35% decrease, $p < .001$).[57] There was concern that increased listing of candidates with higher clinical urgency would result in an increase in waiting list mortality, however there was no observed increase in waiting list mortality,[54–57] likely due to these candidates receiving higher priority under the LAS system.[56]

Impact by Diagnosis With the introduction of the lung allocation score, rankings for candidates in different diagnosis groups changed significantly. Candidates with COPD (group A) ranked lower and candidates with IPF (ILD - group D) higher.[54,57,58] This reflects the fact that candidates with IPF or CF (group C) benefited most from lung transplantation and candidates with COPD benefited the least.[59]

Candidates with PAH (group B) saw an increased transplantation rate[60] but also an increase in waiting list mortality.[60,61] There were few predictors of outcome for PAH candidates[53] and LAS scores may not have been accurate for these candidates.[62] Candidates with PAH were less likely to be transplanted than candidates with IPF and CF, and they were also more likely to die on the waiting list than candidates with COPD and CF.[61]

A small single-centre study (n=45) observed that the introduction of the LAS resulted in significantly lower rankings for group A (COPD) candidates and significantly higher rankings for group D (ILD) candidates, while groups B and C were minimally affected.[58]

A larger study (n = 13,040) calculated the transplant benefit to recipients by diagnosis group, using a variation of an accelerated failure time model instead of the Cox Proportional Hazards (PH) model[59] (an introduction to survival analysis and explanation of the Cox PH model are given in appendix A). Group A recipients had the lowest survival benefit, with only the most seriously ill recipients experiencing a survival benefit. Overall there was no expected survival benefit at 2 years for group A. The greatest potential benefit was with group C (CF) and group D (ILD) recipients. Group C recipients had 54.5% greater survival benefit than group A recipients, and also over 99% of group C recipients experienced a survival benefit. For group D recipients, IPF was the most common indication for transplant. A full summary of benefit by diagnosis group is given in table 2.8. The implementation of LAS also increased the number of recipients aged over 65.[54,59]

Table 2.8: Table showing percentage of patients experiencing a survival benefit from transplant 1, 2 and 3 years after transplant. Original source: Vock et al.[59] COPD = Chronic Obstructive Pulmonary Disease, PAH = Pulmonary Arterial Hypertension/Pulmonary Vascular Disease, CF = Cystic Fibrosis, ILD = Interstitial Lung Disease.

Diagnosis Group	1 Year	2 Years	3 Years
A - COPD	22.1%	39.2%	56.3%
B - PAH	52.9%	64.6%	73.0%
C - CF	95.3%	98.9%	99.7%
D - ILD	90.0%	94.8%	97.7%

A multicentre study assessed the impact of LAS on short-term outcomes post-transplant.[57] This study saw results similar to Vock et al.[59] and Lingaraju et al.:[58] there was a decrease in COPD (group A) recipients and an increase in IPF (ILD - group D) recipients. While the Lingaraju study[58] observed no real change in group B (PAH) and C (CF) recipients, this multicentre study observed a decrease in PAH and CF recipients.[57] One potential weakness of this study is that the size of the waiting list was estimated using data points from only four time intervals, potentially leading to inaccurate estimates of waiting list mortality. This is due to the fact that the size of the waiting list was assumed to be constant throughout the study (equal to the average of the four time intervals). In reality, waiting list sizes fluctuate with time, so the actual percentage of candidates dying on the waiting list at any point in time will differ from this study's calculated percentage.

The diagnosis-specific impacts of the LAS were analysed by Chen et al.[61] This study also observed a decrease in priority for group A (COPD) candidates and an increase in priority for group D (ILD) candidates. However, while waiting list mortality decreased for candidates in diagnosis groups A, C and D, there was no decrease in waiting list mortality for group B (PAH) candidates, and in fact increased significantly. The probability of PAH candidates receiving a transplant also decreased relative to the other diagnoses, with the percentage of PAH candidates being listed decreasing from 6.8% to 3.7%. There was no real impact on post-transplant survival post-LAS, however post-transplant mortality was higher for group B relative to groups A and C pre-LAS.

The impact of LAS on candidates with PAH was investigated by Schaffer et al.[60] This study also observed an increase in waiting list mortality for candidates with PAH despite increased transplant rates.

The number of candidates receiving a transplant with ILD (group D) overtook COPD (group A) with the introduction of the LAS. After five years, waiting list mortality decreased from 500 to 300 per year and the number of lung transplants performed annually doubled.[54]

Introduction of LAS in Europe Historically, lung allocation in Europe was based on the transplant surgeon's judgement of each individual case.[63] However, expectations arose that scarce donor lungs should be allocated in a predictable and justifiable fashion. National allocation according to urgency and waiting time was in place until 2007. The Eurotransplant Lung Allocation Score (ET-LAS) was later developed which took into account additional criteria which aren't in the US version of LAS, such as the use of ECMO.

Implementing LAS in Germany in December 2011 resulted in a 26% reduction in waiting list mortality and also improved survival rates one year post-transplant.[64] As in the US, there was also increased transplant rates amongst IPF (ILD - group D) candidates. Implementation of the LAS diverted allocation away from more stable COPD (group A) candidates towards candidates with IPF, PAH (group B) and CF (group C).[64,65]

Removal of Donation Service Area (DSA) Boundaries

Prior to the US system of allocating in concentric 250nm circles from the donor centre, allocation was based on geographic zones referred to as a donation service area (DSA). One analysis showed that it was possible to increase the chance of receiving a transplant by just over a factor of 2 by strategically listing in a different DSA.[66] Another problem with using DSA boundaries was that candidates with lower clinical urgency would be transplanted while candidates in adjacent DSAs with higher urgency would die while waiting for a transplant.[67]

In November 2017, allocation was changed so that DSA boundaries were no longer used. Instead lungs were offered to candidates within 250nm of the donor centre, then 500nm and so on. This change was triggered by a lawsuit that was filed by a transplant candidate against OPTN, citing that donation service areas being the first unit of allocation was in direct contravention to National Organ Transplant Act (NOTA),[68] the OPTN final rule[43] and sound medical judgement.[69] As with any change in allocation policy, there is always the risk of unintended consequences. One study looked at some of the possible unintended consequences and found that removal of DSA boundaries may have resulted in COPD (group A) candidates having a lower likelihood of receiving a transplant, and also a decline in lung utilisation rates.[70] There was also a slight increase in the donor discard rate, however there was no change in waiting list mortality or recipient characteristics.

A simulation study was also conducted using the Thoracic Simulated Allocation Model (TSAM).[71] The simulation was able to predict general population-level trends: that there would be no difference in waiting list mortality or transplant rates as a result of removing DSA boundaries. However, the simulation made incorrect predictions for specific diagnosis groups: decreased transplant rates for group A (COPD) and increased transplant rates for group D (ILD) candidates did not occur in reality. The simulation did agree with observed results in that group D candidates had the highest transplant rates. While the simulation model incorrectly predicted diagnosis-specific population level trends, a decrease in group

A recipients and increase in group D recipients was observed by Drolen et al.[72]

Drolen at al. compared the impact of this policy change at high and low competition centres.[72] It was observed that there were fewer group A (COPD) recipients post-DSA and that median waiting times for these candidates decreased from 99 days to 70 days. The number of group D (ILD) recipients increased and was near statistical significance, but did not reach the threshold for significance (p = 0.07). Group B (PAH) candidates also had waiting times decrease from 56 days to 28 days. For group C (CF) candidates, median waiting times increased from 58 to 65.5 days, though this was not statistically significant (p = 0.09). There was no change in waiting times for group D candidates. Overall, all candidates had a 9 day decrease in median waiting time (p = 0.001). These results are summarised in tables 2.9 and 2.10.

Table 2.9: Number of recipients by diagnosis group, pre- and post-DSA. Original source: Drolen et al.[72] COPD = Chronic Obstructive Pulmonary Disease, PAH = Pulmonary Arterial Hypertension/Pulmonary Vascular Disease, CF = Cystic Fibrosis, ILD = Interstitial Lung Disease.

Diagnosis Group	DSA era (n = 2336)	Post-DSA era (n = 2435)	P value
A - COPD	723 (31.0%)	679 (27.9%)	.02
B - PAH	112 (4.8%)	135 (5.5%)	.24
C - CF	253 (10.8%)	256 (10.5%)	.72
D - ILD	1248 (53.4%)	1365 (56.1%)	.07

Table 2.10: Median waiting list time (days) by diagnosis group, pre- and post-DSA. Original source: Drolen et al.[72] COPD = Chronic Obstructive Pulmonary Disease, PAH = Pulmonary Arterial Hypertension/Pulmonary Vascular Disease, CF = Cystic Fibrosis, ILD = Interstitial Lung Disease.

Diagnosis Group	DSA era (n = 2336)	Post-DSA era (n = 2435)	P value
A - COPD	99	70	.006
B - PAH	56	28	.0008
C - CF	58	65.5	.09
D - ILD	35	36	.36

A different simulation study by Mooney et al.[67] had some results that complemented the observed results published by Drolen et al.[72] and Lehr et al.[71] Mooney's study also predicted fewer group A (COPD) recipients and a greater number of group D (ILD) recipients post-DSA. The number of group B (PAH) and group C (CF) recipients was also predicted to increase slightly. While Drolen et al. focused on recipient characteristics, Mooney et al. simulated the impact on waiting list mortality. All diagnosis groups had a lower predicted waiting list mortality as a result of removing DSA boundaries, with group D candidates benefiting the most from the change in policy. However this was also a simulation study and these were not observed effects in the population.

One small study (n = 101) did see an increase in waiting list size,[73] however population density varies massively across the United States, therefore this result is not representative of all transplant centres in the US.

Alternative Allocation Algorithms

Extended Criteria Donors One study evaluated at the effect of utilising extended criteria donors (ECD) at Hannover Medical School - a transplant centre located in the Eurotransplant region. ECDs are defined as donor lungs being rejected by three centres due to characteristics of the donor lungs.[65] Due to COPD candidates (group A) having the most stable and predictable condition, they were the target demographic for this algorithm. Group A candidates accounted for 12.6% of recipients with regular allocation, but that increased to 55.6% using this alternative allocation algorithm. Older recipients (age 53.7 ± 11.7 years) also tended to be the recipients of extended criteria donors. There were fewer PAH (Group B) and ILD (group D) recipients with the ECD algorithm as these candidates were not the target demographic. CF (group C) candidates were not significantly affected. Despite using ECD lungs, post-transplant survival was not affected at 90-days or 27 months post-transplant.

Impact of Cold Ischaemia Time The US and UK lung allocation policies allocate based on geography which has the effect of minimising CIT. It was found that CIT beyond six hours does not affect short or long term survival and CIT beyond eight hours had no effect on 1 or 5 year graft survival rates. One unexpected result was that CIT beyond eight hours resulted in greater 5 year survival, however this may be due to lungs with high CIT being reserved for candidates with certain diagnoses.[8]

Informative Censoring One important factor that needs to be accounted for when making survival predictions is informative censoring. Lungs tend to be allocated to candidates with a higher risk of death on the waiting list, and as a result there is a correlation between the risk of death and probability of censoring due to transplant. This correlation between risk of death and risk of censoring is known as 'informative censoring' (or 'dependent censoring').[74] If informative censoring is not taken into account, this results in survival predictions underestimating the risk of waiting list mortality. In the literature two different methods for correcting for informative censoring were mentioned. The most common approach was Inverse Probability of Censoring Weighting (IPCW)[75] but multiple imputation was also used.[75,76] One study showed that using IPCW correction for calculating transplant benefit results in higher predicted benefit for higher urgency candidates.[77] One example given was a 55 year old patient with IPF (ILD - group D): not accounting for informative censoring resulted in overestimating waiting list survival by 220 days. This shows the importance of correcting for informative censoring as that candidate would have been prioritised significantly higher as a result of the correction.

Boundary-less Allocation The use of geographic boundaries, regardless of whether they are DSA-based or concentric rings, can lead to situations where candidates that are lower urgency are prioritised ahead of candidates that are higher urgency simply

because of an arbitrary boundary.[67] A framework for allocating organs without geographic boundaries was outlined by Snyder et al.[78] Allocation would be based on scores that take into account both geographic feasibility and medical priority. This approach would allow candidates with higher clinical urgency to access organs that are located at more distant transplant centres. This concept is taken further by Stewart et al.[79] where a composite allocation score (CAS) was developed and compared to the existing LAS. The CAS was calculated by considering age group, LAS, proximity to donor centre, and type of ABO match. The main benefit explained by Stewart et al. was that a points-based system allows multiple candidate attributes to be taken into account simultaneously, preventing a single attribute (such as age or geographic location) to override all other attributes to determine a candidate's priority.

Use of Single-Lung Transplant One study proposed a change in allocation to increase the number of single lung transplants for COPD (group A) candidates.[80] There was a difference in risk of death post-transplant depending on whether a left lung or right lung was transplanted. In COPD recipients, the non-transplanted lung can experience hyperinflation, which has more of an effect in left-lung transplants than right-lung transplants. This results in left-lung transplants being higher risk for COPD patients, however, the difference in post-transplant survival between a left-lung transplant and a right or bilateral lung transplant (BLT) decreases with recipient age. The authors propose that the size of the donor pool can effectively be increased by allocating single lungs to COPD candidates, especially for candidates aged over 65.

From approximately 1996 until 2000, single lung transplants were more common than double lung transplants. A single centre retrospective analysis (n=339) in the US showed a clear improvement in post transplant survival with BLT compared to single lung transplant (SLT) (hazard ratio = 0.583, p = 0.02), especially for recipients with COPD (group A).[52] A larger national cohort study (n=1997) in the UK from July 1995 to December 2007 observed the same: survival with SLT was significantly worse for COPD recipients compared to BLT. However, for ILD (group D) patients, there was no difference in survival between SLT and BLT.[50]

Additional Diagnosis-Specific Variables There were also additional variables identified that could increase access to transplant for group C (CF) candidates.[81] An updated model was determined to have minimal impact on the rankings of group B (PAH) and D (ILD) candidates. However for 36.8% of group C candidates that died on the waiting list, the updated model would have resulted in an LAS increase of ≥ 5 points which would have increased their access to transplantation without adversely affecting access to transplant for candidates in other diagnosis groups.

Summary

Designing a lung allocation policy requires balancing the competing priorities of access to transplant, waiting list mortality, post-transplant survival and transplant benefit. Depending on the design of the policy, different sub-groups of candidates will either benefit or be disadvantaged. Younger candidates with less serious diagnoses tend to benefit from waiting time based systems, as they can afford to wait for transplant.

The lung allocation score achieved its objectives of giving wider access to transplant across all diagnosis and age groups, while allocating to more clinically urgent candidates that would benefit most from transplant. There are some limitations however: the lung allocation score is only as good as the accuracy of the survival predictions, therefore it is important to ensure all relevant clinical variables are taken into account. It is also important to account for informative censoring as this can drastically overestimate waiting list survival, resulting in lower scores and more limited access to transplant. Finally, geography still has an impact on access to transplant, potentially limiting the benefits of the LAS.

The general direction of lung allocation appears to be towards boundary-less national allocation systems. The composite allocation scoring system being implemented in the US will address the issues of geographic disparity in access to transplant. Whether or not this is successful and becomes the next major step in lung allocation waits to be seen.

2.5 Research and Literature Gap

In this section the research and literature gap will be identified and how this research will address this gap will be described.

2.5.1 Lung Allocation and Analytic Hierarchy Process (AHP) Literature Gap

The use of the analytic hierarchy process (AHP) was a decision that was arrived at independently of the work being completed in the US with the development of the CAS. Upon discovering the OPTN were also planning to utilise the AHP, this was interpreted as positive confirmation of the choice to apply the AHP to lung allocation.

Google scholar, Scopus,[46] Web of Science[47] and PubMed[48] were searched using the search phrase (using syntax appropriate for each website):

```
"lung allocation" AND (AHP OR "analytic hierarchy process")
```

Each website returned between 2 and 10 results. After screening papers not relating to lung transplantation and/or the AHP, the same three papers were returned.

One of the papers was the US CAS presented by Stewart et al.[79] that was discussed in the previous section. The main focus of this paper was comparing the rankings using the proposed composite scores to the existing LAS system. The paper mentions how the OPTN are planning on using the AHP to *inform* the development of a points based system, but the AHP is not being used directly to design or select a lung allocation policy. The second paper was a review article by the same lead author[82] that focuses on many of the same points as the CAS paper. The final paper was an 'Expert Insight' article (an article that focuses on the opinion or perspective of one or more authors, rather than presenting new data) in the the 'Transplantation' journal by Martha Pavlakis[83] which also discusses continuous distribution of lungs and mentions the same points about the AHP being used to inform policy design.

2.5.2 Lung Allocation and Simulation Literature Gap

Another focus of this research will be the use of simulation methods to simulate lung allocation policies. The following search phrase was used on the same four websites:

```
"lung allocation" AND "simulation"
```

The results were manually reviewed and 23 relevant publications were found.

One key article was identified for general organ allocation simulation,[84] organ allocation in general was also looked at in one student's thesis.[85]

The most common use of simulation methods was to compare lung allocation policies, both proposed policies and implemented policies.[58,86–94] Several studies looked at

the impact of geographical allocation rules applied to lung allocation.[95–99] Two articles simulated the impact of the continuous allocation framework for lung allocation in the US.[100,101]

Simulation was also used to predict lung procurement costs,[102] evaluate different data mining techniques for lung transplant,[103] and evaluate the impact of the distribution of organs to different lung transplant centres.[104,105]

2.5.3 Contribution to Lung Allocation Literature

While other papers have been identified that use the AHP for organ allocation (see appendix G.2), there was little literature relating to lung allocation and the AHP, and what was available was all US-based. The limited literature only focused on methods that indirectly inform policy decisions; there did not appear to be anything specific in the published literature about using the AHP to directly design a lung allocation policy, or to use the AHP for selecting a policy from a range of options.

This research will fill that gap by focusing specifically on the application of the AHP to lung allocation for transplantation, and describe a framework/process that can be applied to other allocation problems (i.e., not just lung allocation, or organ allocation). This work will be the first application of the AHP to the UK lung allocation system. The AHP will be used for selecting a policy and/or designing a policy directly, rather than indirectly informing decisions.

While there was more literature relating to lung allocation and simulation, there was nothing UK-specific and the focus of the simulations was often quite narrow. The goal is to create a general simulation engine that can be used to evaluate a wide range of policies and generate custom metrics specific to the UK population.

Only one paper looked at combining simulation and MCDM methods.[94] This looked at the optimisation of policies and applied techniques to a range of problems, lung allocation being one of them. As with many of the other publications this was also US-focused and used SRTR's TSAM. A custom simulation engine will be developed for this research that will allow more flexibility in the types of policies that are simulated compared to using TSAM. The methods and implementation of the simulation engine are also a potential contribution to the literature.

The key contributions of this research to the gap in the literature are:

1. Proposing a national allocation system tailored to the UK population that removes hard boundaries in allocation (see page 10)
2. The required methods for developing a custom lung allocation policy simulation engine
3. Any results generated from simulating national UK lung allocation policies
4. The use of the AHP for selecting and/or designing an allocation policy

5. The novel combination of MCDM and simulation methods for evaluating trade-offs in allocation policy design

6. A re-useable framework/process that can be applied to other allocation problems (i.e., other organs, or scenarios requiring the allocation of scarce resources)

Now that the research gaps with respect to lung transplantation have been identified, the contributions will be developed and described in chapters 4 and 5. However, before getting to lung transplantation, the next chapter explains the methods that were developed as a result of the SARS-CoV-2 pandemic. These methods will lay the foundation for the more sophisticated methods that will be used in future chapters.

Chapter 3

The Clinical Prioritisation Assistance Tool (CPAT)

3.1 Background and Context - The SARS-CoV-2 Pandemic

The SARS-CoV-2 pandemic struck the United Kingdom during the first four months of the work being done for this thesis. There was concern that the increasing number of very sick patients admitted to hospital as a result of COVID-19 would result in demand exceeding supply of critical care resources such as beds and ventilators.

There was little to no literature or data available at the beginning of the COVID-19 pandemic, which led to difficulties in determining the prognosis and predicting the likely outcomes for COVID-19 patients. This in turn made it challenging to triage COVID-19 patients, identify those that are most likely to benefit from treatment, and optimally allocate limited Intensive Therapy Unit (ITU) resources.

When high-demand scenarios such as this occur, the situation can become overwhelming for clinicians to make optimal treatment decisions, leading to sub-optimal care, sub-optimal use of limited ITU resources, loss of life and stress on the individual. There were reports of critical care clinicians experiencing symptoms of post-traumatic stress disorder (PTSD) during the COVID-19 pandemic.[106]

From these considerations it was decided to design a system that had the ability to rank all patients currently admitted to a healthcare setting by risk of mortality. The ranking would not suggest who receives treatment (or which type of treatment) but would aid the clinician in their decision making based on the mortality risk of each patient compared to others within the same healthcare setting.

Development of the Clinical Prioritisation Assistance Tool (CPAT)

In response, a multi-disciplinary team of statisticians, epidemiologists, clinicians, engineers and computer scientists were assembled across Newcastle University and Newcastle upon Tyne Hospitals (NuTH) to address the potential problem of demand for critical care resources exceeding supply.

We enlisted a number of clinicians in order to determine the requirements of a system that could assist with triage decisions during a pandemic. It was determined that software that could be used to deploy expertise at scale which is robust, reliable and auditable would help with prioritisation decisions and help relieve some of the workload from clinicians. A system was developed where an 'electronic policy' could be specified using a combination of clinical expertise and knowledge synthesised from existing literature and data (if available). This centralised policy could then be used nationally if necessary to generate a ranked list of patients in each healthcare setting. Because the policy is informed by relevant experts and the latest available knowledge, this would allow *more consistent* clinical decisions to be made regardless of patient location or the level of experience of the clinician making triage decisions. A four-month project was undertaken to develop this system, named the

Clinical Prioritisation Assistance Tool (CPAT).

Some other scoring systems were proposed and implemented in response to the pandemic, such as the 4C score,[107] however the 4C score took several months to develop and deploy (initial paper published 9 September 2020) compared the time frame in which CPAT was developed and deployed (1 May 2020).[108] The 4C score assigned patients to one of four priority tiers, however, this would result in multiple patients being in the same priority tier and thus still potentially placing a significant cognitive burden on clinicians.

This chapter demonstrates the novel use of 'pairwise comparisons'[109] applied to the healthcare setting - a method where pairs of criteria are compared, and for each pair it must be determined which is more important. Unknown to myself at the time, the novel approach that was developed for CPAT was a simplified version of the widely utilised AHP[36] mentioned in the previous chapter. The simplified methods for CPAT compared to using the AHP were better suited to the rapid decision making that was necessary during the early stages of the pandemic. More specifically, I proposed the use of 'Kahn's Algorithm'[110] for determining the importance of criteria from a pairwise comparison matrix, and Professor Cliff B. Jones suggested a modification to the standard algorithm during the development phase of CPAT.[111]

Designing a Domain-Specific Language (DSL) for Use With CPAT

For CPAT to automatically prioritise (or rank) a list of patients, the allocation policy must be able to run on a computer, more specifically: a server - a specialised type of computer that makes digital services available to multiple users over a network.

General-purpose programming languages can be used for automating processes. Some examples of commonly used programming languages are Java,[112] Python[113] and C++.[114] With these languages being general-purpose they are very complex and it would be unreasonable to expect a clinician to learn a specific programming language to encode their clinical expertise.

In contrast to general-purpose programming languages there are also DSLs. As these languages are designed for a specific purpose, they can be easier to learn than general-purpose programming languages. Some examples of popular DSLs are: Structured Query Language (SQL) for database queries,[115] HyperText Markup Language (HTML) and Cascading Style Sheets (CSS) for web design,[116,117] ladder logic for industrial process control and automation,[118] and LaTeX for generating documents (including this thesis).[119]

Why Use a DSL?

One main distinction between general-purpose programming languages and DSLs is that general-purpose languages focus on *how* to accomplish a task, whereas DSLs focus on *what* needs to be accomplished. For example, using HTML the following can be used to make text appear **bold**:

```
<b>This text is important!</b>
```

The user has simply stated *what* they want: the text to appear bold, they didn't have to specify *how* to draw every pixel on the screen. The goal with CPAT was to create a simple DSL that would allow clinicians to specify which criteria should be used for ranking COVID-19 patients, and CPAT would take care of exactly how they should be ranked by running the DSL on the server.

Handling Incorrect Data

One complication that had to be handled was missing or incorrect patient data as a result of inconsistent data entry; we wanted to avoid patients being unfairly ranked higher or lower due to data input errors. For this reason, the DSL included features to raise 'errors' and 'warnings' to the clinician.

Errors were raised when data fields required by the CPAT policy were missing for a patient. An error results in a patient being removed from the ranked list of patients altogether, with a corresponding message being displayed detailing the data fields that are required to correct the patient data that was input.

Warnings were raised when required data fields were populated but contained a value that was *potentially* incorrect. One example of this was the use of the Clinical Frailty Scale (CFS)[120] for patients aged under 65: the scale was only validated with patients aged 65 or older. The patient could still be included in the ranked list of patients, but a warning would be displayed next to their name indicating the potential mistake. For a detailed explanation of how errors and warnings were encoded see appendix B.1.4.

Application to Lung Allocation

My work on the CPAT project required a pivot away from the focus on lung allocation. However, in doing so the understanding of the methods required for lung allocation were accelerated and the general-purpose nature of the techniques being used for lung allocation were discovered: the original focus of this research was how to allocate a scarce resource (i.e., donor lungs) by rank-ordering a group of individuals waiting for lung transplantation in order of priority (i.e., patients on the active transplant waiting list). From this more abstract view of the problem, the application to the context of the pandemic becomes: how best to prioritise patients affected by COVID-19 for access to limited ITU resources such as ITU beds, ventilators and oxygen?

This chapter lays the foundation for the more advanced application of methods that will be described in chapters 4 and 5. It must be emphasised that **CPAT was never deployed live and was never used for clinical decision making** as the pandemic came under control and thankfully CPAT was never required in a real-world healthcare setting.

3.2 Methods

The core functionality of CPAT is split into three major components: the first component assists in *synthesising* existing knowledge into a ranked set of criteria, the second component *encodes* the ranked criteria into a policy, and the third component *ranks* patients according to the policy and displays the output. A detailed description of the methods behind these components are given in appendix B.

This section deals with the ranking of patients: in the context of this chapter 'ranking' refers to ordering a list of patients from highest priority (or risk) to lowest priority (or risk). In a clinical context this is sometimes referred to as 'named allocation'. To determine a patient's ranking one or more *criteria* need to be used. Criteria are observable/measurable attributes of a patient such as height, age and presence/absence of certain diagnoses.

Synthesising Existing Knowledge

Data from 424 patients admitted to hospital in the north east of England and testing positive for COVID-19 were collected as a test group for evaluating the rankings. In the case of the COVID-19 pandemic, some patient characteristics were more predictive of mortality than others. In order to synthesise the relative importance of several criteria, a novel method of utilising pairwise comparisons[36,109] was created and used (see appendix B.1.1 for a detailed explanation of the novel application of pairwise comparisons in this work).

A pairwise comparison matrix was constructed using Microsoft Excel[121] and several meetings with clinicians were held until agreement was reached on the entries in the matrix. This comparison matrix was then input to the CPAT administration portal (see figure 3.1) and was found to be free of loops, indicating there were no logical contradictions in the comparison matrix (see appendix B.1.2).

Clinicians and epidemiologists combined clinical experience with the (limited) literature at the time of the pandemic to identify criteria that may be predictive of patient mortality. The criteria were then ranked from highest to lowest priority using a variation Kahn's algorithm described in appendix B.1.3.

Encoding the Allocation Policy in a DSL

Now that the criteria were (1) identified and (2) ranked by priority, they could now be encoded in CPAT using a custom DSL. A screenshot of a policy being input using the administration portal is shown in figure 3.2. A detailed description of the CPAT DSL is given in appendix B.1.4. Due to the way in which the DSL was designed, it was possible to calculate a CPAT score for patients according to the policy that was specified. Full details on the score calculation process are given in appendix B.1.5 on page 225.

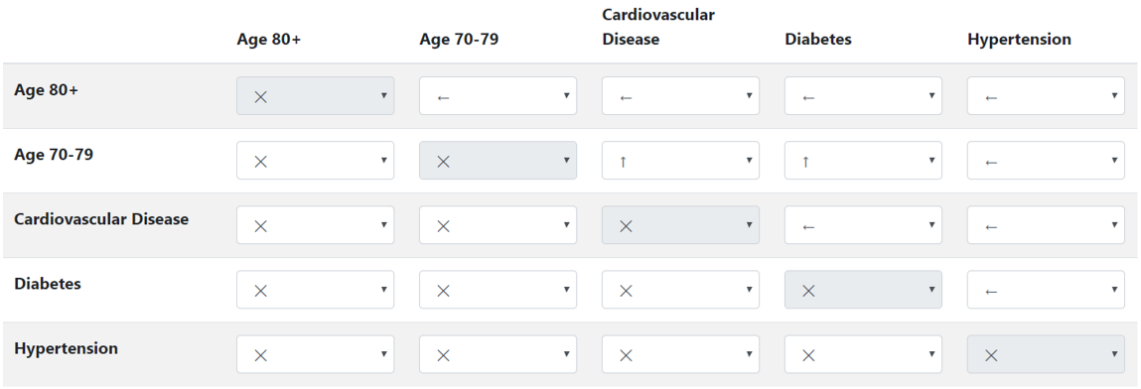


Figure 3.1: A screenshot of the CPAT pairwise comparison tool with some example input. The arrow in each cell points to the criterion that is more important. If a comparison is not possible, an "X" can be input. If there are cycles/loops present in the input, a warning will be displayed to the user so that they can correct their input and ensure logical consistency.

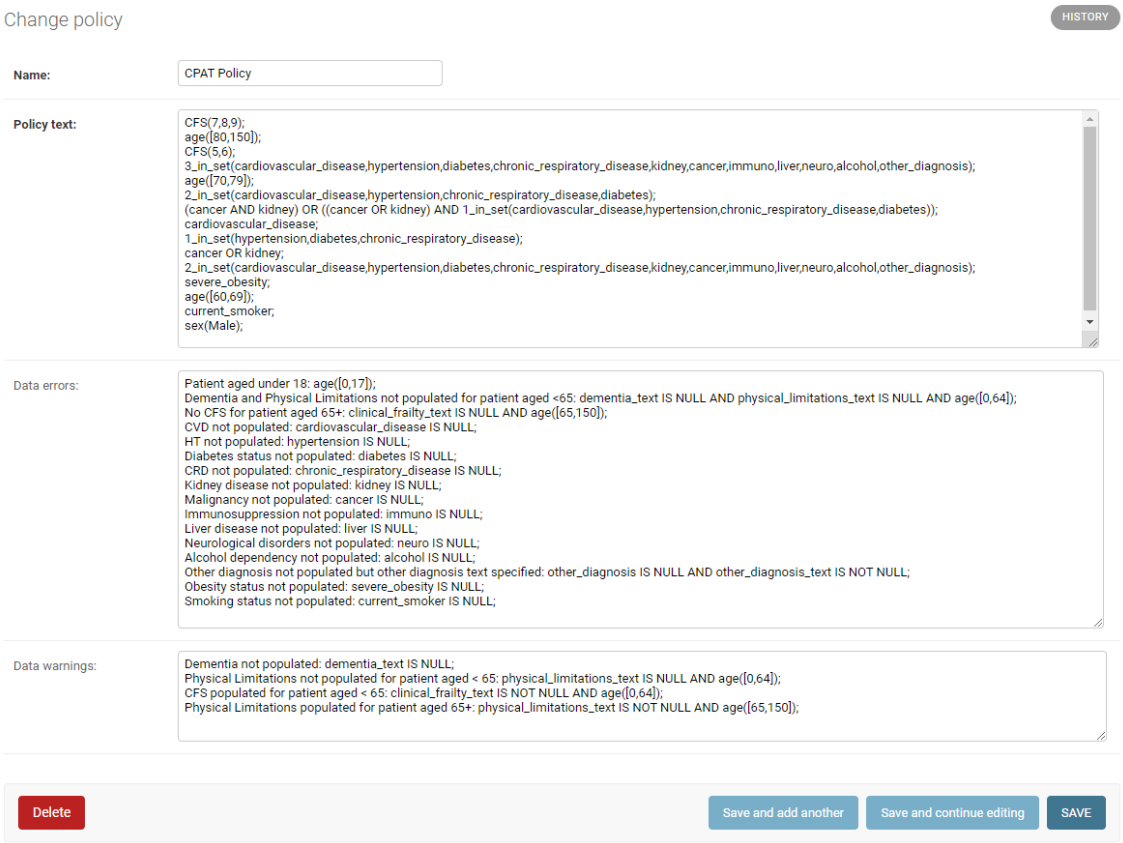


Figure 3.2: The policy, warning conditions, and error conditions were input to the online CPAT administrators portal using the CPAT domain-specific language.

Ranking Patients Using the Allocation Policy

Software was designed and developed which would run on a server, using Python[113] and the Django Representational State Transfer (REST) Framework.[122] Patient data that was stored via the electronic patient record (EPR) system[123] was converted to a format that was accessible by the server software. The CPAT policy written in the custom DSL was input into the CPAT administration portal, this would allow any list of patients input to the server to be ranked by risk of mortality according to the policy (see appendix B.1.5).

When a client connected to the server using a web browser, they were presented with a prioritised list of patients that were prioritised by risk according to the electronic policy that had been input (figure 3.3), along with any error or warning messages that were generated (figure 3.4). The patients were grouped by priority level, where each priority level contained patients with the same priority score.

Evaluating the Predictive Strength of CPAT

Data for these analyses were provided by NuTH which included daily outcome data on 423 patients for 28 days after each patient's first visit to hospital. Outcomes were one of: death, invasive ventilation, non-invasive ventilation, hospitalised (requiring oxygen), hospitalised (requiring medical care), hospitalised (not requiring oxygen/medical care), not hospitalised (requiring home oxygen), and not hospitalised (not requiring home oxygen). To evaluate the predictive strength of CPAT, a univariate Cox model[124] was constructed with the CPAT score being the only variable, and death as the outcome being predicted. The concordance index of this model was calculated to evaluate the ability of CPAT to predict patient mortality. A concordance index of 0.5 indicates no predictive strength (i.e., 50% of predictions are correct and 50% are incorrect, the same as flipping a coin). A concordance index of 1.0 indicates 100% correct predictions, which are extremely unlikely (if not, impossible) in practice.

The impact of each line of the DSL in the policy was evaluated by running the policy from line 1 to line $\{1, ..., N\}$ where N is the number of lines in the policy. The concordance index was then calculated up to each line in the policy.

In addition, the ranked list was split into deciles using the CPAT score and the mortality rate was calculated for each decile.

Because CPAT uses a score to rank each patient and 4C also uses a scoring mechanism, there is a potential for multiple patients to share the same priority level/score. Histograms were generated for CPAT and 4C to visualise the distribution of the number of patients at each level/score, and the mean, standard deviation, minimum and maximum number of patients sharing the same priority level was calculated.

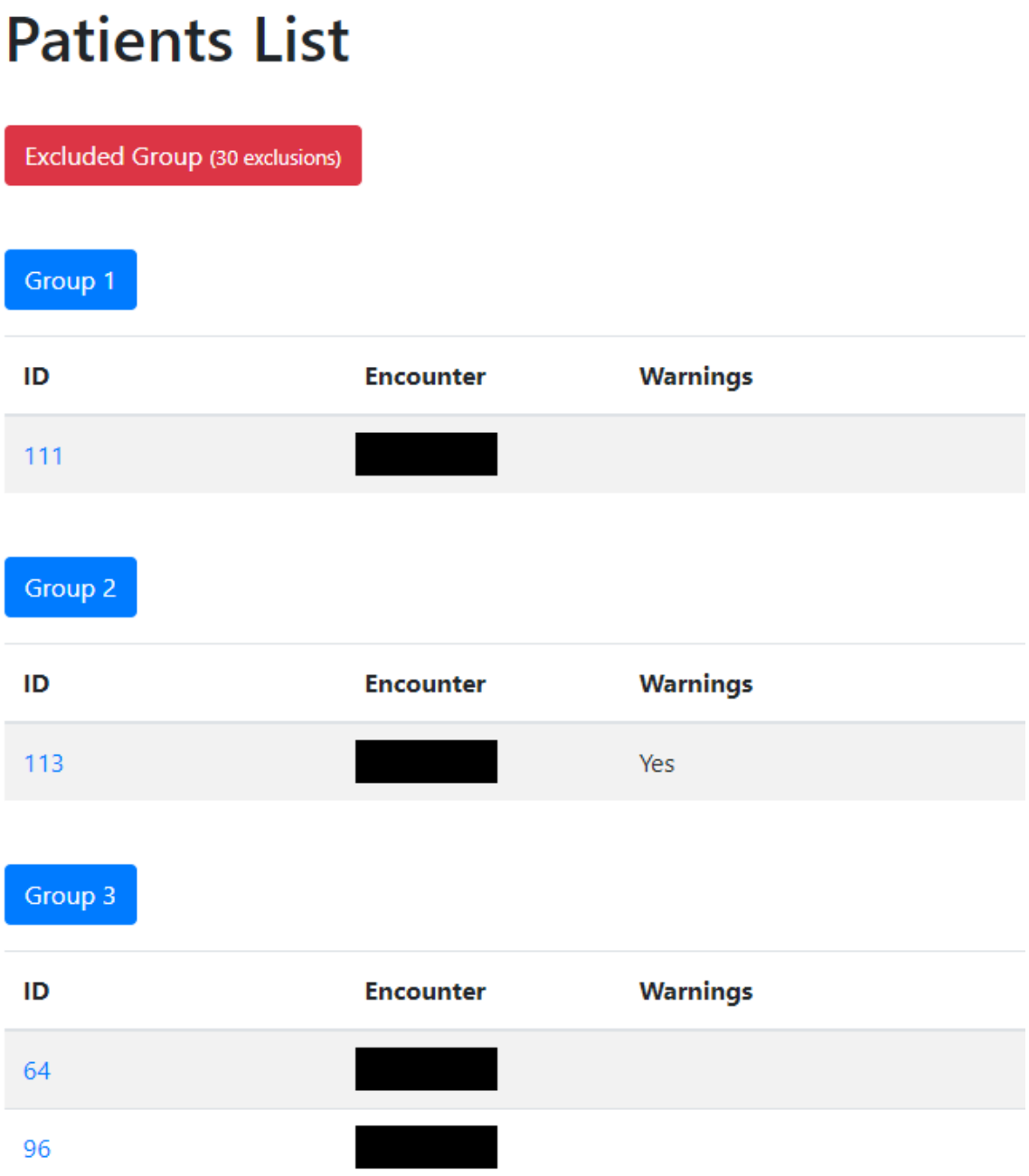


Figure 3.3: A screenshot of an example ranked patient list, with 'Group 1' having the highest risk of mortality, and subsequent groups having lower risk of mortality. Patients at the same priority level will belong to the same group. (Note: potentially identifying information has been redacted.)

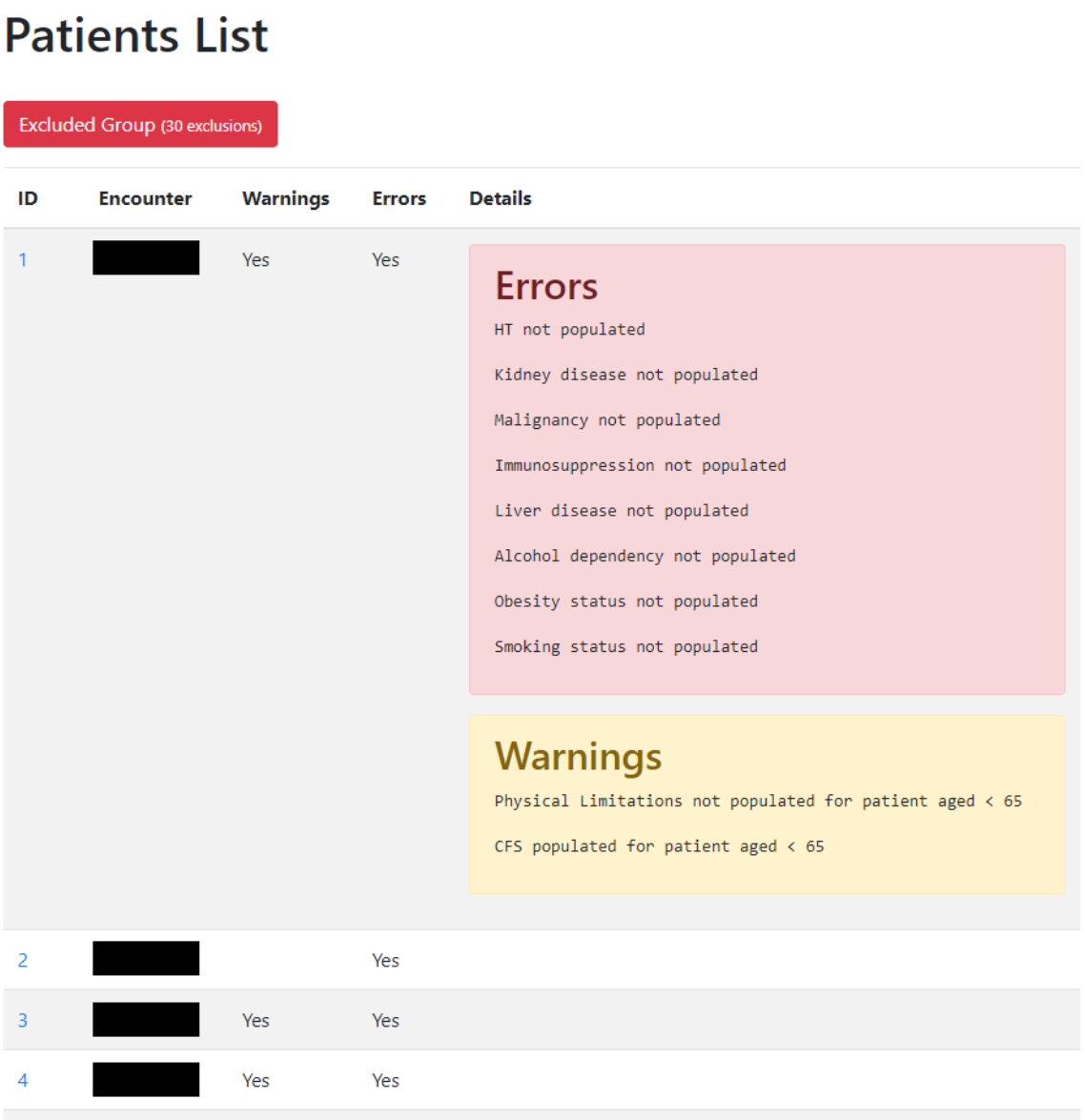


Figure 3.4: A screenshot of some errors and warnings displayed to the end-user to warn of incorrect or potentially incorrect data. (Note: potentially identifying information has been redacted.)

3.3 Results

Ranked Criteria ☿

The result of the pairwise comparisons resulted in the following ranking of criteria:

1. Clinical frailty scale between 7 - 9
2. Age over 80
3. Clinical frailty scale between 5 - 6
4. At least 3 of: cardiovascular disease, hypertension, diabetes, chronic respiratory disease, kidney disease, cancer, immunosuppression, liver disease, neurological disease, alcohol dependency, other relevant comorbidity
5. Age between 70 - 79
6. At least 2 of: cardiovascular disease, hypertension, chronic respiratory disease, diabetes
7. Cancer and kidney disease, or, either cancer or kidney disease and one of: cardiovascular disease, hypertension, chronic respiratory disease, diabetes
8. Cardiovascular disease
9. At least 1 of: hypertension, diabetes, chronic respiratory disease
10. Cancer or kidney disease
11. At least 2 of: cardiovascular disease, hypertension, diabetes, chronic respiratory disease, kidney disease, cancer, immunosuppression, liver disease, neurological disease, alcohol dependency, other relevant comorbidity
12. Obesity
13. Age between 60 - 69
14. Current smoker
15. Male sex

The above criteria were encoded in the CPAT DSL as follows:

Line	CPAT DSL
1	`CFS(7,8,9);`
2	`age([80,150]);`
3	`CFS(5,6);`
4	`3_in_set(cardiovascular_disease,hypertension,diabetes, chronic_respiratory_disease,kidney,cancer,immuno,liver,neuro, alcohol,other_diagnosis);`
5	`age([70,79]);`
6	`2_in_set(cardiovascular_disease,hypertension, chronic_respiratory_disease,diabetes);`
7	`(cancer AND kidney) OR ((cancer OR kidney) AND 1_in_set(cardiovascular_disease,hypertension, chronic_respiratory_disease,diabetes));`
8	`cardiovascular_disease;`
9	`1_in_set(hypertension,diabetes,chronic_respiratory_disease);`
10	`cancer OR kidney;`
11	`2_in_set(cardiovascular_disease,hypertension,diabetes, chronic_respiratory_disease,kidney,cancer,immuno,liver,neuro, alcohol,other_diagnosis);`
12	`severe_obesity;`
13	`age([60,69]);`
14	`current_smoker;`
15	`sex(Male);`

Error conditions for missing data, or inconsistent data entry were input as follows:

Line	CPAT DSL
1	`Patient aged under 18: age([0,17]);`
2	`CVD not populated: cardiovascular_disease IS NULL;`
3	`HT not populated: hypertension IS NULL;`
4	`Diabetes status not populated: diabetes IS NULL;`
5	`CRD not populated: chronic_respiratory_disease IS NULL;`
6	`Kidney disease not populated: kidney IS NULL;`
7	`Malignancy not populated: cancer IS NULL;`
8	`Immunosuppression not populated: immuno IS NULL;`
9	`Liver disease not populated: liver IS NULL;`
10	`Alcohol dependency not populated: alcohol IS NULL;`
11	`Other diagnosis not populated but other diagnosis text specified: other_diagnosis IS NULL AND other_diagnosis_text IS NOT NULL;`
12	`Smoking status not populated: current_smoker IS NULL;`

The warnings for potentially incorrect data input were encoded as:

Line	CPAT DSL
1	`Dementia not populated: dementia_text IS NULL;`
2	`Physical Limitations not populated for patient aged < 65: physical_limitations_text IS NULL AND age([0,64]);`
3	`CFS populated for patient aged < 65: clinical_frailty_text IS NOT NULL AND age([0,64]);`
4	`Physical Limitations populated for patient aged 65+: physical_limitations_text IS NOT NULL AND age([65,150]);`

Patient Rankings

The ranked list is shown in figure 3.5 on page 67 in two halves, with the top half showing the patients at highest risk of mortality. Red indicates that the criterion applies to that patient, green indicates that it does not apply. Outcome data for days 0 through 7, then at 14, 21 and 28 are displayed below each patient.

The patients with the highest risk of mortality were those aged 80+ with a CFS of 7 or higher, followed by patients with a CFS of 7+, then patients aged 80+ and so on. The goal of this particular CPAT policy was to rank patients by risk of mortality with the highest risk patients being at the left hand side of the visualisation on the next page (i.e., top of the ranked list) and lowest risk patients at the right hand side (i.e., bottom of the ranked list).

Looking at the outcome data in the bottom rows, there is a general trend of more deaths for the higher ranked patients (indicated by black squares) and fewer deaths for the lowest ranked patients (indicated by more green and fewer black squares). Note that the highest ranked are older with more co-morbidities and the lowest ranked patients are younger with fewer co-morbidities. A more rigorous analysis of the CPAT policy is given in the next subsection.

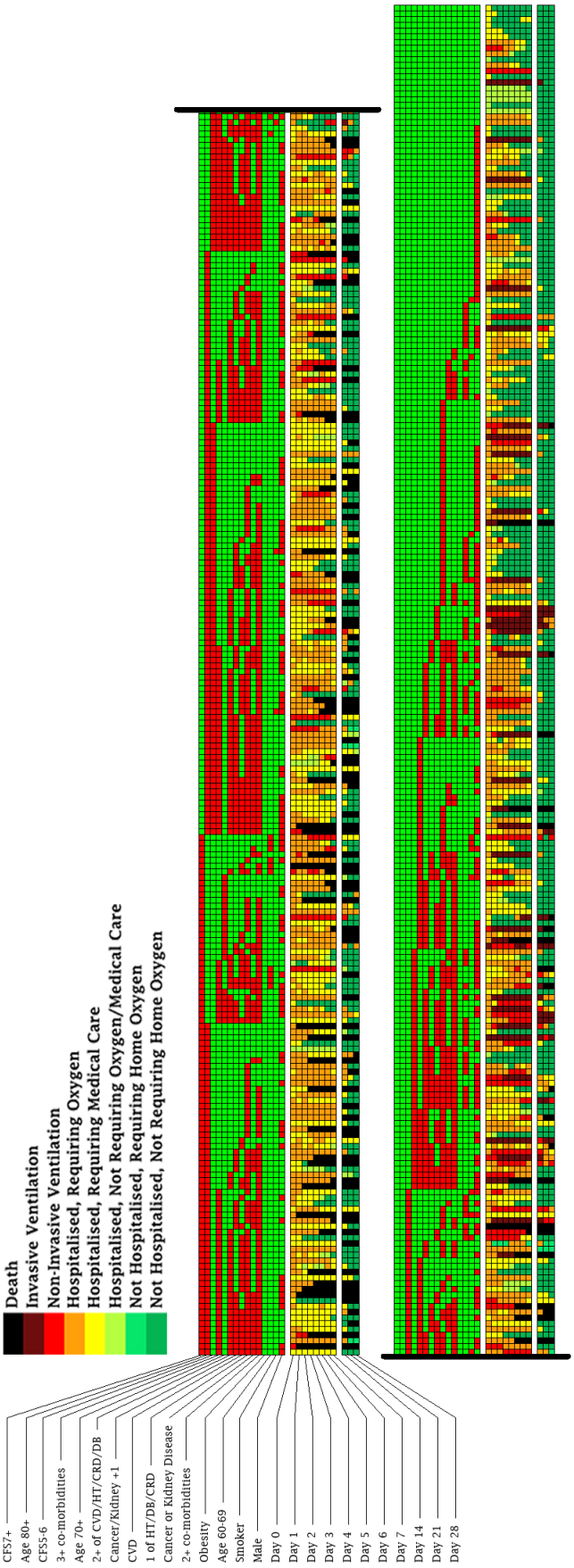


Figure 3.5: Patient characteristics, rankings, and outcome from hospital admission for COVID-19.

3.3.1 Comparison of CPAT and 4C Ranking Distributions

In the methods section, the potential for multiple patients to share the same priority level/score was mentioned. The methods used for CPAT result in N criteria having 2^N possible priority levels. As there were 15 criteria identified for ranking patients, the number of possible priority levels was $2^{15} = 32,768$. The 4C score on the other hand has a score that ranges from 0 to 21 points, resulting in 22 possible priority levels. The number of patients in each priority level/sharing the same score are compared in figure 3.6, and the mean, standard deviation and range of patients sharing the same priority level are shown in table 3.1.

Table 3.1: A comparison of the granularity of patient rankings using CPAT compared to the 4C score. Higher granularity results in fewer patients sharing a priority level and vice versa. CPAT resulted in fewer patients sharing the same priority level and also had a narrower range of patients sharing the same level compared to 4C.

System	Number of Distinct Levels	Mean Patients Per Level (SD)	Range of Patients Sharing Level
CPAT	197	2.2 (2.57)	1 to 24
4C	22	19.2 (13.8)	1 to 40

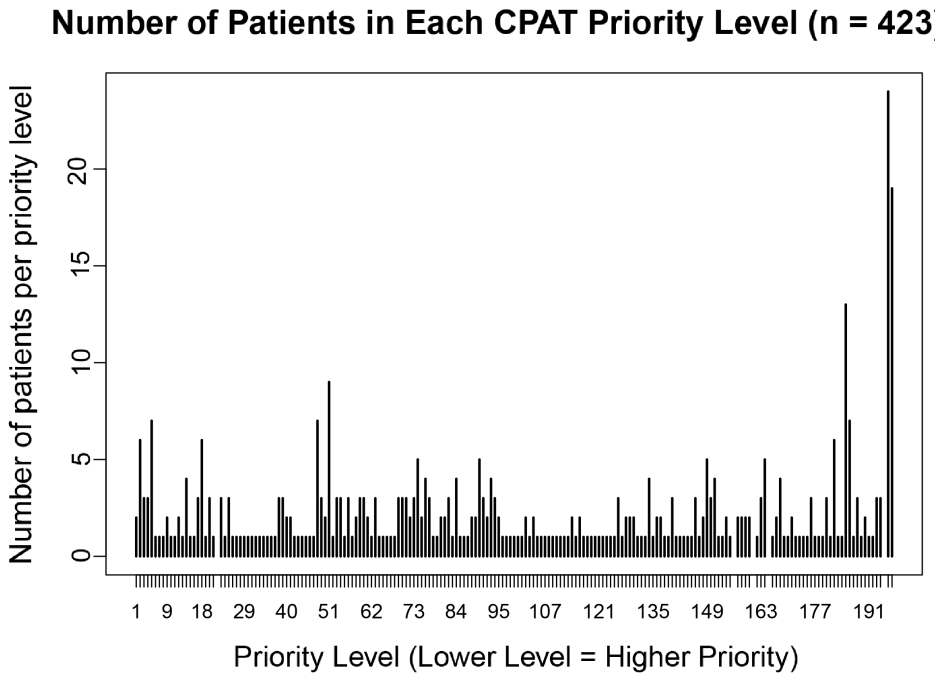
Concordance Index

The overall concordance index was calculated to be 0.73 for the given policy. By comparison, applying the 4C scoring rules to this dataset resulted in a concordance index of 0.76, with the published concordance being in the range of 0.76 to 0.79.[107]

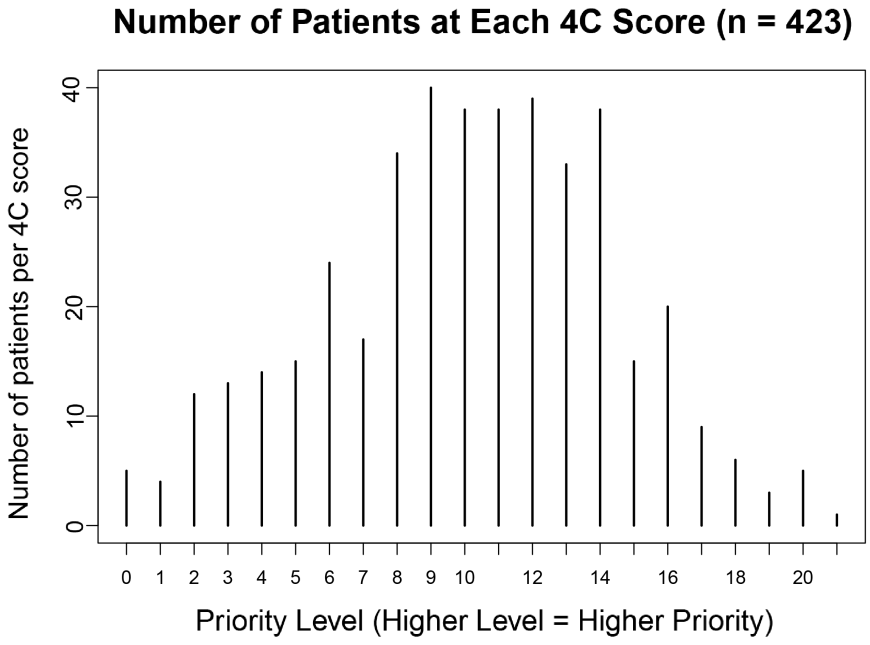
As each line was added to the CPAT policy, the concordance was calculated and the results are shown in table 3.2.

Table 3.2: CPAT concordance index as each line was added to the policy. A concordance index of 0.5 indicates no predictive ability and 1.0 indicates 100% correct predictions.

Line Number	Concordance Index
1	0.59
2	0.69
3	0.72
4	0.72
5 - 14	0.73



(a)



(b)

Figure 3.6: **(a)** Comparison of the number of patients per priority level using the Clinical Prioritisation Assistance Tool (CPAT) compared to **(b)** the 4C score. The CPAT policy resulted in an almost-uniform distribution of patients sharing the same priority level, with a large spike on the right hand side for patients with the lowest risk/scores. The 4C score results in a normally-distributed number of patients sharing the same score, with the largest numbers of patients sharing scores between 8 (intermediate mortality risk) and 14 (high mortality risk).

Mortality by Decile

The mortality percentage for each decile is shown in table 3.3. There was a general decrease in mortality from the highest risk deciles (deciles 1 to 4) to the lowest risk deciles (deciles 6 to 10).

Table 3.3: Mortality percentage by decile using the CPAT policy to rank (n=423) patients. Mortality risk tended to decrease from decile 1 (highest priority/risk) to decile 10 (lowest priority/risk).

Decile	Alive	Dead	Percent Mortality	n
1	22	20	47.6%	42
2	26	16	38.1%	42
3	23	19	45.2%	42
4	27	15	35.7%	42
5	30	12	28.6%	42
6	31	11	26.2%	42
7	37	6	14.0%	43
8	41	2	4.7%	43
9	41	2	4.7%	43
10	43	0	0%	43

3.4 Discussion

Benefits

There are several potential benefits to the approach that CPAT uses for ranking patients by their mortality risk:

Data Independence and Integrity, Rapid Deployment and Adaptability

The first benefit is that CPAT is data independent and so allows existing clinical knowledge and experience to be encoded in the policy. The policy for COVID-19 was developed using the limited literature from China and Italy available very early in the pandemic, with no initial access to data (the data for testing became available after CPAT had been developed). This data independence allows policies to be put into place rapidly, in the space of hours, days or weeks rather than requiring months for data to be collected.

Once data becomes available and new knowledge is discovered and published the policy can be updated, making CPAT adaptable to changes in understanding. As soon as the policy is changed the rankings are **instantly** updated, minimising the delay between expert understanding changing and implementation of the updated policy.

It is also possible to encode in the policy basic data integrity checks and raise a warning or an error. The example of using the CFS on patients aged < 65 is just one example of this feature. If a data field is missing that is critical to the ranking, for example age, then a patient can not be ranked and an error message can be displayed until the field is populated (see appendix B.1.4 for greater detail on the encoding of warnings and errors in the CPAT DSL).

Auditable

CPAT regularly saves audit logs containing a timestamp, a copy of the policy which was in place at that time and a copy of the ranked patient list. This has the potential to ease the burden on clinicians having to justify their decisions, as the audit log will show how their patient ranked relative to all other patients requiring access to limited critical care resources at the time the decision was made.

Expertise at Scale

Another benefit is that multiple experts spanning multiple domains can assist in designing the policy. Once the policy is designed it is input to a central server, and therefore can be implemented in multiple locations instantaneously, allowing the expertise to be deployed at scale. This also allows consistency in decision making as the patient rankings are the same regardless of whether a junior clinician or an expert with many years of experience is deciding on the best course of action for a patient.

Meaningful Rankings

In the results presented, it was shown that there were $2^{15} = 32,768$ possible priority levels that resulted in 423 patients being prioritised into 197 distinct priority levels. The benefit of this large number of priority levels is that there were few patients sharing the same priority level with CPAT, with an average number of 2.2 patients per priority level. When there are only a few patients that need to be considered within each priority level, less mental effort is required to determine the appropriate course of action and allocation of limited ITU resources. In figure 3.6b CPAT resulted in an approximately uniform distribution of patients at each priority level, with the greatest numbers of patients being in the lowest priority levels - patients that are low risk and likely will not require treatment.

By comparison, in figure 3.6b the 4C score resulted in a normal distribution of patients sharing the same score, with an average of 19.2 patients per level. The peak of the normal distribution was centred around patients with a score of 8 to 14, corresponding to 'intermediate' and 'high' risk levels, with 30+ patients per level. This means that the priority levels with the *largest* numbers of patients (and therefore, require more intervention and cognitive effort from clinicians) corresponds to the patients that require the *most attention and careful consideration.*

Given an identical list of patients, the CPAT rankings result in dramatically fewer, more manageable numbers of patients for a clinician to consider at one time compared to 4C.

Statistical Discrimination Ability

The major benefit demonstrated in this paper was that the discrimination ability of CPAT was almost as strong as the well-developed and validated 4C score,[107] while still retaining all the benefits of being data-independent, providing meaningful patient rankings, and rapidly deployable in real time as patients are admitted to the healthcare setting.

Limitations

The challenge of scaling One benefit of CPAT is that is has the potential to be rapidly deployed at scale, however, one limitation of this work is that it was not adopted for widespread use, as a result we did not have to contend with the additional complexities that arise from large-scale deployment of software. The architecture of CPAT centres around a single top-down enforced policy, which also introduces a single point of failure - if the server hosting the policy becomes unavailable, this would impact every hospital attempting to use CPAT. There are also additional performance and scaling issues that arise from potentially having hundreds of hospitals making frequent requests to a single centralised server, as this large demand places a heavy computational load on the server.

These issues can be alleviated by having multiple host servers for redundancy and to spread the computational load evenly, and utilising some sort of caching server such as

Redis.[125] However, this introduces greater complexity to the design of the system, as it will be necessary to ensure the policy is identical across all host servers: when the policy is updated, this change must be reflected across all host servers.

Data quality The meaningfulness and quality of the patient rankings generated by CPAT are strongly dependent on the quality of the data input by the end user (i.e., clinician). An incorrectly input variable (for example, age) can result in patients being ranked significantly higher or lower than they should be, giving an unfair advantage or disadvantage. The exact definitions and units for variables must also be clearly defined and used across all healthcare settings. For example, the CPAT policy made use of the CFS which has been updated over time: before 2007 it was a 7-point scale before becoming a 9-point scale as used here.[126] If different versions were used between healthcare settings then this would result in inconsistency in the data and patient rankings, having the unintended effect of inequitable access to healthcare resources.

Follow-up duration The follow-up data used in these analyses for evaluating CPAT only extended to 28 days after hospitalisation. This may have resulted in the C-statistic (concordance index) for CPAT being 'inflated', however the 4C score did perform similarly on the same dataset (C-statistic of 0.76 compared to 0.73 with CPAT).

Sequential rankings The main limitation of this approach is illustrated in figure 3.7: each line in the policy takes absolute priority over all lines below it. However, as shown in the figure, the patient with a 7+ on the CFS and no co-morbidities is ranked higher than the patient aged over 80 with several co-morbidities. It may not necessarily be the case that a single factor (in this case CFS of 7+) indicates a higher risk of mortality than several lower priority factors. In fact, it was the lower priority patient that died, despite having a lower CPAT ranking.

This sequential approach of ranking patients makes the rankings unstable: a small change in patient characteristics (such as CFS increasing from 6 to 7) can result in a large change in ranking. A stable ranking system would result in small changes in patient characteristics being reflected by small changes in ranking. This is a concrete example of the issues discussed with sequential allocation in sections 2.3.1 and 1.2.1.

In order to allow for more nuanced rankings, it will be necessary to assign weights to each risk factor, thus allowing for the presence of several smaller risk factors to compensate for the absence of larger risk factors. This would then allow the basic techniques developed for CPAT - synthesising clinical knowledge, encoding knowledge in a policy and ranking patients according to a policy - to be expanded and applied to the problem of lung allocation. This will be the primary focus in the next chapter.

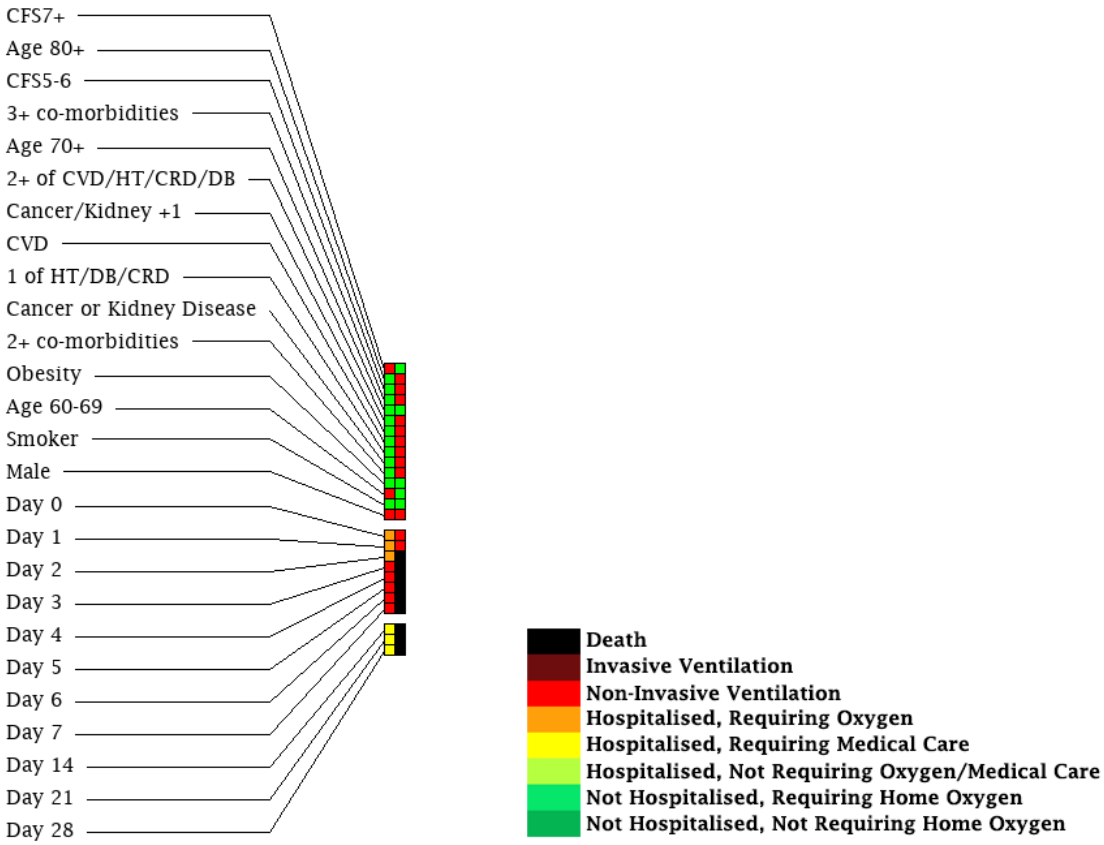


Figure 3.7: Sequential allocation leading to one patient being prioritised solely because of their score on the clinical frailty scale, despite the lower priority patient being aged over 80 with several co-morbidities.

3.5 Summary

The ability of CPAT to rank patients with little or no available data is the single largest benefit it provides. As a result of being data-independent it can be very rapidly deployed at scale. However it must be applied to appropriate scenarios. The types of scenarios CPAT is designed to handle are situations where making a decision or taking some action - even if sub-optimal - is more preferable than waiting to make what might be considered an optimal decision. The COVID-19 pandemic was one of these situations: patients were continuously being admitted to hospital and time was critical. Assessing patients and deciding on the appropriate course of action rapidly was more important than individual clinicians using their limited time and resources to attempt to assess patients in the most optimal order.

As discussed earlier, clinicians attempting to make these optimal decisions on-the-fly poses a large cognitive burden that is further compounded by the stress of constantly making life-or-death decisions. The intended use of CPAT (which luckily was never needed in practice), was to automatically rank all patients entering the hospital by risk of mortality. Clinicians then effectively have a "birds-eye view" of all patients in the system ordered by risk. There may be some patients at the top of the list that are so critically ill that despite best efforts would be unlikely to survive. Nearer the bottom of the list would be patients that are healthy enough that they could be sent home and recover with minimal risk. This would have allowed clinicians to focus their efforts on patients in the middle of the list: patients that are critical but have good chances of survival with appropriate management. This will allow triage on a large scale in real time.

There were plans for CPAT to be developed further by allowing a 'meta-policy' to be defined, comprising multiple sub-policies. Each sub-policy would filter out a subset of the patients depending on the data that was available. For example, one sub-policy could be applied using only the data that is immediately available as patients enter the hospital (such as age, diagnosed comorbidities, smoking status etc.) such as the policy outlined in this chapter. The next sub policy could filter out the patients that are too ill or don't require treatment, and then prioritise them by clinical test results. This process could be repeated for each stage that a patient goes through from admission, to test results to discharge. However, before these ideas could be developed further the demand on hospitals started to decrease, and efforts re-focused on lung allocation.

While CPAT was originally designed for prioritising patients for treatment during the COVID-19 pandemic, the methods are generalisable to any situation where rapid decision making, prioritisation and/or allocation are necessary.

The next chapter will expand on the concepts covered in this chapter, overcoming the limitations described in section 3.4 and describe methods that can be used when data is available to make better informed and more nuanced prioritisation decisions for the allocation of donor lungs to potential recipients.

Chapter 4

Lung Allocation

4.1 Context: The Challenges of Designing Lung Allocation Policies

Designing a lung allocation policy requires deciding on the goals that the policy needs to achieve. These goals are influenced by the subjective values that policymakers, healthcare professionals, clinicians, and other stakeholders hold. Some goals conflict, requiring a trade-off or compromise to be made. One example of this is reducing waiting list deaths by prioritising candidates with higher clinical urgency. This results in reduced waiting list deaths, but potentially results in shorter survival durations post-transplant and unequal access to transplant due to the differing risk of mortality by diagnosis.

Historically allocation was based on waiting time; however this favoured candidates with less clinically urgent conditions, such as COPD as they could afford to wait a prolonged period on the waiting list (see section 2.4 for a detailed history of lung allocation).

Although there are a number of possible allocation goals, the main two goals repeatedly mentioned in the literature are increasing waiting list survival (i.e., reducing the number of annual waiting list deaths) and post-transplant survival (i.e., recipients living for a long period of time after receiving a transplant). The issue of fairness is also discussed in the literature which can have several interpretations, some of which were discussed in section 2.2.1. Some other possible interpretations of fairness could be:

1. Ensuring all candidates have equal access to transplant (i.e., equity of access)
2. Prioritising candidates with the greatest clinical need for transplant (i.e., prioritising the most urgent candidates)
3. Prioritising candidates that will gain the most additional days of life from transplant (i.e., 'net benefit')
4. Using *only* clinically relevant criteria for prioritising candidates and not 'protected characteristics'[127] (or equivalent laws in the country the allocation system is implemented)
5. Aiming to have the same *percentage* of candidates transplanted across all diagnosis groups
6. Aiming to have the same *number* of candidates transplanted across all diagnosis groups
7. Aiming to have the percentage/number of transplanted candidates be *proportional* to the risk of mortality associated with each diagnosis group

This is by no means an exhaustive list but it illustrates how there is no single definition of fairness, and multiple definitions can conflict with each other (for example items 5 - 7

on the list). There is also no single objective way of measuring the fairness of a system; it is possible to calculate transplant rates and outcomes stratified by diagnosis, age, sex etc., however the interpretation of the calculations and how they relate to fairness is still subjective. For these reasons 'fairness' will not be used as a metric in this work.

At the time of writing, in the US and on the Eurotransplant scheme the LAS is used (though this will likely be replaced by the CAS.[100] This is a predictive model that prioritises candidates based on net benefit. Allocation using the LAS can prioritise waiting list survival, post-transplant survival, or a weighted combination of both. The US LAS uses a 2:1 weighting-ratio of waiting list survival to post-transplant survival to prioritise candidates.

The UK currently uses a multi-tiered urgency-based lung allocation policy. Named allocation is implemented for candidates on the super-urgent and urgent schemes. For non-urgent candidates, allocation is based on a free centre choice, where clinicians located at the receiving centre allocate the lungs to the patient they believe has the greatest need.[37]

As discussed in sections 1.1.2, 2.3.1 and 3.4, a multi-tiered (i.e. sequential) policy creates the possibility of two clinically similar candidates receiving vastly different priority rankings due to measurements of lung function falling on different sides of a hard numeric boundary. This leads to the candidates being placed in different tiers and having different access to transplant and therefore doesn't ensure equity of access ('fairness').

The named allocation system for UK super-urgent and urgent candidates uses waiting time and ABO matching and is not based on predictive survival models. The free centre choice is very subjective and depends on who makes the allocation decision at the time of offer, and access to transplant varies depending on the availability of donors within a geographical zone.

To address these potential areas of improvement a new policy can be designed, however a change in allocation policy runs the risk of unintended consequences.

Discrete event simulation (DES) is a technique that can be used to model real-world processes.[128] The goal of this section is to use DES to simulate policies that prioritise candidates using different relative priorities of waiting list and post-transplant survival, allowing the differential impact on candidates and recipients to be evaluated and stratified by diagnosis group, age group and blood group.

More advanced allocation features were also evaluated: risk-adjusted benefit, conditional survival and increased use of SLT for ILD candidates. Additional scenarios were also simulated to evaluate the impact increased donation/utilisation has on the number of waiting list deaths.

Performance metrics were recorded for each policy and scenario, allowing the strengths and weaknesses of different lung allocation policies to be compared and evaluated.

4.2 Methods

4.2.1 Statistical Methods

Datasets

Datasets containing waiting list, post-transplant and donor data were provided by NHSBT[1]. The waiting list population contained data on 4280 candidates listed between 2002 and 2021 and was split 80/20 into a training (n = 3424) and validation (n = 856) cohort. The post-transplant population consisted of 2131 transplant recipients from 2002 – 2021 and was also split 80 / 20 into a training (n = 1705) and validation (n = 426) cohort. Both datasets contained survival durations and censoring indicators. The population was limited to all adult, first-time, lung-only candidates or recipients (i.e., no multi-organ recipients such as heart-lung or lung-liver). The process of cleaning the data is summarised in figure 4.1. For both waiting list and post-transplant survival, patients with a recorded forced expiratory volume over one second (FEV1) or forced vital capacity (FVC) greater than 8 litres were removed from the dataset, as well as any rows in the dataset that contained missing data.

The 80 / 20 split was randomised using the `sample()` function in `R`[129] using a constant seed value of `538` that was generated using the online service `random.org`.[130] This resulted in a random, but repeatable, selection of candidates and recipients for the training and validation cohorts. This reduces the chances of there being any selection bias in any characteristics when assigning individuals to the respective training and validation cohorts, it also makes it possible to test any predictive models trained on the training data against the validation data to prevent over-fitting. There is an implicit assumption that a random split will result in a validation dataset that is representative of the training set. To test this assumption, the Welch Two Sample T-test was used to compare numeric variables between the training and validation datasets, and the Chi-squared test was used to compare categorical variables to ensure all p-values were > 0.05.

Building Cox Models for Waiting List and Post-Transplant Survival

There are a number of methods for building an appropriate Cox model for simulation.[131,132] For this research the methods described by Collett et al.[133] were used.

Two Cox proportional hazards models were built using the training datasets (see figure 4.1) for the purpose of simulating waiting list survival and post-transplant survival. A custom `R` script was created to automate the iterative process of adding statistically significant variables and removing non-significant variables until the model could no longer be improved.[133] The algorithm is described in greater detail in appendix D.1, and for the

[1]The OPTN also provided US cardiothoracic data that was used during the initial development phase of the simulation engine. However, the dataset was not used for generating the results presented in this thesis.

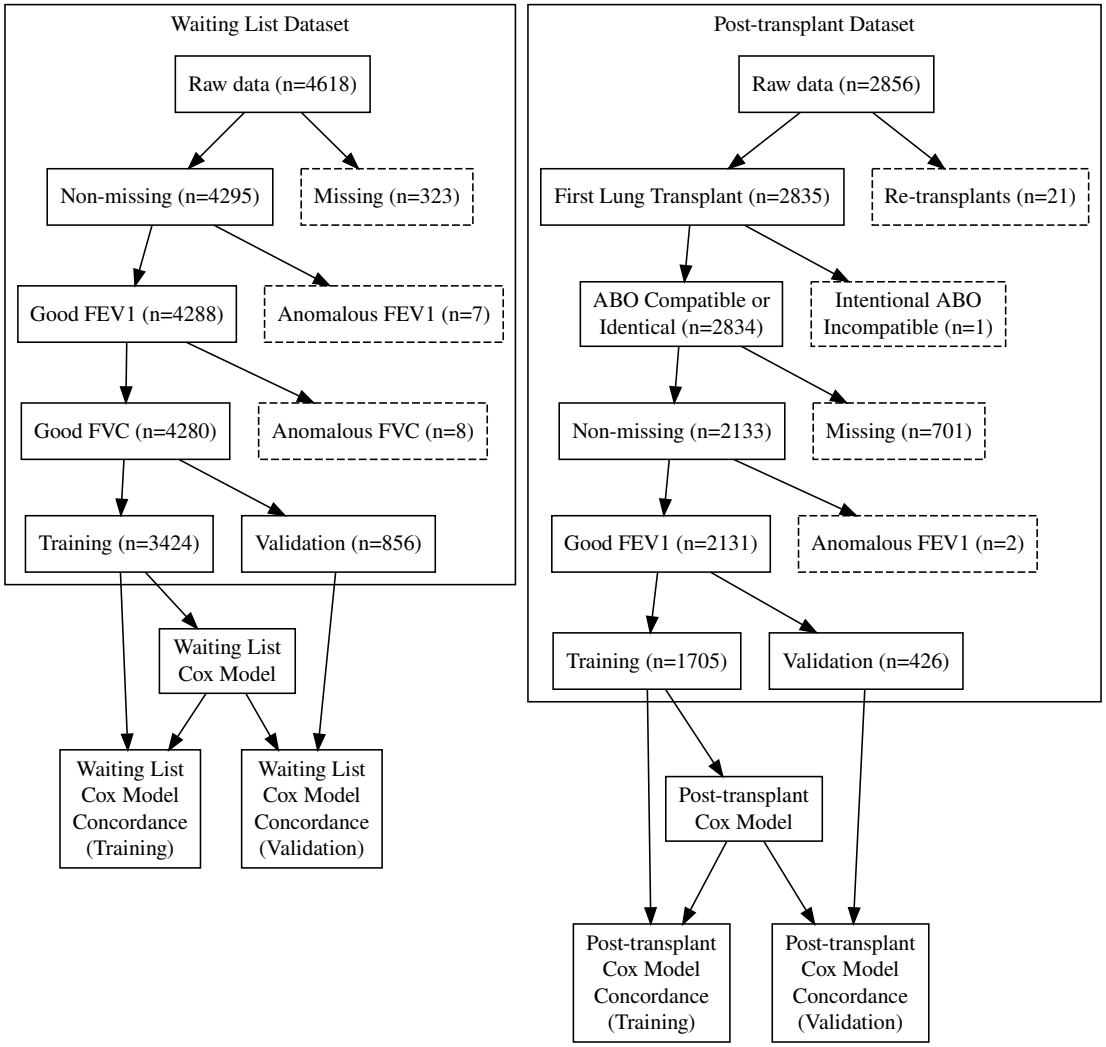


Figure 4.1: Stages of data cleaning for waiting list and post-transplant datasets. Dashed boxes indicate data points that were excluded from analysis. Both Cox models only used the training dataset as input, and the validation concordance indices were calculated using the validation datasets.

NHS-BT lung transplant dataset a p-value threshold of 0.15 was used.

Both waiting list and post-transplant survival durations were capped at 20 years (calculated as $365 \times 20 = 7300$ days). This is significantly longer than what the LAS uses (one year) and CAS (five years), however, the purpose of these models is for simulation and not a prognostic tool. The simulations cover a 20 year period to evaluate the long-term effects of each simulated allocation policy, so any variance in simulated metrics/outcomes is eliminated as they converge towards a mean value. A prognostic tool does not have this flexibility, as it must make accurate predications on a case-by-case basis. By using survival models that are capped at a lower time frame for increases the model's predictive strength/discrimination ability, and is better suited for prognostic use.

The discrimination ability of the survival models was calculated using the concordance

index[134] (also referred to as 'AUC' and 'C-statistic') of the waiting list and post-transplant Cox proportional hazards models. The concordance was calculated using the training datasets and validation datasets (see figure 4.1).

Evaluating Predictive Ability of the NHS-BT UK Lung Allocation Policy

The concordance index of the NHS-BT lung allocation policy was calculated for waiting list and post-transplant survival. This was achieved by calculating a 'score' for each candidate/recipient in the transplant datasets that produce equivalent rankings to the NHS-BT policy.

Table 4.1 shows the scoring system used to predict waiting list survival, and table 4.2 shows the scoring system for post-transplant survival. Note that ABO compatibility is not included in the waiting list scoring system, as this requires comparison to a donor, and is more relevant in the context of post-transplant survival.

The super-urgent and urgent schemes were implemented in the UK in May 2017.[135] To account for this, only candidates listed between 2018 and 2021 were included for these analyses (waiting list n=789, post-transplant n=367). Each candidate was assigned to the same training cohort (waiting list n=619, post-transplant n=302) or validation cohort (waiting list n=170, post-transplant n=65) that was used for the creation of the Cox models for simulation. For each candidate/recipient in the waiting list and post-transplant datasets, the score was calculated according to the relevant table, then a Cox proportional hazards (PH) model was created with the score as the only variable. The concordance index of the Cox models was then calculated for the training and validation cohorts (from years 2018 - 2021) to evaluate the predictive ability of the NHS-BT policy for waiting list and post-transplant survival.

Table 4.1: Scoring system used to approximate rankings from NHS-BT policy to evaluate concordance index for waiting list survival.

Attribute	Score
Super-urgent	4
Urgent, Small Adult	2
Urgent, Adult	1
Non-urgent	0

Generating Realistic Survival Times

Generating realistic survival times was accomplished using the methods outlined by Bender et al.[136] The Cox-Weibull distribution was used for generating post-transplant survival times and the Cox-Gompertz distribution was used for generating waiting list survival times. A detailed explanation of these methods is given in appendix D.2.

To implement these methods it was necessary to calculate the distribution parameters of the Cox-Weibull and Cox-Gompertz distributions. The `FlexSurv` package[137] was used

Table 4.2: Scoring system used to approximate rankings from NHS-BT policy to evaluate concordance index for post-transplant survival.

Attribute	Score
Super-urgent	16
Urgent, Small Adult, ABO Identical	8
Urgent, Small Adult, ABO Compatible	4
Urgent, Adult, ABO Identical	2
Urgent, Adult, ABO Compatible	1
Non-urgent	0

in `R` to fit a Weibull distribution to the waiting list survival dataset and also the post-transplant survival dataset. This package outputs the shape and scale parameters, allowing the survival time formula given by Bender et al.[136] to be used, the calculated values of these parameters are in results section 4.3.3.

Post-transplant survival times were generated using a combination of recipient characteristics, donor characteristics, and type of transplant (i.e., left-lung, right-lung or lung pair) - see appendix D.2.1 for more detail.

The UK dataset was checked for informative censoring (i.e., a correlation between risk of death on the waiting list and probability of receiving a transplant, see page 49 in section 2.4) using methods described by Collett et al.[133] and was found to not have informative censoring present. More detail on the informative censoring methods is given in appendix D.2.2.

Simulating Queueing Processes

The two main processes to simulate are lung transplant candidates being added to the waiting list, and donors becoming available for allocation. Both of these processes can be thought of as queueing processes; there are varying intervals of time and varying numbers of 'arrivals' over time.

To simulate both of these processes, there are two questions that must be answered:

1. How long to wait between events?

2. When an event does occur, how many events should occur?

This was achieved by generating two tables for each queueing process: one containing the frequencies of time between events, and another containing the number of events that occur on the same day mapped to their corresponding frequency (a detailed step-by-step explanation of how this was achieved is given in appendix D.3).

4.2.2 Simulating the Current NHS-BT Policy

There were several challenges when simulating the *current* iteration of the NHS-BT lung allocation policy. The first challenge was that the current policy allows a 'free centre choice' for non-urgent candidates, which is a subjective choice made at the time of allocation. For a detailed simulation, a step-by-step algorithm needs to be implemented for allocation. However, with the data that was available, it was not possible to simulate these subjective allocation decisions in an algorithmic fashion.

The next challenge was that the dataset dates back to 2002, and the allocation policy has changed over time, with the most recent large change occurring in 2017 with the introduction of the 'urgent' allocation scheme. Prior to this time, all candidates in the dataset are labelled as 'non-urgent', even if they would be classed as 'urgent' by the current allocation policy.

A comparison of survival rates pre- and post-2017 showed that the differences in survival were not statistically significant with the introduction of the urgency scheme. In order to simulate the current NHS-BT policy, the following simplified allocation policy was used:

1. All candidates located in same local zonal centre as the donor
2. All other candidates located at other centres, according to the adult lung centre rota

At the beginning of each simulation run, the five transplant centres (labelled A - E) were placed on the rota in a random order. Next, using roulette selection, a random centre was assigned to each donor that became available for allocation, with the probability of each centre being proportional to the transplant volume of each anonymised centre in the dataset.

Each simulation run was initialised with 300 candidates on the waiting list. In order to evaluate how accurately the current NHS-BT lung allocation policy can be simulated, the annual growth rate for the number of donors and listings was set to zero. This decision was made as a result of the roulette wheel selection parameters being calculated from 2002-2019, so the growth in numbers over that period was already "baked-in" to the roulette wheel selection process.

In order to compare the NHS-BT policy against future potential policies it was re-simulated using an annual compound growth rate of 3%. This resulted in two NHS-BT policies being simulated: a 'historic' one to evaluate the accuracy of the simulations, and a 'future' one to evaluate the future performance compared to alternative policies.

4.2.3 The Initial 'Standard' Five Simulated Policies

Five initial policies were simulated over a duration of 20 years, with the number of listings and donor offers increasing by 3% each year. These policies will be referred to as the "standard", "initial", or "bilateral" policies throughout this thesis, to distinguish these policies from the SLT policies that will be described next.

- Waiting list priority (WL)
- 2:1 Waiting list to post-transplant priority (2:1)
- 1:1 Waiting list to post-transplant priority (1:1)
- 1:2 Waiting list to post-transplant priority (1:2)
- Post-transplant priority (PTX)

Each policy determines the relative priority-ratios of waiting list survival to post-transplant survival. A 1:0 ratio would give full priority to waiting list survival (corresponding to 'Waiting list priority (WL)' above), whereas a 0:1 ratio would give full priority to post-transplant survival (corresponding to 'Post-transplant priority (PTX)' above).

Each candidate was assigned randomly generated survival durations for waiting list (WL) and post-transplant (PTX) survival, however to prioritise the waiting list the expected value (EV) of these durations was used. The EV was calculated as the area under the survival curve, integrated up to 20 years.

To increase computation speed, linear predictors (see page 211 in appendix A) were rounded to 3 decimal places and mapped to pre-computed EV's in two separate lookup tables corresponding to waiting list and post-transplant survival (see page 257 in Appendix D.5.14).

The combination of each candidate's WL and PTX survival durations with the priority ratios of the policy determined each candidate's overall rank. Using the same formula given by UNOS[42] the priority score was calculated as:

$$\text{Priority Score} = (\text{PTX Ratio} \times PTX) - (\text{WL Ratio} \times WL)$$

Once every candidate had a priority score assigned, the waiting list was then prioritised from highest to lowest priority score.

Finally, screening criteria were applied to the prioritised list. In the case of the UK lung simulations, ABO matching rules were applied so that the donor-recipient match had to be at least compatible. Height matching rules were also in place, with a maximum height difference between the donor and recipient of 15cm being in place.

To assess the impact of an increased donor pool on the number of waiting list deaths, the WL policy was also simulated with a 5%, 10%, 25% and 50% increase in the number of donors.

Each policy/scenario was simulated 40 times in order to allow the metrics to converge towards a mean value, for a simulated duration of 20 years. For each simulation run, the waiting list was initialised with 300 randomly generated candidates from the NHS-BT dataset. The simulation engine was designed to run multiple independent simulations in parallel, and for this work, eight simulations were run in parallel on an Intel® Core i7 processor (7th generation). It took approximately two to eight minutes to complete all 40 simulation runs spanning 20 years depending on the complexity of the policy.

4.2.4 The SLT (Single-lung Transplant) Policies

The work by Benvenuto et al.[80] inspired an alternative allocation algorithm was evaluated which maximises the use of SLT for ILD (group D) recipients. The algorithm works as follows:

1. Predict net benefit for highest ranked candidate with ILD if transplanted with (left / right) lung only
2. Predict net benefit for next-highest ranked candidate with ILD if transplanted with remaining lung
3. Predict net benefit for highest ranked candidate (any diagnosis) if receiving lung pair
4. If no suitable candidate is identified in step 1 or 2, then allocate the pair
5. Calculate the total net benefit for candidates in steps 1 and 2, if this is greater than the net benefit calculated for the candidate in step 3, then allocate two single lungs, otherwise allocate the pair

This algorithm does not dictate how candidates should be ranked and must be combined with another ranking algorithm. The SLT algorithm was initially combined with the WL policy ('SLT-WL policy') to investigate the impact of the organ offer order, age threshold for SLT, and increased utilisation. The SLT algorithm was also combined with the other policies to investigate the general impact of increasing the use of SLT.

The order of offering impacts which candidate(s) receive which lung(s), the following orderings were simulated:

1. Pair, left lung, right lung
2. Pair, right lung, left lung

The SLT-WL policy was also combined with no age threshold, a minimum age of 55 years and a minimum age of 60 years for receiving SLT.

An important point to consider for these SLT policies is that they simulated "ideal" conditions:

- Left/right lung preferences for ILD candidates were not simulated, it was assumed that all ILD recipients could receive a left or right lung (or pair)
- All donors were assumed to donate a lung pair, resulting in 10.5% more lungs being available for transplant compared to the standard policies (19% of donors in the dataset donated a single lung: $19 \div 2 = 10.5$)

4.2.5 Risk-adjusted Benefit and Conditional Survival

Two additional sets of policies were simulated: risk adjusted benefit (RAB) and conditional survival (CON). The RAB policies aimed to prioritise candidates who had both a high expected net benefit and also a high probability of surviving long enough to realise that benefit (for the theoretical background to this, see page 20 in section 2.2.1). A similar approach to that used by Titman et al.[50] was used. For each candidate on the waiting list, the number of days post-transplant that their post-transplant risk of mortality decreased below their risk of mortality remaining on the waiting list was calculated (the 'equity point' as described by Titman et al.).

Next, the probability of each candidate surviving to the equity point was calculated. The risk-adjusted net benefit was then calculated by multiplying each candidate's net benefit by their probability of surviving to the equity point. This resulted in two net benefit values: the standard net benefit and the risk-adjusted net benefit. Finally, a 'risk-weight' must be decided that gives a percentage weight to the risk-adjusted net benefit value, where 0% uses the standard net benefit value, and 100% uses the risk-adjusted benefit value. RAB policies were simulated using the 1:1 WL:PTX policy, in order to compare the impact of RAB on net benefit. The risk weights simulated were 0%, 25%, 50%, 75% and 100%. These policies were named RAB-0, RAB-25, RAB-50 and so on. The mathematics behind this approach is given in appendix D.4.

The CON policies used an alternative risk-adjustment method, based on the conditional survival formula given by Hieke et al.[138] For each candidate, the probability of surviving another t days, given a candidate has already survived s days on the waiting list, was calculated. Each candidate's expected waiting list survival duration was then multiplied by the previously calculated survival probability. This resulted in candidates with lower expected waiting list survival and lower probability of surviving another t days receiving a higher net benefit score. The details of the calculations used in the simulations are given in appendix D.4. The WL and 1:1 WL:PTX policies were used for simulating the CON policies to evaluate the impact on waiting list deaths and net benefit. The simulated policies predicted conditional survival and surviving an additional 1, 7, 30, 365 and 1095 days. When combined with the WL policy, these policies were named CON-WL-1, CON-WL-7, CON-WL-30 and so on. When combined with the 1:1 policy, they were named CON-1:1-1, CON-1:1-7, and so on.

4.2.6 Policy Performance Metrics and Important Assumptions

Seven performance metrics were used to evaluate each simulated policy:

1. Number of annual deaths on the waiting list
2. Days from listing until transplant
3. Net benefit
4. *Relative* benefit
5. Survival rate 1 year post-transplant
6. Survival rate 5 years post-transplant
7. Annual number of transplants

For each performance metric the mean and standard deviation were calculated for comparison to other policies.

Using the appropriate probability distributions and methods described in section 4.2.1 and by Bender et al.,[136] every simulated candidate was assigned a randomly generated waiting list survival duration (WL), and at the point of transplant, a random post-transplant survival duration (PTX). If the time in the simulation surpassed a candidate's waiting list survival duration, that candidate was removed from the waiting list and the number of waiting list deaths was incremented. To calculate the annual waiting list mortality rate, the number of waiting list deaths for the entire simulation run was divided by the number of years that were simulated.

To calculate the mean number of days from listing until transplant, the following durations per candidate and recipient were summed and divided by the total number of simulated candidates and recipients:

1. Days from listing until death on the waiting list
2. Days from listing until transplant
3. Days from listing until end of simulation run

Assumption #1: Item 3 above is included in the calculation of mean waiting time because there will always be some number of patients still awaiting transplant at the time the simulation ends. Although this may skew the mean waiting time slightly this is an acceptable trade-off as (1) it makes use of all simulated individuals, and (2) all simulated policies will use the same calculation, allowing waiting times to be compared between policies.

'Net benefit' is the expected survival duration on the waiting list subtracted from the expected survival duration post-transplant (in days). This is the expected additional days of life gained from receiving a transplant.

There may be scenarios where net benefit is similar for two candidates, but for a candidate with a lower expected waiting list survival, net benefit represents a greater *proportion* of life gained. To evaluate this, 'relative benefit' was introduced as a metric. An example is given in table 4.3.

Table 4.3: To demonstrate the purpose of calculating relative benefit, this table shows three candidates with identical predicted net benefit, but different relative benefit values. For candidate 'A', an additional 2 years of life represents a three-fold increase in lifespan, whereas for candidate 'B', an additional 2 years of life represents only a 1.5-fold increase in lifespan.

Candidate	Waiting List Survival	Post-transplant Survival	Net Benefit	Relative Benefit
A	1 Year	3 Years	2 Years	3
B	4 Years	6 Years	2 Years	1.5
C	2 Years	4 Years	2 Years	2

When calculating net/relative benefit it was important to consider how waiting list and post-transplant survival durations were generated. One side-effect of using the survival time generating methods described by Bender et al.[136] was that the survival duration could approach infinity. To work around this, all generated survival times were capped at 20 years, calculated as $20 \times 365 = 7300$ days. Any survival data generated from the simulations with a duration over 7300 days had the duration set back to 7300 days and the censoring indicator set to 0 to indicate a death had not been observed. The duration of 20 years was selected as this approximated the longest survival duration observed in the NHS-BT post-transplant dataset and it also minimised the number of capped survival times than if a shorter capping duration were chosen.

This lead to the question of how to calculate net benefit and relative benefit when there was potential for one or both of the survival durations to have been capped/censored? Table 4.4 shows the possible combinations when calculating net/relative benefit metrics.

Table 4.4: There are four possible combinations of censoring/capping for waiting list and post-transplant survival durations generated by the simulation engine. Each of these combinations must be considered when calculating performance metrics for simulated policies.

Combination #	Waiting List Survival	Post-transplant Survival
1	Not Capped	Not Capped
2	Capped	Not Capped
3	Not Capped	Capped
4	Capped	Capped

Assumption #2: If the waiting list survival duration is capped (combinations 2 and 4), exclude that data point from the net/relative benefit calculations.

Assumption #3: If the post-transplant survival duration is capped, it is still acceptable to calculate net benefit, the result is a lower-bound estimate of net benefit.

Net benefit was calculated as $PTX - WL$ where PTX was the post-transplant survival duration in days and WL was the waiting list survival duration in days. Assumption #2 prevents over-estimation of net benefit in cases where WL is capped. Since the true value of WL is greater than the capped value, as the possible values for WL increase, the possible values for net benefit decrease. If only the capped value of WL is used then the upper-bound of net benefit would be calculated, resulting in over-estimation of net benefit.

In the case of assumption #3, using a capped PTX duration to calculate net benefit results in calculating the lower-bound of possible net benefit values. This is acceptable as it does not over-estimate net benefit: it is better to under-estimate net benefit than over-estimate it.

The same assumptions apply to relative benefit. Once combinations #2 and #4 are removed, relative benefit is calculated as: $\frac{PTX}{WL}$

The 1- and 5-year post-transplant survival rate is the total number of recipients surviving to at least 1/5 year(s) after transplant, divided by the total number of recipients that were transplanted. There were two options for the population used for calculating post-transplant survival rates. The first option was to use all simulated recipients, regardless of capping/censoring, and the second option was to use the same subset of recipients used for calculating net/relative benefit.

Assumption #4: Only recipients with uncapped/uncensored waiting list survival durations were used for calculating post-transplant survival rates. The net/relative benefit metrics are measures of post-transplant outcomes, so it makes sense to use the same subset of patients for evaluating post-transplant survival rates.

Assumption #5: Transplant volume is not a measure of post-transplant outcomes, capping/censoring does not apply here, therefore use the count of all transplanted recipients in the calculation. The annual transplant volume is simply the total number of transplants performed in a simulation run divided by the number of years simulated, in the case of this work that was 20 years. Censoring/capping of survival durations was not considered for this metric, and all transplants within the simulation were used in this calculation.

It was important to note that the *average* net benefit per *transplanted* recipient did not give a full indication of how a policy was performing. To fully evaluate how a policy was benefitting the population of lung transplant candidates and recipients, it was necessary to record the *total* net benefit that **would** have been gained, *if* each candidate that died on the waiting list were transplanted.

The potential net benefit that was lost was used to evaluate the opportunity cost of not transplanting lower priority candidates according to the simulated policy. This was then compared to the total net benefit gained, with the difference between realised and unrealised net benefit indicating how the entire population benefitted. Assumption #2 also applies in this case as it relates to the calculation of net benefit, therefore the subset of recipients with uncapped/uncensored waiting list survival durations were used for these calculations.

Simulated post-transplant survival depended on the characteristics of the donor, for simulated candidates dying on the waiting list the characteristics of the donor had to be assumed. **Assumption #6: When calculating the net benefit that was lost, a lung pair from a Cytomegalovirus (CMV) negative, non-diabetic, 170cm tall, DBD donor was assumed.**

One final consideration is that the candidate with the highest priority score is transplanted, even if that score corresponds to a negative net benefit. This also occurs in reality with the LAS, which can be demonstrated by using the equations given in the UNOS guide to calculating the LAS:[42]

$$\text{Raw Score} = \text{PTX} - (2 \times \text{WL})$$

$$\text{LAS} = 100 \times \frac{\text{Raw Score} + 730}{1095}$$

For a net benefit of zero, WL = PTX, and because the raw score gives twice the weight to WL, there are a range of possible LAS values that meet these conditions. The maximum LAS for zero net benefit is 66.67, corresponding to WL = PTX = 0, and the minimum LAS for zero net benefit is 33.33, corresponding to WL = PTX = 365.

This means for LAS scores $\leq$ 33.33, net benefit will always be zero or negative. For scores between 33.33 and 66.67, there is a probability of net benefit being zero or negative: this probability is highest with scores near 33.33, lowest with scores near 66.67, and 50% with a LAS of 50. Above 66.67 net benefit will always be positive (though it could be as low as one day of net benefit). This leads to the final assumption:

Assumption #7: Lungs will be allocated to the highest scoring candidate, even if that score corresponds to a negative (or zero) net benefit. The allocation score is calculated using the expected value (i.e., the average value over multiple samples), so even if the expected net benefit is zero or negative, due to the simulation engine using randomised processes, there is still a chance for a recipient to experience net benefit. This is also a practical consideration: where should the threshold for net benefit be set for allocation? This would require experimenting with different thresholds for each simulated policy, increasing the number of required simulations. There is also the problem of a candidate having a predicted net benefit one day less than the threshold. This would re-introduce hard boundaries to the allocation policy. For these reasons, a threshold for net benefit will not be used.

Assumption #8: The SLT policies assume ideal conditions for donors and recipients. As discussed in section 4.2.4 on page 85, all donors are assumed to donate two lungs, and all SLT recipients are assumed to be able to receive a right or left lung.

4.2.7 Discrete Event Simulation 💻

Overview

The simulation engine developed for simulating UK lung allocation policies will be described at a high level in this section (see figure 4.2). A detailed explanation using flow charts and pseudocode is given in appendix D.5.

The components of the simulation engine were organised as follows:

1. Entry point - The start of the simulation engine. This component is responsible for creating the simulation threads that will run in parallel and aggregating the results over multiple simulation runs
2. Thread - Each simulation runs in its own thread, each thread initialises a unique simulation scenario
3. Listing Process - This is the process of candidates being added to the waiting list
   (a) Roulette selection is used to determine the time delay between listings and also the number of listings that will occur
   (b) A waiting list survival duration will be randomly generated for each candidate
   (c) A waiting list death event will be scheduled - this will be cancelled if the candidate is transplanted
4. Offering Process - This is the process of donors becoming available for allocation to candidates on the transplant waiting list
   (a) Candidates on the waiting list are prioritised according to the allocation policy. For each candidate:
      i. Calculate the linear predictor (LP) for WL and PTX survival using the corresponding survival models
      ii. Map the LP to the **expected** (i.e., mean) survival duration using the corresponding WL or PTX lookup table (see page 84)
      iii. Multiply the expected survival durations by the corresponding priority ratios
      iv. Calculate allocation score (see page 84)
   (b) Candidates are screened using height matching and ABO compatibility rules
   (c) Post-transplant metrics are calculated for the allocated candidate: post-transplant survival duration, net benefit, relative benefit and total waiting time

The Thoracic Simulated Allocation Model (TSAM)

The application of DES to lung allocation will be approached in a similar way to the TSAM used by SRTR. Chapter 2 of the TSAM manual[139] outlines assumptions, processes and events, and text has been used verbatim in table 4.5 that list the TSAM approach and any modifications that will be used in the custom simulation engine.

Table 4.5: Comparison of the Thoracic Simulated Allocation Model (TSAM) by the Scientific Registry of Transplant Recipients (SRTR), and the custom simulation engine developed for this research. **TSAM** entries are taken verbatim from the TSAM manual.[139]

	Feature 1
TSAM	Arrivals of candidates are input to the model with a data file
Custom Engine	Newly listed patients will be simulated using the probability distribution derived from the lung transplant dataset
	Feature 2
TSAM	The initial wait list is input to the model with a data file
Custom Engine	The initial waiting list will be populated with randomly selected patients from the dataset
	Feature 3
TSAM	An entire history of wait-list status changes (to the end of the Allocation Run, death, or removal from the wait-list) must be input to the model for each patient
Custom Engine	Death and removal from the waiting list will both be counted as a waiting list death. Rather than having a fixed sequence of status changes, the time until death/removal on the waiting list will be randomised.
	Feature 4
TSAM	Once a candidate receives an organ, that patient's input stream of status changes no longer applies. If the patient relists, the model assigns a status change history to the patient by randomly selecting a set of status changes from a pool of user-defined histories
Custom Engine	If a patient is allocated lungs their waiting list death event is cancelled. Relisting will not be simulated.
	Feature 5
TSAM	The values of several other parameters are specified in the program or tables and remain constant during a run. These include parameters of the graft failure time distribution and geographic relationships among institutions
Custom Engine	Other values necessary for the simulation will be sampled randomly from appropriate distributions. National allocation is assumed, allocation zones only for current NHS-BT policy.

	Feature 6
TSAM	Organ acceptance: The user defines a calculation used to compute the organ acceptance probability.
Custom Engine	Organ acceptance probabilities will not be simulated.

	Feature 7
TSAM	Post-graft survival: The user defines a calculation which is then used to determine the patient's death date after a transplant.
Custom Engine	Post-transplant survival will be simulated using the same method used for simulating waiting list survival.

	Feature 8
TSAM	Organ arrivals (input-driven)
Custom Engine	Organ arrivals will be randomised using a probability distribution derived from the dataset

	Feature 9
TSAM	Status changes for wait list candidates who have not yet received grafts (input-driven)
Custom Engine	There will only be two possible status changes: death on the waiting list (simulated using randomly generated survival duration) and receiving a transplant (determined by the allocation policy being simulated)

	Feature 10
TSAM	Patient arrivals (input-driven)
Custom Engine	Patient arrivals will be randomised using a probability distribution derived from the dataset

	Feature 11
TSAM	Status changes for relisted graft recipients (sampled from a pool of possible outcomes)
Custom Engine	Relisting will not be simulated

	Feature 12
TSAM	Relisting events of recipients whose grafts fail (sampled by the model)
Custom Engine	Not simulated

	Feature 13
TSAM	Deaths of graft recipients not on the wait list (sampled by the model)
Custom Engine	Post-transplant survival will be determined by a randomly generated post-transplant survival duration

Flowcharts and Pseudocode

A detailed explanation of every step and calculation in the simulation engine is given using a combination of flowcharts and pseudocode in appendix D.5. These explanations describe the low-level detail of exactly how each step in the simulation is performed, without requiring knowledge of a specific programming language. A high-level overview is shown in figure 4.2. For this work, Python 3[113] was used along with the SimPy framework for discrete event simulation.[140]

Two separate processes ran in parallel: the candidate listing process and the donor availability process. The candidate listing process waits a random interval of time before adding one or more candidates to the waiting list, with the randomised interval and number of candidates being sampled from a distribution calculated from UK lung transplant data (see appendix D.3 for an explanation of how this was achieved). Donors also became available for transplant at a frequency and interval that was calculated from UK lung transplant data, when a donor became available the allocation policy was executed.

The allocation policy was coded in Python, allowing a very large degree of flexibility and possible policies to be simulated. The allocation policy screens out candidates that are not ABO compatible with the donor or not within 15cm of the donor's height, then ranks all remaining candidates from highest to lowest allocation score (calculated according to the policy), allocating the donor lungs to the highest ranked candidate.

All candidates added to the waiting list were assigned a randomly generated waiting list survival time using the methods described by Bender et al.,[136] and according to their individual characteristics sampled from the UK lung transplant dataset, the Cox model built for simulating waiting list survival (see page 4.2.1 for a high-level description of this process or appendix D.1 for a detailed explanation), and the parameters that describe the shape of the survival curve for candidates on the active UK lung transplant waiting list. Once a candidate's waiting list survival time elapsed, if they had not yet been transplanted they would be counted as a waiting list death.

When a candidate was transplanted according to the allocation policy, their post-transplant survival duration was calculated, along with net benefit, relative benefit, and total waiting time to transplant.

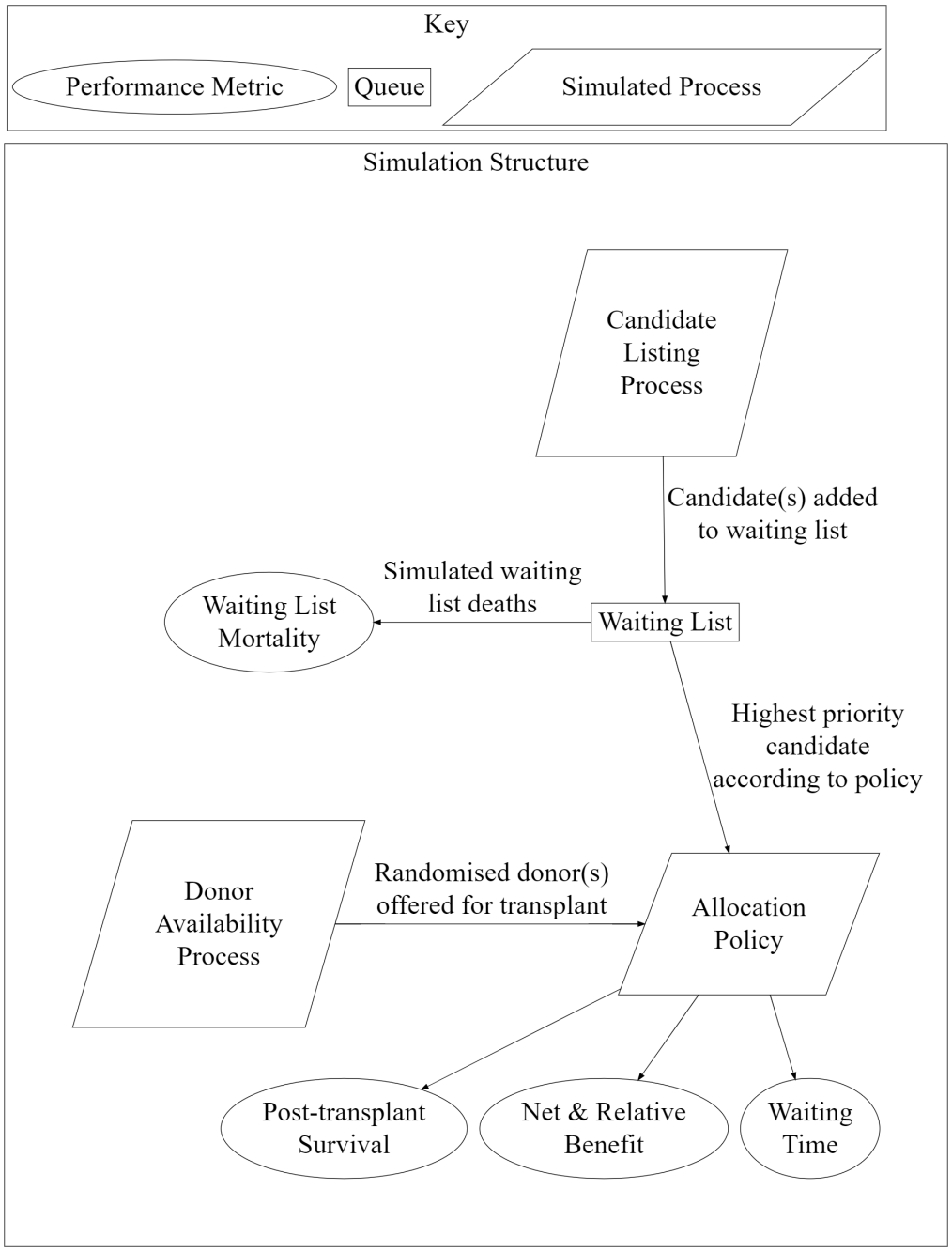


Figure 4.2: High-level overview of major components of the simulation engine. The candidate listing and donor availability processes ran separately in parallel. The candidate listing process would add candidates to the waiting list, using time intervals between candidates being listed that were informed by UK lung transplant data. The donor availability process worked in the same way, and would cause the allocation policy being simulated to select the highest priority candidate from the waiting list. Post-transplant survival, net benefit, relative benefit, and waiting time metrics were then recorded. Candidates not transplanted before their simulated waiting list survival duration elapsed were counted as waiting list deaths.

4.2.8 Simulation Engine Evaluation Methods

There were a number of assumptions and complex interactions that needed to be evaluated in order to ensure the correctness of the simulation engine implementation and the validity of any results generated.

Evaluating Survival Durations

The most important feature of the simulation engine is its ability to generate realistic survival durations. The following sequence of steps were followed to evaluate the accuracy of the survival durations generated by the simulation engine:

1. Split the dataset 80/20 into a training and validation dataset
2. Initialise empty list `diffs`
3. For each `subset_size` from 100, 200, ..., `max_subset_size`:
   (a) Initialise temporary empty list `tmp_diffs`
   (b) Repeat `num_runs` times:
      i. Select `subset_size` individuals from the dataset
      ii. Using the covariates of each individual in the subset, generate a realistic randomised survival duration
      iii. Identify the maximum observed survival duration in the dataset and store in `max_obs`
      iv. Calculate the restricted mean (using `max_obs`) for the simulated (`sim_rmean`) and observed (`obs_rmean`) survival curves
      v. Add `sim_rmean - obs_rmean` to `tmp_diffs`
   (c) Calculate the mean of `tmp_diffs` and add to `diffs`
4. Output table with columns: `subset_size` and `diffs`

This process was repeated for both the training and validation cohorts from the waiting list and post-transplant datasets. For each evaluation run, `num_runs` was set to 1000 and `max_subset_size` was set to the largest multiple of 100 that was less than or equal to the number of rows in the dataset being evaluated. For the waiting list dataset, `max_subset_size` was set to 3100 for the training dataset and 700 for the validation dataset. For the post-transplant dataset, `max_subset_size` was 1600 for the training dataset and 400 for the validation dataset.

Evaluating Simulations of the Current NHS-BT Policy

To evaluate the accuracy of the simulated NHS-BT lung allocation policy, six key measures were used to compare the simulation results to observed results:

1. Waiting list survival restricted mean
2. Post-transplant survival restricted mean
3. 1-year post-transplant survival rate
4. 5-year post-transplant survival rate
5. Number of transplants per year
6. Mean waiting time

The Welch Two Sample T-Test was used with the `R` function `t.test`, to compare the simulated to the observed measures. The goal was to ensure that the distribution of measures were **not** statistically significantly different.

The simulated results and true observations were also compared visually. The proportion of candidates that were transplanted, died on the waiting list, or still waiting at 6 months, 1, 2 and 3 years after listing were recorded. These proportions were plotted and compared against the annual NHS-BT organ activity reports where these same figures are published.

The 1 and 5 year post-transplant survival rates and number of transplants were also reported, these figures were compared to the 2016-2017 and 2017-2018 annual reports on cardiothoracic transplantation published by NHS-BT.[23,141]

4.3 Results

In this section the resulting waiting list and post-transplant survival models will be presented in section 4.3.1, along with the simulation engine validation results using those models (section 4.3.4), where results are compared to those observed in lung transplant datasets provided by NHS-BT and also in published annual organ activity reports.[23,141]

A number of allocation policies and variations on those policies are presented along with their recorded metrics of using the given survival models within the simulation engine. Initially the impact on candidates and recipients as priorities move between prioritising waiting list survival and post-transplant survival is presented in section 4.3.5. Next, the policies are adjusted to increase the number of SLTs for ILD recipients in section 4.3.5, followed by different approaches of adjusting priorities based on conditional survival, opportunity cost and risk of mortality in sections 4.3.5 - 4.3.5.

The metrics of the simulated policies will be presented at a population level and are also stratified by diagnosis group, age group and blood group in section 4.3.5.

4.3.1 Waiting List and Post Transplant Survival Models

Training and Validation Datasets

In section 4.2.1 it was assumed that selecting a random subset of patients would result in a validation dataset that is representative of the training dataset. Numeric variables were compared between the training and validation datasets using the Welch Two Sample T-Test, and categorical variables were compared using the Chi-squared test. All p-values were > 0.05 for both the waiting list and post-transplant datasets. The results are summarised in table 4.6.

Table 4.6: Statistical comparison of training and validation datasets using the Welch Two Sample T-Test for numeric variables and the Chi-squared test for categorical variables. A '-' in the table indicates this variable was not included in the corresponding waiting list/post-transplant model.

Variable	Type	Waiting List p-value	Post-transplant p-value
BMI	Numeric	0.68	0.66
Age	Numeric	0.59	0.56
FEV1	Numeric	0.11	-
FVC	Numeric	0.88	-
Creatinine	Numeric	0.99	0.98
Height Delta	Numeric	-	0.42
Donor Age	Numeric	-	0.69
Donor BMI	Numeric	-	0.86
ABO	Categorical	0.90	0.11
Diagnosis Group	Categorical	0.09	0.69
Sex	Categorical	0.72	-
CMV	Categorical	0.09	-
Diabetes	Categorical	0.31	-
Smoker	Categorical	0.26	-
Home Oxygen	Categorical	0.65	0.88
Donor Type	Categorical	-	0.49
Previous Malignancy	Categorical	-	0.99
Organ Transplanted	Categorical	-	0.86
Donor CMV	Categorical	-	0.68
Donor Diabetes	Categorical	-	0.94

Waiting List Survival Model

The automated process described in section 4.2.1 identified the categorical, linear and non-linear terms to be included in the waiting list survival model. Out of the variables available in the waiting list dataset (see Appendix F), 12 were selected for inclusion in the model (any variables with a high proportion of missing entries or that resulted in the model not converging were excluded). The forest plot in figure 4.3 contains the categorical variables in the model, showing (from left to right): the variable name, the possible categorical values for that variable, the number of individuals in that category (N), the hazard ratio and

confidence intervals (visualised), the hazard ratio and confidence interval values (or a label specifying the value is the reference value) and p-value showing statistical significance.

The residual plots in figures 4.4, 4.5, and 4.6 contain the non-linear terms, fit with restricted cubic splines (RCS). For a detailed explanation of interpreting these models and plots, see the introduction to survival analysis given in appendix A. The non-linear variables were BMI, age, FEV1, FVC and creatinine level at time of listing. The residual plots in figures 4.4, 4.5, and 4.6 show how risk changes with respect to the values of the variables. A value of 0 indicates no change in risk, a positive value indicates a higher risk of mortality and a negative value indicates a lower risk of mortality.

The categorical variables were: ABO (not statistically significant, but included to assess impact by blood group), diagnosis group (group A - COPD - being the reference group), sex, CMV status, diabetes status, smoking history and requirement for home oxygen.

All diagnosis groups were at a statistically significant higher risk of mortality on the waiting list compared to COPD (group A), with ILD (group D) having the highest risk with a hazard ratio of 3.37. For the variable 'sex', males were the reference group with females having a lower hazard ratio of 0.62. History of CMV infection resulted in a slightly elevated risk of mortality with a hazard ratio of 1.14 that was close to statistical significant ($p = 0.075$). Insulin-dependent diabetes carried a higher risk compared to non-insulin dependent diabetes or no diabetes, with a hazard ratio of 1.47 ($p = 0.002$). Having a smoking history resulted in a hazard ratio of 1.56, however the confidence intervals were wide resulting in a p-value of 0.119. Requiring home oxygen was statistically significant ($p < 0.001$) and had a hazard ratio of 1.74.

On the training dataset (n=3424) the C-statistic was calculated as 0.73, and on the validation dataset (n=856) the the C-statistic was 0.66. Calibration plots are also shown for 1-year and 5-year survival in figure 4.7.

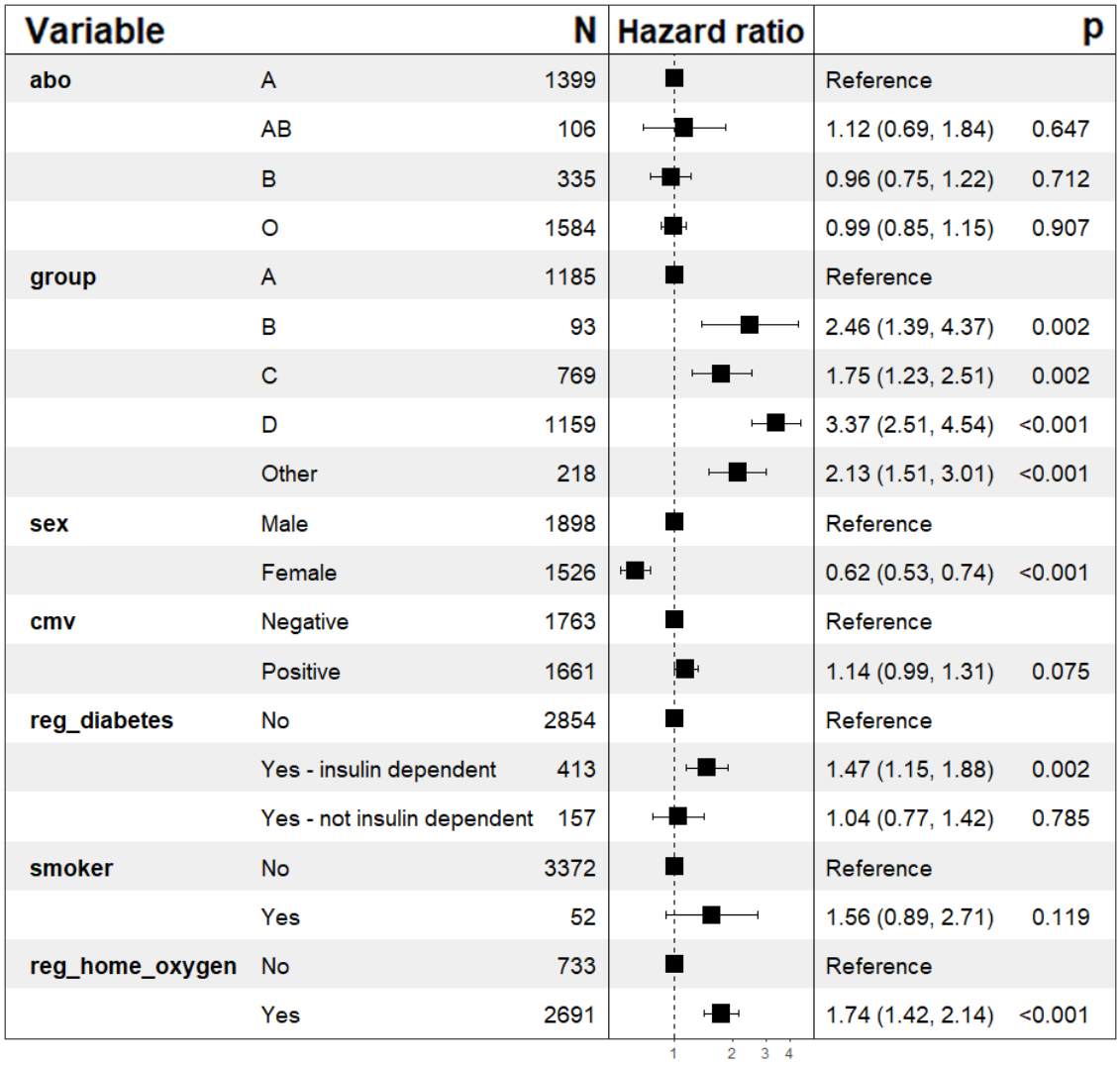
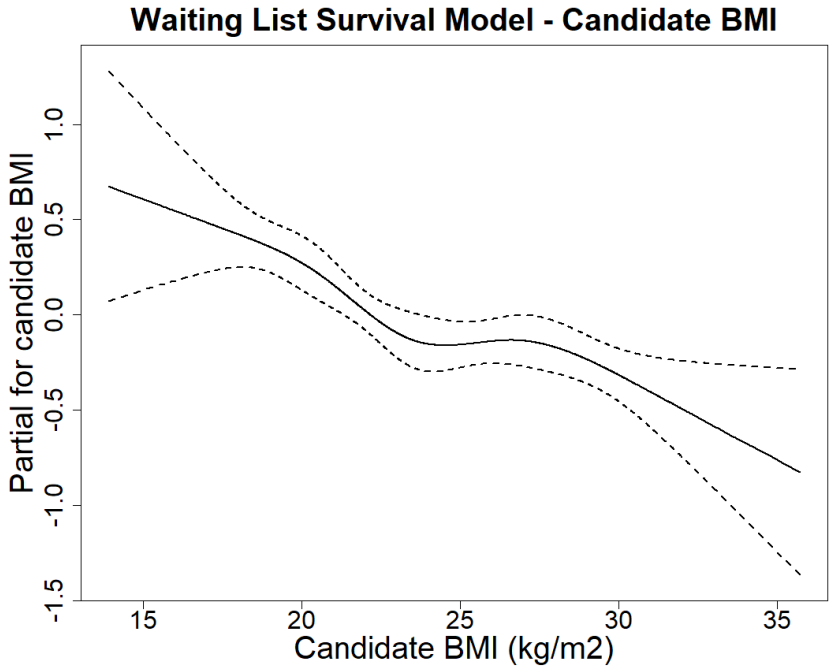
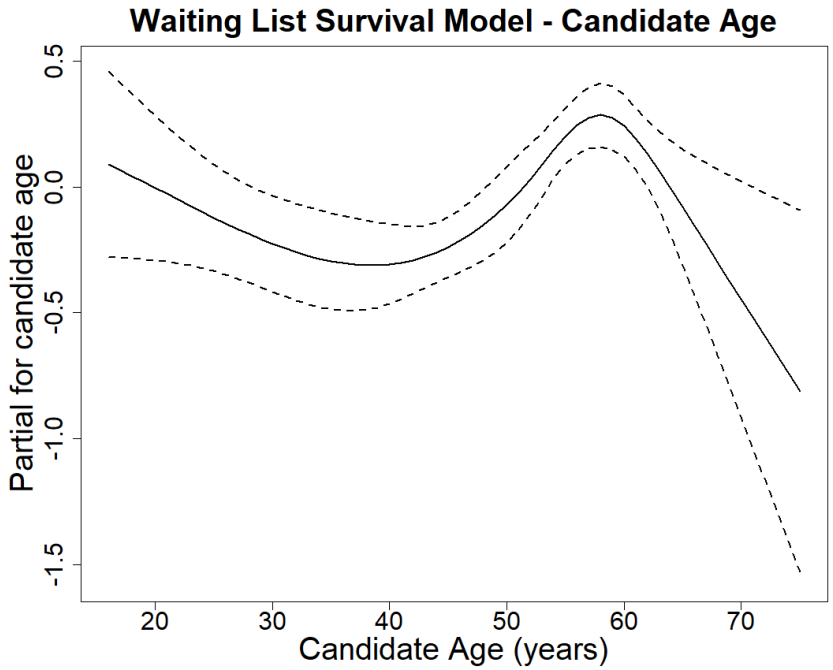


Figure 4.3: Categorical variables: **abo** = blood group, **group** = diagnosis group: A = Chronic Obstructive Pulmonary Disease, B = Pulmonary Arterial Hypertension/Pulmonary Vascular Disease, C = Cystic Fibrosis, D = Interstitial Lung Disease., **sex** = biological sex, **cmv** = cytomegalovirus status, **reg_diabetes** = diabetes status at registration, **smoker** = smoking status, **reg_home_oxygen** = requirement of oxygen at home at time of registration

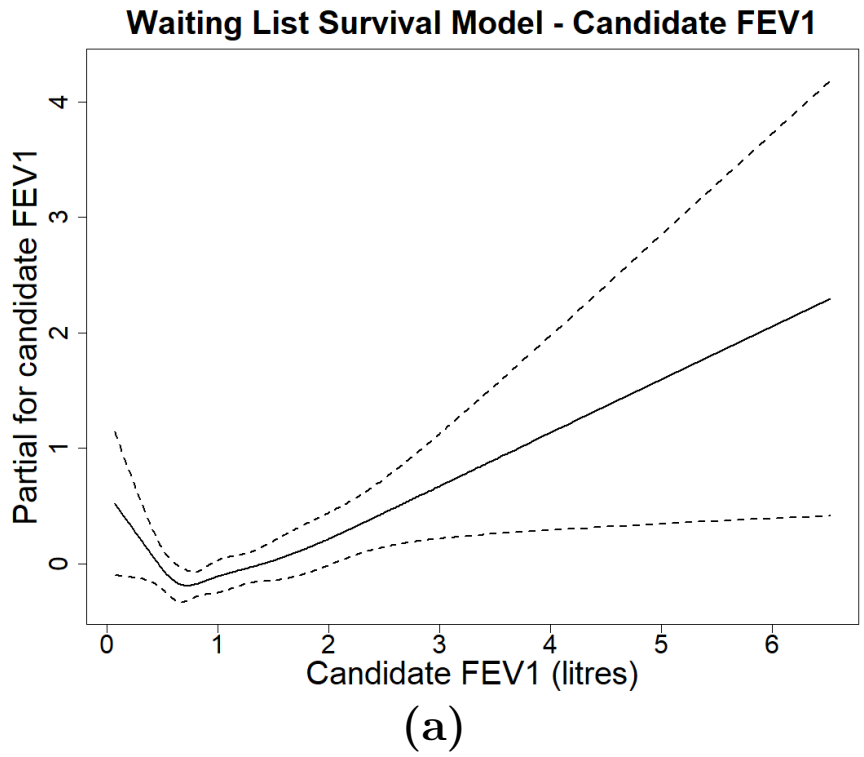


(a)

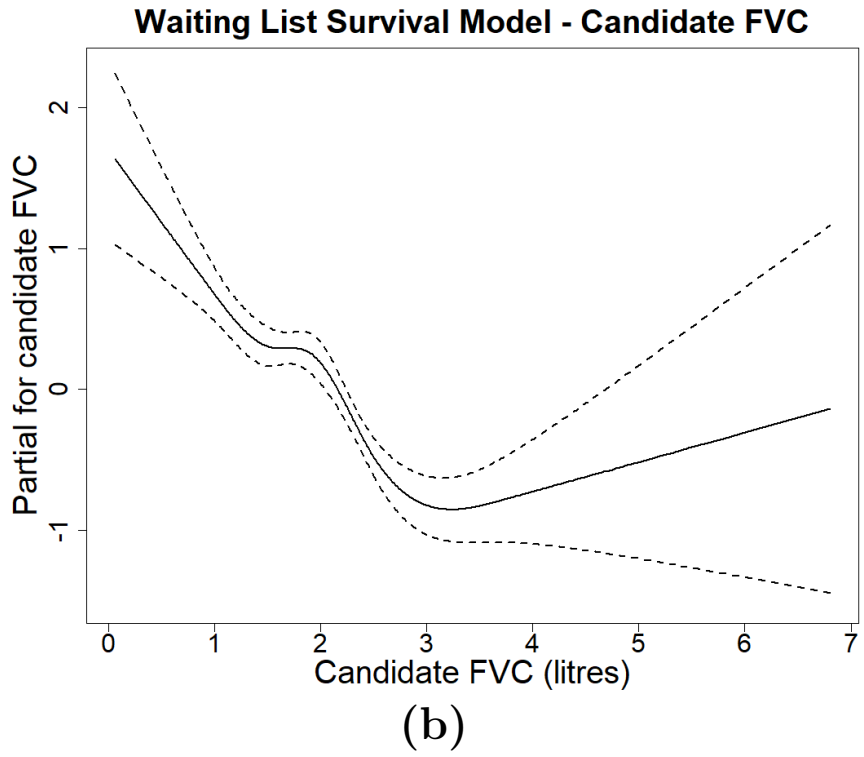


(b)

Figure 4.4: Residual plots of the non-linear terms in the waiting list survival model fit with restricted cubic splines. Positive values on the y-axis indicate a higher risk of mortality, negative values indicate a lower risk of mortality. **(a)** Risk of mortality generally decreased with increasing body mass index (BMI), and increased as BMI decreased below (approximately) 23 kg/m^2. **(b)** Risk of mortality was lowest with a candidate age near 40 years and highest with a candidate age near 60 years. The decrease after 60 years of age was likely due to the relatively low numbers of candidates aged over 60.



(a)



(b)

Figure 4.5: Residual plots of the non-linear terms in the waiting list survival model fit with restricted cubic splines. Positive values on the y-axis indicate a higher risk of mortality, negative values indicate a lower risk of mortality. **(a)** Mortality risk was lowest with an FEV1 of approximately 0.8 litres, with risk increasing with FEV1 values both lower and greater than this value. **(b)** Risk of mortality was increased with FVC measurements less than 2 litres and increased for FVC values greater than 2 litres.

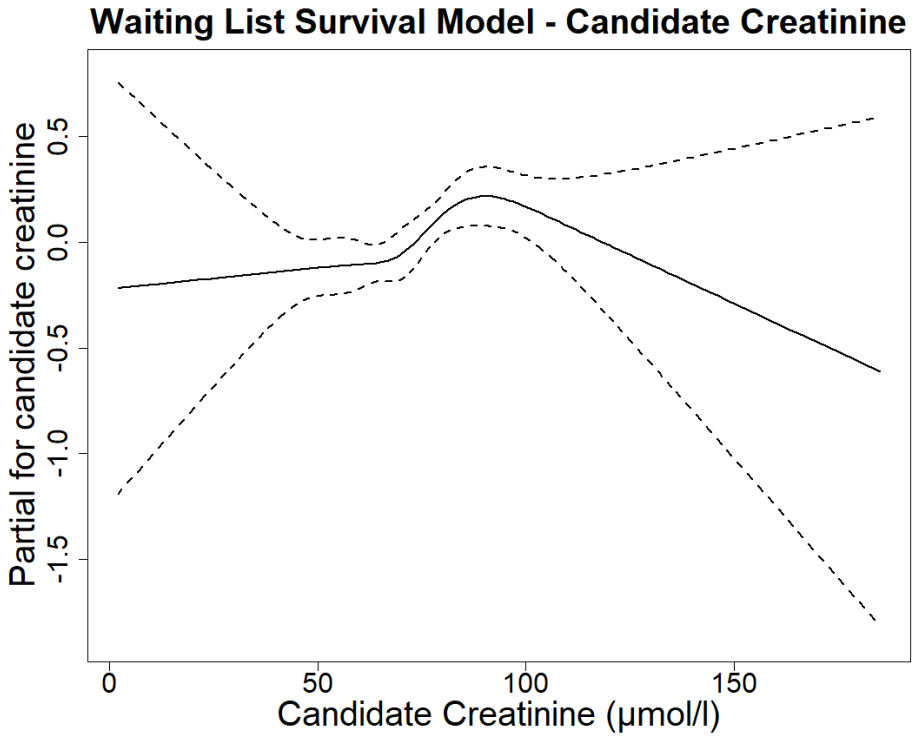
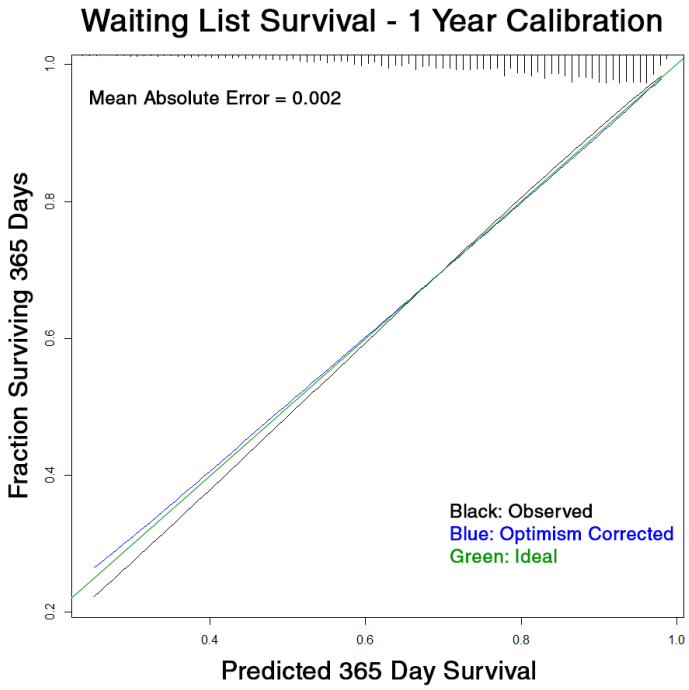
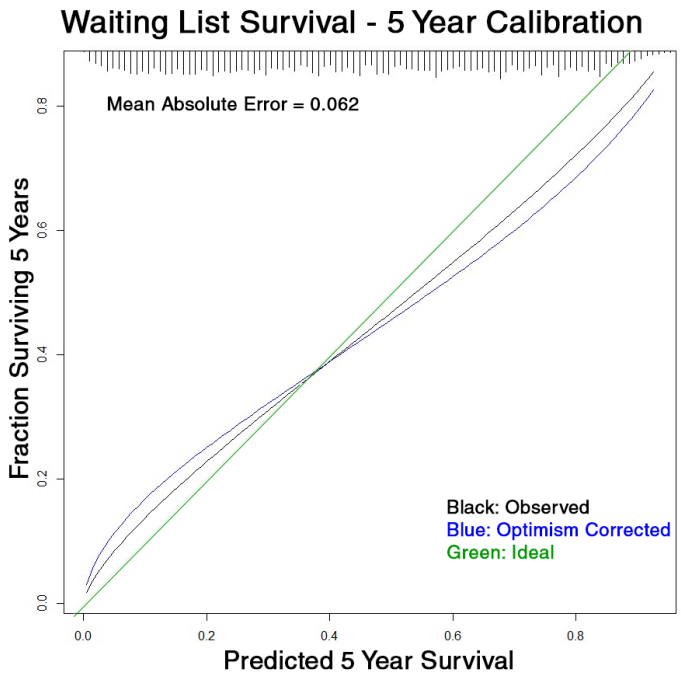


Figure 4.6: Residual plots of the non-linear terms in the waiting list survival model fit with restricted cubic splines. Positive values on the y-axis indicate a higher risk of mortality, negative values indicate a lower risk of mortality. Risk of mortality was lessened with measured creatinine less than approximately 70μmol/litre and heightened above this value.



(a)



(b)

Figure 4.7: **(a)** Calibration at 1 year: mean absolute error = 0.002, and **(b)** 5 years: mean absolute error = 0.062. The green line shows the 'ideal' calibration plot (i.e., perfect predictions), the black line shows how the model actually performed, and the blue line shows how the model performed correcting for over-fitting (or 'optimism'). Mean absolute error is a measure of the average distance between the blue line (corrected) and green line (ideal), with lower values approaching zero indicating better calibration.

Post-transplant Survival Model

Out of the variables available in the post-transplant dataset (see Appendix F), 14 were selected for inclusion in the model (any variables with a high proportion of missing entries or that resulted in the model not converging were excluded). The categorical and linear variables are shown in figure 4.8 and the non-linear variables are contained in figure 4.9. The training dataset for post-transplant survival (n = 1705) had a C-statistic of 0.60, and the validation dataset (n = 426) resulted in a C-statistic of 0.55. Calibration plots at 1 and 5 years post-transplant are shown in figure 4.10.

Variable		N	Hazard ratio	p
abo	A	775	Reference	
	AB	51	1.00 (0.67, 1.49)	0.999
	B	156	1.16 (0.92, 1.47)	0.206
	O	723	0.87 (0.76, 1.01)	0.061
group	A	593	Reference	
	B	46	0.85 (0.53, 1.35)	0.481
	C	460	0.66 (0.49, 0.89)	0.007
	D	400	0.89 (0.74, 1.07)	0.222
	Other	206	0.81 (0.65, 1.02)	0.073
dtype	DBD	1481	Reference	
	DCD	224	1.20 (0.97, 1.48)	0.087
reg_prev_malignancy	No	1667	Reference	
	Yes	38	1.41 (0.93, 2.14)	0.105
reg_creatinine		1705	1.00 (1.00, 1.01)	0.016
reg_home_oxygen	No	408	Reference	
	Yes	1297	1.14 (0.97, 1.34)	0.104
organ	Pair	1390	Reference	
	Left Lung	163	1.51 (1.22, 1.86)	<0.001
	Right Lung	152	1.08 (0.87, 1.34)	0.485
dcmv	Negative	935	Reference	
	Positive	770	1.25 (1.09, 1.43)	0.001
ddiab	No	1633	Reference	
	Yes	72	1.31 (0.95, 1.81)	0.097
dage		1705	1.00 (1.00, 1.01)	0.098
rbmi		1705	1.03 (1.01, 1.05)	0.010
height_delta		1705	0.99 (0.98, 1.00)	0.051

Figure 4.8: Categorical and linear variables in the post-transplant Cox model: **abo** = blood group, **group** = diagnosis group: A = Chronic Obstructive Pulmonary Disease, B = Pulmonary Arterial Hypertension/Pulmonary Vascular Disease, C = Cystic Fibrosis, D = Interstitial Lung Disease., **dtype** = donor type (donation after brainstem death - DBD, or donation after circulatory death - DCD), **reg_prev_malignancy** = history of previous malignancy prior to listing, **reg_creatinine** = creatinine level at time of registration, **reg_home_oxygen** requirement for home oxygen at time of registration, **organ** = donated organ type (right lung, left lung, or lung pair), **dcmv** = donor cytomegalovirus status, **ddiab** = donor diabetes status, **dage** = donor age, **rbmi** = recipient BMI at time of transplant, **height_delta** = donor - recipient height difference (cm)

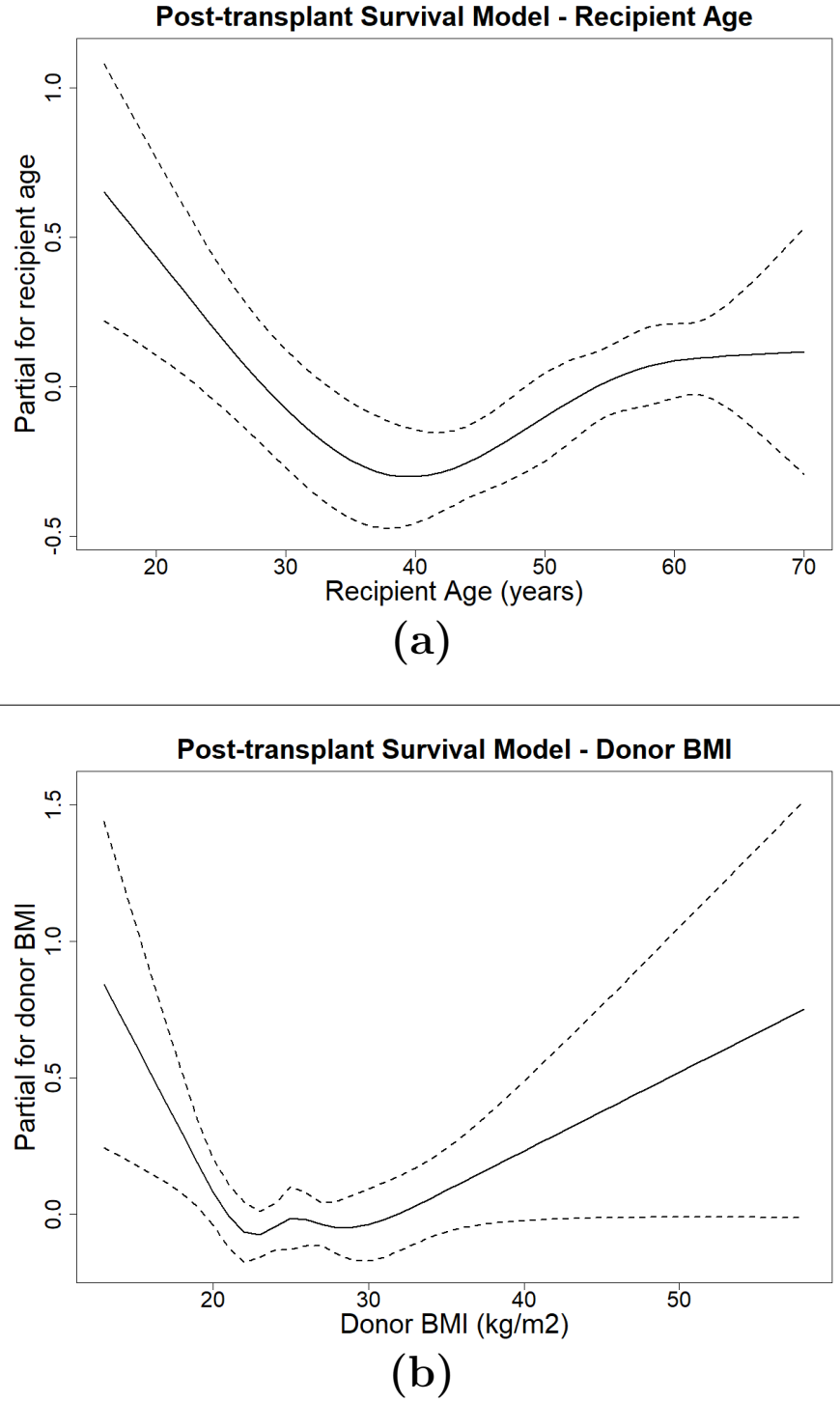
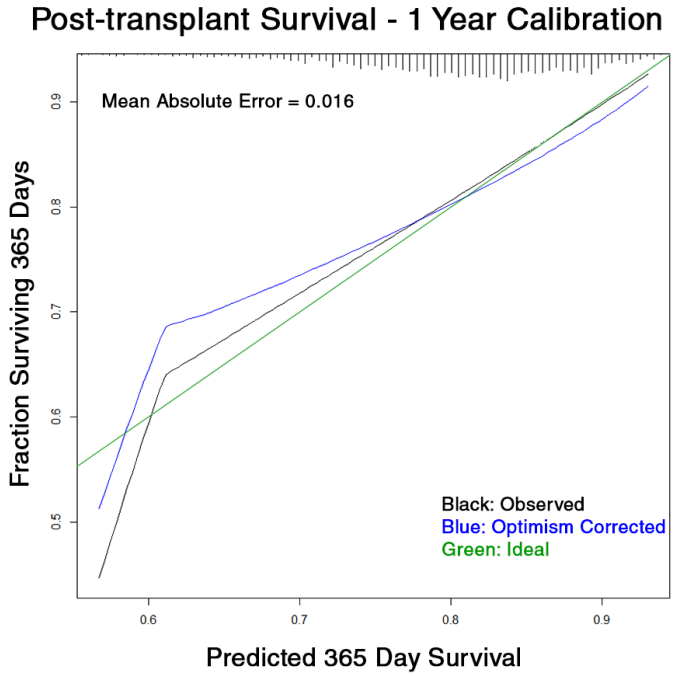
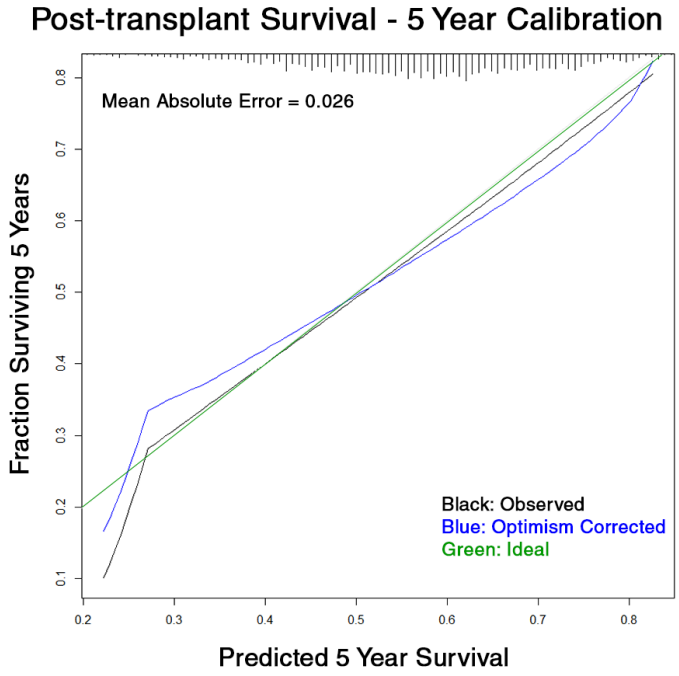


Figure 4.9: Residual plots of the non-linear terms in the post-transplant survival model fit with restricted cubic splines. Positive values on the y-axis indicate a higher risk of mortality, negative values indicate a lower risk of mortality. **(a)** Recipients aged approximately 40 years had the lowest risk of mortality, with risk increasing with both decreasing and increasing age from this point. **(b)** Risk of mortality was increased with BMI less than 20 or greater than 30.



(a)



(b)

Figure 4.10: **(a)** Calibration at 1 year: mean absolute error = 0.016, and **(b)** 5 years: mean absolute error = 0.026. The green line shows the 'ideal' calibration plot (i.e., perfect predictions), the black line shows how the model actually performed, and the blue line shows how the model performed correcting for over-fitting (or 'optimism'). Mean absolute error is a measure of the average distance between the blue line (corrected) and green line (ideal), with values approaching zero indicating better calibration.

4.3.2 NHS-BT UK Lung Allocation Policy Concordance Index

The results of calculating the concordance index (C-statistic/AUC) for the current NHS-BT policy and survival models in this thesis are shown in table 4.7. The dataset was limited to candidates and recipients listed or transplanted in the years 2018 - 2021. The current NHS-BT policy had weak predictive power for both waiting list survival (training cohort = 0.51, validation cohort = 0.56), and post-transplant survival (training cohort = 0.54, validation cohort = 0.52). By comparison, the Cox models in this thesis had much stronger predictive power for waiting list survival (training cohort = 0.71, validation cohort = 0.69), and slightly stronger predictive power for post-transplant survival (training cohort = 0.62, validation cohort = 0.61).

Table 4.7: Comparison of the concordance index (also called C-statistic or AUC) between the survival models generated in this thesis ('Thesis' column) and the NHS-BT UK adult lung allocation policy. Datasets were limited to candidates listed or transplanted in the year 2018 or later.

	Waiting List Survival			Post-transplant Survival		
Cohort	Population (n)	NHS-BT	Thesis	Population (n)	NHS-BT	Thesis
Training	619	0.51	0.71	302	0.54	0.62
Validation	170	0.56	0.69	65	0.52	0.61

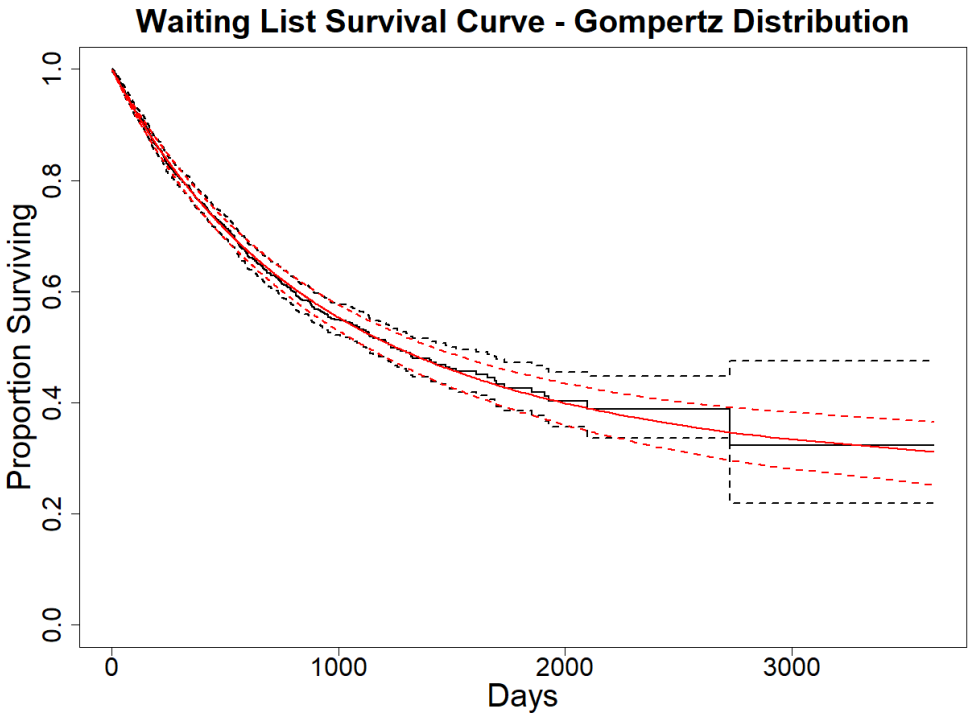
4.3.3 Simulation Parameters

The FlexSurv package in R[137] calculated the parameters for the waiting list survival Cox-Gompertz distribution (scale parameter: λ, shape parameter: α) and post-transplant survival Cox-Weibull distribution (scale parameter: λ, shape parameter: ν), the parameters are shown in table 4.8 and comparisons of modelled and observed survival curves are shown in figure 4.11.

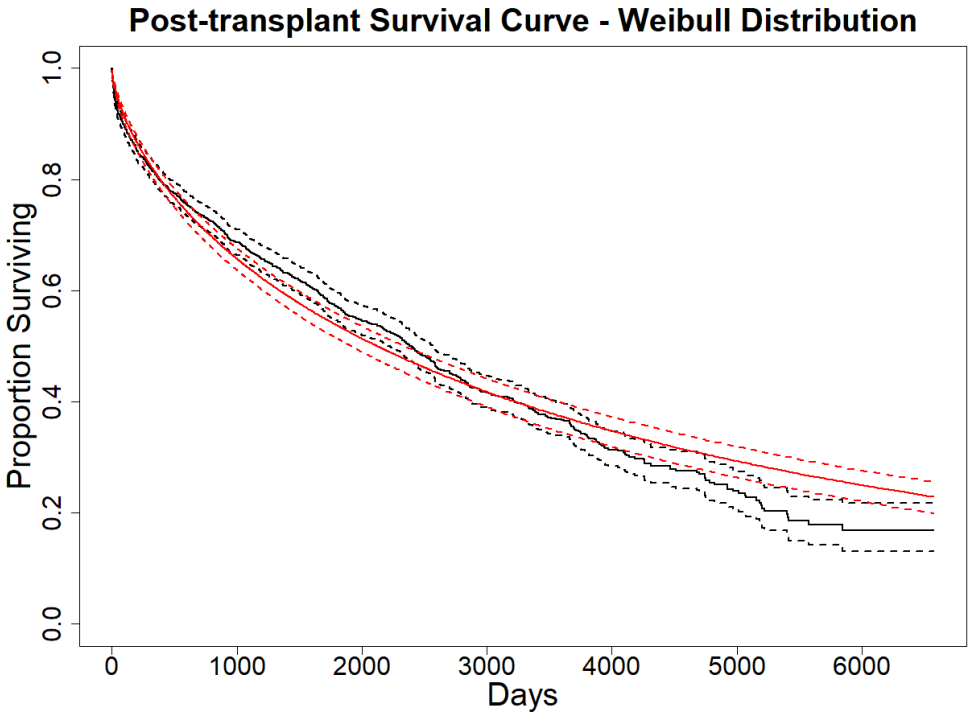
The frequencies of time gaps between candidate listing events along with the frequencies of the number of newly listed candidates are shown in table 4.9, along with the same data for the frequencies and intervals of donor offering events.

Table 4.8: Distribution parameters for waiting list and post-transplant survival. Waiting list survival was modelled using a Cox-Gompertz distribution: λ = scale parameter, α = shape parameter. Post-transplant survival was modelled using a Cox-Weibull distribution: λ = scale parameter, ν = shape parameter.

Population	λ	ν	α
Waiting List	0.0007868	-	-0.0005976
Post-transplant	0.004216	0.6661	-



(a)



(b)

Figure 4.11: **(a)** Gompertz distribution used for simulating waiting list survival durations (red) compared to observed waiting list survival curve (black). **(b)**: Weibull distribution used for simulating post-transplant survival (red) compared to observed post-transplant survival curve (black).

Table 4.9: **Top:** Number of days between new candidate listing events and their frequencies. When a new listing event occurs, the number of new listings and their frequencies are shown in the rightmost columns. **Bottom**: Number of days between donor offering events and their frequencies. When a new offering event occurs, the number of donors being offered and their frequencies are shown in the rightmost columns.

Time between listings (days)	Frequency	Number of new listings	Frequency
1	1008	1	1318
2	357	2	628
3	366	3	276
4	243	4	93
5	157	5	47
6	121	6	21
7	73	7	9
8	28	8	5
9	17	9	3
10	11	10	2
11	10	12	1
12	3		
13	1		
14	1		
15	3		
16	1		
17	1		
18	1		

Time between offers (days)	Frequency	Number of offers	Frequency
1	754	1	1741
2	510	2	393
3	334	3	56
4	156	4	4
5	150		
6	97		
7	56		
8	43		
9	27		
10	20		
11	7		
12	12		
13	15		
14	6		
15	3		
16	2		
17	6		
18	1		
20	2		
26	1		
27	1		

4.3.4 Simulation Evaluation Results

In the hypothesis in section 1.2.2 it is stated: '**If** the current UK lung allocation policy can be simulated and performance metrics measured, then improvements can be identified ...' (emphasis added). The purpose of these results is to demonstrate the ability of the survival models and simulation engine to: (1) - generate realistic survival times and (2) - simulate the current UK lung allocation policy.

(1) was evaluated by using the methods in section 4.2.1 to generate survival durations and plot survival curves to compare with those plot from the dataset. (2) was evaluated by comparing the metrics calculated by the simulation engine and comparing those metrics to those observed in the dataset, and where possible, results published in the NHS-BT annual reports.[23,141]

The evaluation results for the simulated survival durations on the waiting list and post-transplant are shown in this section. The simulated survival durations are compared with the observed survival durations and the differences are tabulated. A negative difference indicates the simulated results under-estimate the survival duration, and positive differences indicate the simulated results over-estimated the survival duration. The tables in this section are condensed, for full results tables see appendix E.

Evaluating Generated Survival Durations

Waiting List Survival - Training Dataset The waiting list survival durations tended to be slightly under-estimated by the simulation technique. With a randomly selected subset of 100 individuals the mean difference was -129 days (an error of approximately 7.2%), however as the random sample size increased the mean difference was closer to -50 days on the training dataset (an error of about 2.8%). For the training dataset comparisons, survival curves are shown in figure 4.12 and appendix E.2. A summary of results are shown in table 4.10.

Waiting List Survival - Validation Dataset When comparing the simulated results to the validation dataset, the difference ranged from 1 to 8 days for a subset size ≥ 200. This represents an error of $< 0.1\%$ to 0.3%. The smallest subset size (100) resulted in under-estimating survival by 29 days, an error of 2.3%. For the validation cohort, survival curves are shown in figure 4.12 and appendix E.2. A summary of results are shown in table 4.10.

Post-Transplant Survival - Training Dataset For post-transplant survival durations, the simulated survival durations were on average 103 - 118 days longer than the observed survival durations (an error of 3.6% to 4.1%) when evaluating using the training dataset. Survival curves for the training dataset are shown in figure 4.13 and appendix E.3, and a summary of results are shown in table 4.10.

Post-Transplant Survival - Validation Dataset For the validation dataset, simulated survival durations were 6 - 35 days longer than observed durations (an error of 0.2% to 1.2%). For the validation dataset, survival curves are shown in figure 4.13 and appendix E.3 and results are shown in table 4.10.

Table 4.10: To validate the accuracy of the simulated waiting list and post-transplant survival times, random subsets of patients were selected and simulated survival durations were generated for each patient. This process was repeated 1000 times for each subset size, and the observed survival times (area under survival curve) were compared to the simulated survival times. This process was completed using the NHS-BT waiting list dataset (training - **(a)**, validation - **(b)**) and post-transplant dataset (training - **(c)**, validation - **(d)**), using both the training and validation cohorts. A negative mean difference indicates the simulation under-estimates survival and vice versa. Survival durations are reported as: mean (standard deviation).

(a) Waiting List Survival - Training Dataset

Subset Size	Simulated Restricted Mean	Observed Restricted Mean	Difference
100	1724 (145.2)	1853 (368.7)	-129
500	1713 (66.7)	1793 (182.6)	-80
1000	1714 (47.2)	1779 (142)	-65
2000	1718 (31.6)	1767 (110.5)	-49
3400	1715 (25.2)	1764 (83.9)	-49

(b) Waiting List Survival - Validation Dataset

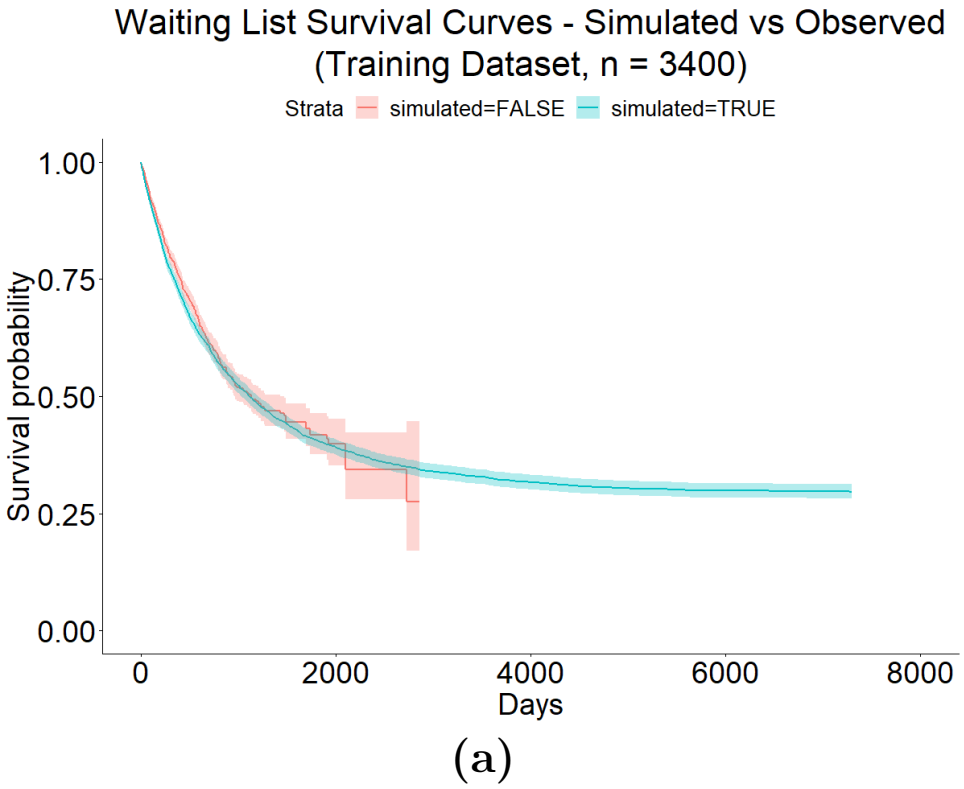
Subset Size	Simulated Restricted Mean	Observed Restricted Mean	Difference
100	1246 (97.1)	1275 (192.6)	-29
200	1249 (68.5)	1244 (131.6)	5
300	1247 (56.9)	1245 (113.3)	2
400	1252 (49.1)	1243 (93.8)	8
500	1252 (42.4)	1251 (83.1)	1
600	1253 (39.4)	1247 (75.2)	6
700	1250 (35.6)	1249 (71.7)	1
800	1250 (33.6)	1245 (67.4)	5

(c) Post-transplant Survival - Training Dataset

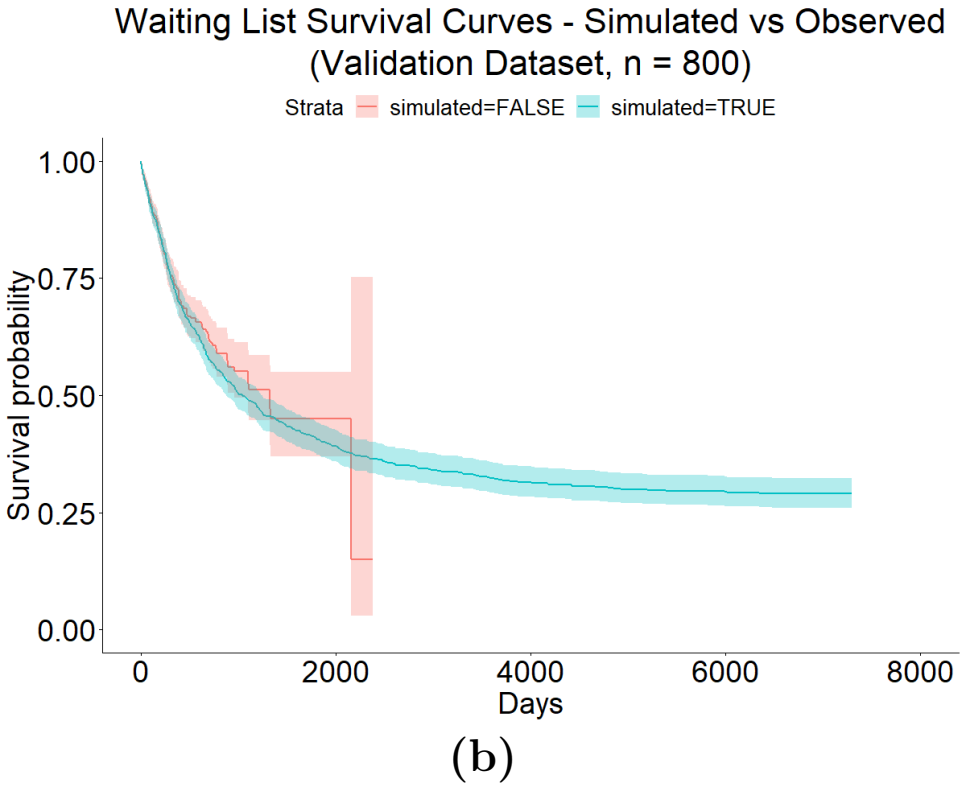
Subset Size	Simulated Restricted Mean	Observed Restricted Mean	Difference
100	2917 (255.3)	2814 (295.8)	103
500	2923 (117.6)	2811 (129)	112
1000	2923 (82.9)	2805 (89.4)	118
1700	2924 (61)	2810 (72.6)	114

(d) Post-transplant Survival - Validation Dataset

Subset Size	Simulated Restricted Mean	Observed Restricted Mean	Difference
100	2930 (263.9)	2924 (305.5)	6
200	2935 (186.1)	2902 (206.7)	33
300	2940 (151.9)	2905 (181.7)	35
400	2935 (129)	2902 (149.9)	33



(a)



(b)

Figure 4.12: Comparison of observed versus simulated waiting list survival curves, using **(a)** a random subset of 3400 lung transplant candidates for the training dataset and **(b)** a random subset of 800 from the validation dataset to plot the survival curves.

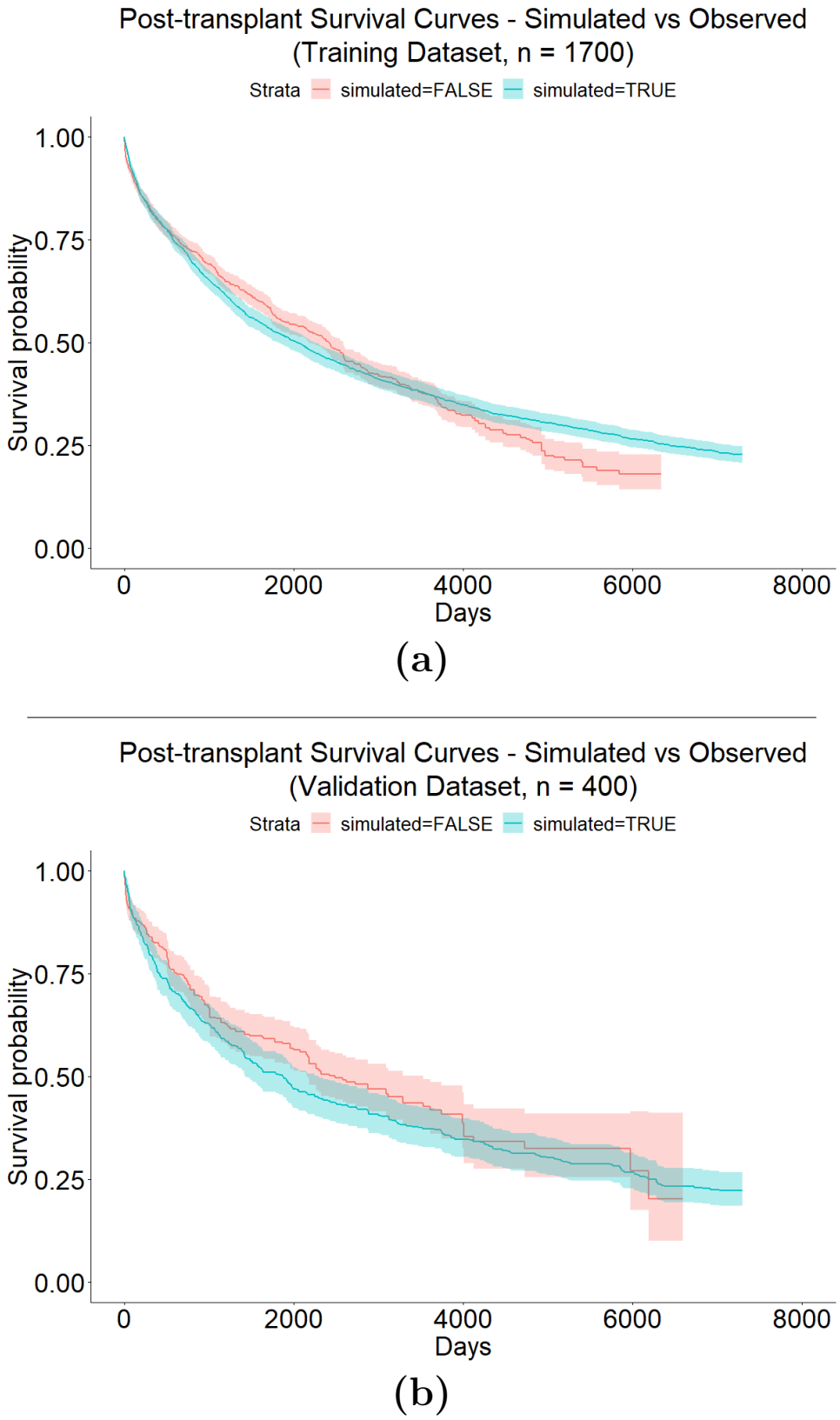


Figure 4.13: A comparison of survival curves for simulated and observed post-transplant survival durations, for **(a)** a random selection of recipients from the training dataset (n=1700) and **(b)** validation dataset (n=400).

Evaluating Simulation Engine Metrics vs Observed Transplant Metrics

The previous section compared simulated survival durations to observed survival durations using only the survival time generation logic in isolation. The simulation engine simulates multiple processes that interact: new candidates being added to the waiting list, donors being offered for transplant and the NHS-BT lung allocation policy.

This section compares the metrics generated from simulating the NHS-BT policy using the simulation engine, to metrics observed in data or published in the annual NHS-BT organ activity reports.

Figure 4.14 shows the Kaplan-Meier (KM) curves comparing simulated to observed survival durations for waiting list and post-transplant survival, using the training and validation datasets. Table 4.11 shows a comparison of simulated and observed data using the restricted means (area under the curve (AUC)) for waiting list and post-transplant survival. A comparison of 90-day, 1- and 5-year post-transplant survival rates are shown in table 4.12 and a comparison of waiting times and number of annual transplants are shown in table 4.13.

Comparison of Simulated and Published Annual Report Metrics

Table 12.2 in the 2016-2017 and 2017-2018 NHS-BT annual reports[23,141] list the mean (95% CI) 1-year post-transplant survival rates as 79.8% (76.7 - 82.6%) and 80.0% (76.8 - 82.8%) respectively. The 95% confidence intervals also overlap with the simulated results (80.2%), giving additional confidence in the simulated results.

For 5-year post-transplant survival the simulated survival rate was 53.5% (53.2% - 53.8%). This is slightly lower than the 5-year survival rates observed in the dataset and what was published in the 2016-2017 annual report.[23] The simulated 5-year post-transplant survival rate did overlap with the rates observed in 2017-18,[141] which reported a rate of 56.9% with 95% confidence intervals spanning from 52.9% to 60.7%.

Table 11.1 in the 2016-2017 and 2017-2018 annual reports[23,141] lists the number of annual transplants as 167 and 207 respectively. These are more difficult to compare to simulated results as there are no confidence intervals and these are reports for a single year, and transplant rates vary by year, making simulating the exact transplant rate year to year difficult. The simulated number of annual transplants (148.8) were very close to those observed in the dataset (149.4) and did not significantly differ ($p = 0.94$).

Simulated outcomes from listing are shown in figure 4.16 and observed outcomes are shown in figure 4.17.

Simulation Engine Survival Durations Compared to Observed Survival

Table 4.11: A comparison of observed and simulated population survival using survival durations generated from the simulation engine.

Population	Simulated Duration	Observed Duration	Difference
Waiting List - Training	1718 (6.2)	1769 (78.3)	-51
Waiting List - Validation	1285 (3.4)	1248 (63.1)	37
Post-Transplant - Training	2989 (7.4)	2810 (71.4)	179
Post-Transplant - Validation	2994 (7.5)	2903 (146.4)	91

Simulation Engine Post-transplant Survival Compared to Observed Rates

Table 4.12: Post-transplant survival rates generated from the simulation engine compared to those observed in the training and validation datasets.

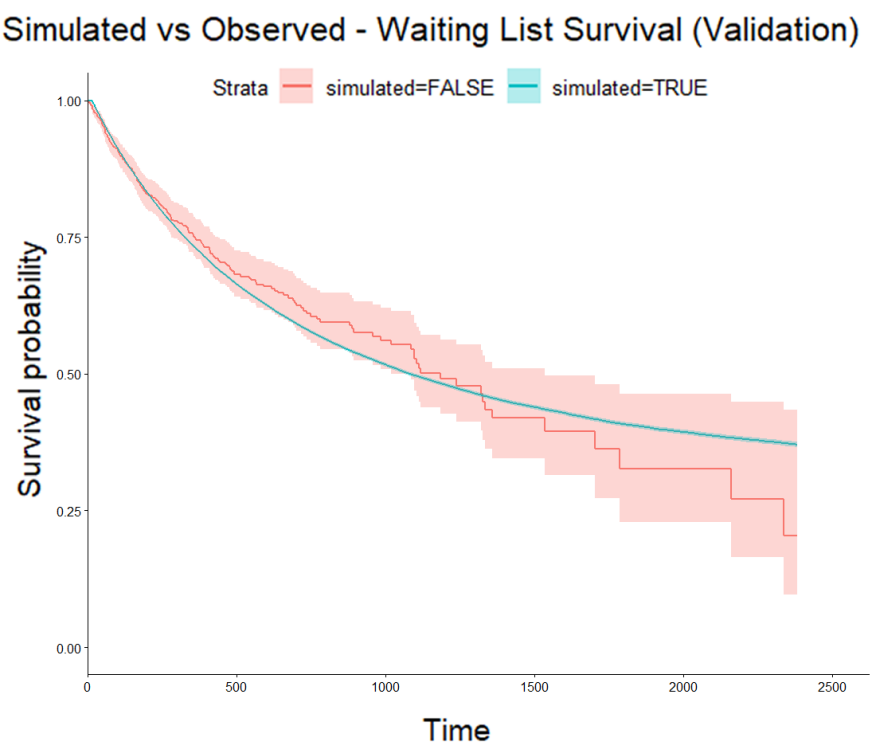
Post-TX Duration	Simulated (95% CI)	Observed - Dataset	Observed - 2016-17	Observed - 2017-18
90-Day(%)	91.7 (91.6 - 91.9)	89.8 (88.5 - 91.1)	90.0 (87.6 - 92.0)	89.4 (86.9 - 91.4)
1-Yr (%)	80.2 (80.0 - 80.5)	80.3 (78.6 - 82.0)	79.8 (76.7 - 82.6)	80.0 (76.8 - 82.8)
5-Yr (%)	53.5 (53.2 - 53.8)	56.4 (54.2 - 58.8)	58.3 (54.2 - 62.2)	56.9 (52.9 - 60.7)

Comparison of Simulated and Observed Transplant Volumes and Waiting Times

Table 4.13: Comparison of simulated and observed transplant volumes and waiting times. Annual transplants were simulated accurately but waiting times were over-estimated by about 60 days.

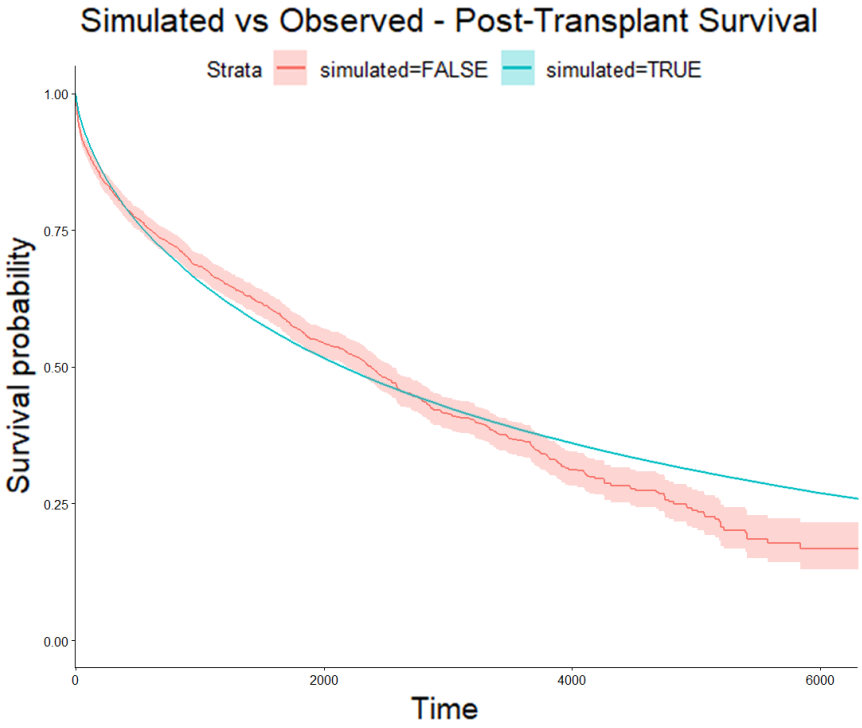
Metric	Simulated	Observed	p-value
Number of Annual Transplants	148.8 ($\pm$2.4)	149.4 ($\pm$30.1)	0.94
Mean Waiting Time (Days)	443.1 ($\pm$30.8)	379.9 ($\pm$434.1)	$<$0.0001

(a)

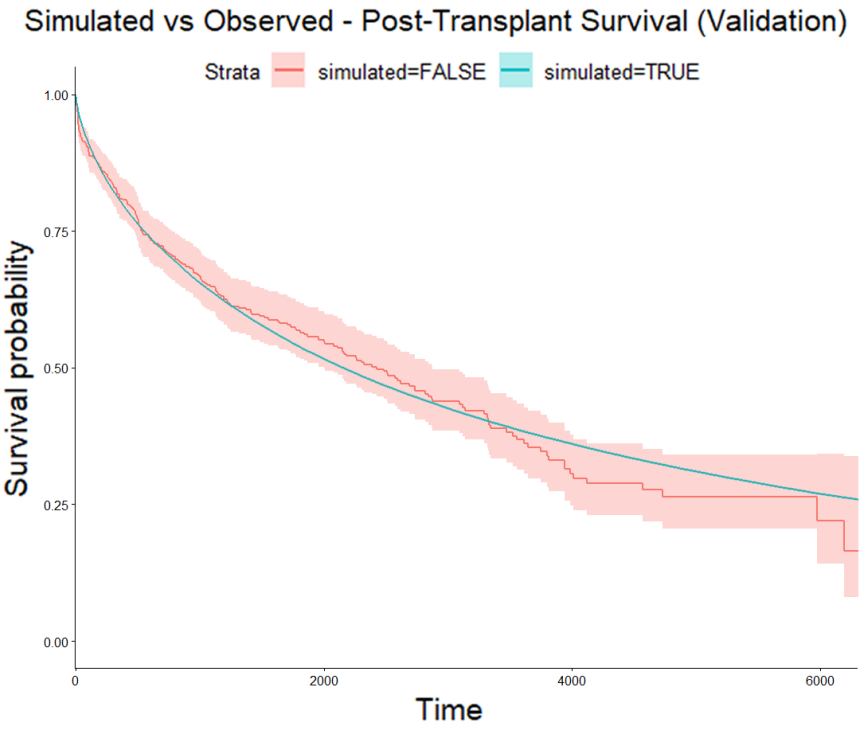


(b)

Figure 4.14: Waiting list survival curves showing population survival from the full simulation engine compared to the training and validation datasets. **(a)** Waiting list survival - training dataset, **(b)** Waiting list survival - validation dataset.

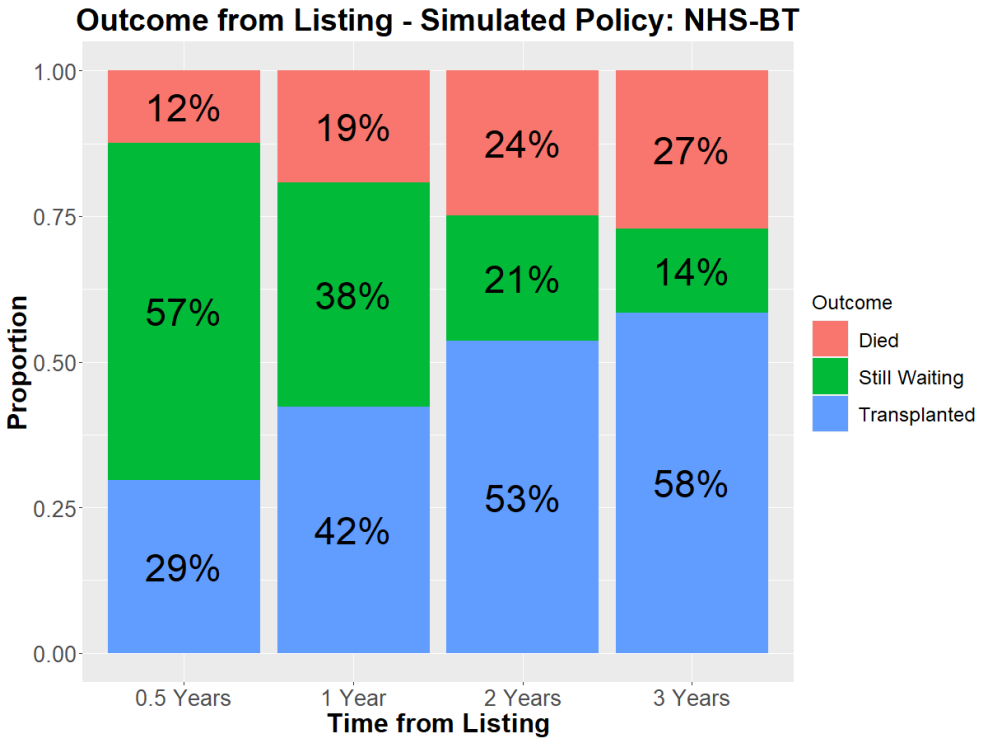


(a)

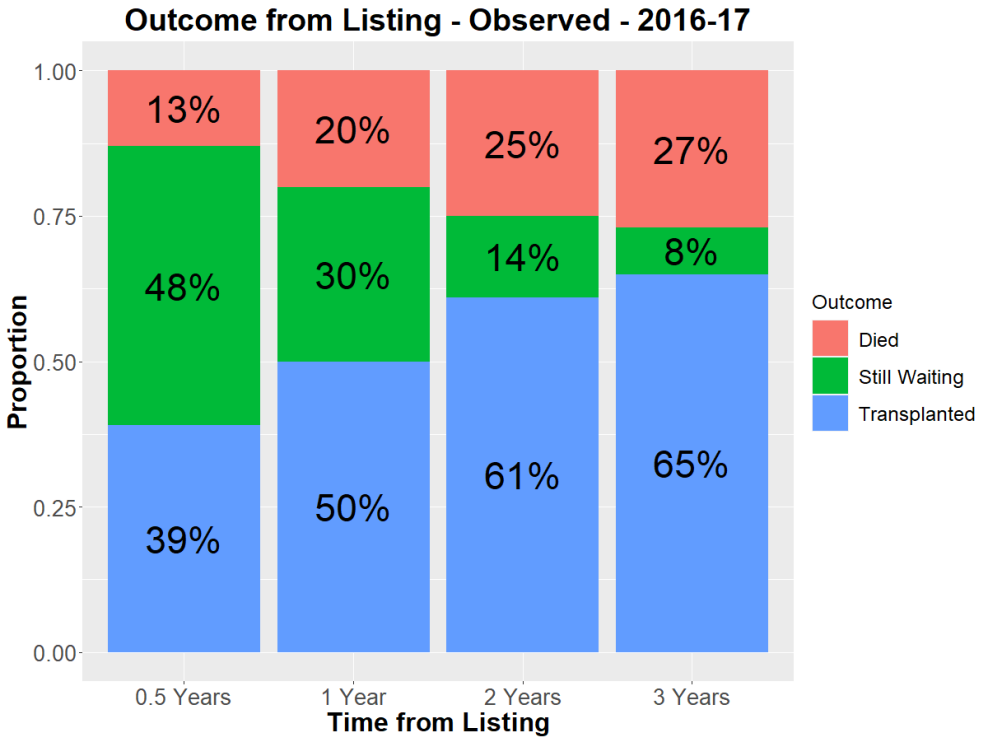


(b)

Figure 4.15: Post-transplant survival curves showing population survival from the simulation engine compared to the training and validation datasets. **(a)** Post-transplant survival - training dataset, **(b)** Post-transplant survival - validation dataset.



(a)



(b)

Figure 4.16: **(a)** Simulated proportions of outcomes from listing. **(b)** Published outcomes from the NHS-BT 2016-17 annual report.[23] Red bars **(topmost bars)** show the percentage of candidates that have died, green bars **(middle bars)** show the percentage still waiting, and blue bars **(bottommost bars)** show the percentage transplanted at 6 months, 1, 2, and 3 years after listing.

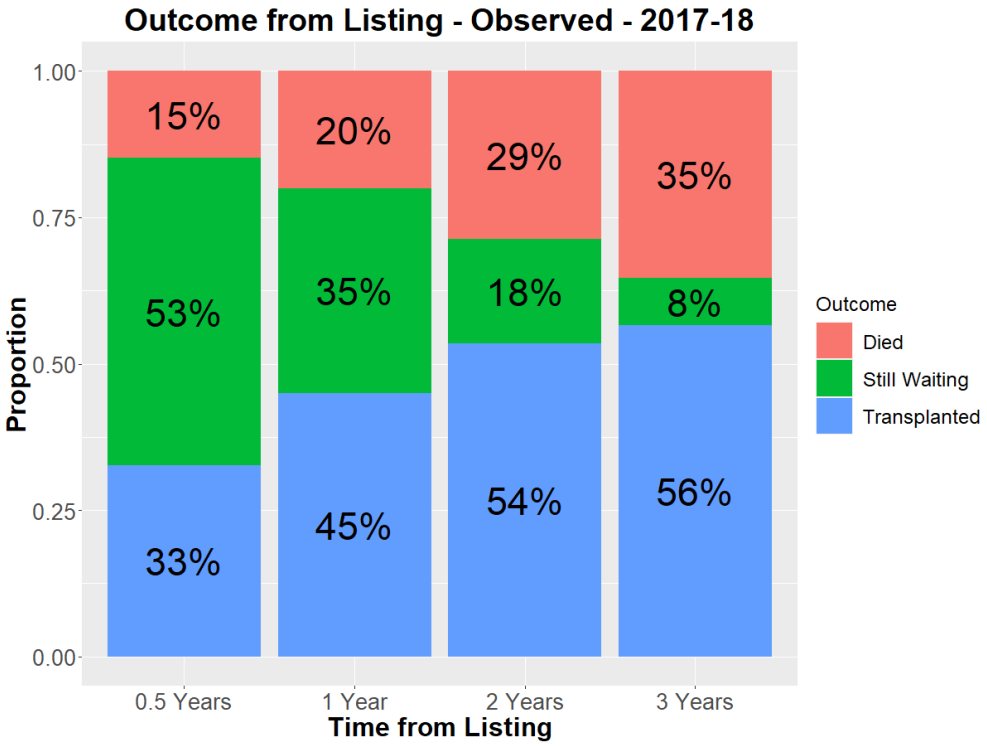


Figure 4.17: Published outcomes from listing from the NHS-BT 2017-18 annual report.[141] Red bars **(topmost bars)** show the percentage of candidates that have died, green bars **(middle bars)** show the percentage still waiting, and blue bars **(bottommost bars)** show the percentage transplanted at 6 months, 1, 2, and 3 years after listing.

4.3.5 Simulation Results

Impact of Prioritising Waiting List Survival vs Post-transplant Survival

At a population level, prioritising waiting list survival tended to result in fewer waiting list deaths, at the expense of lower net benefit, lower post-transplant survival and increased waiting times. On the other hand, prioritising post-transplant survival resulted in increased post-transplant survival rates, increased net benefit and lower waiting times, at the expense of increased waiting list mortality. The results for the current NHS-BT policy and standard WL:PTX policies are shown in table 4.14.

Table 4.15 shows there are statistically significant differences in the number of waiting list deaths between policies, and figure 4.18 contains a violin plot showing the distribution of waiting list deaths for each simulated policy. A violin plot is similar to a 'Box and Whisker' plot, but shows the shape of the entire distribution of values rather than only showing quartiles. This plot also shows the general trend of increasing waiting list deaths as priority shifts towards maximising post-transplant survival.

Table 4.14: Comparison of the five different priority ratios of waiting list (WL) survival and post-transplant (PTX) survival to the current NHS-BT policy. Metrics are reported as: mean (SD).

Policy	Annual Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Annual Transplants	Waiting Time (Days)
NHS-BT	90 (7.4)	1833	6.8	80.2 / 53.3	203 (5.6)	416 (35.2)
WL	46 (4.2)	2238	15.1	80.0 / 53.0	204 (4.1)	544 (33.4)
2:1 WL:PTX	46 (3.8)	2300	15.5	80.5 / 53.9	204 (4.9)	540 (35.1)
1:1 WL:PTX	47 (4)	2376	15.7	81.0 / 55.0	204 (4.3)	538 (30.4)
1:2 WL:PTX	51 (5)	2459	15.7	81.8 / 56.4	204 (4.8)	508 (32.6)
PTX	77 (5.2)	2522	14.1	83.5 / 59.5	204 (4.4)	343 (21.8)

Table 4.15: Significance levels of number of annual waiting list deaths between policies. Key: NS (not significant), · ($p < 0.10$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$)

	WL	2:1	1:1	1:2	PTX
NHS-BT	****	****	****	****	****
WL		NS	NS	****	****
2:1			NS	****	****
1:1				****	****
1:2					****

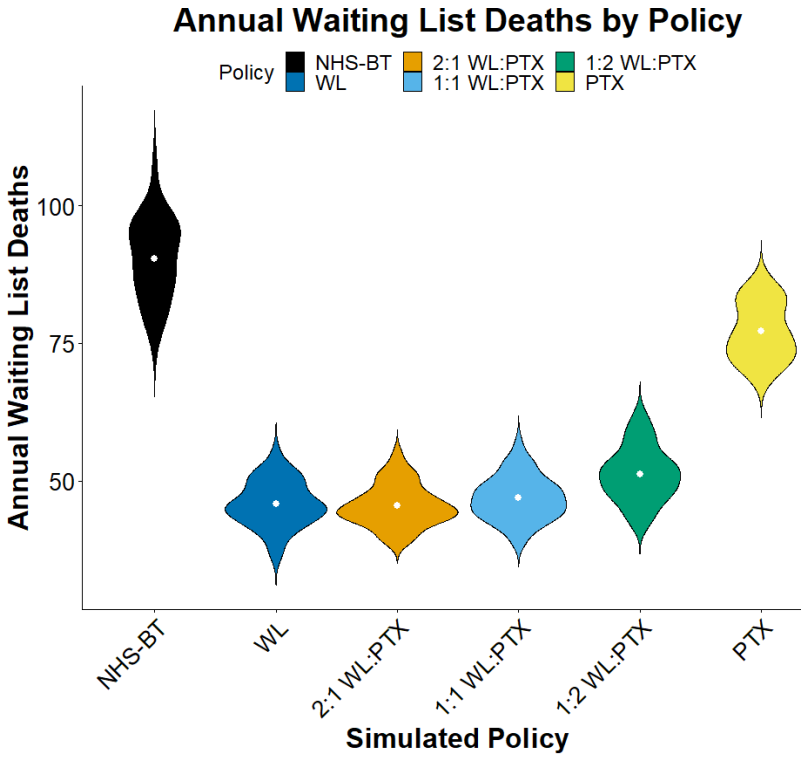


Figure 4.18: Violin plot of the number of annual waiting list deaths as the priority-ratio changes from waiting list survival (WL) to post-transplant survival (PTX) (e.g., the 2:1 WL:PTX policy gives twice as much priority to waiting list survival compared to post-transplant survival.) The width of the plot indicates the relative frequency of the corresponding value on the y-axis from 40 simulations spanning 20 years.

Impact of Increased Utilisation

The impact of increased lung utilisation on waiting list deaths was investigated, using the WL policy to prioritise candidates in each scenario. There was a non-proportional reduction in waiting list deaths compared to the increase in utilisation. The percentage reduction in waiting list deaths was on average 15.5% greater than the corresponding percentage increase in utilisation.

Table 4.16: Percentage reduction in annual waiting list mortality with respect to the simulated percentage increase in lung utilisation. The allocation policy that prioritises waiting list survival was used as a reference to evaluate the decrease in waiting list deaths as utilisation increased. Note the non-proportional reduction in waiting list mortality with respect to the percentage increase in utilisation.

Utilisation % Increase	Annual Waiting List Deaths	% Reduction
WL (Reference)	46	-
WL +5%	41	10.9%
WL +10%	36	21.7%
WL +25%	25	45.7%
WL +50%	12	73.9%

Impact of Increased Single-Lung Transplant for Candidates with ILD

Ordering of Offers Offering the left lung first and right lung first were both simulated using the SLT algorithm combined with the WL policy. When allocating the lung pair all candidates were considered, so only the left/right offering order was simulated. There was no statistical difference in the number of waiting list deaths between the two offering orders (2-tailed t-test, $p \geq 0.41$).

Table 4.17: The order of offering left lung or right lung first for the single-lung transplant policy had no impact on the annual number of waiting list deaths.

Offer Order	Annual Waiting List Deaths
Left, Right	33
Right, Left	32

Impact of SLT Age Threshold The lowest number of waiting list deaths were achieved with no age threshold for ILD recipients to undergo SLT, and the number of waiting list deaths increased as the age threshold increased. Not setting an age threshold resulted in 33 waiting list deaths per year. Setting the age threshold to 55 years or greater increased the number of waiting list deaths to 39 ($p < 0.0001$), and a threshold of 60 years resulted in 44 waiting list deaths per year ($p < 0.0001$ compared to no age threshold, $p < 0.0001$ compared to an age threshold of 55 years).

Table 4.18: Waiting list deaths were minimised with no age threshold for a candidate with ILD to undergo single-lung transplant.

Age Threshold	Annual Waiting List Deaths
None	33
55+	39
60+	44

Combining Single-Lung Transplant with Waiting List / Post-transplant Survival Policies

The SLT algorithm was combined with each allocation policy, and each of the five performance metrics were calculated for comparison. All SLT policies significantly decreased waiting list deaths, with the SLT-1:2 policy resulting in the fewest annual waiting list deaths (31). Compared to the standard 1:2 policy, the reduction in waiting list deaths came at the expense of lower net benefit (2165 days vs 2459 days), lower relative benefit (14.4 vs 15.7), and lower survival rates at 1 and 5 years post-transplant (79.3% vs 81.8% and 52.1% vs 56.4% respectively).

Table 4.19 gives an overview of the performance of the SLT policies, table 4.20 shows the impact on transplant volume, table 4.21 shows which pairs of SLT policies performed significantly differently and figure 4.19 shows a plot of waiting list deaths by SLT policy.

Overall, there was less difference between the SLT policies than with the standard five policies. In terms of waiting list deaths, the SLT-1:2 policy had statistically significantly fewer waiting list deaths compared to the other SLT policies as shown in table 4.21. Maximising the number of single lung transplants resulted in a similar reduction in waiting list deaths as increasing the size of the donor pool by 10%.

Table 4.19: Combining increased single-lung transplant (SLT) for candidates with interstitial lung disease with each of the priority ratios resulted in a large reduction in annual waiting list deaths.

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-yr Post-Tx Survival (%)	Waiting Time (Days)
NHS-BT	90 (7.4)	1833	6.80	80.20 / 53.30	416 (35.2)
SLT-WL	33 (3.6)	2100	14.30	78.80 / 51.20	435 (31.8)
SLT-2:1	32 (3.8)	2155	14.60	79.30 / 52.00	425 (37.1)
SLT-1:1	33 (3.1)	2197	14.60	79.60 / 52.60	423 (29.4)
SLT-1:2	31 (2.8)	2165	14.40	79.30 / 52.10	373 (26.2)
SLT-PTX	34 (3.1)	2073	13.60	78.90 / 51.30	264 (24.2)

Table 4.20: Transplant volume varies depending on the SLT policy in place, with the largest transplant volume occurring with the SLT-PTX policy.

Policy	Transplants Per Year	Number BLT (%)	Number SLT (%)
SLT-WL	230 (4.1)	176.7 (76.5%)	54.3 (23.5%)
SLT-2:1	231 (4.1)	177.1 (76.4%)	54.7 (23.6%)
SLT-1:2	241 (3.7)	166.1 (68.9%)	75.1 (31.1%)
SLT-PTX	257 (4.3)	152.2 (59.7%)	102.6 (40.3%)

Table 4.21: Significance levels of number of annual waiting list deaths between policies, using the Welch two-sample t-test. Key: NS (not significant), · ($p < 0.10$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$)

	SLT-2:1	SLT-1:1	SLT-1:2	SLT-PTX
SLT-WL	NS	NS	**	NS
SLT-2:1		NS	*	*
SLT-1:1			***	NS
SLT-1:2				****

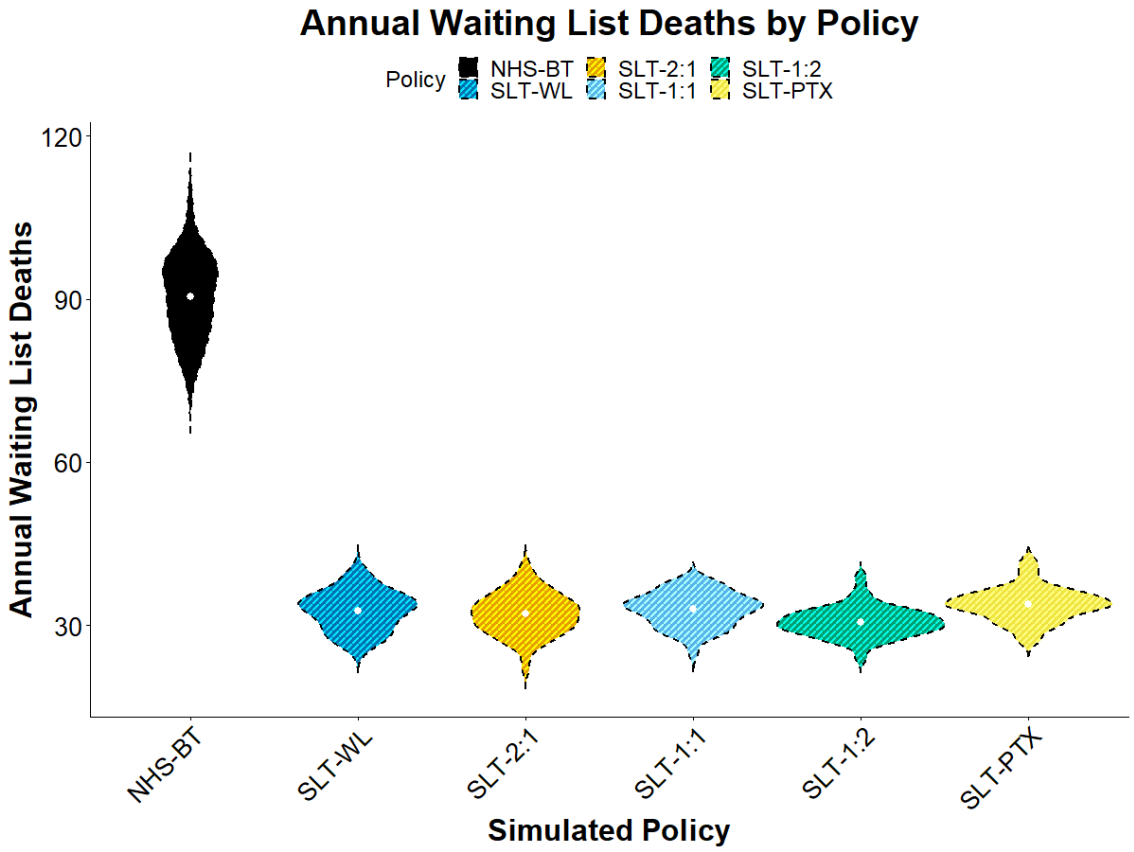


Figure 4.19: Violin plot of the number of annual waiting list deaths as the priority-ratio changes from waiting list survival (WL) to post-transplant survival (PTX) for the policies that preferentially allocate single lungs to candidates with interstitial lung disease (e.g., the SLT-1:2 policy gives twice as much priority to waiting list survival compared to post-transplant survival.) The width of the plot indicates the relative frequency of the corresponding value on the y-axis from 40 simulations spanning 20 years.

Impact of Increased SLT Combined With Increased Utilisation Compared to the standard WL policy, the SLT-WL policy resulted in a 28.3% decrease in annual waiting list deaths. The non-proportional decrease in waiting list deaths was greater than with the WL policy. On average, the percentage reduction in waiting list deaths was 17.6% greater than the percentage increase in utilisation.

Using the SLT-WL policy, a 5% increase in utilisation resulted in a 12.1% reduction in waiting list deaths and a 25% increase resulted in a 48.5% reduction. The full results are shown in table 4.22.

Population Benefit and Opportunity Cost

The standard WL - PTX policies resulted in lower total net life gain compared to the SLT policies. For the standard policies, the 1:2 WL:PTX policy resulted in the highest net life gain of 827 additional years of life gained annually.

Table 4.22: Comparison of the number of annual waiting list deaths between the current NHS-BT policy, the policy that prioritises waiting list survival (WL), and the policy that increases the use of single-lung transplant for ILD candidates that also prioritises waiting list survival (SLT-WL). The SLT-WL policy was also simulated with increased lung utilisation. Note the non-proportional decrease in waiting list deaths in comparison to the percentage increase in utilisation.

Policy	Utilisation Increase	Waiting List Deaths	% Reduction (NHS-BT)	% Reduction (WL)	% Reduction (SLT-WL)
NHS-BT	0%	90	-	-	-
WL	0%	46	48.9%	-	-
SLT-WL	0%	33	63.3%	28.3%	-
SLT-WL	5%	29	67.8%	37.0%	12.1%
SLT-WL	10%	25	72.2%	45.7%	24.2%
SLT-WL	25%	17	81.1%	63.0%	48.5%
SLT-WL	50%	8	91.1%	82.6%	75.8%

The SLT policies resulted in higher overall net life gain compared to the standard policies, despite the less optimal post-transplant outcomes. The highest performing SLT policy was the SLT-1:2 policy with 946 years of additional life gained annually. In terms of minimising net benefit loss and maximising net benefit gain, the SLT policies performed significantly better than the standard policies. As a comparison, the worst performing SLT policy (SLT-WL) resulted in higher net life gain than the best performing standard policy (840 years vs 827 years respectively).

A full summary of results are visualised as a bar graph in figure 4.20.

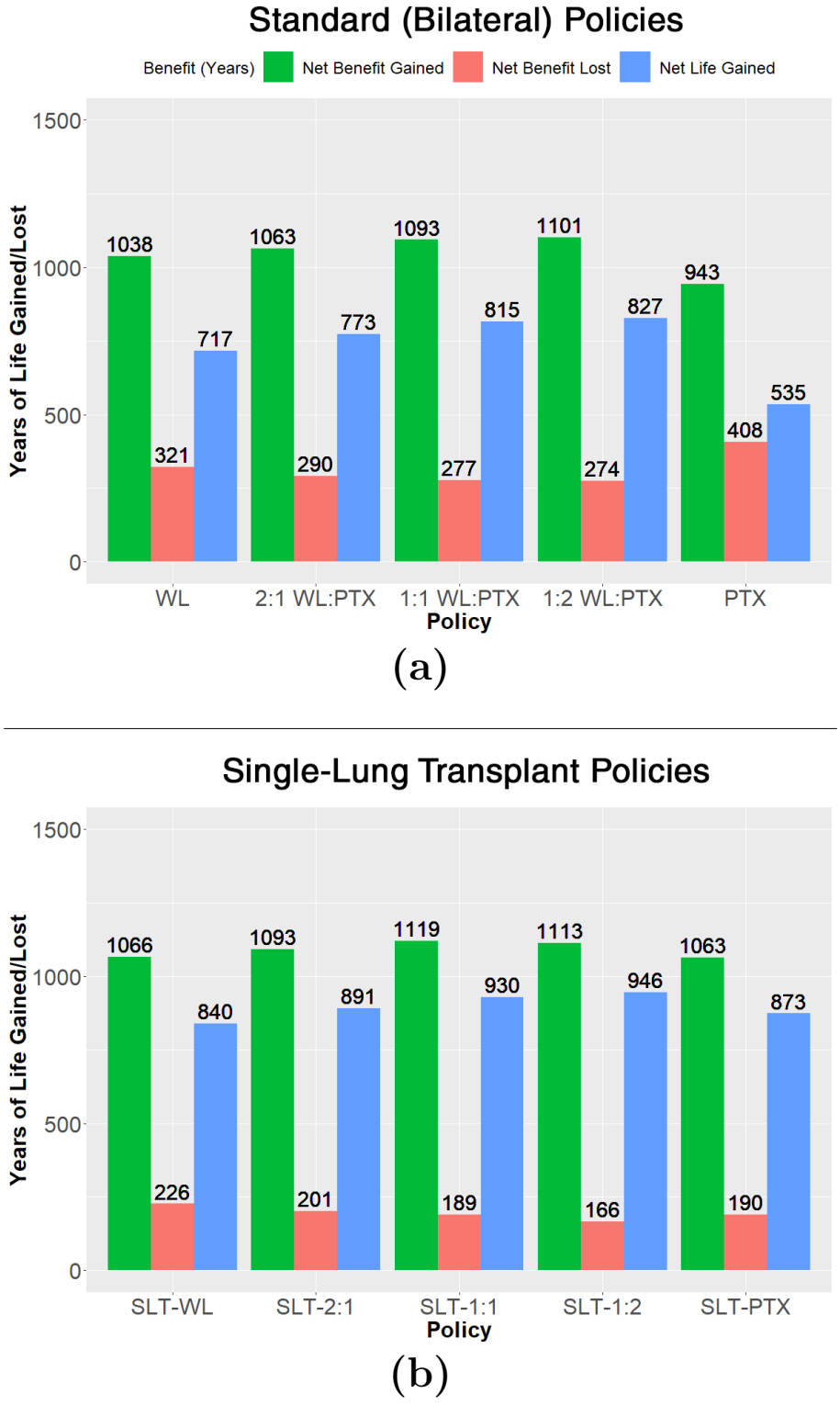


Figure 4.20: A plot showing population-level gain in life-years (i.e., total additional years of life gained from transplant) for transplanted candidates (leftmost/green), potential life-years that were lost due to candidates dying on the waiting list (middle/red) and the net difference in life-years gained (rightmost/blue). **(a)** Results for the policies with a range of priority-ratios ranging from prioritising only waiting list survival (WL) to post-transplant survival (PTX) **(b)** Results for the policies that preferentially allocate single lungs to candidates with interstitial lung disease, combined with the corresponding priority-ratios of waiting list survival (WL) and post-transplant survival (PTX).

Impact of Risk-Adjusted Benefit

Implementing risk-adjusted benefit did not result in a significant difference on any of the metrics. Prioritising patients by probability of reaching the equity point did not have any measurable benefit. Table 4.23 shows how varying the risk-weight didn't significantly change any of the metrics. Additionally, the Welch Two Sample T-Test showed no significant difference between any pair of RAB policies when comparing metrics. Figure 4.21 visualises the population-level survival gain and loss between RAB policies and the standard 1:1 policy.

Table 4.23: Adjusting net benefit by the risk of mortality post-transplant (risk adjusted benefit (RAB)) performance metrics at different risk-weightings compared to the standard 1:1 policy. RAB-0% is identical to the 1:1 policy as 0% weight is given to the risk-adjusted net benefit value. RAB-100% ignores the standard net benefit (difference between waiting list and post-transplant survival durations) and uses only the risk adjusted net benefit. RAB policies between these two take a weighted sum of standard and risk adjusted net benefit.

Policy (Risk Weight %)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
1:1	47 (4)	2376	15.7	81.0 / 55.0	538 (30.4)
RAB-0%	46 (4.5)	2365	15.7	81.0 / 54.9	527 (36.4)
RAB-25%	46 (3.1)	2357	15.6	80.9 / 54.5	534 (23.5)
RAB-50%	46 (3.4)	2375	15.8	81.0 / 54.7	532 (30.4)
RAB-75%	46 (4.2)	2371	15.8	81.2 / 55.0	535 (34.5)
RAB-100%	46 (4)	2371	15.7	80.7 / 54.9	536 (32.4)

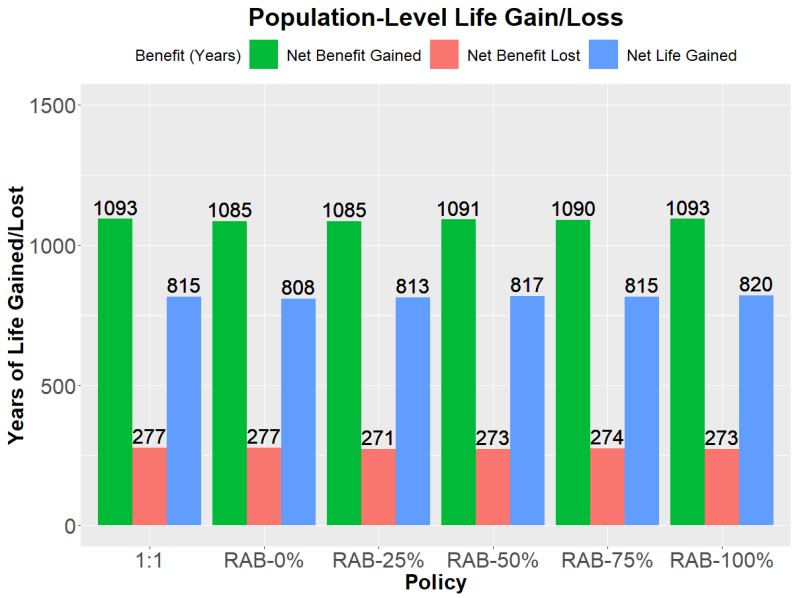


Figure 4.21: Risk-adjusted benefit (RAB) adjusts allocation scores using the risk of a recipient not surviving long enough realise that benefit. Despite adjusting for risk, there was no overall change in life-years gained (i.e., total additional life experienced from all transplant recipients.)

Impact of Conditional Survival

The WL and 1:1 policies were combined with the conditional survival formula in order to assess the impact of allocating based on the risk of mortality on the waiting list over the next day, 7 days, 30 days, 1 year and 5 years.

When combined with the WL policy, comparing waiting list deaths between the CON-WL-365, CON-WL-1 and CON-WL-7 policies resulted in p-values near statistical significance ($p < 0.10$), but did not result in p-values < 0.05 (see table 4.26). There was no statistical significance between waiting list deaths when comparing any of the CON-1:1 policies.

Tables 4.24 and 4.25 show no significant differences in performance metrics when using conditional survival.

Table 4.24: Adjusting allocation scores using conditional survival for an additional 1, 7, 30, 365 or 1095 days on the waiting list did not result in significant differences in performance metrics compared to the standard WL policy.

Policy	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
WL	46 (4.2)	2238	15.1	80.0 / 53.0	544 (33.4)
CON-WL-1	46 (4)	2237	15.2	79.9 / 53.0	546 (31.7)
CON-WL-7	46 (3.8)	2241	15.2	80.0 / 53.1	551 (31.5)
CON-WL-30	46 (3.9)	2207	15.0	79.9 / 52.7	545 (29.9)
CON-WL-365	44 (3.7)	2252	15.3	80.3 / 53.1	540 (33.5)
CON-WL-1095	45 (3.1)	2237	15.2	79.9 / 52.9	541 (27.3)

Table 4.25: Adjusting allocation scores using conditional survival for an additional 1, 7, 30, 365 or 1095 days on the waiting list did not result in significant differences in performance metrics compared to the standard 1:1 policy.

Policy	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1-Yr / 5-Yr PTX Survival (%)	Waiting Time (Days)
1:1	47 (4)	2368	15.70	80.9 / 54.9	536 (32.3)
CON-1:1-1	46 (4)	2352	15.6	80.6 / 54.5	533 (34.7)
CON-1:1-7	46 (4.6)	2367	15.5	80.9 / 54.7	531 (25.7)
CON-1:1-30	46 (3.8)	2370	15.6	81.0 / 54.8	537 (31.7)
CON-1:1-365	47 (4.5)	2361	15.7	80.9 / 54.6	534 (36.9)
CON-1:1-1095	46 (3.9)	2384	15.9	81.1 / 55.2	534 (27.8)

Table 4.26: Significance levels of number of annual waiting list deaths between policies using the Welch two-sample t-test. Allocation scores have been adjusted using conditional survival. Key: NS (not significant), · ($p < 0.10$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$)

	CON-WL-1	CON-7	CON-30	CON-365	CON-1095
WL	NS	NS	NS	NS	NS
CON-WL-1		NS	NS	·	NS
CON-WL-7			NS	·	NS
CON-WL-30				NS	NS
CON-WL-365					NS

Impact by Diagnosis Group

Table 4.27: Impact by policy on candidates and recipients with group A (chronic obstructive pulmonary disease) diagnoses.

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	17 (1.6)	1227	4.6	78.1 / 49.3	526 (52.2)
WL	24 (2)	1642	8.8	78.3 / 50.1	1652 (73)
2:1 WL:PTX	27 (1.8)	1674	8.9	78.6 / 50.1	1746 (76.7)
1:1 WL:PTX	29 (2)	1829	9.5	79.3 / 52.1	1820 (63.3)
1:2 WL:PTX	29 (2.3)	1933	9.8	80.9 / 54.3	1815 (90)
PTX	21 (1.7)	1901	8.9	81.5 / 55.3	981 (68.7)
SLT-WL	20 (2)	1595	8.6	78.2 / 49.6	1450 (83.1)
SLT-2:1	21 (2.3)	1636	8.8	78.6 / 50.1	1531 (103.2)
SLT-1:1	24 (2.2)	1750	8.8	78.9 / 51.4	1615 (90.8)
SLT-1:2	23 (2)	1784	8.7	79.4 / 52.1	1533 (88)
SLT-PTX	23 (2)	1731	7.5	80.5 / 53.0	1077 (105.1)

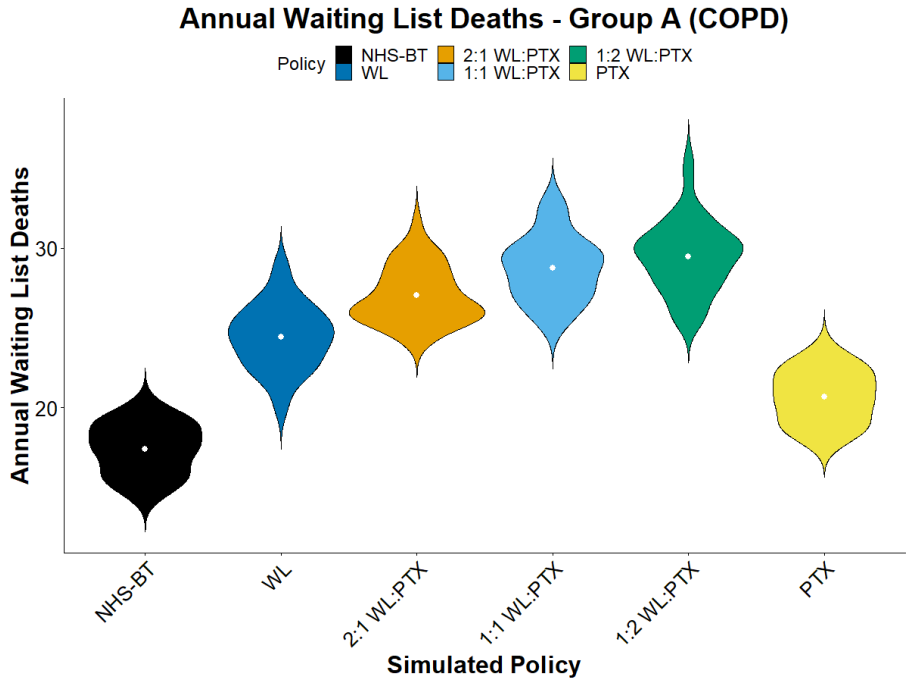
Impact on Group A - COPD Candidates and recipients with group A diagnoses (COPD) tended to benefit as priority shifted away from waiting list survival and emphasised post-transplant survival, though the number of waiting list deaths follows an inverted 'U' shape.

As priority shifted towards post-transplant survival the number of waiting list deaths initially increased from 24 (WL) to 29 (1:1), then decreased to 21 with the PTX policy.

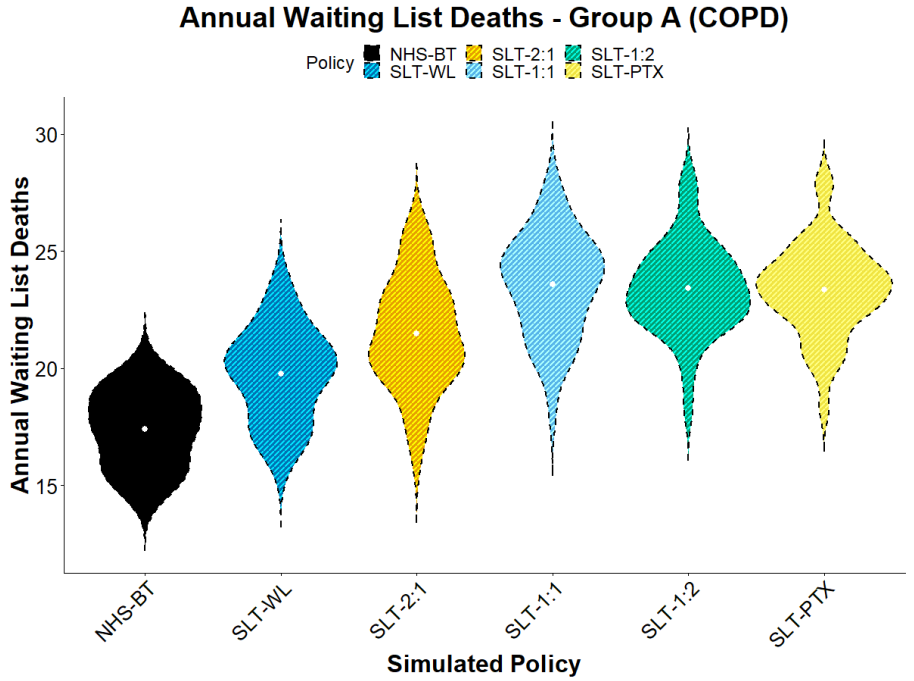
Net benefit increased from 1642 days (WL) to 1933 days (1:2) and then decreased again to 1901 days with the PTX policy. Relative benefit increased from 8.8 (WL) to 9.5 (1:1) and then decreased again to 8.9 with the PTX policy.

1 year post-transplant survival increased from 78.3% (WL) to 81.5% (PTX) and 5 year post-transplant survival increased from 50.1% to 55.3%. Mean waiting time decreased from 1652 days to 981 days.

Implementation of the SLT algorithm for group D (ILD) candidates also resulted in fewer waiting list deaths for group A (COPD) candidates, but also resulted in a slight decrease in post-transplant survival rates, net benefit and waiting time.



(a)



(b)

Figure 4.22: (**a**) Using the standard policies, for group A (COPD) candidates, prioritising post-transplant survival resulted in the fewest waiting list deaths. The WL policy also resulted in a low number of waiting list deaths, with mortality increasing as post-transplant priority increases (up to the 1:2 policy). (**b**) For group A (COPD) candidates, the SLT policies resulted in a reduction in waiting list deaths compared to the standard policies (with the exception of the SLT-PTX policy compared to the PTX policy). Waiting list mortality increased slightly as priority shifted from the SLT-WL policy to the SLT-PTX policy.

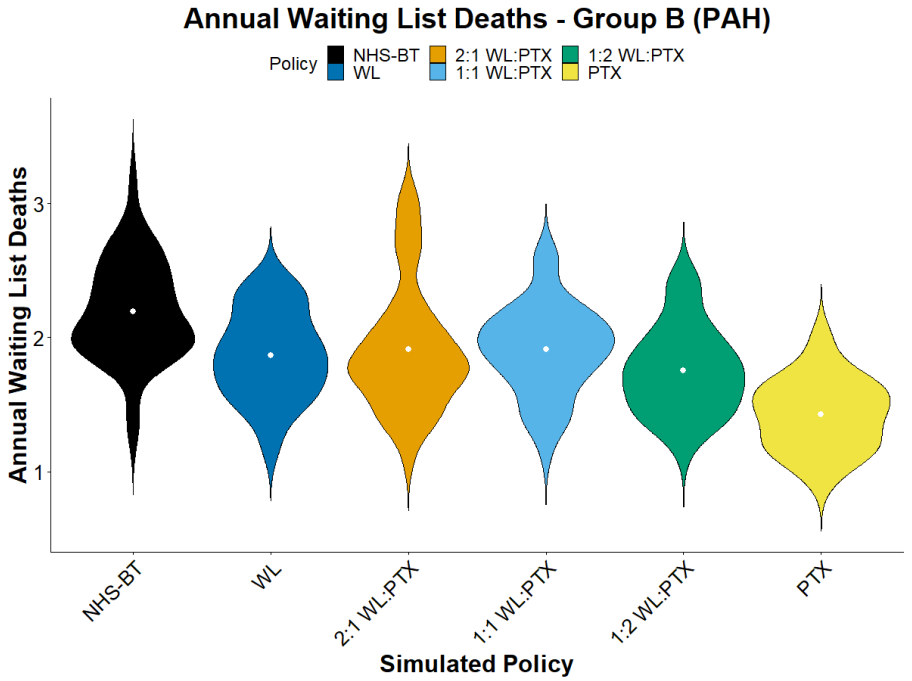
Table 4.28: Impact by policy on candidates and recipients with group B (pulmonary arterial hypertension / pulmonary vascular disease) diagnoses.

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	2 (0.4)	1793	5.8	80.5 / 55.3	473 (59.1)
WL	2 (0.3)	2135	12.5	81.3 / 54.1	1120 (162.1)
2:1 WL:PTX	2 (0.5)	2185	13.0	81.3 / 54.8	1092 (119.3)
1:1 WL:PTX	2 (0.3)	2300	13.7	81.8 / 57.4	1080 (139.5)
1:2 WL:PTX	2 (0.3)	2357	12.7	82.5 / 55.7	920 (172)
PTX	1 (0.3)	2371	12.9	83.0 / 59.6	305 (63.5)
SLT-WL	1 (0.4)	2072	13.4	78.9 / 52.6	901 (131.6)
SLT-2:1	1 (0.3)	2234	13.3	81.5 / 55.0	883 (154.8)
SLT-1:1	1 (0.2)	2202	12.5	81.7 / 55.9	789 (133.8)
SLT-1:2	1 (0.2)	2296	12.2	81.9 / 56.5	592 (117.6)
SLT-PTX	2 (0.4)	2238	10.9	82.6 / 58.3	396 (74.6)

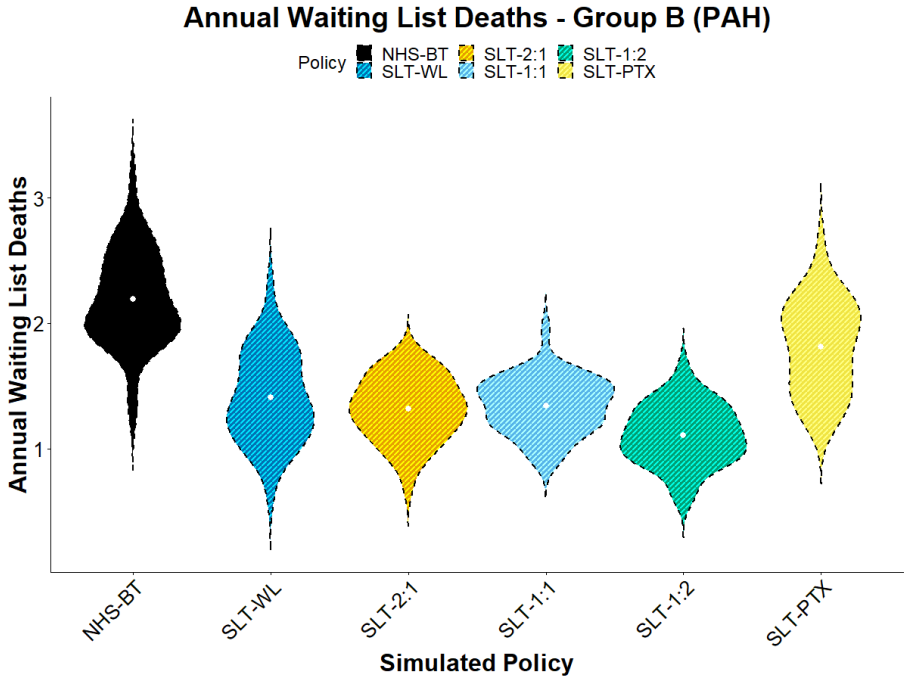
Impact on Group B - Pulmonary Vascular Disease/PAH Candidates with group B diagnoses (PAH) tended to benefit as priority shifted to maximising post-transplant survival. It is worth noting that there were relatively fewer group B candidates than in the other three diagnosis groups.

The number of waiting list deaths was low for all policies, ranging from 1 to 2. As priority shifted from the WL policy to the PTX policy, net benefit increased from 2135 days to 2371 days. One year post-transplant survival rates increased from 81.3% to 83.0% and give year post-transplant survival rates increased from 54.1% to 59.6%.

Introduction of the SLT algorithm for ILD candidates had a marginal impact on the number of waiting list deaths and both net benefit and post-transplant survival rates decreased slightly.



(a)



(b)

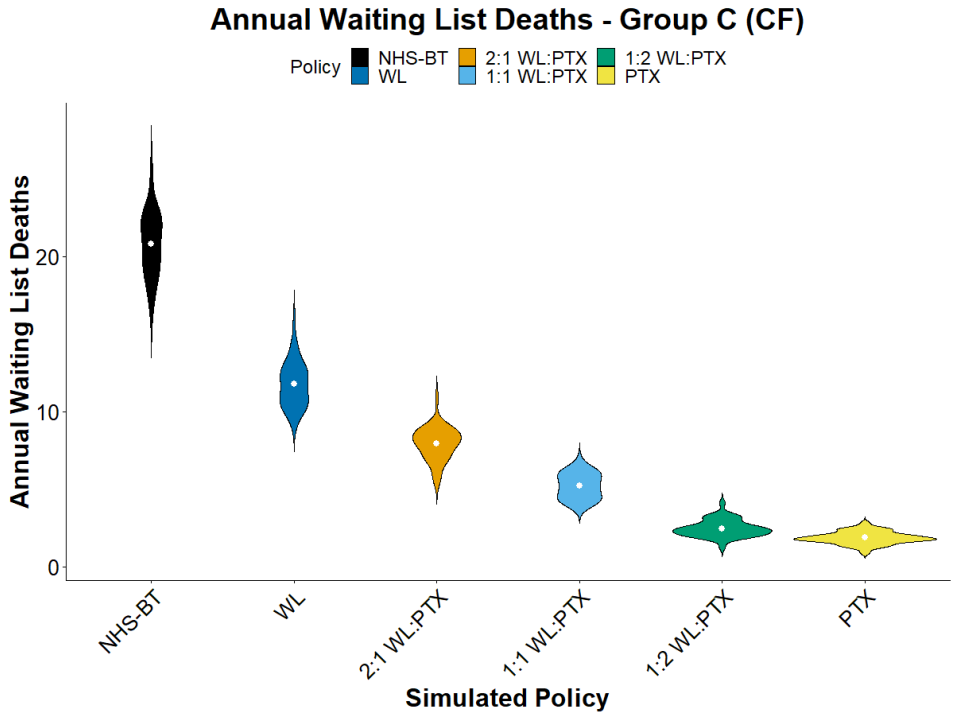
Figure 4.23: **(a)** For candidates with pulmonary vascular disease/PAH, increasing post-transplant priority generally resulted in decreasing waiting list deaths, though the difference was relatively small compared to other diagnosis groups. **(b)** For candidates with pulmonary vascular disease/-PAH, the SLT policies resulted in a slight decrease in annual waiting list deaths, though the difference was not very substantial.

Table 4.29: Impact by policy on candidates and recipients with group C (cystic fibrosis) diagnoses.

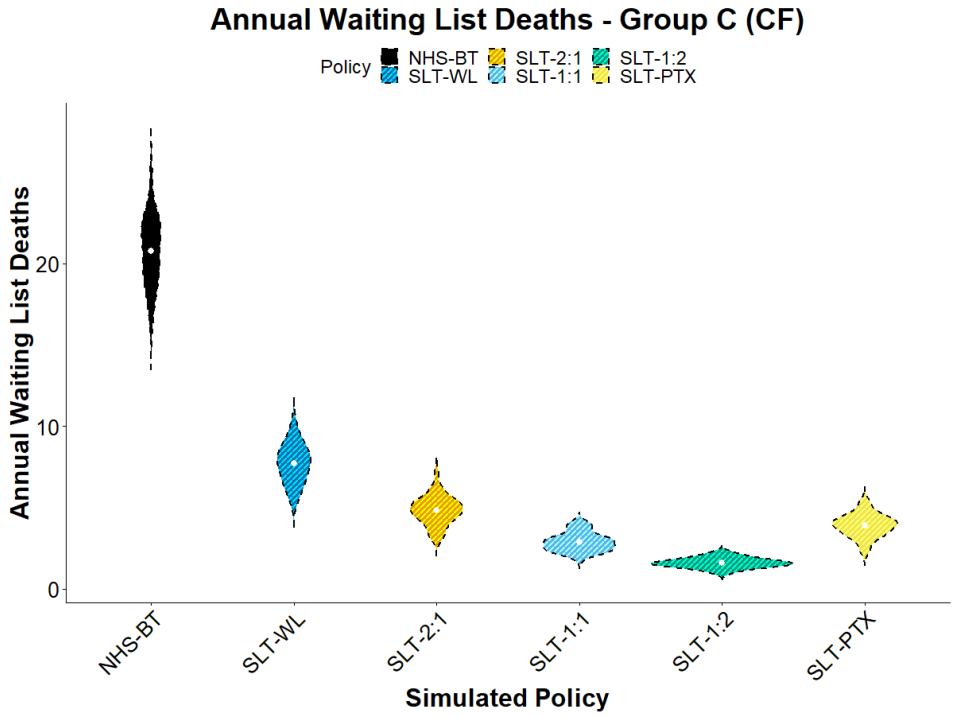
Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	21 (2.1)	2601	7.8	85.8 / 64.5	436 (39)
WL	12 (1.5)	2825	15.3	84.8 / 62.6	683 (57.2)
2:1 WL:PTX	8 (1.1)	2884	16.0	85.6 / 63.6	486 (55.7)
1:1 WL:PTX	5 (0.8)	2908	16.1	85.9 / 64.3	322 (38.4)
1:2 WL:PTX	2 (0.5)	2951	16.7	86.0 / 64.9	118 (22.4)
PTX	2 (0.4)	2955	16.3	86.1 / 65.0	32 (4.1)
SLT-WL	8 (1.2)	2860	15.6	85.0 / 63.3	494 (46.8)
SLT-2:1	5 (0.9)	2901	16.1	85.6 / 64.1	317 (49.1)
SLT-1:1	3 (0.6)	2913	16.2	85.7 / 64.5	199 (36.4)
SLT-1:2	2 (0.3)	2897	16.4	85.5 / 64.0	69 (13.5)
SLT-PTX	4 (0.7)	2922	15.4	86.3 / 65.3	63 (11.5)

Impact on Group C Diagnoses - Cystic Fibrosis For group C candidates the WL policy resulted in the highest number of waiting list deaths (12). The other policies (2:1, 1:1, 1:2 and PTX) resulted in a lower number of waiting list deaths, ranging from 8 with the 2:1 policy to 2 with the PTX policy. All policies had similar net benefit, relative benefit and post-transplant survival rates. The PTX policy resulted in the shortest waiting times.

Implementing the SLT policies for ILD candidates resulted in the number of waiting list deaths decreasing from 12 with the WL policy to 8 with the SLT-WL policy. The number of waiting list deaths decreased slightly for the other policies, except for SLT-PTX which resulted in the number of waiting list deaths increasing from 2 to 4. There was also a slight decrease in net benefit, ranging from a few days to approximately 50 days. Overall group C candidates and recipients were not adversely impacted with the introduction of the SLT policies.



(a)



(b)

Figure 4.24: **(a)** For group C (CF) candidates, the PTX policy resulted in the fewest waiting list deaths, changing the priority ratio towards waiting list survival resulted in increasing waiting list deaths, with the largest number of waiting list deaths resulting from prioritising waiting list survival (WL). **(b)** For group C(CF) candidates, the SLT policies also resulted in a reduced number of waiting list deaths with the lowest number occurring with the SLT-1:2 policy.

Table 4.30: Impact by policy on candidates and recipients with group D (interstitial lung disease) diagnoses.

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	45 (3.6)	1743	7.7	77.9 / 48.6	317 (23.2)
WL	5 (0.9)	2119	16.6	78.2 / 49.4	98 (13.4)
2:1 WL:PTX	6 (0.9)	2145	16.7	78.3 / 49.7	115 (16.8)
1:1 WL:PTX	8 (1.3)	2193	16.8	78.6 / 50.4	147 (24.3)
1:2 WL:PTX	15 (2.3)	2244	16.2	79.2 / 51.2	204 (27.1)
PTX	50 (3.3)	2397	14.7	81.5 / 55.3	432 (27.2)
SLT-WL	2 (0.4)	1874	15.4	75.8 / 45.7	45 (10.4)
SLT-2:1	2 (0.5)	1902	15.4	76.1 / 46.0	56 (13.5)
SLT-1:1	3 (0.5)	1898	15.2	76.1 / 46.0	71 (12.6)
SLT-1:2	3 (0.5)	1813	14.7	75.5 / 44.7	51 (9.4)
SLT-PTX	1 (0.3)	1680	14.3	74.2 / 42.8	14 (1.3)

Impact on Group D Diagnoses - Interstitial Lung Disease Candidates with group D (ILD) diagnoses had the lowest waiting list mortality and waiting times with the WL policy (5 per year and 98 days respectively), however they also had the lowest net benefit (2119 days) and post-transplant survival (78.2% at one year and 49.4% at 5 years).

There was a large increase in waiting list deaths as post-transplant survival was prioritised, increasing from 5 per year (WL) to 50 per year with the PTX policy.

Net benefit was maximised with the PTX policy (2397 days) and also post-transplant survival (81.5% at one year and 55.3% at five years).

Introducing the SLT policies resulted in significantly lower waiting list deaths for group D candidates, which were the target group for the SLT policies. The SLT-PTX policy almost eliminated waiting list deaths in group D, with an average of 1 waiting list death per year (SD: 0.3) and an average waiting time of 14 days.

Unlike with the standard policies, increasing post-transplant priority did not result in increasing waiting list deaths. As post-transplant priority increased, net benefit decreased from 1874 days (SLT-WL) to 1680 days (SLT-PTX). Post-transplant survival ranged from 74.2% to 76.1% at one year and 42.8% to 46.0% at five years.

The reduction in waiting list deaths for group D candidates also resulted in a decrease in net benefit and post-transplant survival rates. The PTX policy resulted in a net benefit of 2397 days and survival rate of 81.5% one year post-transplant. In comparison, the SLT-PTX policy resulted in a net benefit of 1680 days (a 717 day decrease) and a post-transplant survival rate of 74.2% (a 7.3% decrease).

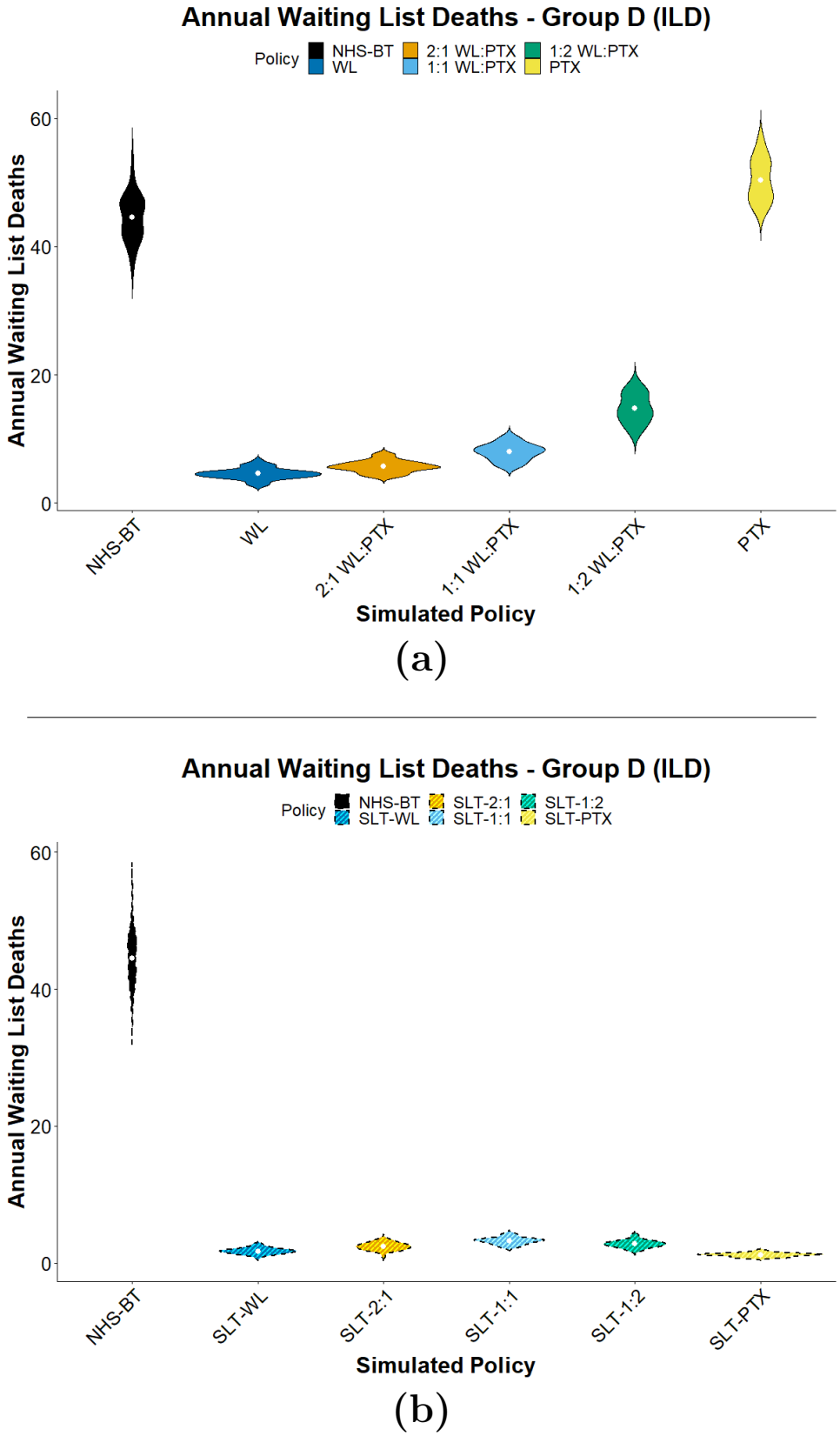


Figure 4.25: **(a)** For group D (ILD) candidates, the WL policy minimised the number of waiting list deaths. Increasing post-transplant survival priority resulted in increasing waiting list deaths. **(b)** For group D (ILD) candidates, the SLT policies significantly decreased waiting list deaths, with the lowest numbers occurring with the WL and PTX policies.

Results Stratified by Age Group

Table 4.31: Impact by policy on candidates and recipients aged 16 to 30

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	17 (1.8)	2298	7.4	83.7 / 59.8	424 (35.6)
WL	7 (0.8)	2552	15.4	82.5 / 58.2	553 (47.8)
2:1 WL:PTX	6 (0.8)	2593	15.7	83.3 / 59.0	433 (52.2)
1:1 WL:PTX	5 (0.7)	2628	15.8	84.0 / 59.8	346 (40.3)
1:2 WL:PTX	3 (0.6)	2674	16.0	84.1 / 60.7	201 (28.6)
PTX	4 (0.6)	2665	15.0	84.2 / 60.9	88 (12.4)
SLT-WL	5 (0.7)	2537	15.5	82.4 / 58.2	396 (49.9)
SLT-2:1	3 (0.7)	2580	15.7	83.1 / 59.0	297 (47.4)
SLT-1:1	3 (0.5)	2606	15.9	83.5 / 59.5	214 (36.9)
SLT-1:2	2 (0.4)	2598	15.7	83.2 / 59.3	126 (23.8)
SLT-PTX	5 (0.8)	2606	14.3	84.2 / 60.7	123 (17)

Ages 16 to 30 Candidates and recipients in this age group tended to benefit with increased post-transplant priority. The number of waiting list deaths in this age group decreased from 7 with the WL policy to 3/4 with the 1:2 / PTX policies respectively.

Net benefit was similar between all policies, ranging from 2552 days with the WL policy to 2674 days with the 1:2 policy (range: 122 days). Relative benefit was similar between policies, ranging from 15.0 to 16.0.

Post-transplant survival increased from 82.2% at one year and 58.2% at five years with the WL policy to 84.2% and 60.9% respectively.

Candidates and recipients in the 16 to 30 age group were not adversely affected by the SLT policies. There was a slight decrease in waiting list deaths (except in the case of the SLT-PTX policy). There was also a slight decrease in net benefit (approximately 30 days on average) and post-transplant survival (approximately 0.3% at one year and 0.4% at five years).

Table 4.32: Impact by policy on candidates and recipients aged 31 to 40

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	8 (0.9)	2683	7.9	86.7 / 66.4	470 (49)
WL	6 (0.9)	3025	15.30	86.3 / 65.1	1007 (78.7)
2:1 WL:PTX	4 (0.6)	3095	17.2	86.9 / 67.0	810 (87.8)
1:1 WL:PTX	3 (0.6)	3104	16.6	86.5 / 67.1	610 (69.7)
1:2 WL:PTX	1 (0.4)	3161	17.5	87.7 / 68.1	351 (57.8)
PTX	0 (0.1)	3107	17.4	87.5 / 67.4	32 (8.8)
SLT-WL	4 (0.7)	3042	16.9	86.0 / 65.8	794 (79.9)
SLT-2:1	3 (0.5)	3059	16.0	86.6 / 66.3	587 (76.4)
SLT-1:1	2 (0.3)	3030	16.3	86.4 / 66.6	448 (59)
SLT-1:2	1 (0.3)	2976	16.4	86.1 / 65.8	258 (36.1)
SLT-PTX	0 (0.1)	2924	15.8	86.1 / 65.3	52 (18.8)

Ages 31 to 40 Candidates and recipients aged 31 – 40 tended to benefit with increasing post-transplant priority. As PTX priority increased there were fewer waiting list deaths in this age group (decreasing from 6 per year to 0 per year) and drastically reduced waiting times (1007 days down to 32 days). Net benefit in this age group was much higher than the 16 - 30 age group, with net benefit ranging from 3025 (WL) to 3161 days (1:2).

Post-transplant survival rates increased from the WL policy to the 1:2 policy, with a slight decrease in survival rates with the PTX policy. One and five year survival rates increased from 86.3% and 65.1% with the WL policy to 87.7% and 68.1% with the 1:2 policy.

This age group was not majorly affected by the implementation of the SLT policy. There was a slight reduction in waiting list deaths with the SLT-WL and SLT-2:1 policies. There was a slight decrease on the order of 100 days in net benefit, and approximately a 1% reduction in post-transplant survival at 1 year and 2% reduction at 5 years.

Table 4.33: Impact by policy on candidates and recipients aged 41 to 50

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	13 (1.2)	2057	6.8	83.0 / 57.8	457 (43.5)
WL	9 (1)	2565	15.1	83.1 / 59.4	1141 (70.4)
2:1 WL:PTX	9 (1)	2571	15.5	83.1 / 59.0	1104 (75.3)
1:1 WL:PTX	8 (1)	2643	15.6	83.7 / 60.2	1073 (52.6)
1:2 WL:PTX	7 (0.7)	2664	15.8	83.9 / 60.1	959 (69.8)
PTX	5 (0.6)	2625	14.4	84.2 / 61.5	267 (39.5)
SLT-WL	7 (0.7)	2455	14.9	82.4 / 58.0	988 (74.6)
SLT-2:1	6 (0.9)	2470	15.3	82.6 / 58.1	952 (83.9)
SLT-1:1	6 (0.8)	2438	14.3	82.3 / 57.4	912 (69)
SLT-1:2	5 (0.6)	2382	14.6	82.2 / 56.7	724 (72.6)
SLT-PTX	3 (0.5)	2247	13.3	81.2 / 55.5	283 (43.7)

Ages 41 to 50 The 41 to 50 age group had the highest net benefit with the 1:2 policy (2664 days), and had fewer waiting list deaths as post-transplant priority increased. The WL policy resulted in 9 annual waiting list deaths per year for this age group on average, whereas the PTX policy resulted in 5. Post-transplant survival was maximised with the PTX policy (84.% at one year and 61.5% at five years).

Waiting times also decreased sharply with increasing post-transplant survival, decreasing from 1141 days on average for the WL policy down to 267 days with the PTX policy.

Implementation of the SLT policies resulted in a slight decrease in waiting list deaths for this group, and also a decrease in net benefit of approximately 200 - 300 days. One year post-transplant survival rates decreased by about 3% and five year rates decreased by 1% to 6%. Overall this group did not benefit from implementation of the SLT policies.

Table 4.34: Impact by policy on candidates and recipients aged 51 to 60

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	38 (3.1)	1506	6.5	77.4 / 48.3	411 (37.8)
WL	17 (1.6)	2014	15.5	77.9 / 48.9	796 (54.3)
2:1 WL:PTX	19 (1.7)	2053	15.6	78.4 / 49.3	861 (46.7)
1:1 WL:PTX	22 (1.8)	2113	16.2	78.3 / 50.0	929 (43.6)
1:2 WL:PTX	27 (2.6)	2175	15.6	79.1 / 51.2	971 (59.4)
PTX	46 (3.4)	2147	12.10	81.2 / 54.4	757 (48.7)
SLT-WL	12 (1.6)	1768	13.7	75.8 / 45.7	665 (51.7)
SLT-2:1	14 (1.9)	1807	13.9	76.2 / 46.1	715 (57.4)
SLT-1:1	16 (1.5)	1859	14.3	76.4 / 46.5	780 (51.4)
SLT-1:2	17 (1.6)	1811	13.8	76.0 / 45.8	763 (51.3)
SLT-PTX	20 (1.7)	1707	13.3	75.4 / 44.7	620 (56.3)

Ages 51 to 60 While younger age groups tended to benefit from increased post-transplant priority, this age group benefited from increased waiting list priority. The number of annual waiting list deaths were 17 with the WL policy, increasing to 46 with the PTX policy. The WL and PTX policies also resulted in the lowest waiting times (796 and 757 days respectively).

Net benefit was highest with the 1:2 policy, with an average of 2175 days of life gained.

The PTX policy resulted in the highest post-transplant survival rates (81.2% at one year and 54.4% at five years) and the WL policy resulted in the lowest (77.9% and 48.9%).

Implementing the SLT policies resulted in a noticeable reduction in waiting list deaths: 17 down to 12 for the SLT-WL policy and 46 down to 20 with the SLT-PTX policy. Net benefit also decreased by 250 - 450 days with the SLT policies. Post-transplant survival at one year decreased from 81.2% with the PTX policy to 75.4% with the SLT-PTX policy. Waiting times also decreased by 100 to 200 days on average.

Table 4.35: Impact by policy on candidates and recipients aged 61+

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	15 (1.5)	1378	6.2	76.2 / 45.7	401 (34.9)
WL	6 (0.7)	1815	13.90	76.8 / 46.6	710 (45.1)
2:1 WL:PTX	7 (0.7)	1855	14.1	76.3 / 46.60	809 (59.8)
1:1 WL:PTX	9 (0.9)	1930	13.8	77.4 / 47.80	875 (64.7)
1:2 WL:PTX	12 (1.4)	1982	13.4	77.4 / 48.20	946 (56)
PTX	23 (1.5)	1940	10.4	79.6 / 52.2	928 (62.4)
SLT-WL	4 (0.6)	1647	12.9	75.1 / 44.0	605 (48.1)
SLT-2:1	5 (0.6)	1687	13.3	74.9 / 44.2	686 (60)
SLT-1:1	6 (0.8)	1731	13.1	75.5 / 45.2	732 (49.8)
SLT-1:2	6 (0.6)	1672	12.8	74.8 / 43.6	739 (57.7)
SLT-PTX	6 (0.8)	1484	12.0	73.1 / 40.9	594 (64.3)

Ages 61+ For candidates and recipients aged 61+ the WL policy resulted in the fewest annual waiting list deaths (6), but this policy also resulted in the lowest net benefit (1815 days) and survival one year post-transplant (76.8%). It also resulted in the shortest waiting times: 710 days compared to 928 days with the PTX policy.

Net benefit was maximised with the 1:2 policy (1982 days) and post-transplant survival was maximised with the PTX policy (79.6% at one year and 52.2% at five years).

Implementing the SLT policies resulted in fewer waiting list deaths for every policy, with the largest decrease resulting from the SLT-PTX policy (6 down from 23 with the PTX policy). Net benefit decreased by 150 - 450 days depending on the policy, and post transplant survival rates also decreased by 1.5% to 6.5%, with the SLT-PTX policy having the largest decrease in post-transplant survival (73.1% down from 79.6% with the PTX policy). Waiting times were also reduced to 605 days with the SLT-WL policy (compared to 710 days with the WL policy), and 594 days with the SLT-PTX policy (compared to 928 days with the PTX policy).

Results Stratified by Blood Group

Table 4.36: Impact by policy on candidates and recipients with blood group A

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	26 (3)	1918	7.6	80.1 / 53.0	256 (32.4)
WL	12 (1.8)	2190	15.0	79.7 / 52.8	637 (66.2)
2:1 WL:PTX	12 (1.8)	2231	15.2	80.2 / 53.4	655 (65.2)
1:1 WL:PTX	14 (1.8)	2308	15.5	80.7 / 54.3	657 (51)
1:2 WL:PTX	16 (2.2)	2345	15.2	81.0 / 54.8	655 (68.6)
PTX	26 (2.8)	2412	14.1	82.8 / 58.1	394 (47.1)
SLT-WL	8 (1.6)	2056	14.4	78.7 / 51.1	485 (70)
SLT-2:1	8 (1.6)	2114	14.4	79.1 / 51.7	487 (81.8)
SLT-1:1	9 (1.3)	2142	14.5	79.4 / 52.1	506 (59)
SLT-1:2	9 (1.4)	2096	14.1	79.3 / 51.5	438 (52.1)
SLT-PTX	8 (1.4)	2004	13.4	78.5 / 50.5	231 (46.1)

ABO - A Trends were not as strong when stratifying by blood group compared to age group and diagnosis group.

Waiting list deaths were minimised with the WL policy (12 per year) and increased to 26 per year with the PTX policy. Net benefit was maximised with the PTX policy (2412 days) and decreased as priority moved towards waiting list survival (2190 days with the WL policy).

Post-transplant priority resulted in the highest survival rate at one year (82.8% compared to 79.7% with the WL policy) and five years (58.1% compared to 52.8% with the WL policy). The lowest waiting times were achieved with the PTX policy: 394 days compared to 637 days with the WL policy.

Implementation of the SLT policies resulted in fewer ABO-A waiting list deaths, with the range of waiting list deaths decreasing from 12 - 26 with the standard policies to 8 - 9 with the SLT policies. The SLT policies resulted in lower net benefit, with an average net benefit 100 - 400 days lower than with the standard policies.

Post-transplant survival at one year was approximately 2% - 4% lower with the SLT policies and 2% - 8% lower at five years. Average waiting times decreased by approximately 150 days on average with the SLT policies.

Table 4.37: Impact by policy on candidates and recipients with blood group B

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	8 (0.9)	1410	5.7	76.5 / 46.7	342 (53.5)
WL	3 (0.6)	1738	12.0	76.2 / 45.8	494 (89.1)
2:1 WL:PTX	3 (0.5)	1757	12.2	77.0 / 46.6	574 (85.2)
1:1 WL:PTX	4 (0.8)	1839	12.4	76.9 / 47.7	673 (99.8)
1:2 WL:PTX	6 (0.9)	1928	12.5	77.6 / 48.9	756 (93.9)
PTX	9 (1)	2009	10.8	80.2 / 52.9	555 (84)
SLT-WL	2 (0.4)	1592	11.2	74.3 / 44.2	366 (78.2)
SLT-2:1	2 (0.6)	1637	11.8	75.4 / 45.1	412 (84.5)
SLT-1:1	3 (0.5)	1704	12.7	75.6 / 45.8	488 (86.8)
SLT-1:2	3 (0.6)	1685	12.0	75.6 / 45.2	482 (100)
SLT-PTX	3 (0.6)	1648	11.6	75.4 / 44.7	292 (61.3)

ABO - B The trends with ABO-B candidates and recipients were similar to ABO-A: waiting list mortality increased as post-transplant priority increased (from 3 annual waiting list deaths to 9). Net benefit was highest with the PTX policy (2009 days) and post-transplant survival rates were also highest with the PTX policy: 80.2% at one year and 52.9% at five years.

As with ABO-A, the SLT policies resulted in a reduction in waiting list deaths (down to 2 - 3 per year) but also a decrease in net benefit (200 - 350 day decrease) and post-transplant survival (1% to 5% at one year and 1.5% to 8% at five years).

Table 4.38: Impact by policy on candidates and recipients with blood group AB

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	1 (0.2)	1950	8.5	79.9 / 52.2	142 (22)
WL	0 (0.2)	2156	15.1	79.6 / 53.5	225 (106.3)
2:1 WL:PTX	0 (0.2)	2024	13.9	78.8 / 50.9	268 (96.8)
1:1 WL:PTX	1 (0.2)	2184	14.8	79.8 / 53.0	286 (89)
1:2 WL:PTX	1 (0.2)	2202	14.5	79.8 / 53.9	308 (105.1)
PTX	1 (0.3)	2090	12.0	80.7 / 53.1	124 (50.5)
SLT-WL	0 (0.1)	2004	13.1	79.5 / 50.2	135 (49.2)
SLT-2:1	0 (0.1)	2055	14.0	79.0 / 51.0	158 (72.4)
SLT-1:1	0 (0.1)	2075	13.5	78.7 / 51.4	191 (80.9)
SLT-1:2	0 (0.1)	1992	14.0	77.4 / 49.3	136 (54.5)
SLT-PTX	0 (0.1)	1930	12.2	78.4 / 50.6	73 (28.2)

ABO - AB There were relatively few ABO-AB candidates and recipients compared to the other blood groups, as a result there were no strong trends. The standard policies resulted in 0 to 1 annual waiting list death for ABO-AB candidates and the SLT policies resulted in 0 waiting list deaths with a standard deviation of 0.1.

Table 4.39: Impact by policy on candidates and recipients with blood group O

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
NHS-BT	55 (3.9)	1824	5.7	81.6 / 56.1	731 (66.2)
WL	31 (2.3)	2446	16.2	81.3 / 55.3	1141 (52.9)
2:1 WL:PTX	30 (2.1)	2551	16.8	82.0 / 56.6	1085 (65.2)
1:1 WL:PTX	28 (2.3)	2605	16.8	82.6 / 57.9	1047 (46.2)
1:2 WL:PTX	29 (2.6)	2719	16.9	83.7 / 60.0	928 (49)
PTX	41 (2.4)	2793	15.0	85.2 / 63.1	595 (43.3)
SLT-WL	23 (2.2)	2293	15.2	80.0 / 53.4	950 (44.5)
SLT-2:1	22 (2.2)	2341	15.5	80.4 / 54.1	907 (53.2)
SLT-1:1	21 (1.7)	2386	15.4	80.9 / 54.9	857 (44.3)
SLT-1:2	19 (1.4)	2362	15.4	80.4 / 54.6	754 (40.4)
SLT-PTX	22 (2)	2260	14.3	80.2 / 53.8	575 (50.2)

ABO - O In terms of absolute numbers, waiting list mortality was highest for ABO-O candidates. Waiting list deaths followed an approximately 'U'-shaped distribution, with the fewest waiting list deaths being achieved with the 1:1 policy (28 per year). Waiting list mortality increased in both directions to 31 per year with the WL policy and 41 per year with the PTX policy.

Net benefit increased as priority shifted towards PTX survival, with the PTX policy achieving the highest net benefit of 2793 days compared to 2446 days with the WL policy.

One year post-transplant survival also increased with increasing PTX priority, from 81.3% with the WL policy to 85.2% with the PTX policy. Five year post-transplant mortality increased from 55.3% to 63.1%.

The SLT policies resulted in a reduction in waiting list deaths and waiting times. Waiting list deaths decreased from 28 per year with the 1:1 policy to 21 per year with the SLT-1:1 policy, the aproximate 'U'-shaped distribution remained the same.

Net benefit decreased by 150 - 500 days depending on the policy, one year post-transplant survival rates decreased 1% - 5% and five year survival decreased 2% - 9%.

4.4 Discussion

4.4.1 Summary of Results

While the simulation results didn't identify a single policy that outperformed all others across all metrics for all candidates and recipients, they did identify key trade-off decisions and general trends. At a population level, prioritising waiting list survival resulted in the lowest waiting list mortality, which increased as priority shifted towards post-transplant survival. This lower mortality rate however also came at the cost of lower net benefit and post-transplant survival rates, and also higher average waiting times.

Net benefit increased as prioritising post-transplant survival increased, with the PTX policy resulting in the highest net benefit. This may be due to the PTX policy selecting patients with longer post-transplant survival durations, increasing the value on the left hand side of the net benefit equation:

$$\text{Net benefit} = \text{Post-transplant Survival} - \text{Waiting List Survival}$$

It could be expected that candidates with high expected post-transplant survival would also have higher expected waiting list survival, however, if the increase in post-transplant survival is of a larger magnitude than the increase in waiting list survival, overall net benefit would be higher.

While prioritising waiting list survival resulted in lower waiting list mortality overall, it resulted in higher waiting list mortality for diagnosis groups A (COPD) (to a degree, though the 1:1 and 1:2 policies maximised WL mortality for this group), B (PAH) and C (CF). Group D (ILD) experienced the highest reduction in waiting list mortality from prioritising waiting list survival: a 90% decrease from 50 annual waiting list deaths to 5. Due to the relatively large proportion of group D candidates (34%), the outcomes for this group may have skewed the overall population-level results for waiting list mortality.

Net benefit was maximised for all diagnosis groups with increasing priority on PTX survival, with either the 1:2 or PTX policies maximising net benefit.

Implementation of SLT policies resulted in a very large reduction in waiting list deaths, on the order of 30% - 50% compared to the WL and NHS-BT policies respectively. This large decrease in mortality is due to the increased number of group D (ILD) candidates receiving single lungs and therefore being removed from the waiting list at twice the rate of the standard policies. The total number of waiting list deaths was similar between all SLT policies, ranging from 31 to 34 per year. Average waiting times also decreased by approximately 100 days. These benefits did however come at the cost of lower net benefit and post-transplant survival.

SLT policies resulted in reduced waiting list mortality across all diagnosis groups, with the largest reduction being with group D candidates which were the target demographic for these policies. All diagnosis groups had reduced net benefit with the introduction of

the SLT policies, with group D also experiencing the largest decrease in net benefit.

Applying an age threshold for ILD recipients to be eligible for SLT resulted in increased waiting list mortality as shown in table 4.18. The most likely explanation for this is that the age threshold reduced the number of recipients that could receive a SLT, resulting in fewer candidates being transplanted and more candidates remaining on the waiting list, and thus contributing to waiting list mortality.

All groups tended to have reduced post-transplant survival rates at one and five years, with group D (ILD) again experiencing the largest decrease. All diagnosis groups experienced reduced waiting times with the SLT policies.

Overall, the introduction of SLT policies resulted in a large reduction in waiting list mortality, primarily benefiting group D, at the cost of slightly lower net benefit and post-transplant survival rates for groups A (COPD), B (PAH) and C (CF), and relatively larger decreases in these metrics for group D.

Up to age 50, candidates and recipients tended to benefit from post-transplant survival being prioritised, however this changed at ages 51+ where prioritising waiting list survival became more beneficial in terms of reducing waiting list deaths. Across all age groups the highest net benefit was achieved with the 1:2 policy.

Post-transplant survival rates tended to be similar within age groups, regardless of the allocation policy. However, the 16 - 50 age group tended to have higher post-transplant survival rates and a narrower range in survival rates than the 51+ age group.

Table 4.40: Post-transplant survival tends to decrease with age, independent of the allocation policy. The variation in post-transplant survival rates between policies increases in the 51+ age group.

Age Group	Lowest 1-Yr PTX Survival (%)	Highest 1-Yr PTX Survival (%)	Range (%)
16 - 30	82.5	84.2	1.7
31 - 40	86.3	87.7	1.4
41 - 50	83.1	84.2	1.1
51 - 60	77.9	81.2	3.3
61+	76.3	79.6	3.3

The impact of the SLT policies was less noticeable in the 16 - 50 age group. All age groups tended to have reduced waiting list mortality across all SLT policies, with the only exception being the SLT-PTX policy for ages 18-30.

Net benefit followed the same trend across age groups as post-transplant survival: the reduction in net benefit with the implementation of SLT policies resulted in smaller decreases in the 16 - 50 age group, with a larger reduction in the 51+ age group. Absolute net benefit also tended to decrease with age.

Trends across blood groups were less clear. In general, the waiting list priority policy resulted in the fewest waiting list deaths with ABO-O being the only exception, following more of a shallow 'U'-shaped distribution of waiting list deaths as priority shifted towards

post-transplant survival. The highest net benefit across blood groups tended to result from the PTX policy, with ABO-AB being the only exception with the 1:2 policy maximising net benefit.

Post-transplant survival rates increased with increasing post-transplant priority across all blood groups, with ABO-O patients having the highest post-transplant survival rates and also the highest net benefit on average.

Implementation of the SLT policies reduced waiting list deaths across all blood groups and SLT policies. As with other stratifications of the results, the trade-offs were the same: the reduced number of waiting list deaths also resulted in lower net benefit and post-transplant survival.

Attempts to adjust survival or net benefit calculations by using conditional survival (CON) or risk-adjusted benefit (RAB) didn't significantly change any of the performance metrics.

4.4.2 Handling Protected Characteristics

Protected characteristics such as ethnicity, age, sex, and so on must be considered very carefully for inclusion in an allocation score. By including such characteristics there is risk of unfair bias/unintentional discrimination, lawsuits, and other legal issues. For the models in this work, age and sex were used as part of a larger model that predicts risk of mortality, and allocation is performed with respect to the balance of waiting list and post-transplant mortality risk.

There was however a purposeful decision to not include ethnicity in the models. The first reason is that variables such as sex or age have clear definitions, whereas ethnicity is a vastly simplified proxy variable for a much more complex biological reality. The second reason is the risk of a potential 'feedback loop' in allocation:

1. Individuals with a specific ethnicity 'E' have poor post-transplant outcomes due to the difficulty of matching to a suitable donor due to biological reasons related to their ethnicity
2. Survival models are trained on this data, identifying ethnicity 'E' as having higher post-transplant risk
3. Allocation scores based on the survival models assign candidates with ethnicity 'E' lower scores
4. As a result of lower scores, candidates with ethnicity 'E' have access to lower quality donor lungs, and therefore poor post-transplant outcomes

This same risk of a feedback loop can also be applied to transplant centres, for example, if one centre tends to transplant higher risk recipients, or for other reasons has lower

expected post-transplant survival. Including transplant centre as a variable in a survival model will reinforce the poor outcomes at that centre.

Given these considerations, there were two possibilities of how to handle these variables. The first was to include them in the survival models, but for allocation the reference value would be used regardless of ethnicity/transplant centre. This has the advantage of including additional variables that are statistically significant in the model, without directly using protected characteristics. However, due to the way regression models work, the inclusion or exclusion of variables impacts the risk coefficients of other variables (see Appendix D.1 for more on this). This still introduces the risk of indirect bias/discrimination if there is a correlation between ethnicity and other variables used in the model. For this reason the second option was used for this research: simply exclude these variables from survival models and the allocation score. This way, allocation is not influenced at all by ethnicity, transplant centre, or other protected characteristics.

There are also frameworks and methods being developed in the field of artificial intelligence/machine learning to ensure fairness in models that are developed.[142] These methods could be incorporated into the methodologies presented in this work as potential future work.

4.4.3 Alignment With / Contribution to the Literature

The results in this chapter align with some results published in the literature, but also differ from certain publications. This subsection reviews the similarities and differences in the survival models that were built, methods used in simulation, and simulated outcomes. Several unique aspects of the methods described in this thesis are highlighted and their contributions to the literature are discussed.

Survival Models and Variable Selection

Variable Selection There was agreement between variables used in the lung allocation score[143] and the survival models created in section 4.3.1. For waiting list survival both models used diagnosis group, diabetes status, age at listing and FVC at listing. For post-transplant survival both models used diagnosis group, creatinine and age. One difference is that the models in this thesis used restricted cubic splines for some continuous variables, allowing a non-linear relationship between the variable and risk of mortality (see appendix A for an explanation of RCS).

One contribution this thesis makes is demonstrating how to combine recipient and donor variables to predict post-transplant survival (see appendix D.2.1). Currently the LAS does not take any donor variables into consideration (such as DBD/DCD donor and CMV status), the organ transplanted (left/right/pair), or donor-recipient interactions such as donor-recipient height difference. Accounting for the interaction of the donor with the recipient results in candidate rankings that dynamically adapt to the characteristics of the

donor and each potential recipient, leading to allocation that is more patient-centred.

Hazard Ratios - Diagnosis Group The hazard ratios for each diagnosis group differed between the waiting list survival model in this thesis and the model in the UNOS guide to calculating LAS.[42] The model in the UNOS guide shows that relative to group A (COPD), group B (PAH) had the highest risk of mortality on the waiting list, followed by group C (CF) and finally group D (ILD). However in this work, that ordering was almost reversed: group D had the highest risk of mortality (relative to group A), followed by group B and finally group C.

Hazard ratios by diagnosis group also differed for post-transplant survival. Table 4.41 shows a comparison of the hazard ratios presented in this chapter and a study comparing BLT and SLT.[52] The hazard ratios in the results section show groups B and D have similar hazard ratios with group D having a slightly higher hazard ratio than group B, however, group B has over 3 times the hazard ratio of group D in.[52] Similar to the waiting list survival model, the post-transplant model generated from this work results in the diagnosis groups having opposite orderings of risk.

Table 4.41: Post-transplant hazard ratios by diagnosis group differ between the results presented in this thesis for the UK population and those by Chang et al. for the US population.[52] COPD = Chronic Obstructive Pulmonary Disease, PAH = Pulmonary Arterial Hypertension/Pulmonary Vascular Disease, CF = Cystic Fibrosis, ILD = Interstitial Lung Disease.

Diagnosis Group	Hazard Ratio (Chang et al.)	Hazard Ratio (Section 4.3.1)
A - COPD	Reference	Reference
B - PAH	3.37	0.85
C - CF	0.83	0.66
D - ILD	1.03	0.89

There are a number of potential reasons for these differences, the most likely being the differences in characteristics, surgical procedures and patient management between the US and UK lung transplant populations. These population differences will lead to different survival rates and result in different hazard ratios being calculated. The choice of variables for inclusion in a Cox model will also result in different hazard ratios being calculated. The hazard ratio is calculated from the regression coefficients in the model, and the regression coefficient for one variable assumes the values of all other variables are unchanged. Therefore, adding or removing variables from a model influences the coefficients of all variables in the model, so even with identical data sets, different variable choices can lead to different hazard ratios being calculated for the same variables.

Hazard Ratios - Single vs Bilateral Lung Transplantation This work investigates the impact of utilising SLT for recipients with ILD, similarly, Benvenuto et al.[80] calculated the hazard ratios for left-single, right-single and bilateral lung transplantation, but with

COPD recipients being the target demographic. In the results published by Benvenuto et al.[80] right-single lung transplantation was used as the reference group, requiring the hazard ratios in this work to be adjusted for comparison.

In this work, the post-transplant survival model (figure 4.8 on page 106) showed that there was no statistical difference between right-single and bilateral lung transplantation, but left-single lung transplantation had a statistically significantly higher risk compared to BLT. Their results for COPD recipients showed that left-single lung transplant had a hazard ratio of 1.19 (unadjusted model) or 1.24 (adjusted model) compared to right-single lung transplant.

Converting the hazard ratio in this work to be relative to right-single lung transplant, the resulting hazard ratio is HR = 1.51/1.08 = 1.398. For BLT, their work shows a hazard ratio of 0.81 (unadjusted) or 0.88 (adjusted) relative to right-single lung transplant. In this work, the hazard ratio is calculated as: HR = 1/1.08 = 0.9259. While the exact hazard ratios differ, and despite the focus being recipients with COPD rather than ILD, the trend of transplanting only the left lung resulting in the highest risk of mortality matches between this work and the work by Benvenuto et al.[80]

A retrospective analysis of single vs bilateral lung transplantation found a similar trend,[52] though it did not differentiate between left-single and right-single lungs. Using SLT as the reference group, they calculated the hazard ratio of BLT as 0.583. This hazard ratio is similar to this work when comparing left-single lung transplant to BLT: HR = 1/1.51 = 0.662.

Simulation Methods

The largest contribution of this work is the development of a novel lung allocation policy simulation engine. The purpose of the engine is similar to TSAM[139] but there are key differences between TSAM and the simulation engine developed for this thesis (discussed in section 4.2.7).

The most similar published literature to the work completed in this thesis is the work by Valapour et al. to predict the potential impact of replacing the LAS in the US with the CAS,[100] which would be a large change to the US allocation system. In the same way, the work in this chapter was to predict the impact of changing UK lung allocation from a sequential centre-based policy to a national score-based policy.

To investigate the impact of the ratio of WL to PTX survival they used two different priority-ratios: 2:1 and 1:1. The impact of this ratio was also investigated in this chapter, and the range of ratios was expanded to cover the full spectrum from 1:0 WL:PTX to 0:1 WL:PTX, with the ratios 2:1, 1:1, and 1:2 in between the two extremes. The methods were also similar when evaluating the impact of variations in allocation rules. Their work varied the weight of 'placement efficiency' (a metric that considers the proximity of the donor to the recipient and travel costs) for each WL:PTX ratio, resulting in six scenarios being simulated.

This work also simulated variations in allocation rules for a total of 46 scenarios:

1. 5 scenarios of 'standard' policies utilising BLT (section 4.3.5)
2. 4 additional scenarios using the standard WL policy with increased donor utilisation (section 4.3.5)
3. 5 scenarios where SLT was used for patients with ILD (section 4.3.5)
4. 4 additional scenarios where the SLT-WL policy was simulated with increased donor utilisation (section 4.3.5)
5. 1 additional scenario where the ordering of offers were reversed (right lung first instead of left lung first) for SLT policies (section 4.3.5)
6. 2 additional scenarios to evaluate the impact of setting an age threshold to receive SLT (section 4.3.5)
7. 5 scenarios using a 1:1 WL:PTX ratio combined with risk-adjusted benefit with various weights applied to the risk factor (section 4.3.5)
8. 5 scenarios of using conditional survival combined with the WL policy (section 4.3.5)
9. 5 scenarios of using conditional survival combined with the 1:1 policy (secton 4.3.5)

Another difference, and also contribution is the simulation engine for this work allowed multiple simulations to run in parallel, allowing 40 simulation runs to be completed per scenario, compared to the 10 simulation runs using TSAM.

Simulation Results - Waiting List Mortality

Geographic Boundaries Geographic boundaries limit access to lung transplant, potentially causing unnecessary waiting list mortality and likely resulting in sub-optimal allocation (as discussed in section 1.1.2). For the US CAS simulations[100] a 36% to 47% decrease in waiting list mortality was predicted with the removal of hard geographic boundaries. In addition, the *greatest decreases* in waiting list mortality corresponding with placement efficiency having the *lowest weight* - indicating a correlation between prioritising geographic proximity and increased waiting list deaths.

Despite different populations, policies and scenarios being simulated, the removal of geographic boundaries had the same impact in the UK simulations. Table 4.14 on page 122 shows that centre-based allocation with the NHS-BT policy resulted in 90 annual waiting list deaths, whereas using a national score-based system resulted in 46 to 47 annual waiting list deaths (depending on whether a 2:1 or 1:1 ratio was used, respectively). This represents a 48% to 49% decrease in waiting list mortality which is similar to the US simulations (36-47%). However, the proportion of the decrease that can be attributed to removing

geographical boundaries and the proportion attributed to to using a LAS-style allocation score can't be derived from these results.

Waiting List Mortality by Diagnosis Group The two major changes to UK lung allocation that are simulated in this work are: (1) changing from centre-based to a national named allocation system, and (2) using an allocation score instead of a tier-based system. In the results published by Lingaraju et al.[58] the impact of the LAS in the US is evaluated, and the changes of ranking on the waiting list for each diagnosis group are given:

Table 4.42: Changes in rankings by diagnosis group comparing the pre-LAS and post-LAS eras in the US, as given by Lingaraju et al.[58] COPD = Chronic Obstructive Pulmonary Disease, CF = Cystic Fibrosis, ILD = Interstitial Lung Disease.

Diagnosis Group	Pre-LAS Ranking	Post-LAS Ranking	Description
Group A - COPD	9.3	14.2	Lower
Group C - CF	8.7	9.8	Slightly Lower
Group D - ILD	10.2	5.3	Much Higher

The results presented in this chapter don't explicitly show changes in rankings, however this can be inferred by making the assumption that higher rankings result in lower waiting list mortality, and then looking at changes in waiting list mortality compared to the NHS-BT policy. Table 4.27 on page 131 shows waiting list mortality increased for group A (COPD) candidates, indicating a lower ranking, which is in agreement with the summary in table 4.42. Tables 4.29 and 4.30 show almost 90% reductions in waiting list mortality for group C (CF) and group D (ILD) candidates depending on the WL:PTX ratio simulated. However, when comparing results using the same 2:1 ratio as the LAS, there was a 60% decrease for group C and 85% decrease for group D. This indicates much higher rankings for group D candidates which agrees with table 4.42, however the results for group C differ, indicating a substantial increase in rankings for group C candidates as well.

The US CAS simulations[100] predicted a 74% to 76% decrease (46 down to 11-12) in waiting list mortality for group C candidates, and the results in this thesis predicted a similar 62% to 76% decrease (21 down to 5-8). For group D, the US simulations predicted a 42% to 56% decrease (278 down to 123-161) in waiting list mortality, whereas the simulations in this chapter predicted a much larger 82% to 87% reduction (45 to 6-8). Their results for group A predicted an unchanged waiting list mortality rate, whereas the results from this work (table 4.27 on page 131) shows a 59% to 71% increase from 17 deaths per year to 27-29. However, it is worth noting that the changes that were being simulated for the US[100] focused on replacing the *existing* LAS scoring system with CAS and the removal of geographic boundaries, whereas this work focused on simulating the *introduction* of a scoring system, with the removal of geographic boundaries being a secondary focus.

A UK study[50] of the survival benefit of lung transplant for different diagnosis groups presented similar waiting list mortality rates to those shown in this chapter. Their results

span 11 years from 1995 to 2006, so for comparison the annual waiting list mortality rates in this section were multiplied by 11 and displayed in table 4.43. The simulation engine also assumed a 3% annual growth rate in population and started with 300 candidates on the waiting list, resulting in a difference in absolute numbers. To compare waiting list mortality rates, the percentage of total mortality by diagnosis group is also shown in the table. The mortality rates for groups B (PAH) and C (CF) were very similar between the results published by Titman et al.[50] and this work. The main difference appears to be that the simulation engine predicted a lower percentage of group A (COPD) candidates dying on the waiting list (underestimate of 8.5% from observed), and a higher percentage waiting list mortality for group D (ILD) candidates (overestimate of 10.4% from observed).

Table 4.43: A comparison of UK waiting list mortality by diagnosis group from the results published by Titman et al.[50] and the results in this chapter. COPD = chronic obstructive pulmonary disease, BE = bronchiectasis, PH = pulmonary arterial hypertension/pulmonary vascular disease, CF = cystic fibrosis, ILD = interstitial lung disease.

Diagnosis	Waiting List Mortality - Titman et al.[50] (% Total)	Waiting List Mortality - Thesis (% Total)
A - COPD and BE	184 (28.5%)	187 (20%)
B - PAH	30 (4.65%)	22 (2.35%)
C - CF	157 (24.34%)	231 (24.7%)
D - ILD	274 (42.5%)	495 (52.9%)

Waiting List Mortality by ABO The simulations in this work were used to evaluate the impact of switching from a centre-based allocation system to a national allocation system. The US simulation study[100] evaluated the potential impact of removing the concentric ring based geographical boundaries in the US and using a CAS instead. However, this study only used WL:PTX ratios of 1:1 and 2:1, so the same ratios in this work will be used for comparison. When looking at the differential impact by blood group, the authors stated:

> Declines in waitlist deaths were more pronounced among type O candidates than other blood types.

The results in this work likewise predicted a large decline in waiting list mortality for ABO-O candidates as a result of removing geographical boundaries: a decrease from 55 with the NHS-BT policy to 28 with the 1:1 WL:PTX policy (a 49% decrease) to 30 with the 2:1 policy (a 45% decrease) (see table 4.39).

However, unlike the results presented by Valapour et al.,[100] the simulations in this work predicted large decreases in waiting list mortality across all blood groups (except ABO-AB, due to the extremely small numbers of candidates with this blood group). For ABO-A candidates, waiting list mortality was 26 per year with the NHS-BT policy, and decreased by 46 - 54% using national allocation (see table 4.36). For ABO-B candidates,

the NHS-BT policy resulted in an average of 8 annual waiting list deaths, which decreased to 3 - 4 (50 - 62% decrease) using national allocation (see table 4.37 on page 145).

Waiting List Mortality by Age Group The work by Valapour et al. shows waiting list mortality by age group,[100] the following quote from the paper aligns exactly with the results in this chapter:

> "[...] simulated waitlist deaths declined for all age groups."

Tables 4.31 to 4.35 show that compared to the simulated NHS-BT policy, all age groups experienced a decrease in waiting list mortality (compared to either the 2:1 or 1:2 WL:PTX policy). This is summarised in table 4.44.

Table 4.44: All age groups experienced a decrease in waiting list mortality when simulating the 1:1 WL:PTX and 2:1 WL:PTX policies compared to the simulated NHS-BT policy.

Age Group	Simulated NHS-BT Policy Waiting List Deaths	Simulated Waiting List Deaths (Policy)	% Decrease
16-31	17	5 (1:1)	70.6%
31-40	8	3 (1:1)	62.5%
41-50	13	8 (1:1)	38.5%
51-60	38	19 (2:1)	50.0%
61+	15	7 (2:1)	53.3%

As with the other comparisons to this study, the changes being modelled are different but the overlap in results can still be considered a 'soft' validation of the results.

Simulating SLT: DES and Markov Models In a similar way that this work investigated the impact of SLT on recipients with ILD in the UK, another study used a Markov model[144] to predict the impact of SLT on recipients with COPD in the US.[91] Although the focus in that study was on recipients with COPD, recipients with IPF (ILD) could also undergo SLT in the simulations. The Markov model was used to simulate the probabilities of candidates transitioning between different 'states'. For example, if a candidate is on the waiting list and is transplanted, they have transitioned from the 'waiting list' state to the 'transplanted' state. This is a more abstract approach to simulating the problem compared to DES, which simulates the underlying mechanisms of a system.

To simulate the interval between donors for the Markov model, the authors used UNOS data to stratify donors by blood group, height, and region. The median donor interval for ABO-A and ABO-O donors was then used. This is a simplifying assumption for their model, and differs from the roulette selection used in this thesis that randomises the donor interval according to the probability distribution derived from data.

Rather than simulating survival times for candidates, the daily probability of being removed from the waiting list was calculated by dividing the number of individuals removed from the waiting list for any reason other than transplantation by the waiting time. This

was performed for the five most common lung transplant diagnoses, but did not take into account other candidate variables. The calculated daily probability was then modelled as a beta distribution[145] that the Markov model sampled from. This approach assumed candidates that were removed from the waiting list due to their condition deteriorating died on the waiting list. The number of waiting list deaths was then calculated as a constant proportion (85%) of candidates awaiting transplant. This is another simplifying assumption, and this method of simulating waiting list mortality is very different from the methods described in this thesis.

Post-transplant survival was modelled using ISHLT registry data. First the median PTX survival was calculated for IPF/ILD recipients receiving SLT, and recipients with cystic fibrosis, pulmonary hypertension, sarcoidosis and COPD receiving BLT. Then these median PTX survival durations were used per-diagnosis in the model. The method of determining PTX survival for COPD with SLT was more involved, but described in the supplementary material published by Munson et al.[91]

Although the methods used by Munson et al. were very different from the methods in this work, the result of increasing the use of SLT matched the results in this chapter. They predicted a 6.7% increase in transplant recipients (increasing from 758 to 809) and a 21.1% reduction in waiting list deaths (decreasing from 199 to 157). The results in tables 4.19, 4.14 and 4.20 show an increased transplant volume of 12.7% to 26.0% and a decrease in waiting list mortality of 28.3% to 55.8% when comparing each SLT policy to its corresponding 'standard' policy. The results in this chapter suggest larger increases in transplant rates and decreases in waiting list mortality for the UK population. These differences are due to different assumptions, simulation methods and populations, however the conclusion is the same: using SLT increases transplant rates and reduces waiting list mortality.

The next comparison is post-transplant outcomes: Munson et al. evaluated PTX survival using the total number of years lived post-transplant for the entire population, and calculated a difference of 9 years: $4,586$ years for the SLT policy and $4,577$ years for the BLT policy. One downside of only using total PTX survival is that the SLT policies result in increased transplant volumes, so more recipients will contribute to total survival, but the duration of post-transplant survival for recipients receiving SLT is unknown. When comparing policies, more information can be gained by evaluating recipient-level and population-level metrics. This chapter focuses on the average impact per recipient, however population survival was also considered in section 4.3.5.

When the confidence intervals for the difference between these two policies were calculated the range was from -34 years to $+54$ years. If the total PTX survival is assumed to be the mean of $4,586$ and $4,577$ ($4,581.5$ years), then the percentage difference ranges from 0.74% to 1.2%. For comparison, the *absolute* differences in 1- and 5-year PTX survival in this chapter were calculated. The smallest variation was 1.2% when comparing the 1-year PTX survival rate of the SLT-WL and SLT-2:1 policies to their corresponding WL

and 2:1 policies. The largest difference was 8.2% when comparing the SLT-PTX policy to the PTX policy.

One consideration to make is that this work simulated a range of priority-ratios, whereas the results published by Munson et al. compared SLT and BLT using data derived from the LAS era which used a 2:1 WL:PTX ratio. Looking at the 2:1 WL:PTX policy, the difference was 1.2% at 1-year and 1.9% at 5-years post-transplant. The results presented in this chapter suggest a slightly larger decrease in post-transplant survival (1.2% and 1.9% at 1- and 5-years respectively) when utilising SLT than predicted by the US results. The simulations in this thesis and also those published by Munson et al. predict a relatively small decrease in post-transplant survival compared to the much larger decrease in waiting list deaths. A summary of these comparisons is shown in table 4.45.

Table 4.45: Comparison of 1- and 5-year post-transplant survival rates between single-lung (for ILD recipients) and bilateral lung transplant policies for each priority-ratio.

Priority-Ratio	1-Year PTX Survival Standard / SLT	5-Year PTX Survival (Standard/SLT)	% Difference (1-/5-yr PTX)
1:0 (WL)	80% / 78.8%	53.0% / 51.2%	1.2% / 1.8%
2:1 WL:PTX	80.5% / 79.3%	53.9% / 52%	1.2% / 1.9%
1:1 WL:PTX	81% / 79.6%	55% / 52.6%	1.4% / 2.4%
1:2 WL:PTX	81.8% / 79.3%	56.4% / 52.1%	2.5% / 4.3%
0:1 (PTX)	83.5% / 78.9%	59.5% / 51.3%	4.6% / 8.2%

The final comparison to make between the SLT and BLT policies is how they impacted candidates and recipients with diagnoses other than the target population for SLT (group D (ILD) for this thesis, and groups A (COPD) and D for Munson et al). The key takeaway regarding the use of SLT in allocation is the same for the results presented here and the results presented by Munsol et al.,[91] they state (emphasis added):

> "[...] this study demonstrates that by *prolonging* waiting times for *every patient* listed below a single lung transplant recipient and a patient with COPD, often by several donor cycles, a policy of BLT for COPD increases the risk of waitlist mortality for potential single and bilateral recipients with many different diseases. Indeed, it was found that **a policy of SLT in the base model resulted in an absolute reduction in the risk of waitlist mortality of 4.2% among all listed patients.**"

The same conclusion can be drawn from the results presented in this chapter, but applied to ILD (group D) candidates instead of COPD (group A) candidates. The SLT-1:2 policy resulted in the fewest waiting list deaths, representing a 39.2% reduction compared to the 1:2 policy using only BLT. Although SLT was only targetted at group D recipients, group A experienced a 20.6% reduction, group B a 50% reduction (though this was from 2 to 1 due to the relatively few recipients in group B), group C remained unchanged, and

group D experienced an 80% reduction (from 15 with the 1:2 policy to 3 with the SLT-1:2 policy).

One explanation for this is that two group D candidates are removed from the waiting list whenever a single donor is allocated for SLT. This results in group D candidates being removed from the waiting list (due to transplant) at a higher rate than candidates in other diagnosis groups. This results in less competition for subsequent donor lungs for all candidates on the waiting list, resulting in an overall decrease in waiting list mortality.

In summary, the work in this thesis used a completely different set of methods applied to a different population than used by Munson et al.,[91] but arrived at the same conclusion that the use of SLT results in an overall reduction in waiting list mortality, an increase in transplant volume, and only a marginal change in post-transplant outcomes.

Simulated SLT PTX Outcomes Compared to Reported US and UK Outcomes There is a published OPTN/SRTR report on lung transplantation in the US that includes comparisons of post-transplant outcomes utilising SLT and BLT.[19] The report shows that BLT resulted in a higher 5-year post-transplant survival rate of 63.1% compared to SLT with a 5-year survival rate of 53.3%.

BLT Using table 4.14 as a comparison of population-level outcomes for BLT, the results in this chapter show 5-year PTX survival rates ranging from 53% with the WL policy to 59.5% with the PTX policy. If the same 2:1 WL:PTX priority-ratio as the LAS in the US is used for comparison, the simulated 5-year PTX survival rate was 53.9% which is still almost 10% lower than the 63.1% in the report. However, this could be a result of population differences between the US and UK, as the simulation engine utilised a dataset of UK patients. The UK NHS-BT organ specific reports do report lower mean 5-year PTX survival for BLT, ranging from 56.6% in 2022[26] to 58.3% in 2019[21] which is 4.8% to 6.5% lower than the OPTN/SRTR report.

With these considerations, the simulation engine does still appear to under-estimate 5-year PTX survival by about 3%. However, when considering the 95% confidence intervals included in the NHS-BT reports, the range of 5-year post-transplant survival rates for BLT in the UK ranged from a lower bound of 52.5% in 2022 to an upper bound 62.1% in 2019. With the confidence intervals in mind, the simulated range of mean 5-year PTX survival rates are within the reported 95% confidence intervals.

SLT The 5-year PTX survival rates for the SLT policies are given in table 4.19. The simulated 5-year PTX survival rates ranged from 51.2% with the SLT-WL policy to 52.6% with the SLT-1:1 policy, which is much closer to the reported 53.3% survival rate in the US.[19] For the UK population, NHS-BT reported mean 5-year survival for UK patients receiving SLT ranging from 38.2% in 2022 to 42.9% in 2020.[21,24–26] This is approximately 10% lower than the simulated results and the US OPTN report. 95% confidence intervals

for SLT recipients ranged from 27.7% in 2022 to 52.5% in 2020. The confidence intervals between reported outcomes and mean simulation outcomes do overlap, but the simulated results are near the upper 95% confidence limit and exceed it by 0.1% (52.6% simulated vs the maximum reported 95% CI of 52.5%).

It is difficult to determine the precise reason for the differences in post-transplant survival for SLT recipients between the simulation engine and reported outcomes. One possibility is the simulated SLT algorithm selects for candidates with higher expected PTX survival due to the allocation score including PTX survival as one of the variables in the calculation. However, if this were the case, the SLT-WL policy would have much lower 5-year PTX survival than the other policies, as they will all give some weight to PTX survival in the score calculation.

The most plausible explanation appears to be the difference between the observed survival curve and the Weibull distribution that was fit to the population, shown in figure 4.11b on page 110. The figure shows the fit of the curve for post-transplant survival isn't as close as with waiting list survival, and at 5 years (1825 days) the fitted curve is lower than the observed survival curve. However, the method of generating survival times needs to be considered: the fitted curve is 'shifted' up or down according to the calculated hazard ratio of each simulated transplant recipient, then a random number is generated between 0 and 1. The survival duration is then determined by calculating the survival duration that results in the proportion of the population surviving being equal to the randomly generated number. At approximately 3000 days, the fitted curve crosses above the observed curve, resulting in higher post-transplant survival durations being more likely to be generated than what was observed. The effect of SLT is modelled using the hazard ratio, so this slight over-estimation of post-transplant survival durations will also apply in the case of SLT, resulting in slightly higher 5-year PTX survival rates than those observed.

Simulation Results - Post-transplant Survival

One large point of discrepancy between the results in this chapter and published results are the post-transplant survival rates published in the OPTN/SRTR annual data report.[19] The differences in 1- and 5-year post-transplant survival rates by diagnosis group are shown in table 4.46. This chapter predicts 1-year PTX survival rates that are approximately 10% lower for group A (COPD), 7% lower for group C (CF), and 10% lower for group D (ILD). 5-year survival rates are approximately 10% lower for group A, 5% lower for group B, 3% lower for group C, and 10% lower for group D.

Because the OPTN report is for the US population, simulated 1- and 5-year post-transplant survival rates were instead compared to the UK population using the NHS-BT annual reports from 2018 to 2022.[21,24–26,146] The results of this comparison on shown in table 4.47 and have been visualised in figures E.9 to E.14 in appendix E.4 - note that group B (PAH) is not included as this diagnosis was grouped into the "other" category in the reports. The simulated range of 1-year post-transplant survival rates was within the

range of observed rates for all diagnosis groups. For 5-year post-transplant survival, the simulated range overlapped the lower end of the observed range for group A (COPD), was within the intervals but on the higher end of the range for group C (CF), and was also on the higher end of the range for group D (ILD). However, despite being on the higher end of the range for group D recipients, the simulated 5-year post-transplant survival rates were more centred for observations pre-2022.

Table 4.46: Differences between simulated 1-year and 5-year post-transplant survival rates (PTX) shown in this chapter (UK population) and those published in the US OPTN/SRTR 2020 annual data report.[19] COPD = Chronic Obstructive Pulmonary Disease, PAH = Pulmonary Arterial Hypertension/Pulmonary Vascular Disease, CF = Cystic Fibrosis, ILD = Interstitial Lung Disease.

Diagnosis group	Reported 1-year PTX survival	Simulated 1-year PTX survival	Reported 5-year PTX survival	Simulated 5-year PTX survival
A - COPD	91%	78 - 81%	62.4%	49 - 53%
B - PAH	82%	78 - 83%	59.5%	52 - 59%
C - CF	92%	84 - 86%	66.6%	63 - 65%
D - ILD	88%	78 - 81%	57.8%	42 - 55%

Table 4.47: Comparison of simulated 1-year adn 5-year post-transplant survival rates (PTX) by diagnosis group compared to the 2019-23 NHS-BT annual reports. The range is reported as the minimum and maximum observed rates between 2018-22. COPD = Chronic Obstructive Pulmonary Disease, CF = Cystic Fibrosis, ILD = Interstitial Lung Disease.

Diagnosis group	Reported 1-year PTX survival	Simulated 1-year PTX survival	Reported 5-year PTX survival	Simulated PTX survival
A - COPD	73.1 - 88.2%	78.1 - 81.5%	51.2 - 65.5%	49.3 - 55.3%
C - CF	74.5 - 92.3%	84.8 - 86.1%	52.3 - 69.8%	62.6 - 65.0%
D - ILD	70.1 - 87.3%	77.9 - 81.5%	37.4 - 58.6%	48.6 - 55.3%

4.4.4 Simulation Benefits and Limitations

Accuracy of Survival Times

In section 4.3.4 the simulated survival times differed from observed survival times by as little as 0.1% to as much as 7.2%. The accuracy of the simulated survival times tended to increase as the number generated survival times increased, and with both waiting list and post-transplant survival the accuracy was higher with the validation dataset rather than the training dataset. The simulation engine generates hundreds of thousands of survival times per simulation run, so using the percentage error from the highest subset sizes the error for waiting list survival times is likely around 3% and for post-transplant survival 4%.

Generated waiting list survival durations tended to be lower than observed survival, and generated post-transplant survival durations tended to be higher than observed. One possible contributor to differences between observed and generated survival durations is

the 'flat tail' in the survival curve for the observed data. This results in the survival curves crossing due to the simulated survival durations being sampled from a distribution and generally being 'smooth' as opposed to step-wise. The area between the the simulated survival curve and flat tail of the observed curve will contribute to some degree to the differences between generated and observed survival times.

While the survival models for waiting list and post-transplant survival had reasonable calibration, the discrimination ability of the post-transplant model was weaker. The waiting list survival model had a higher C-statistic of 0.73 on the training dataset and 0.66 on the validation dataset, compared to the post-transplant survival model with a C-statistic of 0.60 on the training data and 0.55 on the validation data (see pages 100 and 106). While the C-statistic for post-transplant survival may be considered weak, it does represent an improvement compared to the existing NHS-BT policy: 0.62 on the training cohort compared to 0.54 with the NHS-BT policy, and 0.61 on the validation cohort compared to 0.52 with the NHS-BT policy (results in section 4.3.2).

The problem of poor discrimination for post-transplant outcomes is not unique to this research, but has also been a topic of discussion for the US LAS, which uses a post-transplant survival model with similar discrimination abilities.[147] Informal conversations with other statisticians have also revealed that predicting post-transplant survival for other organs is also a challenge, and similar weak C-statistic measures have been observed when predicting post-transplant survival for other organs.

Alternative Assumptions

One limitation with this simulation study (and simulation studies more generally) is that simplifying assumptions must usually must be made, otherwise the system is generally too complex to be computable in any reasonable amount of time. It is important that any assumptions made should guide how results are interpreted - results from a simulation study should not simply be followed blindly (i.e., a simulated allocation policy should never just replace an existing policy solely on the basis of the simulation results).

In section 4.2.6 eight key assumptions were made, with assumptions 2 - 4 determining how net benefit and post-transplant survival should be calculated. There are many other assumptions that could be made, one set of assumptions will be discussed briefly here.

Survival was capped at 20 years for both waiting list and post-transplant survival, this was to minimise the capping of survival times and approximate the longest post-transplant survival duration observed in the dataset. However, nobody on the waiting list survived or waited 20 years and it is reasonable to assume that if a candidate were to live 20 years on the waiting list, they should not have been listed in the first place.

A shorter capping duration for waiting list survival of 5 years could be used instead. In terms of actual waiting list survival observed in the dataset (including censored results), only 1.5% of candidates had a survival duration $\geq$ 5 years. This results in an implicit assumption that in the case where both waiting list and post-transplant survival are capped

then the net benefit was 15 years (5475 days). Survival durations $\geq$ 20 years are possible, and one study showed 16.4% of recipients survived to at least 20 years post-transplant.[148] Assuming that if a candidate is correctly listed they have 5 or fewer years to live on the waiting list means that the assumed 15 years of net benefit is a *lower* bound estimate, rather than an *upper* bound estimate as discussed in section 4.2.6. This difference is important: there is no longer a need to exclude data points that have a capped waiting list survival duration (assumption #2). With these alternative assumptions the entire dataset generated from the simulations can be used when calculating net benefit, post-transplant survival rates, and other post-transplant related metrics.

These alternative assumptions were applied to the same datasets used to generate the results in this chapter, resulting in additional simulated PTX outcomes being included in the calculation of metrics in the 'Alternative' columns in table 4.48. It is interesting to observe the similarity between the two sets of results: the alternative assumptions included additional simulated outcomes that change the net benefit, relative benefit and PTX survival calculations. The largest difference in net benefit was a 110 day (4.4%) decrease, and the largest decrease in relative benefit was 4.0 (28.4%) - both for the PTX policy. The largest difference in 1-year PTX survival was an increase of 0.9%, and for 5-year PTX survival an increase of 1.6% - both for the SLT-PTX policy. These differences likely have minimal (if any) impact on any clinical meaning or interpretation, so despite the necessity of having to make simplifying assumptions, there was no major difference (in this case) when changing the assumptions.

The results for group D (ILD) were also re-calculated using these alternative set of assumptions, to see if the impact of the SLT policies was the same. These results are summarised in table 4.49, and show only marginal changes in performance metrics.

Table 4.48: Comparison of post-transplant outcomes between original assumptions used in this thesis, and a potential alternative set of assumptions.

Policy (WL:PTX)	Net Benefit	Relative Benefit	1-Yr Post-Tx Survival (%)	5-Yr Post-Tx Survival (%)
	Original / Alternative	Original / Alternative	Original / Alternative	Original / Alternative
NHS-BT	1833 / 1838	6.8 / 4.8	80.2 / 80.3	53.3 / 53.7
WL	2238 / 2232	15.1 / 12.9	80.0 / 80.1	53.0 / 53.2
2:1 WL:PTX	2300 / 2312	15.5 / 13.2	80.5 / 80.8	53.9 / 54.4
1:1 WL:PTX	2376 / 2404	15.7 / 13.3	81.0 / 81.4	55.0 / 55.8
1:2 WL:PTX	2459 / 2484	15.7 / 13.0	81.8 / 82.3	56.4 / 57.3
PTX	2522 / 2412	14.1 / 10.1	83.5 / 83.4	59.5 / 59.4
SLT-WL	2100 / 2107	14.3 / 12.0	78.8 / 79.2	51.2 / 51.9
SLT-2:1	2155 / 2174	14.6 / 12.1	79.3 / 79.8	52.0 / 52.9
SLT-1:1	2197 / 2230	14.6 / 12.2	79.6 / 80.2	52.6 / 53.7
SLT-1:2	2165 / 2199	14.4 / 11.7	79.3 / 80.1	52.1 / 53.6
SLT-PTX	2073 / 2080	13.6 / 10.5	78.9 / 79.8	51.3 / 52.9

Table 4.49: Comparison of the impact of the single-lung transplant (SLT) policies on candidates and recipients with group D (ILD) diagnoses using the original set of assumptions and alternative set of assumptions (marked with an '*').

Policy (WL:PTX)	Waiting List Deaths	Net Benefit (Days)	Relative Benefit	1/5-Yr PTX Survival (%)	Waiting Time (Days)
SLT-WL	2 (0.4)	1874	15.4	75.8 / 45.7	45 (10.4)
SLT-2:1	2 (0.5)	1902	15.4	76.1 / 46.0	56 (13.5)
SLT-1:1	3 (0.5)	1898	15.2	76.1 / 46.0	71 (12.6)
SLT-1:2	3 (0.5)	1813	14.7	75.5 / 44.7	51 (9.4)
SLT-PTX	1 (0.3)	1680	14.3	74.2 / 42.8	14 (1.3)
SLT-WL*	2 (0.4)	1912	14.20	76.10 / 46.10	45 (10.4)
SLT-2:1*	2 (0.5)	1934	14.20	76.30 / 46.50	56 (13.5)
SLT-1:1*	3 (0.5)	1932	14.10	76.40 / 46.50	71 (12.6)
SLT-1:2*	3 (0.5)	1841	13.50	75.80 / 45.10	51 (9.4)
SLT-PTX*	1 (0.3)	1685	13.10	74.30 / 42.90	14 (1.3)

General Purpose Applicability

The methods described in this thesis have been applied to simulating lung allocation, but can be re-used and applied in other problem domains, as evidenced in chapter 3 where simplified versions of the techniques used in this chapter were applied to triage during the COVID-19 pandemic.

For scenarios requiring survival models, the automated process described in section 4.2.1 can be applied to any survival dataset containing sufficient data. The algorithm is separate from any simulation engine code, enabling the code to be re-used for other problems besides lung allocation.

The simulation engine can also be used (potentially with some modification) for problems related to limited resource allocation, due to the fact that it is paramaterised in multiple ways:

- Coefficients describing the *shape* and *scale* of survival curves
- Coefficients in the Cox Proportional Hazards models for simulating survival durations
- Tables of time durations between events, their corresponding frequency, and frequencies of the number of events
- Datasets containing the necessary variables for simulation
- Parameters to customise each simulation, such as number of simulation runs, simulation duration, population growth rate etc.

In the case of UK lung allocation, there will be one set of parameters to describe waiting list survival and a second set of parameters to describe post-transplant survival. If these

techniques were to be applied to a difference organ (e.g., liver) or a different population (e.g., US lung allocation) a different set of parameters would be required in each case.

This parameterisation 'de-couples' the simulation logic from the problem being simulated; so to apply the techniques to a different organ or allocation problem requires configuring the engine with the necessary parameters and providing the needed datasets as input. If specific logic relating to the new problem is required then additional code can be written and included in the simulation engine, or if only minor modifications are required the existing code can be edited.

4.4.5 A Brief Discussion on the Acceptability of Trade-offs

The allocation policy can *only dictate which candidates should be prioritised* for transplant, the performance metrics used for evaluating policies (waiting list mortality rates, net benefit per patient, post-transplant survival rates) are an *indirect* result of the demographics of candidates chosen for transplant.

When evaluating trade-offs, it is not only differences in performance metrics that should be considered - the impacts on different groups of candidates and recipients (both positive and negative) also need to be considered. The acceptability of a trade-off is subjective (however there will be methods of quantifying acceptable trade-offs discussed in chapter 5), in general, the question to be answered is: *does the benefit to group 'X' outweigh the negative impact on group(s) 'Y'?* The general principal is to **maximise benefit to one group** while **minimising negative effects on other group(s)**.

4.4.6 Is Lung Allocation Really Zero-Sum?

Lung allocation (and organ-allocation in general) can be seen as zero-sum, that is to say: one person's gain is another's loss. Allocating lungs to one candidate (or in the case of the SLT algorithm; two candidates) means that other candidates will not receive a transplant. However, this is not strictly true - take the following example, at time $T = 0$:

1. Candidate A: Waiting list survival duration: 30 days
2. Candidate B: Waiting list survival duration: 45 days
3. Candidate C: Waiting list survival duration: 60 days
4. Candidate D: Waiting list survival duration: 90 days

Let's now assume that compatible donors will arrive at times $T = 25$, $T = 40$, $T = 55$, and $T = 80$. If the first donor is allocated to Candidate A, the second donor to Candidate B, and so on, then all four candidates will receive a transplant without dying on the waiting list. However, if the donors are allocated in the reverse sequence, first to Candidate D, then to Candidate C and so on, then Candidate A and Candidate B will have died before

the offers at $T = 55$ and $T = 80$ become available. In both scenarios the waiting list and availability of donors are identical, the only difference is the **allocation policy**. This simple example demonstrates two principles:

1. One candidate's gain is not necessarily another candidate's loss
2. The degree to which allocation can be considered zero-sum is determined by the allocation policy

The definition of a candidate's 'gain' has been loosely defined: if 'gain' is interpreted as 'receiving a transplant', it is not necessarily true that just because one candidate received a transplant that other candidates can't receive a transplant. If 'gain' is interpreted as 'net benefit', then just because one recipient gains (for example) 2000 additional days of life doesn't mean that another recipient loses 2000 days of net benefit.

In reality, the system is more complicated than the example shown since candidates can be added or removed from the waiting list, compatibility between donors and recipients varies, there is risk of primary graft dysfunction (PGD), the exact survival duration on the waiting list is unknown, and the availability and characteristics of donors varies with time. However, it can still be argued that even with these considerations lung allocation is still not strictly zero-sum.

4.4.7 Clinical Applicability

One limitation of this work is that the results in this section are not directly clinically applicable. The datasets used for generating these results spanned 20 years (from 2002 to 2022), and there have been a number of changes in allocation and practice in that time (see section 2.4). Indications for lung transplant have changed, alternative treatments have become available, and from 2020-2022 Covid-19 has had a large impact on lung transplantation. For real-world clinical implementation a more recent cohort of candidates and recipients should be used for survival modelling.

There is also another limitation in the datasets themselves: there are no indicators of functional status, and crucial variables such as use of a ventilator or ECMO have a high proportion of missing entries in the dataset (approximately 50%). To include these variables in the allocation score, first data would have to be collected and made available on functional status for survival modelling. Next, there would also need to be institutional changes put in place to ensure that variables that are vital for predicting survival are being consistently and correctly input at the time of registration and transplant to ensure high quality data are available for research.

4.5 Conclusion

Overall, waiting list mortality trends in the opposite direction of the benefit metrics and post-transplant survival rates. Increasing one metric means having to compromise on another.

Tables 4.50 and 4.51 summarise the costs and benefits of changing the priority of waiting list and post-transplant survival. The tables are interpreted as follows: the current policy is identified in the leftmost column, and the alternative policy is identified in the topmost row. The cell where the the row and column intersect is split into two sections labelled 'Benefits' and 'Costs'.

For example, if the current policy was the 2:1 WL:PTX policy and it was replaced with the PTX policy, post-transplant survival at 1 year would increase by 3% and at 5 years 5.6%, average net benefit would increase by 222 days and waiting times would decrease by 197 days. This would come at the cost of 31 additional waiting list deaths per year.

The table can also be read in the opposite direction, so if the PTX policy was currently in place and was to be replaced with the 2:1 WL:PTX policy, the benefits would become costs and vice versa, and the $+/-$ signs would flip. Using the previous example, there would be 31 fewer waiting list deaths per year, at the cost of post-transplant survival decreasing by 3% and 5.6% at one and five years, net benefit decreasing by 222 days and waiting times increasing by 197 days on average.

The single-lung transplant policies showed great potential to decrease waiting list mortality, however this needs to be weighed up against the reduction in net benefit and post-transplant survival. One major limitation when simulating these policies is that left/right lung preferences were not modelled. In reality this would reduce the number of viable candidates for a single lung transplant, resulting in more frequent bilateral lung transplants and a lower reduction in waiting list deaths. Despite this, the ability of the SLT policies to have an effective impact of increasing the size of the donor pool has been clearly demonstrated.

The key question that needs to be answered is: *is the reduction in benefit and post-transplant outcomes justified by the decrease in waiting list deaths?*

This question needs to be answered not only at a population level, but for each sub-population of lung transplant candidates and recipients, as each group is impacted differently by changes to allocation policy.

Finally, if the answer to the above question is "yes", the next question that must be answered is: *which ratio of prioritising waiting list survival to post-transplant survival represents the optimal trade-off between minimising waiting list mortality and maximising benefit to recipients?*

Techniques for deciding on optimal trade-offs exist, and are the subject of the next chapter.

Allocation Policy Costs and Benefits

Table 4.50: Summary of trade-offs between pairs of policies. The 'benefits' and 'costs' rows show the positive and negative consequences of replacing the policy in the leftmost column ('original policy') with the policy in the topmost row ('new policy'). Acronyms: WLD = Annual Waiting List Deaths, NB = Net Benefit, PTX-1 = 1-Year Post-transplant Survival Rate, PTX-5 = 5-Year Post-transplant Survival Rate, WT = Waiting Time.

Original Policy ↓	New Policy →	WL	2:1 WL:PTX	1:1 WL:PTX	1:2 WL:PTX	PTX
NHS-BT	Benefits	WLD -44 NB +405	WLD -44 NB +467 PTX-1 +0.3 PTX-5 +0.6	WLD -43 NB +543 PTX-1 +0.8 PTX-5 +1.7	WLD -39 NB +626 PTX-1 +1.6 PTX-5 +3.1	WLD -13 NB +689 PTX-1 +3.3 PTX-5 +6.2 WT -73
	Costs	PTX-1 -0.2 PTX-5 -0.3 WT +128	WT +124	WT +122	WT +92	
WL	Benefits		NB +62 PTX-1 +0.5 PTX-5 +0.9 WT -4	NB +138 PTX-1 +1 PTX-5 +2 WT -6	NB +221 PTX-1 +1.8 PTX-5 +3.4 WT -36	NB +284 PTX-1 +3.5 PTX-5 +6.5 WT -201
	Costs			WLD +1	WLD +5	WLD +31
2:1 WL:PTX	Benefits			NB +76 PTX-1 +0.5 PTX-5 +1.1 WT -2	NB +159 PTX-1 +1.3 PTX-5 +2.5 WT -32	NB +222 PTX-1 +3 PTX-5 +5.6 WT -197
	Costs			WLD +1	WLD +5	WLD +31
1:1 WL:PTX	Benefits				NB +83 PTX-1 +0.8 PTX-5 +1.4 WT -30	NB +146 PTX-1 +2.5 PTX-5 +4.5 WT -195
	Costs				WLD +4	WLD +30
1:2 WL:PTX	Benefits					NB +63 PTX-1 +1.7 PTX-5 +3.1 WT -165
	Costs					WLD +26

Allocation Policy Costs and Benefits

Table 4.51: Summary of trade-offs between pairs of policies, where 'SLT' indicates policies that attempt to utilise single lung transplant for recipients with interstitial lung disease (ILD) in cases where this results in greater overall net benefit, compared to transplanting a single recipient with a lung pair. The 'benefits' and 'costs' rows show the positive and negative consequences of replacing the policy in the leftmost column ('original policy') with the policy in the topmost row ('new policy'). Acronyms: WLD = Annual Waiting List Deaths, NB = Net Benefit, PTX-1 = 1-Year Post-transplant Survival Rate, PTX-5 = 5-Year Post-transplant Survival Rate, WT = Waiting Time.

Original Policy ↓	New Policy →	SLT-WL	SLT-2:1	SLT-1:1	SLT-1:2	SLT-PTX
NHS-BT	Benefits	WLD -57 NB +267	WLD -58 NB +322	WLD -57 NB +364	WLD -59 NB +332 WT -43	WLD -56 NB +240 WT -152
	Costs	PTX-1 -1.4 PTX-5 -2.1 WT +19	PTX-1 -0.9 PTX-5 -1.3 WT +9	PTX-1 -0.6 PTX-5 -0.7 WT +7	PTX-1 -0.9 PTX-5 -1.2	PTX-1 -1.3 PTX-5 -2
SLT-WL	Benefits		WLD -1 NB +55 PTX-1 +0.5 PTX-5 +0.8 WT -10	NB +97 PTX-1 +0.8 PTX-5 +1.4 WT -12	WLD -2 NB +65 PTX-1 +0.5 PTX-5 +0.9 WT -62	PTX-1 +0.1 PTX-5 +0.1 WT -171
	Costs					WLD +1 NB -27
SLT-2:1	Benefits			NB +42 PTX-1 +0.3 PTX-5 +0.6 WT -2	WLD -1 NB +10 WT -52	WT -161
	Costs			WLD +1		WLD +2 NB -82 PTX-1 -0.4 PTX-5 -0.7
SLT-1:1	Benefits				WLD -2 WT -50	WT -159
	Costs				NB -32 PTX-1 -0.3 PTX-5 -0.5	WLD +1 NB -124 PTX-1 -0.7 PTX-5 -1.3
SLT-1:2	Benefits					WT -109
	Costs					WLD +3 NB -92 PTX-1 -0.4 PTX-5 -0.8

Chapter 5

Allocation Policy Selection Using the Analytic Hierarchy Process (AHP)

5.1 Background

Chapter 4 demonstrated that there are several goals in any lung allocation policy that could be optimised: reducing waiting list mortality, increasing net/relative benefit, increasing post-transplant survival and reducing mean waiting time. However, these goals are potentially conflicting, meaning that designing an allocation policy to maximise one goal means having to compromise on one or several other goals.

The next question that must be answered is which goals are most important? This is a subjective decision that will likely differ from person to person. There may be differences in opinion between candidates, recipients, and clinicians, or even differences within each group of candidates, recipients, and clinicians. Candidates on the waiting list and patients that have received a transplant may have differing opinions, there could also be differences between candidates, recipients, and their family members. Within the group of clinicians there may be disagreement between cardiothoracic surgeons, physicians, nurses and other healthcare professionals.

The goals that are prioritised will be specific to the population the allocation system is being designed for. Priorities are likely to differ between countries due to societal differences and impacted by the prevalence and quality of private or public healthcare available in each country.

In order to select an 'optimal' lung allocation policy it is vital to understand the goals and values across the entire lung transplant community. In this chapter a process will be demonstrated which can quantify the subjective decisions as to which allocation goal(s) are judged to be most important. This process is called the Analytic Hierarchy Process[36] and has been utilised in many industries for many different purposes (see the review in appendix G.2).

The AHP is a more advanced version of the pairwise comparison techniques used in chapter 3. For CPAT the participants completed a pairwise comparison matrix containing entries that decided *which* of each pair of criteria were more important. The AHP takes this a step further by also having each participant specify *how much more important* they feel one criterion is than another. The AHP also has methods to ensure the responses received are *consistent*, or in other words; make logical sense, have not been filled at random, and do not contain logical contradictions.

In practice, the AHP can be utilised by surveying key stakeholders (in this case, candidates, recipients, and professionals involved in the care of patients undergoing lung transplantation), and having them compare allocation goals such as 'reduce waiting list deaths', 'increase post-transplant survival' and 'increase net benefit'. For each pair of goals, the participant decides which (if any) they feel is more important, and specifies the relative importance using a verbal and/or numeric scale. The survey responses can then be converted to numeric entries in a pairwise comparison matrix and the resulting weight of each goal (a numeric value between 0 and 1, with higher values indicating greater

importance) can then be determined.

There are approaches other than pairwise comparisons that could be used, for example participants could be asked directly their percentage preference for each of the goals. There are a few issues with this approach. The first is that there is no mechanism to check for consistency, for example if a patient's preferences are split 20%/25%/55% between three options, there is no way to know if they just chose a random option, whether they understood the question or if there were any inconsistencies in their reasoning for choosing those values. The other problem is the general difficulty of answering a question of this style. It is much easier to say option 'A' is much more important than option 'B', or options 'C' and 'D' are equally important, compared to trying to decide on a percentage preference for one of the options, essentially giving 100 possible responses for an option.

Another possible approach is rank-ordering multiple options. This style of question is more intuitive than selecting a percentage, especially when there are only a few options to rank, however the weight of each preference can not be determined.

The AHP isn't the only option to calculate weights from pairwise comparisons, there are options such as the probabilistic Bradley-Terry model[149] and Borda count.[150] However, the AHP has been much more widely used, has a lot of literature supporting it (see review in G.2), provides an entire framework for decision making, identifies responses that are low quality, and is relatively straightforward to implement computationally.

The overall goal of this chapter (and thesis more generally), is to combine the weighted goals with the relative performance of each simulated policy, in order to identify which policies are most desirable (or 'optimal') for each stakeholder. It must be emphasised that *these are general-purpose techniques being applied specifically to UK lung allocation*. It is possible to 'pivot' these techniques to other domains to assist with clinical decision making in other contexts (as illustrated in chapter 3 with COVID-19).

5.2 Methods

5.2.1 The Analytic Hierarchy Process

The AHP was first proposed by Thomas L. Saaty in the 1970's.[36] The AHP captures the three principles of problem solving:[151]

1. Decomposition - achieved by starting with a goal and repeatedly breaking the goal into criteria and sub-criteria. This process is continued until all factors relating to the overall goal have been included.

2. Comparative Judgements - achieved via a pairwise comparison process. This process is repeated for every level in the hierarchy. Within each level, every pair of criteria are compared and decision makers are asked: "When making a decision (with respect to the goal) which of the two criteria is more important, and by how much?"

3. Synthesis of Priorities - achieved by combining (1) and (2) to calculate **weights** of each priority

Comparative judgements use a verbal scale that corresponds to a numeric scale,[152] and is shown in table refsaatyverbalscale. Intermediate values can be chosen, for example '4' would correspond to 'moderate to strong importance'.

Table 5.1: Saaty's verbal scale for comparing pairs of criteria and corresponding numeric values.[152]

Value	Definition
1	Equal Importance
3	Moderate importance of one over the other
5	Strong importance of one over the other
7	Very strong importance of one over the other
9	Extreme importance of one over the other

The results of the comparisons are recorded using a pairwise comparison matrix. Each entry in the matrix contains the judgement of how much more (or less) important the criterion in the row is compared to the criterion in the column. A detailed mathematical explanation of how to calculate weights from a pairwise comparison matrix and how to ensure *consistency* in comparative judgements is given in appendix H.

5.2.2 Application of the AHP to Evaluating Lung Transplant Patient Preferences 💻

A range of allocation policies were simulated and evaluated in chapter 4 with varying priorities of waiting list and post-transplant survival, and with the option of prioritising single-lung transplant for ILD (group D) recipients. These simulation results (see results starting on page 122) illustrate the key trade-offs between different allocation policies, however, the choice of which is considered to be the 'best' policy is a subjective decision.

The AHP allows many subjective comparisons to be made by multiple individuals, and then outputs *objective* measures (i.e., weightings) of the relative importance of multiple criteria (e.g., the relative importance of reducing waiting list mortality compared to increasing post-transplant survival). This makes it possible to survey lung transplant patients (both recipients and those on the waiting list), their family members and also family members who have lost a relative while waiting for a transplant. This will allow for a range of perspectives to be collected.

Allocation Goals Survey Design The first step taken to design the survey was to decide which criteria will be compared. The number of survey questions (q) scales with the number of criteria (n): $q = \frac{n(n-1)}{2}$. There were 5 metrics used for evaluating allocation policies which would result in 10 comparisons, however two of the metrics - net benefit and relative benefit - are very similar measures. Informal discussions with clinicians, statisticians and the CTAG patient group chair revealed that the concept of 'net benefit' was more easily understood than 'relative benefit'. The survey can be simplified by excluding relative benefit, reducing the number of criteria to 4 and only requiring 6 questions. The four criteria to be used for the survey would then be:

1. Reduce waiting list mortality
2. Increase benefit to patients
3. Increase survival duration after transplant
4. Reduce waiting time for transplant

The first question of the survey required participants to rank these four goals from highest to lowest priority by rearranging the four goals on screen. Although this approach can't be used to calculate weights, it is a helpful 'sense check' to compare the directly ranked goals against the (indirectly) calculated weights. This question was also placed first with the intention being to start the participant thinking about the relative importance of each goal before moving on to complete the pairwise comparisons.

The next step in the survey design was to present the options clearly for use with the AHP. The AHP uses a 1 to 9 scale to compare each pair of criteria, which extends in both directions and would result in 17 options per question. The verbal scale that

Saaty gives only assigns descriptions to the values 1, 3, 5, 7 and 9 on the scale, this would reduce the number of options to 9 per question, however this can be reduced further. Each comparison can consist of two questions, the first would ask which of the two criteria is more important, or if they are of equal importance, resulting in 3 options. If the criteria are of equal importance then the second question can be skipped, otherwise 4 options from the verbal scale: 3, 5, 7 and 9 corresponding to 'moderate', 'strong', 'very strong' and 'extreme' relative importance can be presented, resulting in a maximum of 7 options per question.

As part of a service evaluation exercise, the surveys were circulated via the CTAG patient involvement group via email and social media, and internally within NHS-BT. This was a voluntary survey where individuals could participate if they chose to and clicked the link directing them to the survey hosted by Microsoft Forms.[153] A range of perspectives from clinicians, candidates on the waiting list, lung transplant recipients, and family members of candidates requiring lung transplant were collected. Results were summarised at a population level and also by demographic (for example, how do preferences for candidates on the waiting list differ from those who have received a transplant?) To aggregate multiple results into a decision on the most preferable policy there were two options:

1. Calculate the geometric mean across all comparison matrices, use the resulting comparison matrix to calculate weights for each of the four goals to select a policy

2. Calculate weights for each individual, then determine the most desirable policy at an individual level and count this as a "vote". Track the number of votes each policy receives from each demographic.

The major shortcoming with option 1 is that opposing opinions cancel each other out, for example the geometric mean of 9 and $\frac{1}{9}$ is 1. If the geometric mean is used in cases like this, the conclusion would be that the two criteria were of equal importance, rather than showing there is a split in opinion. In the same case, option 2 would count one "vote" for the policy that agrees with one side of the split in opinion, and one vote for the other, giving a more representative measure of opinion. For these reasons, option 2 was decided as the best approach for this use case.

The survey for candidates/recipients/family members is included in appendix I, and for clinicians in appendix J.

5.2.3 Visualising Survey Results with Ternary Diagrams 💻

Ternary diagrams are a useful way to visualise 3-dimensional data in 2 dimensions[154] and in this case can be used to visualise the preferences of clinicians, candidates and recipients with respect to allocation goals. Each survey response can be plotted on the diagram allowing a visual comparison of the similarities and differences in opinion between patients (i.e., candidates and recipients) and clinicians (for an example of this, see figure 5.2). Ternary diagrams can be used providing two conditions hold:

1. The measurement along each dimension is normalised into the 0 to 100 range
2. Any valid point in the 3-dimensional space **must** result in the sum of values along all three axes summing to 100

Weights generated from the AHP are suitable for this purpose, as they sum to 1 so can easily be scaled into the 0 to 100 range, thus meeting both of the above conditions.

One limitation is that only 3 variables can be visualised at once. However, of the 4 criteria used for the AHP survey, 'waiting time' is the only one not relating to the survival of lung transplant patients. The three remaining criteria relate to (1) survival on the waiting list, (2) additional survival from transplant and (3) survival after transplant. These will be the three criteria visualised using ternary diagrams.

Each point on the ternary diagram can be written as a vector, where WL is the weight given to reducing waiting list mortality, NB is the weight given to increasing net benefit, and PTX is the weight given to maximising 1-year post-transplant survival. As a result of how ternary diagrams are constructed and interpreted, the property $WL + NB + PTX = 100$ holds true for all points on the diagram:

$$\begin{pmatrix} WL \\ NB \\ PTX \end{pmatrix}$$

Next, the metrics for each policy are normalised in the range 0 - 1, with the number of waiting list deaths being multiplied by -1, as lower values equate to more desirable policies. Let $P_{n,m}$ refer to the nth policy and metric m, $\min(P_{.,m})$ refer to the minimum value of metric m on any of the policies, and $\max(P_{.,m})$ refer to the maximum value of metric m on any of the policies. The results are all normalised using the formula:

$$|P_{n,m}| = \frac{P_{n,m} - \min(P_{.,m})}{\max(P_{.,m}) - \min(P_{.,m})}$$

This results in there always being at least one policy where $|P_{.,m}| = 0$ and $|P_{.,m}| = 1$ for each metric, corresponding respectively to the worst and best performing policies on that metric.

The simulation results can then be represented as a matrix:

$$\begin{pmatrix} P_{1,WL} & P_{1,NB} & P_{1,PTX} \\ P_{2,WL} & P_{2,NB} & P_{2,PTX} \\ P_{3,WL} & P_{3,NB} & P_{3,PTX} \\ P_{..,WL} & P_{..,NB} & P_{..,PTX} \\ P_{n,WL} & P_{n,NB} & P_{n,PTX} \end{pmatrix}$$

To assign a colour to each point on the ternary diagram, the overall performance score is calculated using the matrix-vector product of the results matrix and the vector of ternary diagram weights:

$$\text{Performance Score} = \begin{pmatrix} P_{1,WL} & P_{1,NB} & P_{1,PTX} \\ P_{2,WL} & P_{2,NB} & P_{2,PTX} \\ P_{3,WL} & P_{3,NB} & P_{3,PTX} \\ P_{..,WL} & P_{..,NB} & P_{..,PTX} \\ P_{n,WL} & P_{n,NB} & P_{n,PTX} \end{pmatrix} \begin{pmatrix} WL \\ NB \\ PTX \end{pmatrix}$$

Finally, the index for the policy with the highest performance score is mapped to a colour for that point on the ternary diagram. As this process is repeated, the areas corresponding to the most desirable policy given a range of priority weights for each metric can be visualised, this can be seen in figure 5.1.

5.3 Survey Results

This section presents a summary of the survey results, identifies the policies that most closely align with the survey results and calculates the trade-offs between those policies.

5.3.1 Demographics

Two surveys were circulated to collect transplant candidate, recipient and family member opinions (n = 100) and clinician opinions (n = 62). The demographics of candidate, recipient and family member respondents are shown in table 5.2 and clinical respondents in table 5.3.

Table 5.2: Number and percentage of allocation goal survey participants by demographic category.

Category	Number (%)
Candidate on the active waiting list	7 (7%)
Family member of candidate on active waiting list	1 (1%)
Lung transplant recipient	74 (74%)
Family member of lung transplant recipient	7 (7%)
Family member of a candidate who was on the waiting list, but passed away before receiving a transplant	1 (1%)
Other	10 (10%)

Table 5.3: Number of percentage of allocation goal survey clinical participants by demographic category.

Category	Number (%)
Transplant Surgeon	12 (19.4%)
Transplant Physician	14 (22.6%)
Transplant Recipient Co-ordinator	14 (22.6%)
Specialist Nurse in Organ Donation	9 (14.5%)
Transplant Nurse / Nurse Practitioner	3 (4.8%)
Governance/Administration/Policymaker	2 (3.2%)
Other	8 (12.9%)

5.3.2 Results of Question 1: Ranking Allocation Goals

The results of the four allocation goals that were ranked by candidates/recipients/family members and clinicians are shown in tables 5.4 and 5.5 respectively. The tables show the percentage of respondents that placed each goal as first, second, third, or fourth choice, with first being highest priority and fourth lowest.

Table 5.4: Percentage of candidates, recipients and their family members (n=100) that ranked the four allocation goals in each position. First choice = highest priority, fourth choice = lowest priority.

Goal	Average Rank	% 1^{st} Choice	% 2^{nd} Choice	% 3^{rd} Choice	% 4^{th} Choice
Reduce waiting list deaths	1	38.0%	26.1%	20.7%	15.2%
Increase post-transplant survival	2	34.8%	25.0%	19.6%	20.7%
Increase net benefit	3	10.9%	29.3%	31.5%	28.3%
Decrease waiting time	4	16.3%	19.6%	28.3%	35.9%

Table 5.5: Percentage of clinicians (n=62) that ranked the four allocation goals in each position. First choice = highest priority, fourth choice = lowest priority.

Goal	Average Rank	% 1^{st} Choice	% 2^{nd} Choice	% 3^{rd} Choice	% 4^{th} Choice
Reduce waiting list deaths	1	37.7%	31.1%	18%	13.1%
Increase post-transplant survival	2	29.5%	19.7%	31.1%	19.7%
Increase net benefit	3	23%	21.3%	31.1%	24.6%
Decrease waiting time	4	9.8%	27.9%	19.7%	42.6%

5.3.3 Pareto Set of Simulated Policies 💻

In section 2.2.3 (on page 25) the concept of the 'Pareto set' was introduced, and this same concept can be applied to the results of the simulated policies. In the case of the survey, ternary diagrams were generated for the three metrics related to survival: waiting list survival, post-transplant survival and net benefit. Table 5.6 shows the simulated results for the initial five policies (page 122) and SLT policies from section 4.3.5 (page 124), along with the three key metrics (note: 1-year post-transplant survival or 5-year post-transplant survival can be used in this process, the results are identical). Four of the policies were excluded from the Pareto set due to at least one policy performing equally or better on all metrics.

Table 5.6 shows that out of the five initial policies (WL, .., PTX) only four need to be considered. Three of the five SLT policies were eliminated, resulting in only two SLT policies and overall 6 out of the 10 remaining to be considered.

5.3.4 Ternary Diagrams 💻

Each point on a ternary diagram represents a potential combination of preferences; in this case the importance of reducing waiting list deaths, increasing net benefit, and increasing post-transplant survival. The ternary diagrams in this section are colour-coded to show which policy (according to its relative performance compared to the other simulated policies) most closely aligns with the three preferences at every point. The colour-coding of the diagram is generated as a result of the simulations, and individual points that are

Table 5.6: The standard five policies and SLT policies and their corresponding survival metrics. Four policies can be removed from consideration as there are other policies that perform better on all metrics.

Policy	Annual Waiting List Deaths	Net Benefit (Days)	1-Yr PTX Survival (%)	Eliminated
WL	46	2238	80.00	X
2:1 WL:PTX	46	2300	80.50	
1:1 WL:PTX	47	2376	81.00	
1:2 WL:PTX	51	2459	81.80	
PTX	77	2522	83.50	
SLT-WL	33	2100	78.80	X
SLT-2:1	32	2155	79.30	X
SLT-1:1	33	2197	79.60	
SLT-1:2	31	2165	79.30	
SLT-PTX	34	2073	78.90	X

plotted on the diagram correspond to individual responses to the allocation goals survey, thus combining the results of the simulations and AHP in a single diagram.

The sum of the preferences must always equal 100%, this adds additional constraints to the range of possible values for the preferences: each apex of the ternary diagram represent 100% preference being assigned, but has an area of the ternary diagram of zero. This results in the 2:1 and 1:1 WL:PTX policies being eliminated from the Pareto set. The four policies remaining for consideration were: 1:2 WL:PTX, PTX, SLT-1:1 and SLT-1:2.

Figure 5.1 shows the ternary diagram with each area coloured according to which policy aligns with the preferences at each point on the diagram.

The ternary diagram in figure 5.2 shows the preferences for every candidate/recipient who responded to the survey. The responses tended to cluster near decreasing waiting list deaths (SLT-1:2) and increasing post-transplant survival (PTX).

Figure 5.3 shows the preferences for every clinician who responded to the survey. The clinician responses were more equally distributed than candidate/recipient responses, with clusters near each corner of the diagram, demonstrating a wide spread in opinion. When the skew in candidate/recipient demographics is considered (that is, 75% of respondents had already received a transplant compared to the 7% on the active waiting list), this may explain why there is a more even spread among clinicians by comparison.

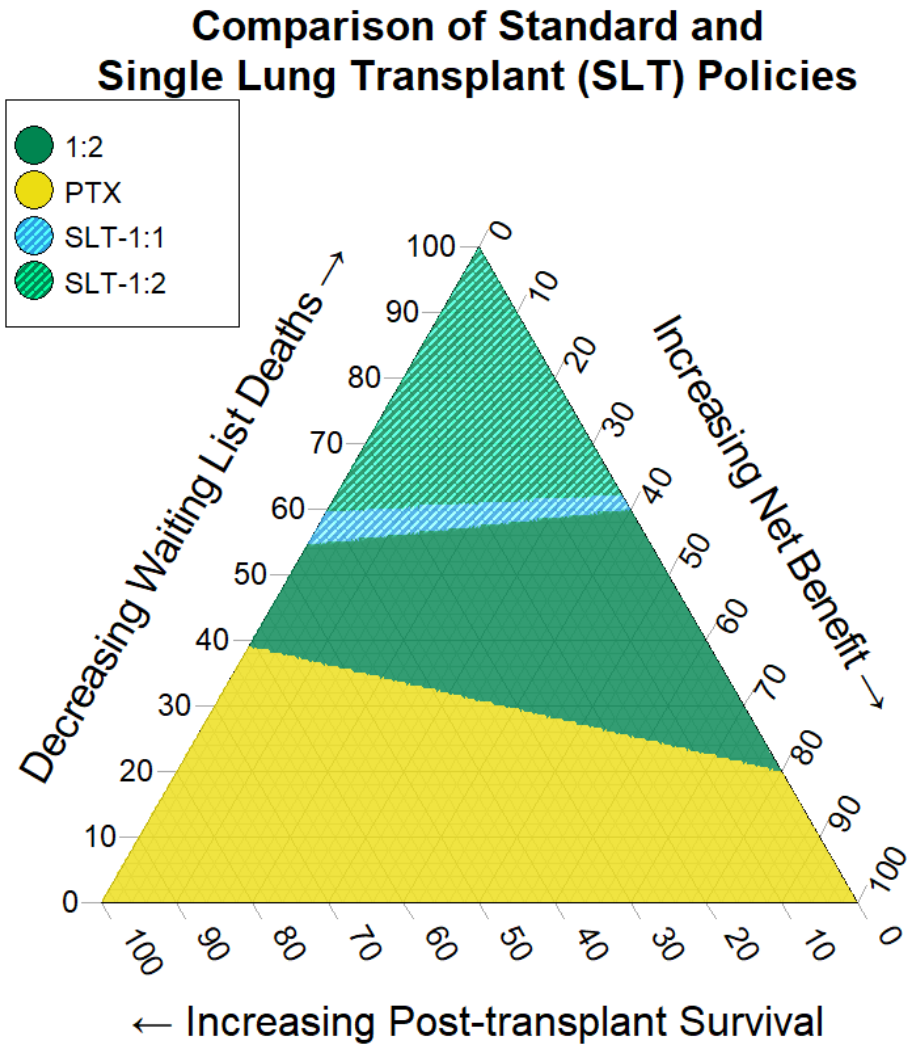


Figure 5.1: Ternary diagram showing which policy (colour coded) most closely aligns with preferences at each point. Each point on the diagram sums to 100%, with the percentage preference of each of the three goals shown on each axis (decreasing waiting list deaths, increasing net benefit, increasing post-transplant survival). Arrows for each goal point in the direction of increasing preference, with 100% corresponding to the highest preference and 0% to lowest.

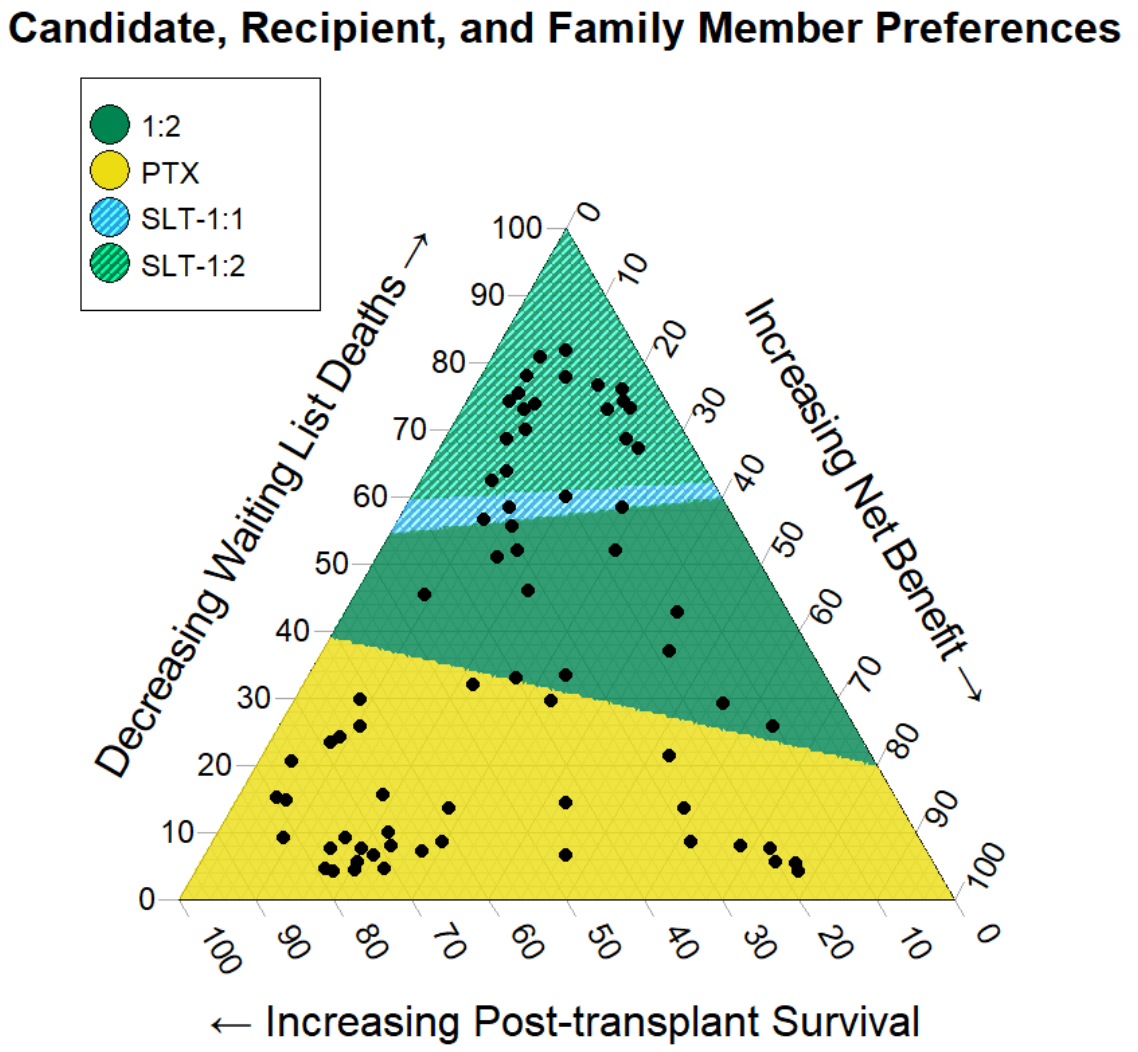


Figure 5.2: Ternary diagram showing distribution of candidate, recipient, and family member preferences relating to lung allocation goals. The survey results for each respondent are plotted with a solid circle. Each response corresponds to the percentage preference of each of the three goals shown on each axis (decreasing waiting list deaths, increasing net benefit, increasing post-transplant survival). Arrows for each goal point in the direction of increasing preference, with 100% corresponding to the highest preference and 0% to lowest.

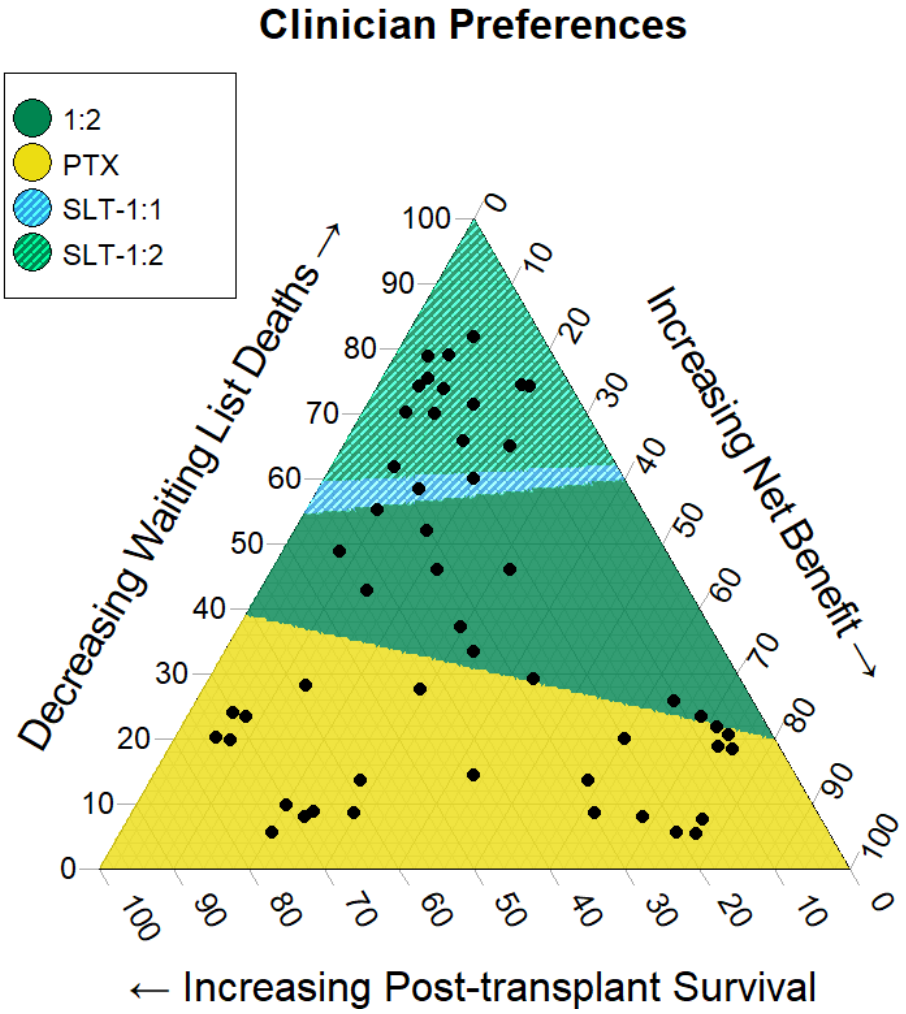


Figure 5.3: Ternary diagram showing distribution of clinician preferences relating to lung allocation goals. The survey results for each clinician respondent are plotted with a solid circle. Each response corresponds to the percentage preference of each of the three goals shown on each axis (decreasing waiting list deaths, increasing net benefit, increasing post-transplant survival). Arrows for each goal point in the direction of increasing preference, with 100% corresponding to the highest preference and 0% to lowest.

5.3.5 Survey Results

Table 5.7 shows the number of candidate/recipient/family member preferences and clinician preferences that aligned with each of the simulated policies (i.e., the number of individual points that fall within each coloured region of the ternary diagram).

Table 5.7: Number of candidate/recipient/family member opinions and clinician opinions aligning with each of the simulated policies.

Policy	Number of Aligned Candidate/Recipient Preferences (%)	Number of Aligned Clinician Preferences (%)	Total (%)
1:2	15 (15%)	13 (21.0%)	28 (17.3%)
PTX	52 (52%)	30 (48.4%)	82 (50.6%)
SLT-1:1	4 (4%)	4 (6.5%)	8 (4.9%)
SLT-1:2	29 (29%)	15 (24.1%)	44 (27.2%)

The policy that aligned with the greatest number of overall preferences was the PTX policy. However, it is important to note that the majority of candidate/recipient respondents were lung transplant recipients (74%) and relatively few were on the active waiting list (7%). It was expected that this skew in the demographics would also result in a bias/skew in the results generated from this survey, however this may not have been the case and will be discussed later in this section.

It is interesting to note the similarity in percentages comparing candidate/recipient preferences to clinician preferences. There was a larger divide of opinion *within* groups than *between* groups. The main differences were that a slightly larger percentage of clinicians aligned with the 1:2 policy compared to candidate/recipients, and a lower percentage of clinicians aligned with the SLT-1:2 policy.

5.3.6 Trade-offs

The trade-off tables starting on page 168 can be condensed into a single table shown in figure 5.8. This table shows the costs and benefits of replacing one policy with another. It is interesting to observe that the policy that aligned with the greatest number of survey participants (PTX) also outperformed the simulated existing policy (NHS-BT) on every metric. Comparing the PTX policy to the second most aligned policy (SLT-1:2) reveals the following trade-offs between the two policies (shown in **bold** in the table):

1. The SLT-1:2 policy would result in 46 fewer waiting list deaths per year.
2. The PTX policy would result in an additional 357 days of net benefit.
3. The PTX policy would result in a 4.2% increase in 1-year post-transplant survival rates and a 7.4% increase in 5-year post-transplant survival.
4. The PTX policy would reduce the mean waiting time by 30 days.

Table 5.8: Trade-offs between the four policies evaluated by the AHP survey technique and the simulated existing policy (NHS-BT). The results in **bold** show a comparison of the policy that aligned with the greatest number of survey participants (PTX) and the second-greatest number (SLT-1:2). Acronyms: WLD = Waiting List Deaths, NB = Net Benefit, PTX-1 = 1-Year Post-transplant Survival Rate, PTX-5 = 5-Year Post-transplant Survival Rate, WT = Waiting Time

Original Policy ↓	New Policy →	1:2 WL:PTX	PTX	SLT-1:1	SLT-1:2
NHS-BT	Benefits	WLD -39 NB +626 PTX-1 +1.6 PTX-5 +3.1	WLD -13 NB +689 PTX-1 +3.3 PTX-5 + 6.2 WT -73	WLD -57 NB +364	WLD -59 NB +332 WT -43
	Costs	WT +92		WT +7 PTX-1 -0.6 PTX-5 - 0.7	 PTX-1 -0.9 PTX-5 -1.2
1:2 WL:PTX	Benefits		NB +63 PTX-1 +1.7 PTX-5 +3.1 WT -165	WLD -18 WT -85	WLD -20 WT -135
	Costs		WLD +26	NB -262 PTX-1 -2.2 PTX-5 -3.8	NB -294 PTX-1 -2.5 PTX-5 -4.3
PTX	Benefits			WLD -44	**WLD -46**
	Costs			NB -325 PTX-1 -3.9 PTX-5 -6.9 WT +80	**NB -357** **PTX-1 -4.2** **PTX-5 -7.4** **WT +30**
SLT-1:1	Benefits				WLD -2 WT -50
	Costs				NB -32 PTX-1 -0.3 PTX-5 -0.5

5.4 Survey Results Discussion

The survey and analysis techniques used in this section demonstrate that there was no consensus in opinion on what the most important goals of a lung allocation should be. There was greater variance in opinion *within* the population of candidates/recipients and clinicians that were surveyed than *between* populations.

It is worth reviewing the significance of the results generated in this chapter by taking a high-level look at the process required to produce the results in table 5.8.

First, ten policies were simulated: five with a range of allocation priorities from WL to PTX (section 4.3.5) and an additional five where SLT was used for recipients with ILD (section 4.3.5).

Applying the concept of the Pareto set (section 2.2.3) resulted in narrowing down the ten policies to six. Ternary diagrams were then colour-coded to visualise which policy would be the most desirable at each point on the diagram; where every point on the diagram represents a possible combination of preferences of decreasing waiting list deaths, increasing post-transplant survival, and increasing net benefit (page 179). This further reduced the number of policies under consideration to four (two policies effectively had an area of zero on the ternary diagram).

The AHP was used to survey clinicians, transplant candidates, recipients and their family members (page 174), making it possible to plot every survey response on ternary diagrams. The number of responses in each colour-coded area were then totalled, showing that the PTX policy aligned with the greatest number of candidates/recipients and clinicians (52% and 48.4% respectively, 50.6% overall). The SLT-1:2 policy aligned with the second greatest number of candidates/recipients and clinicians (29% and 24.1% respectively, 27.2% overall). This suggests that a real-world implementation of these methods *may* want to implement a policy with a priority-ratio that prioritises post-transplant survival, such as the 1:2 WL:PTX or the PTX policy.

Finally, table 5.8 shows the trade-offs (i.e., costs and benefits) between the four remaining policies from this process, including the trade-offs between the first- and second-most 'popular' policies. Thus, this table **displays the main trade-offs between the policies that (1) align with the greatest number of candidate/recipient/family member preferences and clinician preferences and (2) are the highest performing on the survival metrics of interest.**

It will be necessary to survey additional candidates, recipients and family members with greater participation amongst those on the active waiting list. When visualising the difference in opinion between candidates on the active waiting list (figure 5.4) there was still a spread of opinion. Contrary to what was expected, opinion doesn't exclusively cluster around decreasing waiting list deaths, though opinion does appear to skew somewhat towards reducing waiting list deaths amongst candidates on the waiting list.

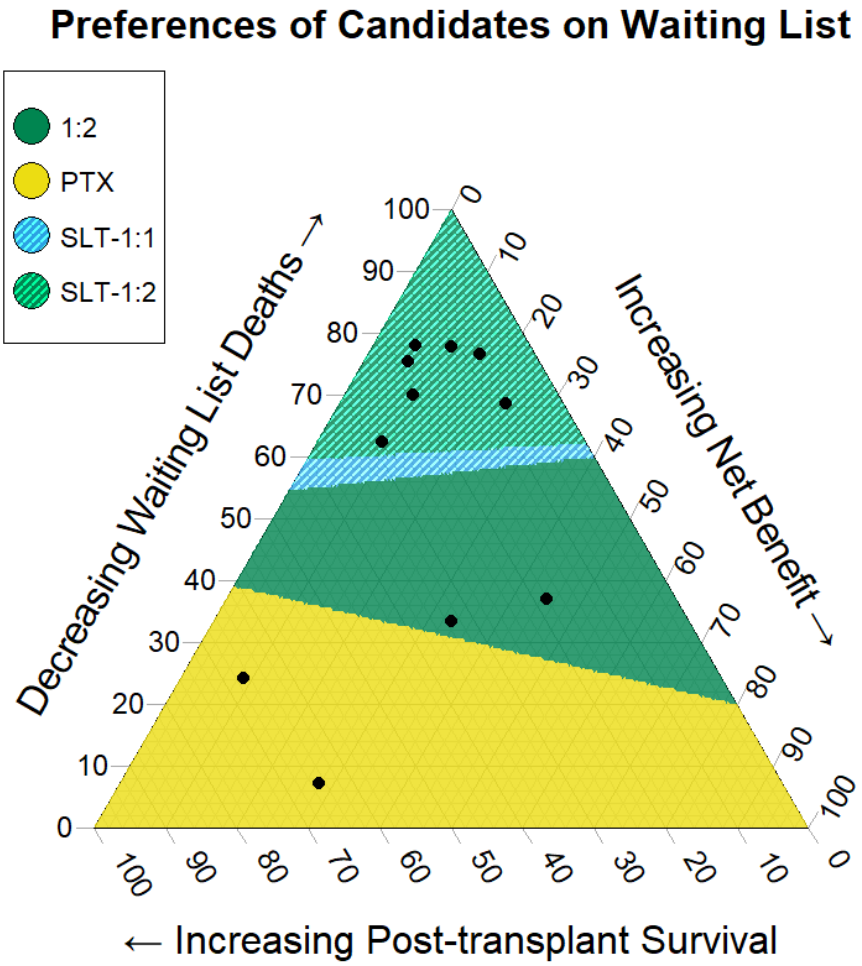


Figure 5.4: Visualisation of opinions for candidates on the active waiting list. The survey results for each respondent are plotted with a solid circle. Each response corresponds to the percentage preference of each of the three goals shown on each axis (decreasing waiting list deaths, increasing net benefit, increasing post-transplant survival). Arrows for each goal point in the direction of increasing preference, with 100% corresponding to the highest preference and 0% to lowest.

Participants were surveyed without being informed of the simulation results. When looking at the trade-off table in figure 5.8, despite the PTX policy aligning with the greatest number of preferences, when comparing this to the second most frequently selected policy (SLT-1:2), the reduction in waiting list deaths is very substantial. A reduction of 46 waiting list deaths per year may more than compensate for the small reduction in net benefit (-357 days), 1-year post-transplant survival (-4.2%) and 5-year post-transplant survival (-7.4%).

Any additional surveys should clearly explain the trade-offs between each of the policies, this could lead to a higher degree of consensus between candidates/recipients and clinicians, and help identify which trade-offs are considered reasonable, and thus identifying a policy that balances costs and benefits in line with candidate/recipient and clinician opinion.

Chapter 6

Conclusion and Next Steps

6.1 Summary

This section summarises the range of topics covered in the main body of this thesis.

First, the lung allocation policies in the US, UK, Europe, Australia, New Zealand and Scandinavian countries were reviewed. One common feature between all the allocation policies that were reviewed was that they used sequential allocation, which requires hard boundaries to be specified. The main problem with hard boundaries is that they can lead to candidates with similar levels of disease severity (and therefore similar risks), having different levels of access to transplantation. These problems were revisited in chapter 3, where a sequential allocation system to prioritise hospitalised COVID-19 patients was developed (CPAT). CPAT was developed in response to the COVID-19 pandemic; this allowed the development of methods for allocating ITU resources to patients affected by COVID-19. The methods developed for CPAT were a simplified, more limited version of the methods that were used for lung allocation in chapter 5. By the end of chapter 3 the limitations of the simplified methods became clear.

After the CPAT project concluded and a better understanding of the requirements and potential limitations of automated allocation systems had been established, focus returned to lung allocation. Next, the history of lung allocation policies was reviewed. Early policies allocated lungs according to waiting time, however this biased allocation towards candidates with less serious lung diseases. Over time, allocation systems were changed to take into account the clinical urgency of each candidate, with the most notable being the LAS (see page 45).

At a high level the concept of the LAS is quite straight forward: predict how long a candidate is expected to survive on the waiting list and with a transplant, with the 'net benefit' being the difference between these two predicted durations. Allocation can then be based on prioritising candidates with the highest expected net benefit. This concept was used to overcome the limitations of sequential allocation and was used throughout chapter 4 to develop a simulation engine to evaluate the impact of different priority-ratios of waiting list to post-transplant survival.

Although the concept of net benefit is straightforward, the introduction of priority-ratios (i.e., weights determining the relative importance of WL and PTX survival) and the use of SLT for recipients with ILD, the number of possible trade-offs grew rapidly. It then became necessary to use an appropriate method to determine which policy (out of the many possible options) was 'optimal' (i.e., aligns with candidate, recipient and clinician beliefs on what an optimal allocation system should achieve). The AHP was determined to be the appropriate method for this purpose.

Next, the AHP was reviewed in detail, which is a framework that can be used to assign weights to various criteria by surveying and aggregating expert opinion. In the literature there were examples of using the AHP to design an allocation policy directly, by assigning weights to various clinical variables. However, the goal of the work in this thesis pivoted to

using the AHP to evaluate the relative performance of each policy, rather than designing the policy directly.

Chapter 4 addressed the shortcomings of the sequential approach used in chapter 3, and investigated a number of potential allocation policies using a range of performance metrics. At the end of chapter 4 the key output of this work is summarised in the form of a number of trade-off tables, showing how changing priority from waiting list urgency to post-transplant survival affects different groups of candidates and recipients. The trade-off tables summarise results at a population level, and are also stratified by diagnosis group, age group and blood group.

Chapter 5 ties everything together by demonstrating how the AHP can be used to identify which allocation policy(s) most closely align with the goals and values of lung transplant candidates, recipients, and their family members.

6.2 Revisiting the Original Hypothesis, Goals, and Contributions

6.2.1 Hypothesis

Finally, let's revisit the original hypothesis in this thesis, stated in chapter 1.2 on page 11:

> Improvements to the current UK lung allocation policy can be made by use of survival analysis and simulation techniques. If the current UK lung allocation policy can be simulated and performance metrics measured, then improvements can be identified by using statistical techniques to compare the current and potential alternative policies. In addition to this, if the current allocation policy is sub-optimal with respect to the performance metrics of interest, then there should exist at least one alternative policy that performs better according to the metrics of interest.

This can be broken down into several logical statements that can each be proven or falsified individually:

A. The UK lung allocation policy can be simulated

B. Performance metrics can be measured from the simulated UK lung allocation policy

C. Improvements can be identified using statistical techniques

D. There exists at least one alternative policy that performs better according to the metrics of interest

E. The current allocation policy is sub-optimal

This can be represented in propositional logic as follows:

$$(A \wedge B \Rightarrow C) \wedge (D \Leftrightarrow E)$$

This is interpreted as: "If statements A and ($\wedge$) B are true, then ($\Rightarrow$) statement C should be true. If A and/or B are false, then nothing can be said about statement C, however, if statements A and B are true and statement C is false, then the hypothesis has been proven false. In addition ($\wedge$), if statement D is true then E should be true, or if D is false then E should be false and vice versa ($\Leftrightarrow$). If D is true and E is false or vice versa then the hypothesis has been proven false."

This means that if all statements A - E can be demonstrated to be true, then the hypothesis has been proven true.

Starting with statement A, section 4.2.2 (page 83) describes the approach taken to simulate the NHS-BT policy and section 4.2.8 (page 96) describes the methods used to

ensure the simulated results match what is observed in reality. Section 4.3.4 (page 112) shows a comparison of observed and simulated results, showing that the simulation engine successfully simulates the current NHS-BT policy, thus proving A to be **True**.

For statement B, section 4.2.6 (page 87) describes the performance metrics that were measured, and section 4.3.5 (page 122) shows tables of performance metrics recorded for each policy and scenario. Section 4.3.5 also contains results showing statistically significant differences in waiting list mortality between policies. This shows that B and C are also **True**.

Statement D can be proven **True** with the existence of one policy that performs better on one of the performance metrics. The first table in section 4.3.5 contains five policies that perform better than the NHS-BT policy on at least one metric. The SLT policies in section 4.3.5 also perform significantly better in terms of reducing waiting list deaths. These sections show that statement D is **True**, thus demonstrating that E is **True**, or in other words, the current NHS-BT policy is sub-optimal with respect to at least one performance metric. This also falsifies the null hypothesis (page 11): "There is no statistically significant difference in performance metrics between the current UK lung allocation policy and any alternative simulated policy."

Finally, the initial statement in the hypothesis is supported: "Improvements to the current UK lung allocation policy can be made by use of survival analysis and simulation techniques."

6.2.2 Goals

In section 1.2.3 (page 12) the goals for this thesis were outlined. Each goal will now be evaluated to assess the degree to which it was completed.

> **Goal 1:** Design and implement a simulation engine to predict the impact of different lung allocation policies according to specific performance metrics

This was completed successfully, the methods are described in section 4.2.7 and appendix D. A number of performance metrics were collected: annual waiting list deaths, average net benefit per recipient, average relative benefit per recipient, 1- and 5-year post-transplant survival rates, average waiting time for transplant, annual transplant volume and life-years gained and lost. These results are presented and discussed throughout section 4.3.

> **Goal 2:** Quantify the goals and values that the lung transplant community (i.e., patients, clinicians and other stakeholders) believe should be part of an ideal allocation policy

This was completed successfully, however follow-up surveys will be required to gather a wider and more representative sample of the lung transplant community. The methods

for designing the survey are described in section 5.2 and the results are presented in section 5.3.

> **Goal 3:** Using the results generated from goal (2), identify which potential policy most closely aligns with the goals and values of the lung transplant community

This was also completed successfully as described and presented in sections 5.2.3 and 5.3.5. As previously discussed, the survey was conducted with participants not being aware of the trade-offs between the various policies. An additional follow-up survey that presents trade-offs could be completed as future work.

> **Goal 4:** Using goal (1) and the results from goal (3), compare the current and proposed policies using performance metrics of interest

Table 5.8 on page 185 shows the trade-offs between the simulated NHS-BT policy and the four top performing policies. Interestingly, the PTX policy outperformed the NHS-BT policy on all performance metrics, and this was also the policy that aligned with the greatest number of survey participants.

> **Goal 5:** Ensure policies are equitable: all candidates should be prioritised based solely on the same clinical criteria, it should not be possible to unfairly influence candidate rankings, and rankings should not be skewed to benefit or disadvantage specific groups of candidates

This goal is somewhat more subjective than the others, but can still be evaluated. Candidates were prioritised using an allocation score that is derived from survival models for waiting list and post-transplant survival. As a result of this, all (simulated) candidates were prioritised using the same *solely clinical* criteria.

In terms of how easily the rankings could be manipulated: for both the waiting list and post-transplant survival models, all variables were observable/measurable, or would be recorded in a candidate's medical record, such as being diagnosed with diabetes, prior malignancy or requiring home oxygen.

> **Goal 6:** Ensure policies are auditable: it should be possible to justify allocation decisions and understand the exact reasoning that was undertaken at the time of allocation

This goal was completed for both CPAT and the lung allocation policies. For the CPAT system, audit logs were saved any time a prioritised list of patients was requested from the server. The audit log contained a snapshot of all patients and their condition (comorbidities, demographic data such as age and sex) along with the date and time the log was generated.

The design of the lung allocation policies themselves are auditable: if a snapshot of the waiting list was to be taken any time a match-run was performed, it would be possible to justify allocation decisions by calculating allocation scores for every candidate on the waiting list at the time of the snapshot. This is due to the policies being deterministic, and not probabilistic or relying on random processes as discussed in section 2.2 on page 20.

> **Goal 7:** Ensure policies are transparent to candidates: it should be clear to candidates how their position on the allocation rankings was determined

This goal was mostly completed. Though not shown in the main body of this thesis, it would be possible to visually show a candidate how their allocation score was calculated, an example of this is given in figure 6.1.

The graph would be interpreted from left-to-right: the first column shows which attributes contribute to increasing waiting list risk, with the size of the bar indicating the relative magnitude of the risk. The next column shows attributes that decrease waiting list risk. The next two columns show attributes that increase and decrease post-transplant risk (note that these include donor variables such as the lung(s) being donated, and the height difference between the donor and the candidate).

The last three bars show the total waiting list risk (i.e., total attributes decreasing risk subtracted from total attributes increasing risk), total post-transplant risk, and finally net benefit (the difference between waiting list risk and post-transplant risk). The scores are based on the coefficients from the Cox models: the coefficients have simply been multiplied by 100 and rounded down to the nearest integer. The units are intentionally abstract, since mapping scores to predicted survival durations on the waiting list and post-transplant, then showing these predictions to candidates could potentially cause undue distress.

The purpose of figure 6.1 is just to illustrate that the design of the allocation policy makes this visualisation possible, however, the visualisation of candidate scores was not a focus of this work, and no feedback has been received as to the clarity and ease of interpretation for lung transplant candidates. Future research could focus on designing a visualisation for candidates that is easy to understand.

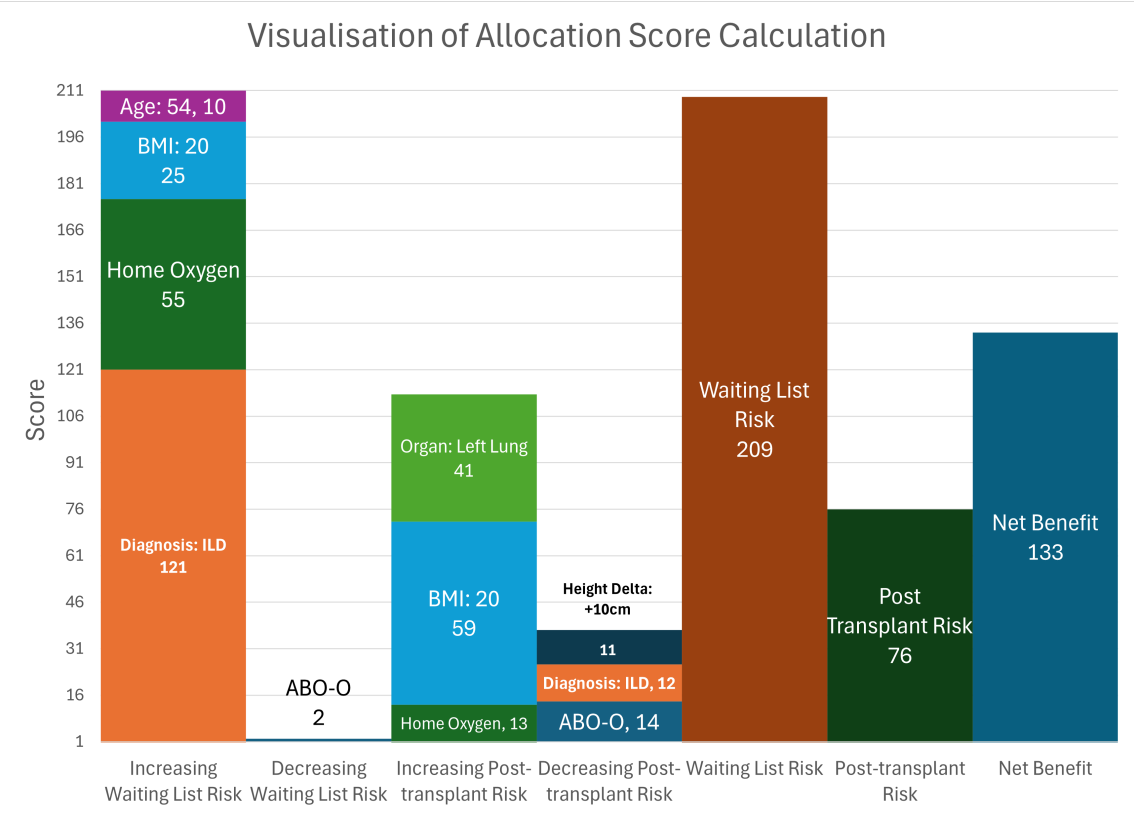


Figure 6.1: An example allocation score visualisation for a 54 year old candidate on home oxygen with ILD, a BMI of 20, blood type O, being offered a left lung from a donor that is 10cm taller. Columns from left-to-right are attributes that: increase risk of mortality on the waiting list, decrease risk on the waiting list, increase post-transplant risk, decrease post-transplant risk. The final three columns are: overall waiting list risk, overall post-transplant risk and net benefit.

6.2.3 Contributions

In section 1.2.4 (page 13) the potential contributions to the literature (and lung transplantation more specifically) were outlined. They are revisited here along with additional contributions that came about as a result of this work.

> **Contribution 1:** Development of predictive models for waiting list and post-transplant survival for the UK lung transplant population

These predictive models were successfully created using the automated process described in section 4.2.1 and appendix D.1, the models are shown in section 4.3.1. An additional contribution as a result of this, is the utilisation of these models to create an allocation score using the same methodology as the LAS, but applied to the UK population.

Another unique contribution is the ability of the allocation score to take donor variables and the combination of donor and candidate variables (such as height difference) into

consideration. At the time of writing, no other lung allocation system in the world does this, including the CAS in the US.

> **Contribution 2:** Development of a novel simulation engine to assess the impact of lung allocation design decisions on different groups of lung transplant candidates

> **Contribution 3:** Identification and quantification of the goals and values of the UK transplant community

> **Contribution 4:** Development of a framework that combines contributions (1), (2) and (3) to identify the lung allocation policy that most closely aligns with the goals and values of the UK transplant community

Contributions 2 - 4 correspond to goals 1 - 3 that have already been discussed and successfully achieved.

> **Contribution 5:** A general-purpose simulation engine that can simulate the allocation of any limited resource (assuming sufficient data is available)

The simulation engine that was developed was tailored specifically to simulating UK lung allocation, however this 'tailoring' was achieved by specifying a set of input parameters and providing the necessary datasets to the simulation engine.

By providing different datasets and using a different set of input parameters, the simulation engine can be used to evaluate allocation policies from other domains, thus making the simulation techniques described in this thesis general purpose.

> **Contribution 6:** A general-purpose framework for comparing allocation policies and selecting the most desirable policy

The methods in chapters 4 and 5 could be considered a description of a general-purpose framework for clinical decision making/allocation of scarce resources. The overall high-level process has been condensed and summarised in the section that follows.

Next are some additional contributions that were not originally outlined at the start of this thesis. The first is the Clinical Prioritisation Assistance Tool (CPAT) developed as a result of the COVID-19 pandemic. The methods that were developed were (unknown at the time) a simplified version of the methods used for the AHP. These methods included a novel combination of a modified version of Kahn's algorithm, loop detection algorithms and lexicographic sorting. This contribution can be summarised as a novel framework for emergency decision making in scenarios with limited or no available data or literature.

6.3 A Framework for Decision Making

The techniques used throughout this thesis are not specific to the allocation of lungs within the UK. These methods could be used for other countries, other organs and to other health resource allocation problems in general. The aim of this section is to outline a high level framework that can be followed and customised for specific use cases.

6.3.1 Step 1 - Identify Main Processes and Interactions

The first step is to identify *what is being simulated*. A general outline of processes and their interactions is given in chapter 4 for lung allocation, however this template can also be used for other organs. From a more abstract point of view a model of processes and interactions is very similar to a 'Stock and Flow' model.[155] A 'stock' is a generic container that holds an abstract quantity, and a 'flow' describes how quantities move between stocks. A flow starts at a source, can pass through one or more intermediate points, and ends at a sink.

This is illustrated in figure 6.2 with an example of a simplified economy. Funds flow from *sources* 'Employer Funds' and 'Bank Funds' into the *stock* 'Employee Bank Account', and eventually to the *sinks* of 'Purchases' and 'Government Funds'. The rate of flow of funds into the employee bank account are controlled by their salary, with the income tax rate determining the amount of salary being diverted to government funds. The employee bank account also earns interest, and the amount of interest earned is determined by the interest rate and the account balance. Funds in the employee bank account stock then flow to the 'Purchases' sink, and the rate of flow is determined by the account balance (i.e., as the balance of the account approaches zero, the amount of spending decreases). Finally, the sales tax rate determines the rate that funds spent on purchases are diverted back to government funds.

In the case of the lung simulation model, the two flow sources are the patient listing process and the donor availability process. Patients flow into the waiting list stock, and flow out to either the 'Waiting List Mortality' sink, or to the intermediate 'allocation policy' node, where the flow of donors combines with the flow of patients.

The flow terminates with the calculation of the post-transplant survival, benefit and waiting time metrics.

Identifying the main processes and how they interact will inform which data is necessary to collect/access, and also the structure of the simulation (see page 95 for an example of this applied to lung allocation).

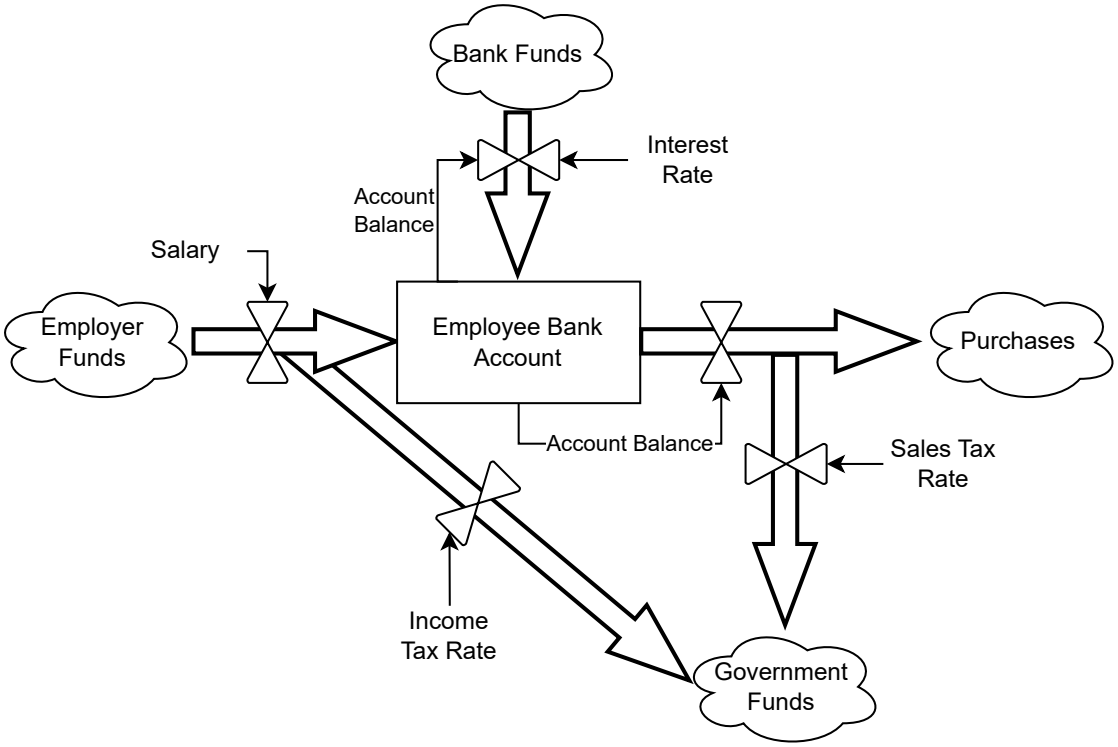


Figure 6.2: Stock and flow diagram example using a simplified economy. Sources are 'Employer Funds' and 'Bank Funds'. Sinks are 'Government Funds' and 'Spending'. 'Employee Bank Account' is a stock, and the quantity of funds in the account depend on the net difference between the flow of funds into and out of the account.

6.3.2 Step 2 - Data Integrity

The next step is to ensure the quality and integrity of any data used for decision making / simulation.

It is generally useful to plot continuous values on histograms and tabulate discrete/-categorical values to ensure they are within an expected range.

It is also necessary to decide how missing fields should be handled, some options are: exclude any data points that contain missing values, use a default value, infer the value from other fields (e.g. you can calculate BMI from height and weight even if the BMI field isn't populated) or use techniques such as multiple imputation. The data can then be filtered to only contain valid rows that are usable for analysis.

Formal methods can be used to assist in the process of ensuring data integrity. Libraries are available that allow invariants (i.e., properties that must always hold) to be specified per row, column or cell in a Comma Separated Values (CSV) file.[156] For example, a column invariant could ensure that no value is a pre-specified number of standard deviations from the mean value. This could highlight situations where a height should be input in centimetres, but instead has been input in feet. A row invariant could ensure no values are missing, and a cell invariant could cross-reference values, for example, ensuring an input

BMI is within a certain range given a height and weight. If a CSV file is verified using formal methods then all desired properties (invariants) are guaranteed to hold.

6.3.3 Step 3 - Extract Probability Distributions

The rate of 'flow' between 'stocks' in the model needs to be informed by observed data in order to approximate reality. Where possible, parameters describing probability distributions should be calculated. For survival / time-to-event data, some options are the exponential, Weibull and Gompertz distributions. For arrival frequencies/time between events, a Poisson process/distribution or exponential distribution can be used. Roulette selection as demonstrated in this thesis is also an option if no suitable probability distribution can be found.

6.3.4 Step 4 - Identify Performance Metrics

In order to decide between multiple alternatives, performance metrics can help identify the strengths and weaknesses of each option. It is important not to have too many metrics, but equally there should be enough to distinguish differences in performance between options. In the example given in this thesis, 'net benefit' and 'relative benefit' were essentially different measures of the same metric, so 'relative benefit' could be dropped without impacting the overall decision making process.

6.3.5 Step 5 - Simulate and Evaluate

Next, the system should be simulated, including all scenarios / options being considered. The performance metrics for each option should be normalised in the range 0 to 1, with 0 corresponding to the worst performing option on that metric and 1 corresponding to the best. This should result in there being a 'best' and 'worst' option for every metric.

6.3.6 Step 6 - Survey and Decide

The final step is to identify the key stakeholders with respect to the system being modelled, then use the AHP to identify and decide on the most desirable option.

The weights from the AHP survey combined with the normalised performance metrics from the previous step will identify the most desirable option for each stakeholder that participated in the survey.

The overall decision could be decided using the option that was most frequently identified as being the most desirable (similar to a majority vote), or ternary diagrams could be used to identify a set of weights that represents the best overall compromise between all stakeholders.

This process will likely **not** be linear, proceeding from steps 1 through 6 in sequence, arriving at the final result in step 6. As you work through the steps, assumptions may

be found to be incorrect, alternative simulation scenarios may be proposed, additional datasets may be needed, different performance metrics may be required and so on. This will necessitate returning to earlier steps and working through the process again. Overall this will likely be an iterative process that requires repeating steps many times.

6.4 Research Limitations and Future Work

Survival Models

This research makes use of waiting list and post-transplant survival models and a simulation engine that approximates the real-world processes of candidate listing for lung transplantation, donor offering, donor lung allocation, waiting list survival and post-transplant survival. As discussed in earlier sections, a model is a lower-resolution, more abstract view of a real-world system. The predictions, trends and trade-offs contained in this thesis are subject to a number of assumptions that must be taken into consideration if any real-world implementation of this research were to take place. The quality of the results generated from these techniques is limited by the quality of the data used for simulation, as such, any real-world applications should use the highest quality datasets available.

In section 4.4.4 the difficulty of predicting post-lung transplant outcomes was discussed. Future research should aim to improve predictions for post-transplant survival and identify data that is not currently being collected that is predictive of post-transplant outcomes. Research has already been completed comparing the Cox PH model to other survival models such as accelerated failure time (AFT) models, and concluded that there was no major difference in predictive ability between the types of model used.[157,158] There may also be individual psychological factors that impact post-transplant survival. Depression and use of anti-depressants post-transplant has been associated with an increased risk of mortality and graft loss,[159,160] and adherence to therapy has also been associated with post-transplant survival.[161]

One possible approach to increasing the concordance index of the models used for simulation is to stratify the population by diagnosis group. A future iteration of this work should look into developing separate survival models for each of the diagnosis groups and calculating the C-statistic for each group. This may result in a higher overall C-statistic for the population.

In section 4.4.3 the differences between simulated 5-year PTX rates and those observed and published by NHS-BT[21,24–26,146] were discussed. It was concluded the differences were due to an imperfect fit between the parameterised survival curve and the observed survival curve generated from the PTX survival dataset. Future work should focus on methods that result in a better fit of survival curves used for simulation, such as the Royston-Parmer flexible parametric model.[162]

Target Population

The next limitation is that this research only looked at the adult, first time, lung-only recipient population. This work can be extended by expanding the dataset to include paediatric patients and donors, re-transplants, and multi-organ recipients such as heart-lung and lung-liver patients. Currently, once a recipient has been allocated within the

simulation no further steps are taken for that individual. With appropriate data on re-transplant rates, and also the risk of developing different grades of PGD, the simulation engine can be extended so that there is a risk of developing complications post-transplant requiring the recipient to be added back onto the waiting list. This would make it possible to compare policies that handle re-transplants differently; should candidates requiring another transplant be prioritised in the same way as a first-time candidate? Or should they be lower priority since they have already "had their chance" at a transplant? Or is the fact they are experiencing a potentially life-threatening complication with their lung graft cause for re-transplanted candidate to be higher priority? What impact would this have on other candidates on the waiting list?

Because the simulations in this thesis generate a waiting list survival duration at the time a simulated candidate is added to the waiting list, the risk of mortality on the waiting list remains constant from the time of listing. In reality a candidate's "trajectory" on the waiting list could change, and this should also be modelled to improve the accuracy of the simulations.

For example, recent advances in treatment for cystic fibrosis patients has recently resulted in their frequent removal from the waiting list, as the disease-specific treatment is so effective they no longer require a lung transplant (see pages 16 and 18).

Conversely, prioritising candidates only once their clinical status has deteriorated may not make the best use of limited donor lungs and could be resulting in poorer outcomes post-transplant. Another extension to this work is to develop models that predict the changes in a candidate's clinical condition as time progresses on the waiting list, and evaluate the impact of allocation policies that take changing risk into account.

The datasets used in this work have a limited number of rare patient characteristics. For example, pulmonary arterial hypertension is relatively rare compared to the other diagnoses, and very tall or very short patients are also rare, given that they are at the tail ends of a normal distribution. The combination of these characteristics means there are few (or zero) patients in the dataset with these rare combinations of attributes. The result of this is that it is not possible to predict the impacts of different allocation policies on these patients. Just because these attributes are rare and don't occur in the dataset does not mean they can't occur at some point in the future. To overcome this limitation, the attributes of patients used in the simulations must be expanded beyond what is contained in the dataset. There are two ways this could be achieved:

The first is to check for any correlations between variables (for example, FVC and FEV1 are correlated), and if only a few variables are correlated then it may be possible to only model survival with uncorrelated variables. Next, each variable used in the survival models can have an appropriate distribution fit to it (e.g. normal, exponential). Then, instead of sampling random rows from a transplant dataset, patient attributes can be generated independently and randomly. This will result in a wider range of patient characteristics being simulated, overcoming the limitations previously outlined.

The second method is to make use of machine learning, more specifically generative adversarial network (GAN)s.[163] This approach is better suited to large datasets, where there may be more complex correlations between variables. A GAN takes training data as input, and through a training process generates new data that isn't contained in the original training dataset, but has the same underlying structure. This would allow 'new' transplant patients to be added to the datasets used for simulation, with the patient characteristics still being realistic. The UK transplant datasets may not contain enough data for this approach, however it may be possible to aggregate multiple datasets worldwide in order to have sufficient data to utilise GANs.

Additional Simulation Scenarios and Policy Performance Metrics

In section 4.4.3 (page 153) the simulations performed in this thesis for the UK population were compared to the simulations performed by Valapour et al.[100] for the US population. While the percentage reduction in waiting list deaths was similar, there was difficulty in determining how much of of the decrease was due to the removal of geographic boundaries, and how much was due to the use of survival models and an allocation score. To help answer this question, future work could compare the NHS-BT policy with geographic boundaries to one without, and comparing each of the score-based policies with and without geographic boundaries.

In section 4.4.3 the outcomes of SLT vs BLT are compared between this work and published simulation results using a Markov model in the US.[91] One difference between the two methodologies is that the US study simulated SLT for candidates with ILD as well as candidates with COPD, whereas only candidates with ILD were eligible for SLT in the simulations performed in this thesis. Additional simulations could be performed to evaluate the outcomes of SLT for COPD as well as ILD candidates.

There are additional metrics that could be used for evaluating the performance of policies. One interesting metric would be to calculate the probability of a recipient benefiting from transplant. The net benefit metrics that were calculated in this thesis are an average of a probability distribution. Due to the randomisation that is inherent in the simulation engine, there are instances where simulated recipients receive a negative net benefit, due to their simulated post-transplant survival duration being less than their simulated waiting list survival duration. To calculate the probability of benefit, the calculation would be:

$$\%\ \text{Probability of Benefit} = \frac{\text{Number of simulated recipients with net benefit} \geq 1\ \text{day}}{\text{Total number of simulated recipients}}$$

This could be taken further by calculating the probability of achieving at least x years of net benefit, for example 1 or 5 years. The results could be stratified by diagnosis group, age group, blood group etc. to identify the characteristics of patients most likely to benefit from transplant.

Order of Offering Left and Right Lungs with the SLT Policies for ILD Patients

In section 4.3.5 the ordering of offers was evaluated to see if there was any difference between offering the left lung first or the right lung first. On page 106 it is shown that compared to a BLT, transplanting a single left lung had a hazard ratio of 1.51 (CI: 1.22 to 1.86, p <0.001), and transplanting a single right lung had a hazard ratio of 1.08 (CI: 0.87 to 1.34, p = 0.485).

The simulation results showed no statistical difference in mean net benefit per recipient between allocating left-lung-first or right-lung-first (page 124). The SLT algorithm that was implemented to generate the results in section 4.3.5 compares the net benefit from BLT for a single recipient to the sum of net benefit for transplanting two ILD recipients (page 85) with a strict left-then-right ordering. However, the SLT policy doesn't have to strictly decide on always allocating the left lung first or right lung first.

In practice, at an individual level it **does** matter which recipient receives the left lung and which receives the right lung. If the SLT algorithm was implemented for real-world allocation, this could result in clinicians listing their patients as only able to receive a right lung in order to avoid them being allocated a left lung which has an expected lower net benefit. However, this would be at the cost of a higher risk of death on the waiting list. Alternatively the ordering of offers could be decided on a case-by-case basis.

An Alternative SLT Allocation Algorithm

One alternative SLT algorithm could be the following:

1. Calculate net benefit for recipients of any diagnosis receiving a BLT
2. Calculate the total net benefit for ILD recipients using the ordering: left, then right
3. Calculate the total net benefit for ILD recipients using the ordering: right, then left

These options could then be ranked by total net benefit in descending order, this results in six possible scenarios:

1. BLT $\geq$ left-first $\geq$ right-first
2. BLT $\geq$ right-first $\geq$ left-first
3. Left-first $\geq$ BLT $\geq$ right-first
4. Left-first $\geq$ right-first $\geq$ BLT
5. Right-first $\geq$ BLT $\geq$ left-first
6. Right-first $\geq$ left-first $\geq$ BLT

In some scenarios, the ordering of offers is either pre-determined or does not need to be considered: in scenarios 1 and 2, a BLT on a single patient would be performed. Scenario 3 dictates the SLT offering should be left-then-right, and scenario 5 dictates the offering should be right-then-left. This leaves only scenarios 4 and 6 to be considered.

The ordering can then be determined using a number of possible methods, including some from the field of decision theory: 'maximin', 'maximax', and 'minimax regret', all of which are described in greater detail in appendix K.

Method 1: Use Total Net Benefit With this approach the total net benefit is compared between left-first and right-first allocation. If left-first > right-first, then the left lung is allocated first. If right-first > left-first, then the right lung is allocated first. The final scenario is left-first = right-first, in this case, the ordering could be determined at random (though random processes in allocation may be undesirable, see pages 24 and 25), or one of the methods to follow could be used.

Method 2: Use the Maximin Strategy The maximin strategy[164,165] would choose the offer ordering (left/right/both lungs) that maximises the lowest net benefit a recipient will receive compared to all other offering orders (this concept was discussed in section 2.2.1 under 'Rawlsian Ethics').

Method 3: Use the Maximax Strategy The maximax strategy[165] chooses the offer ordering that **maxi**mises the **max**imum net benefit received by any recipient. Note that this method does not necessarily maximise the total net benefit as in method 1, due to the fact that one recipient may lose more net benefit than the other recipient gains, even if one of the recipients has the highest possible net benefit they could individually receive.

Method 4: Use the Minimax Regret Strategy The minimax regret strategy[165] chooses the offer ordering that **mini**mises the **max**imum 'regret' experienced by each recipient depending on the offer ordering. The amount of regret experienced by each recipient is the difference between the net benefit they received with the chosen order of offering (left/right/both lungs), and the maximum possible net benefit they would have received with alternative orderings.

Combining Methods If this approach were to be implemented in a real-world allocation algorithm, multiple methods could be used to determine the ordering of offers. This would only be necessary in scenarios 4 or 6 as shown on page 204.

This would require determining the relative value of each approach and ranking them from highest to lowest, this could be determined using the comparison matrices such as those used for CPAT in chapter 3. For example, it could be decided the highest priority

is to maximise the minimum net benefit, then to minimise maximum regret, then to maximise total net benefit, and finally to maximise maximum benefit.

The SLT algorithm could then operate as follows:

1. Identify which ordering scenario is present
   (a) If BLT results in the highest total net benefit, then allocate the lung pair
   (b) If left-first or right-first results in the highest total net benefit, and BLT results in the second-highest net benefit, then allocate in the order that results in the highest total net benefit
   (c) If BLT results in the lowest net benefit, then:
      i. Use the ordering as determined by Method 2: Maximin. If choices are identical, then move to next method
      ii. Use the ordering as determined by Method 4: Minimax Regret. If choices are identical, then move to next method
      iii. Use the ordering as determined by Method 1: Total Net Benefit. If choices are identical, then move to next method
      iv. Use the ordering as determined by Method 3: Maximax. If choices are identical, then move to the final method
      v. Use an appropriate random number generator to randomly decide allocation order

It is unlikely in practice that a situation will arise that results in the necessity to decide on the allocation order at random. However, should this situation arise it is important to use a suitable random number generator. Many random number generating functions in software make use of pseudorandom number generation. A pseudorandom number generator is initialised with a seed value, and then generates a sequence of (apparently) random numbers, however, if the same seed value is used then the exact same sequence of numbers will be generated.

One option is to initialise the random number generator using Unix time, which is the number of seconds elapsed since 00:00:00 Coordinated Universal Time (UTC) on 1 January 1970. This would allow for auditing of allocation decisions, as the timestamp could be input to the random number generation function, which should output the same pseudorandom number and result in the same offer ordering being chosen.

Another option is to use a true random number generator; these require a source of entropy such as computer mouse movements, keyboard typing intervals or disk I/O operations. While this ensures no bias in the random numbers, it is not possible to repeat the generation process for auditing purposes.

Other potential extensions of this work are to apply it to other organs, other prioritisation problems, or to lung transplant patients from other countries. The same processes

could be repeated as outlined in this thesis, and trade-offs/trends can be identified. It will then be possible to identify if the same trends show up in other transplant populations.

6.4.1 Final Words on Real-World Implementation

Through the course of completing the work for this thesis I was able to begin collaborating with the Cardiothoracic Advisory Group (CTAG) within NHS-BT, who are responsible for overseeing the UK lung allocation policy (meeting agendas and minutes where I was present can be found at[166–168]). This has presented the opportunity to potentially have this research guide the implementation of future UK lung allocation policies.

Before discussing further, it is worth recounting the principle of 'Chesterton's fence', named after the English philosopher G. K. Chesterton and published in his 1929 book "The Thing":[169]

> In the matter of reforming things, as distinct from deforming them, there is one plain and simple principle; a principle which will probably be called a paradox. There exists in such a case a certain institution or law; let us say for the sake of simplicity, a fence or gate erected across a road. The more modern type of reformer goes gaily up to it and says, "I don't see the use of this; let us clear it away."
>
> To which the more intelligent type of reformer will do well to answer:
>
> "If you don't see the use of it, I certainly won't let you clear it away. Go away and think. Then, when you can come back and tell me that you do see the use of it, I may allow you to destroy it."

Applying this principle to the UK lung allocation system, the reasoning behind the design of the current system should be understood - even if the reasoning is flawed or outdated - before replacing it with a new system. It may not be possible to fully understand the reasoning behind all design decisions, however, an attempt should at least be made to retrospectively understand the potential reasons for historical policy design decisions (i.e., "go away and think").

The results in chapter 4 have demonstrated that the existing allocation policy has very weak predictive ability compared to the models developed in this thesis (page 109), and simulations have shown the use of a national allocation score significantly outperforms the existing policy in terms of reducing waiting list survival and increasing net benefit (pages 122 and 125).

The potential benefits may be perceived to outweigh any potential risks or unintended consequences from replacing the existing allocation system with a national score-based one, and the change could be justified if all stakeholders agree they are happy to take on the risk to experience the benefits. However, Chesterton's fence still applies: the benefits are framed in the context of waiting list survival, post-transplant survival, and net benefit.

The original reasoning behind the existing policy may have been in the context of equity of access to transplant, ethical considerations, operational/practical limitations, budget constraints, legal responsibilities, or some other context that has not been considered.

The two largest changes proposed by this work are the removal of geographic boundaries, and removing the strict condition that ABO-identical candidates are prioritised ahead of ABO-compatible candidates. As discussed previously (section 1.1.2 on page 8), the ABO-identical rule may be in place to allow ABO-O candidates greater access to transplant. Centre-based allocation may be in place to reduce the costs of transporting lungs across the UK, minimise CIT, or an attempt to prevent differing access to transplant based on a candidate's location.

These would represent large changes to the UK lung allocation system, and any change (especially large changes) to a complex system introduces the risk of unintended consequences; there are many potential complex interactions that could result in emergent behaviour that is difficult or impossible to predict before the change is introduced.

If a real-world implementation of this work were to be pursued, it would require engagement from all stakeholders, ensuring they understand the potential risks and benefits of changing the existing allocation system, and agreement to implement the new system. Continuous monitoring of candidate and recipient outcomes would be required to ensure there are no unexpected or adverse consequences.

Appendix A

An Introduction to Survival Analysis

Survival analysis is a branch of statistics that focuses on analysing and predicting survival durations and outcomes. In a more general sense, survival analysis focuses on analysing "time-to-event" data, which does not necessarily have to relate to the mortality of individuals. These same techniques can be used to analyse the reliability of consumer hardware or machinery in a factory, where the time-to-event data specifies the time between hardware/machinery failures.

The specific challenge that survival analysis deals with is censoring. For each individual in a dataset, a survival duration can be recorded, however not all individuals in the dataset have necessarily died. The survival duration in some cases gives a minimum survival duration, or in other words, we know that individual X has survived at least t days, but have no further data beyond that point in time. If we know an individual has survived from time t_0 to time t without dying, that data point is said to be right-censored.

One method for dealing with censored data is to simply ignore it and only use data points where an event has been observed, however there are some problems with this approach. The most immediate problem is that potentially useful information is being discarded. Ideally an analysis would make use of all data that is available. The other problem is that excluding data points with no event recorded results in over-estimating mortality rates, biasing any models to predict shorter survival durations than would be observed in reality.

To correctly account for censoring, survival datasets contain two values per observation: the first is the observed survival duration in some unit (days, weeks, months, years etc.) and the second is a censoring indicator. The censoring indicator is typically set to 1 to indicate an event was observed, and 0 to indicate an event has not yet been observed. In addition to these two variables, any other variables that could potentially be related to mortality can be included.

One basic way of analysing the survival of a population using right-censored survival data is to use KM survival curves. The specifics of how the curve is calculated is beyond the scope of this thesis, however a basic overview will be given here. The KM curve is constructed by taking into account the number of individuals that are "at risk" and the number of observed events at each point in time. The KM curve also takes the censoring status of each individual into account. The x-axis is time, starting at 0 and increasing up to the longest observed survival duration in the dataset. The y-axis is the proportion of individuals surviving to that point in time.

Regardless of the population being analysed, all KM curves start at 100% survival at time 0. The survival curve then decreases monotonically (i.e., remains the same or decreases, but never increases) with time as more events are observed and a lower percentage of the population remains alive.

Some useful quantities can be calculated from a KM curve, for example the time at which 50% of the population is alive gives the median survival time. The area under the curve from time t_0 up to some arbitrary time x (referred to as the restricted mean) gives the mean survival duration of the population up to time x.

Along with the main survival curve, 95% upper and lower confidence intervals can also be calculated. These can be plotted as a shaded area on the graph, or as dotted lines above and below the survival curve. Typically, the confidence intervals widen with time, as fewer and fewer observations are remaining in the dataset as time progresses. The confidence intervals can also be used to visually compare the survival of different populations. For example, if you wanted to analyse the efficacy of a treatment, you could plot the survival curves and confidence intervals for a control population and a treatment population. If there is a large amount of overlap of the confidence intervals then the survival between populations is not statistically significant, or in other words the treatment has not been observed to be effective at reducing mortality between the control and treatment populations.

Rather than comparing visually, there are statistical techniques that can be used to calculate a p-value to determine if there is a statistically significant difference in survival between populations. One statistical test for survival data is the log rank test, which uses the null hypothesis that there is no statistically significant difference between the survival curves between two populations. A p-value ≤ 0.05 indicates that there is a statistically significant difference between population survival.

For certain subsets of the population survival rates may differ. For example, mortality may be higher for older individuals, or for individuals with certain medical conditions. For few variables of interest, plotting separate survival curves may be practical, however, as the number of variables increases this becomes impractical.

One of the most used survival models is the Cox PH model. The Cox PH model was first described by Sir David R Cox in 1972 in his paper "Regression models and life-tables".[124] In order to make use of the Cox PH model the proportional hazards assumption

must hold. This assumption states that the relative hazard between two populations remains constant. In relation to the KM survival curves, if the survival curves between two populations intersect then the proportional hazards assumption has been violated.

If the proportional hazards assumption holds then the model can be parameterised as follows:

$$h(t) \times \exp(X \cdot \beta')$$

Where $h(t)$ is the hazard function, which takes time t as an input and returns the instantaneous hazard rate at that time. If the hazard function is integrated, you have the cumulative hazard function, $H(t)$, which returns the total accumulated risk up to time t of an event occurring. The survival function $S(t)$ is related to the cumulative hazard function as follows:

$$S(t) = \exp(-H(t))$$

Cox observed that if the proportional hazards assumption holds, then $h(t)$ does not need to be explicitly defined. The benefit of this is that the Cox PH model can be used to analyse the hazard ratio between different populations, without having to estimate or compute the underlying hazard function.

In practice, a survival dataset is taken, containing survival durations and censoring indicators, as well as any patient-specific variables of interested, referred to as covariates. The vector X refers to the values of the covariates for an individual. The Cox PH model then estimates the coefficients for each of the covariates, where the coefficients are the vector β shown in the formula above. The coefficient vector is sometimes shown as β', the $'$ symbol indicates the vector has been *transposed*, which means the vector is re-written so that every row becomes a column:

$$\beta = (1, 2, 3)$$

$$beta' = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

As with the KM survival curves, useful quantities can be calculated using a Cox PH model. The first is the LP, referred to as the LP, or in some cases X-Beta. To calculate the linear predictor, the dot product of patient covariates (vector X) and model coefficients (vector β) is taken. The dot product is calculated by simply multiplying each element in X with the corresponding element in vector β and summing the result.

An example is given below for a 50 year old male with a BMI of 32:

$$\text{Patient Attributes} = X = (\text{Age}, \text{Sex: Female}, \text{BMI})$$

$$X = (50, 0, 32)$$

$$\text{Cox Model Coefficients} = \beta = (0.02, -0.25, 0.015)$$

$$\text{LP} = X \cdot \beta'$$

$$\text{LP} = \begin{pmatrix} 50 & 0 & 32 \end{pmatrix} \cdot \begin{pmatrix} 0.02 \\ -0.25 \\ 0.015 \end{pmatrix} = (50 \times 0.02) + (0 \times -0.25) + (32 \times 0.015)$$

$$\text{LP} = 1 + 0 + 0.48 = 1.48$$

This value is the linear predictor: an LP of 0 means there is neither an increase or decrease in risk, a negative LP indicates a lower risk and a positive LP indicates a higher risk of mortality. The hazard ratio (hazard ratio (HR)) can be calculated as follows:

$$\text{HR} = \exp(\text{LP})$$

Continuing the example from above, the hazard ratio would be:

$$\text{HR} = \exp(1.48) = 4.4$$

This means that this individual has 4.4 times the risk of experiencing mortality compared to the reference level, but what is the 'reference level'?

Each variable in a Cox PH model can either be categorical (i.e., discrete) or continuous. Some examples of categorical variables are biological sex (Male/Female), presence of a certain medical condition (Yes/No) or smoking status (Never smoked/Occasional Smoker/Daily Smoker/Heavy Smoker). Some examples of continuous variables are age, height, BMI and so on.

For every variable, a reference value must be specified. In the case of categorical variables, one of the categories must be chosen to be the reference level. In the case of biological sex, if the reference level was specified as "Male", then the coefficient for biological sex gives the log-hazard ratio for "Female" compared to the reference level of "Male". For continuous variables such as age, one value can be specified as the reference value, and the coefficient for the age variable tells you the log-hazard ratio per unit difference from the reference age.

In some cases, there may not be a linear relationship between a continuous variable and risk of mortality. For example, extremely low and extremely high values for BMI can

result in a higher risk of death, with values in the middle resulting in a relatively lower risk of death. There are several approaches for dealing with this, however the approach chosen for the work completed in this thesis was to use RCS.

RCS allow non-linear relationships to be modelled between a continuous variable and the model's beta coefficient. To use RCS a number of "knots" need to be specified, in this example, 3 knots will be used at ages 20, 40 and 55. RCS also include a linear term, so in this case, four beta coefficients need to be calculated for a single continuous variable. For this example, the coefficients will be as follows:

$$\begin{gathered}\text{Linear} : -0.05 \\ \text{Knot } 20 : -0.00001 \\ \text{Knot } 40 : 0.0002 \\ \text{Knot } 55 : 0.00035\end{gathered}$$

The linear term is fairly straightforward to calculate, it is simply -0.05 multiplied by the individual's age in years. The knot terms are more involved and are calculated using the following formula:

$$\text{Coefficient} * \max(\text{value} - \text{knot value},\ 0)^3$$

The max function returns the maximum of two values, so in the example of 'Knot 20' given above, if an individual is aged 20 or younger then the entire term evaluates to 0. Here is an example calculation for a 15 year old:

$$\max(\text{value - knot value},\ 0)^3 = \max(15\ -\ 20,\ 0)^3 = \max(-5,\ 0)^3 = 0^3 = 0$$

This allows the coefficients to apply to specific ranges of values, by evaluating to zero if the input value is less than or equal to the knot value. The next example shows the calculation of an entire RCS term for an individual aged 47:

$$\begin{aligned}&\text{Linear term} = -0.05 * 47 = -2.35 \\ &\text{Knot } 20 = -0.00001 * \max(47\text{–}20,\ 0)^3 = -0.00001 * 27^3 = -0.19683 \\ &\text{Knot } 40 = 0.0002 * \max(47\text{–}40,\ 0)^3 = 0.0002 * 7^3 = 0.0686 \\ &\text{Knot } 55 = 0.00035 * \max(47\text{–}55,\ 0)^3 = 0.00035 * \max(-8, 0)^3 = 0.00035 * 0^3 = 0\end{aligned}$$

Summing all the above terms results in the RCS term being calculated as -2.47823. It is also possible to plot the entire function for a range of ages (see figure A.1).

Throughout this thesis, wherever you see "RCS", this refers to a continuous variable that has been fit with a restricted cubic spline, allowing a non-linear relationship to exist between the covariate and the risk of mortality.

It is also possible to measure the discrimination ability of a Cox PH model by calculating a measure known as the concordance index or C-statistic. This is a measure of how

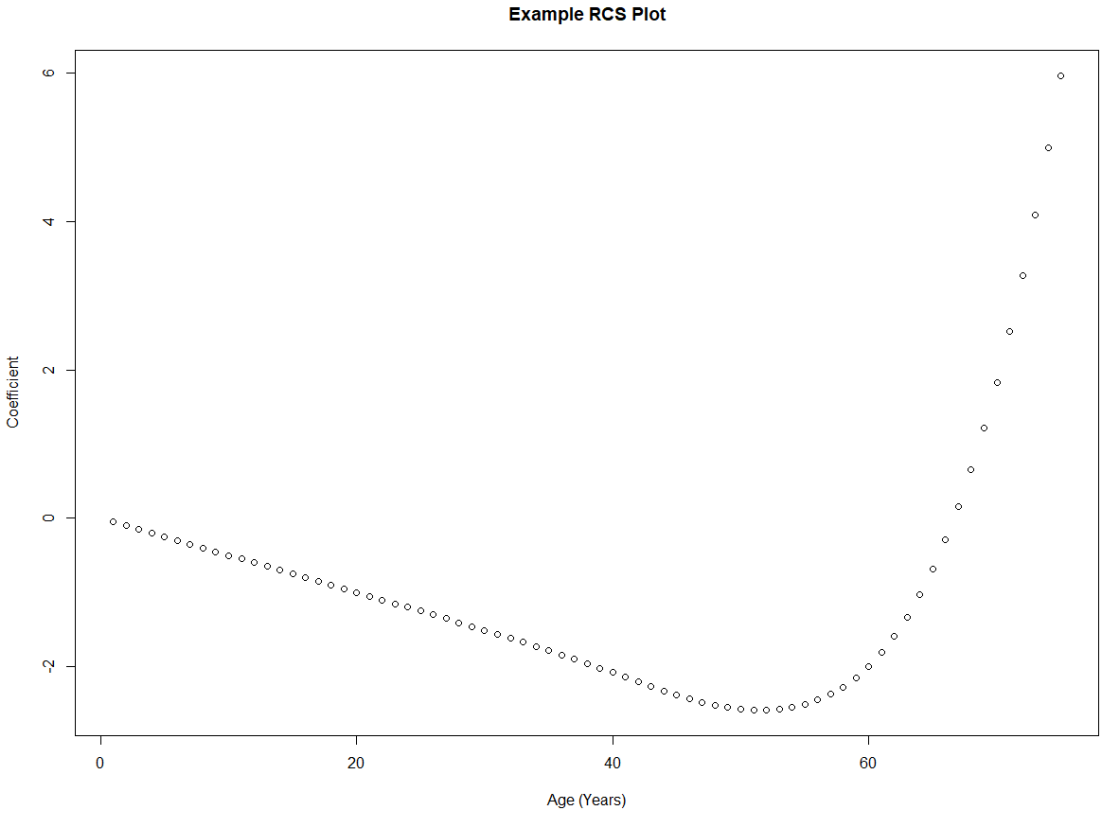


Figure A.1: Example restricted cubic spline allowing a non-linear relationship between age and risk of mortality.

well the model can predict the relative risk of mortality between any two patients. The calculation of the C-statistic is analogous to presenting pairs of patients to the model, and noting which patient the model predicts is at a higher risk of mortality. A tally is kept of how many predictions the model predicts correctly, and how many incorrectly. The ratio of correct predictions to the total number of predictions is the C-statistic. If a model is no better than random chance (i.e. flipping a coin to decide which patient is higher risk), then the C-statistic will be 0.5. If the model makes perfect predictions, then the C-statistic will be 1, and if the model makes a wrong prediction every time, the C-statistic will be 0.

The model's calibration ability can also be evaluated, which is the agreement between the model's predicted probabilities of an event occurring and the actual observed rate of an event occurring. Calibration curves for a specific point in time can be generated by repeatedly selecting a random subset of the population, observing the survival rate up to the specified point in time, and recording the predicted survival rate. This will result in a range of predicted probabilities and observed survival rates, which can be plot with the predictions on the x-axis and the observed rates on the y-axis. If the model is perfectly calibrated, then a diagonal line will run from the bottom left corner to the top right corner,

meaning that the model's prediction of the probabilities of survival perfectly match the observed probabilities of survival. One measure of calibration is the Mean absolute error (MAE). This is the mean deviation of the calibration curve from a perfect diagonal, with an MAE of 0 indicating a perfectly calibrated model (at that time point).

Each individual variable in the model can also be evaluated in terms of the statistical significance of the estimated coefficients. The 95% confidence intervals for each coefficient is calculated, and is often shown on forest plots as a horizontal line spanning the range of the confidence intervals for the hazard ratio for that variable. If the confidence intervals overlap zero for the coefficients (or 1 for the hazard ratio), then the estimate of the coefficients is not statistically significant. The further the confidence intervals are from 0, the more likely the estimate of the coefficient is statistically significant. The default test used in 'R' for evaluating Cox model coefficients is the Wald test and outputs a p-value per coefficient.

Appendix B

CPAT Ranking Methods 💻

B.1 CPAT Policy Design Process

The algorithms in this appendix can be used to rank a list of *alternatives* against a set of *criteria*. An *alternative* refers to one possible option from a set of options. In this context an alternative is a single patient (selected from a larger list of patients) that might be prioritised based on risk to access limited resources for treatment of COVID-19 related symptoms. The *criteria* refer to the means by which each patient is compared, for example age, sex, and presence of certain co-morbidities.

B.1.1 Component 1 - Pairwise Comparisons and Tournament Graphs

Pairwise comparisons can be used to rank a set of criteria. A pairwise comparison matrix can be constructed where each row and each column are mapped to the criteria of interest. Within each cell of the matrix, the criterion in the row is compared to the criterion in the column and an arrow is inserted into the cell pointing to the more important criterion. In some cases it may be difficult to compare two criteria, in these cases the cell can be left empty.

To illustrate this process, a set of five criteria will be used: A, B, C, D, E. The pairwise comparison matrix is mirrored across the major diagonal, and the cells along the major diagonal all compare a criterion to itself, therefore only the cells to the right of the major diagonal are of interest.

$$\begin{pmatrix} & A & B & C & D & E \\ A & X & & & & \\ B & X & X & & & \\ C & X & X & X & & \\ D & X & X & X & X & \\ E & X & X & X & X & X \end{pmatrix}$$

In this example, the matrix has been populated as follows (a '.' is a cell where a comparison could not be made):

$$\begin{pmatrix} & A & B & C & D & E \\ A & X & \leftarrow & . & . & \leftarrow \\ B & X & X & \leftarrow & . & . \\ C & X & X & X & \leftarrow & . \\ D & X & X & X & X & \uparrow \\ E & X & X & X & X & X \end{pmatrix}$$

This matrix corresponds to the adjacency matrix of a graph, and because each criterion is compared to every other criterion this is a special case referred to as a *tournament graph* ([36] - final paragraph of section 7). The corresponding graph for the above matrix is shown in figure B.1.

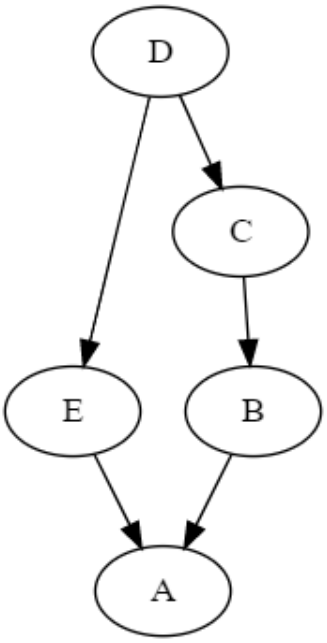


Figure B.1: A directed graph generated from an adjacency matrix.

This graph can be used to determine the relative importance of each of the criteria, but first the graph must be determined to be free of loops.

B.1.2 The Problem with Loops

For any two nodes in the graph that are connected by a directed edge, the arrow points to the criterion which is of greater importance. If a loop exists in the graph then a logical contradiction has occurred. Take the case where $A \leftarrow B$, $B \leftarrow C$ and $C \leftarrow A$, it is impossible to determine which of the three criteria is most important. The corresponding graph is shown in figure B.2.

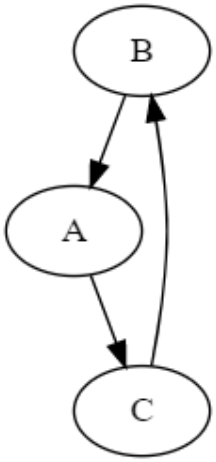


Figure B.2: An example of a loop creating a logical contradiction

Starting at any node and following the directed edges results in returning to the same node. In this example there is no way to order A, B and C.

When a loop is detected in the graph, this highlights the fact that there has been an inconsistency in the comparisons the decision maker(s) have made. In this situation the comparisons within the loop must each be re-evaluated and corrected until no loops are found in the graph.

Once all loops have been removed, the resulting graph is a Directed Acyclic Graph (DAG).[170] DAGs are commonly used in computer science and statistics and can be used in a number of fields for a range of problems such as scheduling and data compression. In the case of CPAT the DAG will be used to determine the relative importance of criteria used for ranking patients.

B.1.3 Topological Sorting

The DAG created in the previous section contains a number of directed edges, showing which criteria are more important than others. However, to make use of the DAG it is necessary to determine which criteria are of lowest importance, which are of higher importance, and the relative importance of criteria in between. This can be accomplished by performing a topological sort.

A topological sort requires a loop-free (or cycle-free) graph, which by definition a DAG meets this requirement. Applying a topological sort to a DAG will result in a list of nodes

(i.e. criteria) ranked in order of importance. One algorithm for performing a topological sort on a directed graph is Kahn's algorithm,[110] which proceeds as follows:

1. Remove a node from the graph that has no incoming edges and place it in the list of ranked nodes
2. Remove all edges that were connected to the node identified in step 1
3. Repeat until all nodes from the graph are in the list of ranked nodes

One issue with the standard algorithm is step 1: any node that has no incoming edges can be chosen, and without any rules on which to select the algorithm is non-deterministic. A non-deterministic algorithm has the potential to generate different orderings of the criteria given the same tournament graph. One solution to this problem is rather than having a prioritised list of individual nodes, instead a prioritised list of *sets of nodes* are generated.[111] The above algorithm was adjusted as follows:

1. Remove *all nodes* from the graph that have no incoming edges, *add them to a set* and *add this set* to the list of ranked nodes
2. Remove all edges that were connected to the *nodes* identified in step 1
3. Repeat until all nodes from the graph are *contained in sets* within the list of ranked nodes

The processes for the same example graph is shown in figure B.3. Once Kahn's algorithm has successfully been applied to the criteria of interest and a hierarchy of relative importance is established, the next step is to encode the hierarchy using CPAT's custom DSL.

B.1.4 Component 2 - Encoding CPAT Policies using a DSL

In the previous section, the relative importance of criteria were identified using Kahn's algorithm. The next step is to instruct the CPAT software to automatically rank a list of patients using the criteria of interest.

CPAT encodes a policy using an ordered list of factors, with the most important factor being placed on the first line and the least important on the last line. The relative importance of factors were determined in section B.1.3. Factors are logical expressions which can be evaluated to **True** or **False**. CPAT supports the basic boolean operators: **AND**, **OR** and **NOT** as well as some more advanced logical statements.

Bracketed Ranges Bracketed ranges can be used to test if a value falls within a range, for example:

```
Age([85,120])
```

Selects all individuals with an age between 85 and 120 (inclusive).

N in Set If a patient has several boolean (i.e. True/False) fields the 'N in Set' statement can be used to evaluate if at least N of those fields evaluated to 'True'. For example:

```
3_in_set(Diabetes,Malignancy,Hypertension,CardiovascularDisease)
```

Selects all individuals with at least 3 of the diagnoses specified between the parentheses.

Option Fields Option fields can be selected if a patient has a field which can have multiple options, for example 'Smoking Status' could have the possible values: 'Never', 'Light', 'Moderate', 'Heavy'.

To select only the individuals with a 'Moderate' or 'Heavy' smoking status, the following statement would be used:

```
SmokingStatus(Moderate,Heavy)
```

An Example Policy

The previously mentioned features can be combined using boolean operators to create more sophisticated policies. Below is an example of a policy encoded using CPAT's DSL, with patients aged 80+ being the most important factor, and the patient's sex being the least important factor.

```
age([80,150]);
2_in_set(cardiovascular_disease, hypertension, diabetes);
cancer OR kidney_disease;
sex(Male);
```

Errors and Warnings

The syntax for errors and warnings is identical, however there are two separate input fields in the user interface and the results of errors and warnings are different.

The general syntax is:

```
error/warning message: factor;
```

Where any error or warning message can be written before ':' and a factor encoded in CPAT's DSL is written on the right hand side. If an error is raised for a patient, that patient is removed from the ranking list and the error message is given next to their name.

A warning is used to display a message, but still allows a patient to be ranked. These can be used to highlight situations where there *may* be an inconsistency in the data that has been input for the patient, but the currently entered value can still be used for ranking.

An example of an error could be for a system which is not designed to prioritise paediatric patients, in this case the following would raise an error:

```
Patient aged under 18: age([0,17]);
```

An example of a warning is the use of the CFS for patients aged 65 years or younger. The scale is designed for use with patients aged over 65, so the patient can still be ranked, but a warning will be displayed for a clinician to review and confirm the data entered is correct.

```
Use of CFS on patient aged <= 65: age([0,65]) AND cfs IS NOT NULL;
```

Once a policy has been encoded, it can be used to prioritise a list of patients, this process will be explained in the next section.

Lexing and Parsing CPAT DSL

To interpret the custom CPAT DSL, the policy needs to go through two processes: lexing and parsing. Lexing is the process of performing pattern matching on the raw text and converting the text to a series of tokens. Pattern matching is accomplished using regular expressions, for a full list of regular expressions used for CPAT, see appendix C.

The series of tokens output by the lexer are then interpreted and executed by the parser. In the case of CPAT the DSL is parsed into a series of custom SQL queries, where each line of the policy corresponds to a single query.

Each query selects a (possibly empty) subset of patients from the database and adds the weight of the current coefficient to their total score. The algorithm in pseudocode is shown below:

```
# Split the policy into individual lines
policy_lines = extract_policy_lines(policy)
num_factors = count(policy_lines)
current_power = 2 ^ (num_factors - 1)

# Map patient id's to a score
patient_scores = {id |-> 0 for all patient id's}

foreach policy_line in policy_lines {
query = parse(policy_line)  # Parse the policy to an SQL query

# Select all patients which match factors in current line of policy
patients_matching = select_matching_patients_from_DB(query)

# For each matching patient, increment their score by the weight
# of the current line in the policy
foreach patient_id in patients_matching {
patient_scores[patient_id] = patient_scores[patient_id] + current_power
}

# The next line will have half the weight of the current policy line
current_power = current_power / 2
}
```

This algorithm will take the current policy and a list of patients as input, and output a mapping of patients to CPAT scores.

B.1.5 Component 3 - Lexicographic Sort

The previous steps all lead up to this final step: ranking patients in order of risk of mortality. The first step determines which criteria should be used and which criteria are most and least important. The second step encoded the ordered criteria using a DSL for automation. This final step orders patients using the criteria specified using the DSL.

Patients are ranked using a lexicographic approach, which is the same approach used for ordering words in alphabetical order in a dictionary. Figure B.4 shows an example of checking the words 'Apple', 'Banana' and 'Bandana' are in alphabetical order.

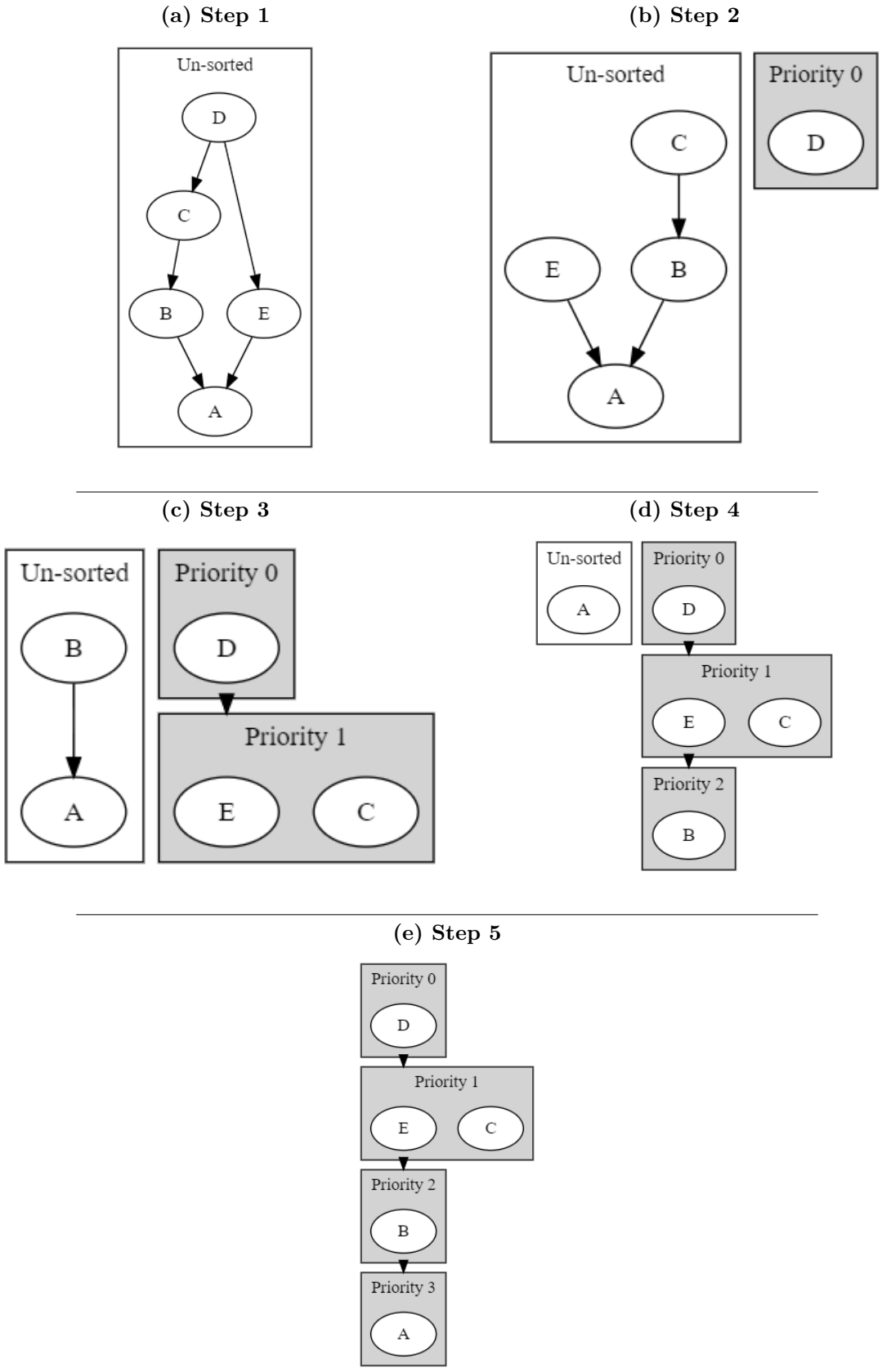


Figure B.3: Modified Kahn's algorithm: given a set of criteria (in this case labelled **(a)**-**(e)**) and their relative importance (arrows point in the direction of increasing importance), a prioritised list of sets is generated. Priority 0 is lowest priority and priority 2 is highest priority.

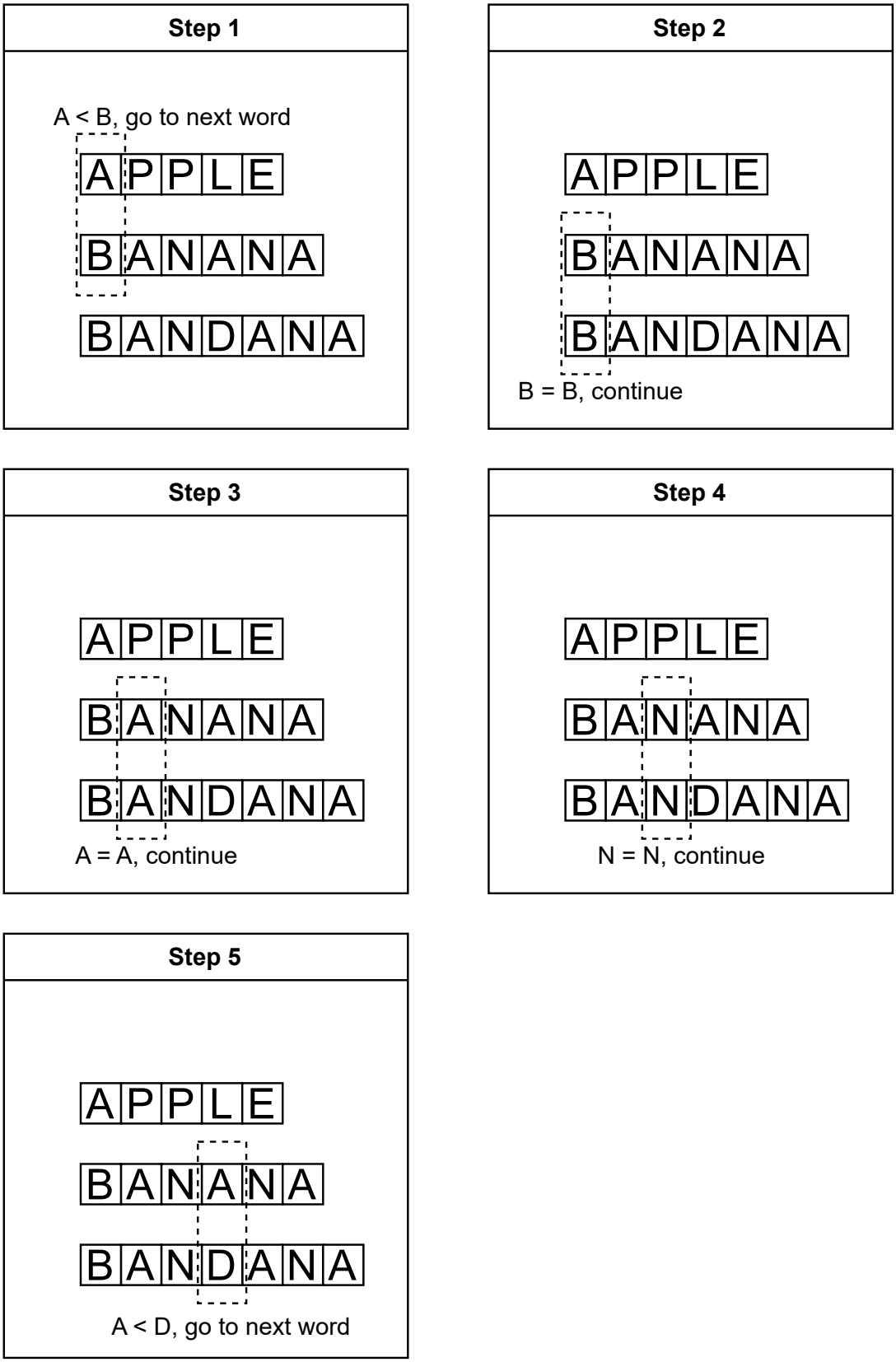


Figure B.4: Algorithmic example of checking the lexicographic (i.e., alphabetical) order of three words.

To place words in lexicographic order the following must be decided: which direction should words be evaluated in? (i.e., left-to-right or right-to-left?) and which letters are 'higher priority' than others? (i.e., the alphabet).

The process (in the English language) begins with the left-most letter, with 'A' being the highest 'priority' and 'Z' the lowest. If two words begin with the same letter, then the next letter to the right is evaluated. This process is repeated until the list of words is in alphabetic/lexicographic order.

With CPAT, the policy for ranking patients is the 'word', and each line of the policy is a 'letter', however instead of being evaluated left-to-right, it is evaluated from top-to-bottom (i.e., starting on the first line of the policy, and moving down to subsequent lines).

The first line in the policy contains the most important factor and subsequent lines are in decreasing order of priority. Consider two words beginning with different letters: `APPLE` and `BANANA` from the previous example. 'Banana' begins with a 'B', and no matter what the following letters are, it will never be placed before 'Apple' when being sorted in alphabetical order. This constraint can be enforced mathematically as follows: for a policy with N lines in total, the nth line of the policy has a weight of 2^{N-n}.

Using the example policy in section B.1.4, the policy coefficients can be expressed as:

$$\beta = (8, 4, 2, 1)$$

For each individual patient, each line of the policy will evaluate to **True** or **False** for that patient. This can be encoded as '1' and '0' respectively. For a male patient aged 85 with diabetes and hypertension, the covariate vector would be:

$$X = (1, 1, 0, 1)$$

The vector X can be thought of as a 'word' with an alphabet containing two 'letters': 1 and 0. The score can then be calculated (and lexicographic order enforced) as follows

$$X\beta' = (1, 1, 0, 1) \begin{pmatrix} 8 \\ 4 \\ 2 \\ 1 \end{pmatrix} = 13$$

A score can then be calculated for each patient, then the list can be sorted by CPAT score in descending order, placing highest priority patients at the top of the list, and thus, the list of patients can be said to be in lexicographic order. Patients with the same score can be grouped, resulting in a ordered list of prioritised groups of patients. Patients within the same group are at the same priority level, this is equivalent to sorting a list of words into lexicographic order where there are duplicate words - patients with identical characteristics will evaluate to the same score and same ranking.

Appendix C

CPAT Regular Expressions

Table C.1: Regular expressions used to parse the CPAT DSL.

Regex	Explanation
`(\w+)\s+IS\s+NULL`	Check if data field is null
`(\w+)\s+IS\s+NOT\s+NULL`	Check if data field is not null
`(\d+)_in_set\(([\w\,\!]+)\)`	Parse n_in_set statement
`(\w+)\(([\w\,]+)\)`	Parse 'option' statement
`(\w+)\s*\((\[\S+\])\)`	Parse bracketed range of values
`(\w+)`	Parse 'binary'/boolean statement
`(AND)`	Logical 'AND'
`(OR)`	Logical 'OR'
`(NOT)`	Logical 'NOT'

Appendix D

Description of Lung Simulation Methods

D.1 Iterative Cox Model Building Algorithm

This algorithm was implemented in R as described in.[133] The algorithm iterates over forward selection and backward elimination steps until the model can not be improved further. At each step of the algorithm, two nested models are compared using the likelihood ratio (LR) test to evaluate if there is a significant difference in the goodness of fit. If a dataset is particularly large (such as those seen in the OPTN datasets), a p-value threshold of 0.05 can be used. For smaller datasets, higher p-value thresholds can be used if few variables are added to the model. In the case of the relatively smaller UK dataset, a p-value threshold for the LR test of 0.15 was used.

The algorithm proceeds as follows:

1. Start with an empty model
2. Identify all variables that when added to the model individually, pass the LR test
3. Add all variables from step 2 to the model
4. Remove each variable in turn, check if it still passes the LR test, if not remove from model
5. Go back to step 2, unless model is unchanged from previous iteration, in which case terminate

The process is outlined in greater detail in figures D.1 to D.3.

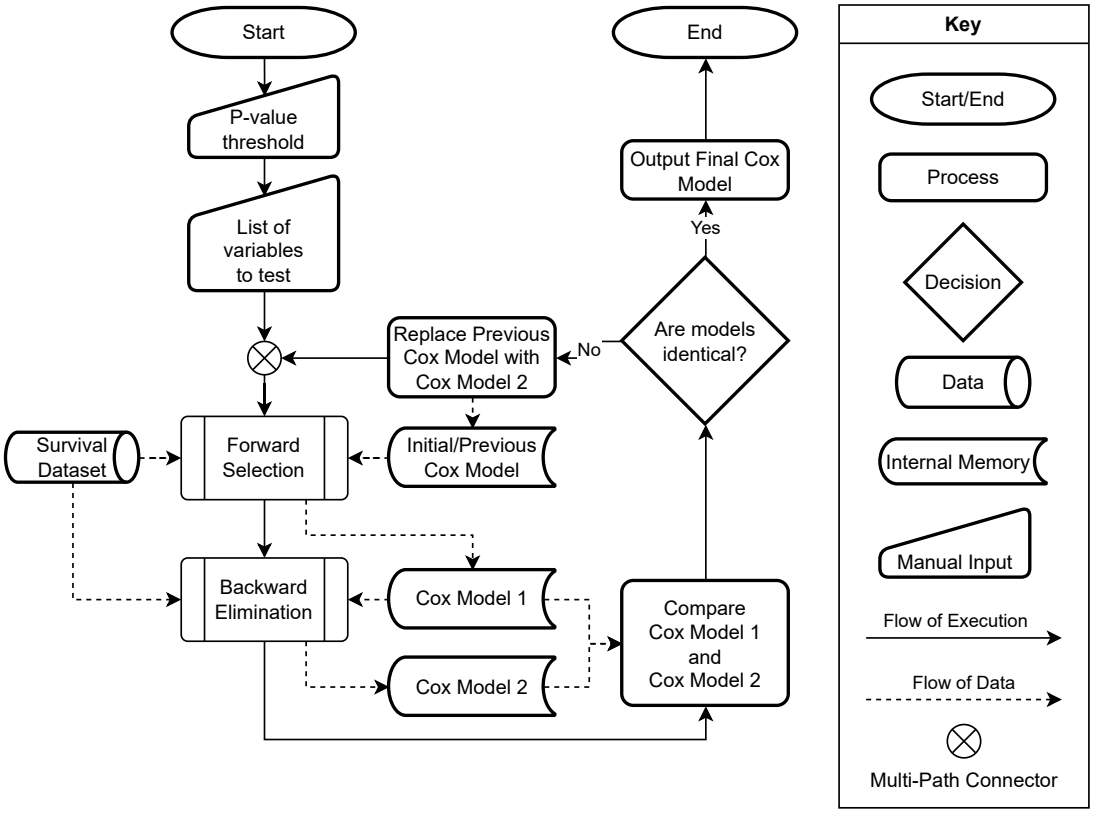


Figure D.1: High-level overview of the iterative algorithm for creating a Cox Proportional Hazards model through repeated forward selection and backward elimination steps.

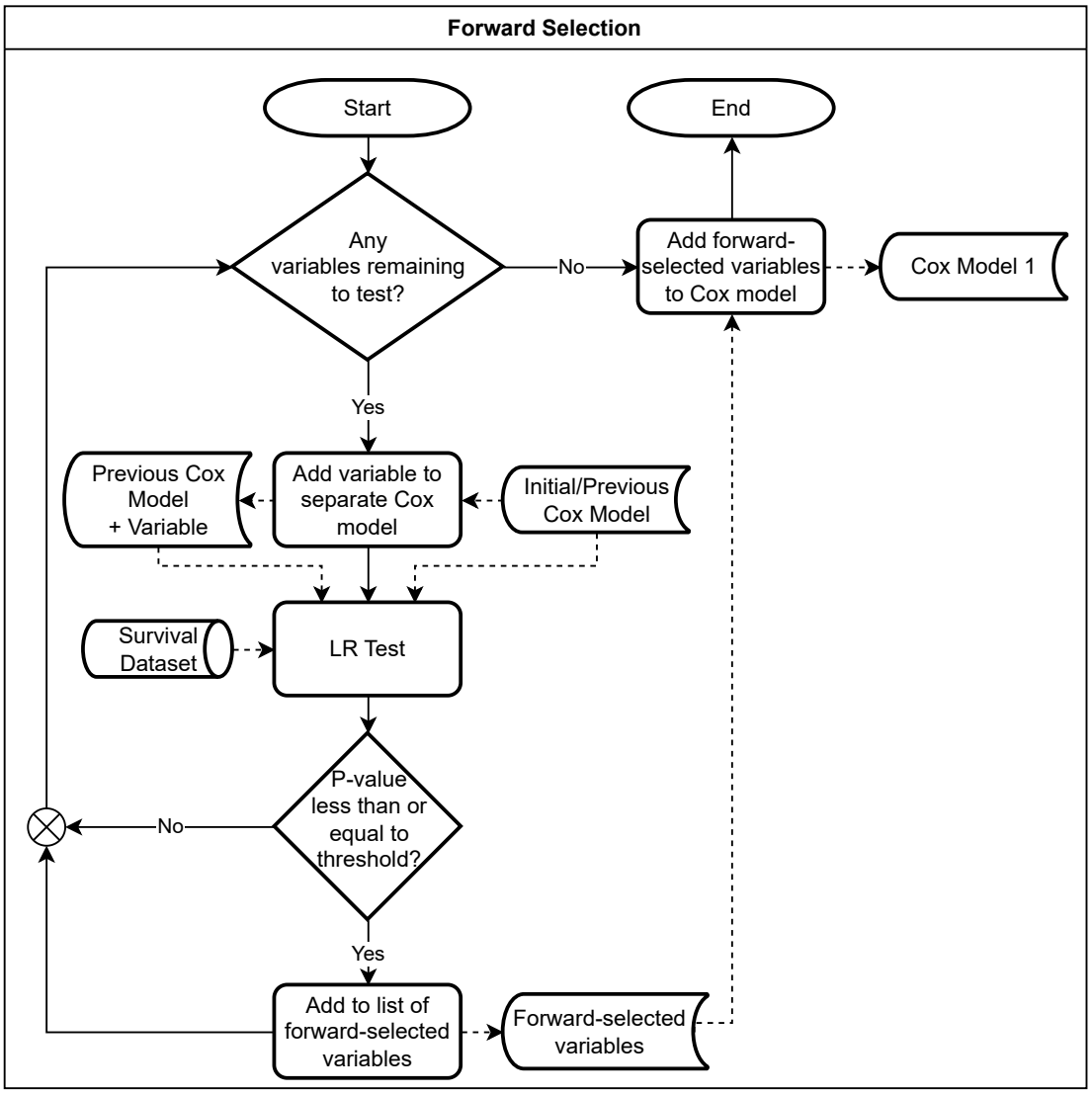


Figure D.2: The forward selection process compares the initial (typically empty) model or model from the previous iteration to a 'parent' model with just a single variable added. All variables that pass this forward selection process are subsequently added to the Cox model.

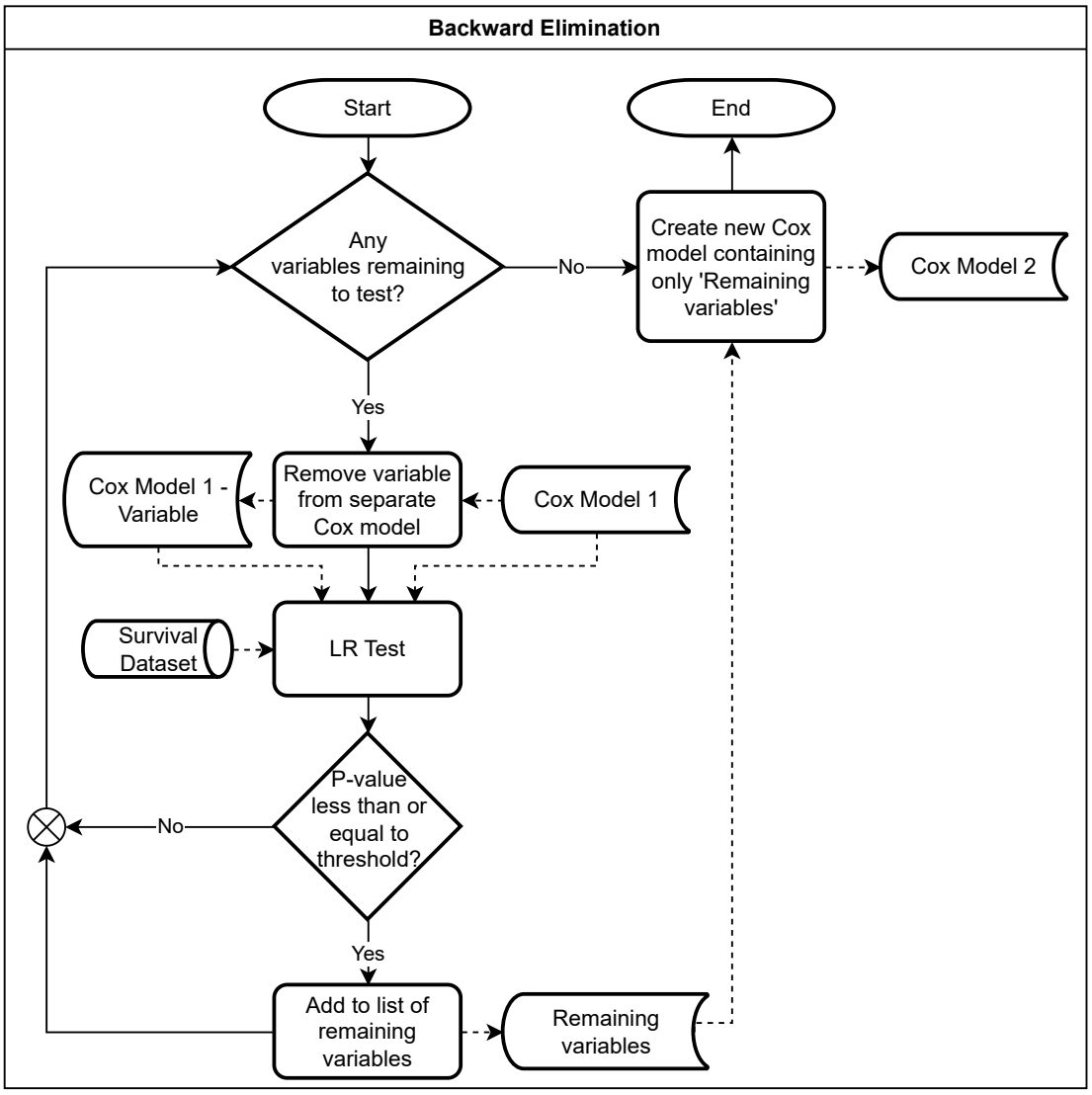


Figure D.3: The backward elimination process compares the model output from the forward selection process to a model with a single variable removed, the LR test is then repeated and any variables that no longer pass the test in the presence of other variables are removed.

D.2 Generating Realistic Survival Times

The Cox-Weibull survival duration generation formula as given in[136] is shown below:

$$T = \left(-\frac{\log(U)}{\lambda \exp(\beta' x)} \right)^{1/\nu}$$

and the Cox-Gompertz survival duration formula is:

$$T = \frac{1}{\alpha} \log \left[1 - \frac{\alpha \log(U)}{\lambda \exp(\beta' x)} \right]$$

U is a uniform random number generated in the range $(0, 1]$, this introduces randomness into the process so that even if two patients have the same characteristics, a different survival time will likely be generated.

β is the vector of coefficients from the Cox model being used to generate survival times and x is a vector of patient covariates obtained from an appropriate lung transplant dataset (a more detailed explanation is contained in section A).

The final parameters correspond to the *shape* (ν / α) and *scale* (λ) parameters that describe the shape of the Weibull distribution fit to the survival data.

D.2.1 Accounting for Donor Characteristics

Post-transplant survival is determined by the combination of recipient characteristics, donor characteristics, and type of transplant (left-lung, right-lung or lung pair).

To facilitate this, variables were categorised as those unique to the patient, those unique to the donor, and 'dynamic' variables that must be determined at the time of transplant within the simulation engine.

Within the simulation, at the time of transplant the $\beta' x$ term is split into three separate terms for patient (pat), donor (don), and dynamic (dyn) variables:

$$\beta' x = \beta'_{pat} x_{pat} + \beta'_{don} x_{don} + \beta'_{dyn} x_{dyn}$$

This allows the simulation engine to combine patient and donor characteristics and add additional years of life to patient's initial ages at listing, decide on left/right/bilateral lung transplant and also simulate the impact of accrued waiting time on post-transplant survival.

D.2.2 Accounting for Informative Censoring

Allocation systems often prioritise patients with higher degrees of clinical urgency. As a result of this, patients that are more likely to die are also more likely to have their waiting list survival time censored due to that patient receiving a transplant. In situations where the risk of censoring is correlated with the risk of mortality, 'informative censoring'

is present and must be accounted for. If informative censoring is not accounted for, the Kaplan-Meier survival curve for waiting list survival will over-estimate waiting list survival. This can have implications for patient rankings due to the waiting list survival not being estimated correctly.

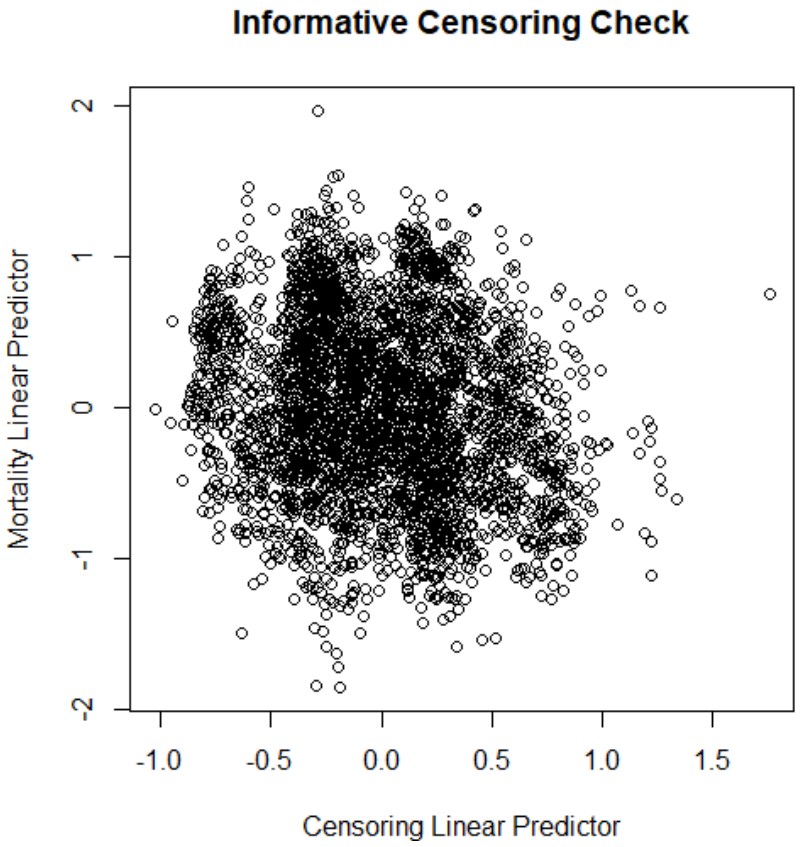
To visualise the degree of informative censoring, a technique from[133] can be used. A Cox model is built using whichever techniques are appropriate and the linear predictor value for mortality is calculated for each patient. Next, the censoring indicators are inverted: a '0' becomes a '1' and vice versa. A second Cox model is then built with these inverted censoring indicators and a second set of linear predictor values are calculated corresponding to the risk of being censored. These two sets of linear predictor values can be plotted against each other to visualise the degree of informative censoring.

Two plots were generated comparing the degree of informative censoring in the OPTN dataset to the NHS-BT dataset and are shown in figure D.4. For the UK/NHS-BT dataset, there is no correlation between the linear predictors indicating no informative censoring, however, possibly due to the use of the LAS in the US, the OPTN dataset shows strong information censoring.

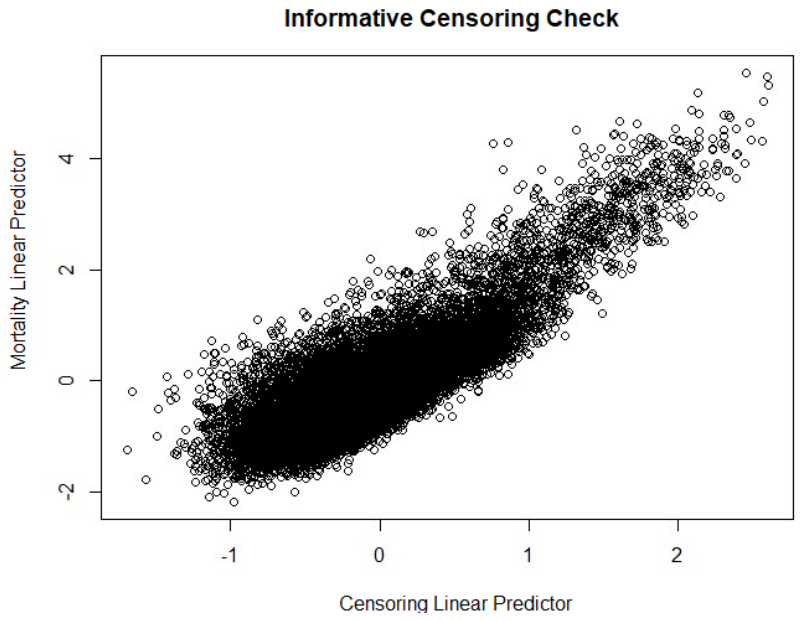
As lung allocation in the UK is the focus of this work and informative censoring is not present it is not necessary to adjust for it. However, if informative censoring is present, a technique called IPCW can be used. For a full explanation on performing these calculations, see.[171]

Applying IPCW will change the coefficients in the Cox model, and the mean of the coefficients will no longer be centred around zero. When generating waiting list survival times the mean offset (μ_o) must be subtracted from the linear predictor, and the equation given in[136] must be adjusted as follows:

$$T = \left(-\frac{\log(U)}{\lambda \exp(\beta' x - \mu_o)} \right)^{1/\nu}$$



(a)



(b)

Figure D.4: **(a)** Informative censoring check on the UK lung transplant dataset. Risk of mortality is not correlated with risk of censoring, so no informative censoring is present in this dataset. **(b)** Informative censoring check on the US lung transplant dataset. Risk of mortality **is** correlated with risk of censoring, so informative censoring is present.

D.3 Simulating Queueing Processes

To accurately simulate the queuing processes involved with UK lung transplantation it was necessary to generate two tables from the UK transplant dataset:

1. A table of time gaps between events (i.e., the time between donors becoming available and the time between patients being added to the waiting list)
2. A table of event frequencies (i.e., when a donor does become available on a particular day, how many donors become available on the same day?)

Poisson processes[172] can be used for modelling queueing processes, however in this case 'roulette wheel selection' was used for simulating the processes of patient listing and donor offers. Roulette wheel selection[173] was used to generate random time gaps between events and also to decide on the number of events.

D.3.1 Generating Time Gaps

To generate a time gap, a random number is generated, which conceptually can be thought of as a random point on a roulette wheel. Each number on the roulette wheel corresponds to the time gap, and the size of each segment corresponding to each number is proportional to the frequency of that time gap being observed in data.

To generate the roulette wheel, the following process is followed:

1. Order event timestamps in chronological order
2. Starting at the first timestamp, tabulate the time gap to the next event
3. Use the table generated in step 2 to generate a frequency table of time gaps

This process is illustrated in the following example:

Event Timestamp	Time Gap
1	1
2	0
2	1
3	1
4	2
6	1
7	1
8	0
8	3
11	2
13	-

Table D.1: For practical considerations, the '0' time gap is not modelled. Roulette selection would generate a time gap of '1' 62.5% of the time, a gap of '2' 25% of the time, and a gap of '3' 12.5% of the time.

Time Gap	Frequency
0	2
1	5
2	2
3	1

D.3.2 Generating Event Frequencies

Once a time gap has been randomly generated, the number of events that will occur must be generated as well.

A similar process is followed to generate a frequency table:

1. Order event timestamps in chronological order
2. Tabulate the frequency of each timestamp
3. Tabulate the frequency of the frequencies of the table generated in step 2

Using the same timestamps as in the previous examples, the frequency table from step 2 would be:

Timestamp	Frequency
1	1
2	2
3	1
4	1
6	1
7	1
8	2
11	1
13	1

The frequency of frequencies table would then be:

Frequency	Frequency of Frequency
1	7
2	2

This would result in a single event occurring 77.8% of the time, and two events occurring 22.2% of the time.

D.3.3 Roulette Wheel Selection Algorithm

Entries from table D.1 will be used to illustrate the roulette wheel selection algorithm.[173]

The algorithm proceeds as follows:

1. Generate an additional cumulative frequency column
2. Generate a uniform random number between 1 and the total cumulative frequency (inclusive)
3. Starting at the lowest time gap, iterate over the cumulative frequency column and find the value that is greater than or equal to the randomly generated number from step 2
4. Use the corresponding time gap in the simulation

Time Gap	Frequency	Cumulative Frequency
1	5	5
2	2	7
3	1	8

Table D.2: Frequency and cumulative frequencies of observed time gaps in data, used for roulette selection

For example, given table D.2, the following steps would be followed:

1. Generate a random number in the range $[1, 8]$, for this example, say '6'
2. Look at the cumulative frequency corresponding to a time gap of '1'. $5 < 6$ so continue.
3. Look at the cumulative frequency corresponding to a time gap of '2'. $7 \geq 6$, so a time gap of '2' is generated.

D.4 Risk-adjusted Benefit and Conditional Survival Policy Methods

D.4.1 Risk Adjusted Benefit

A method was developed to adjust the calculated net benefit for the amount of risk undertaken to achieve that benefit, using similar concepts from.[50]

Immediately post-transplant, the risk of death is higher compared to remaining on the waiting list, however the difference in risk decreases over time. Eventually the post-transplant risk of death will become equal to the risk of death without a transplant, this will be referred to as the 'equity point'.

When prioritising patients, the number of days until the cumulative hazard of death post-transplant equals the cumulative hazard of death remaining on the waiting list is calculated. This requires the the patient's linear predictors for waiting list and post-transplant survival to be calculated (LP_{WL} and LP_{PTX} respectively), and also the shape (ν) and scale (λ) parameters for the waiting list and post-transplant survival functions. The equity point (EP) can then be calculated as follows:

$$\text{EP} = \exp\left(\frac{\text{LP}_{WL} + \text{LP}_{PTX} - \log(\lambda_{WL}) + \log(\lambda_{PTX})}{\nu_{WL} - \nu_{PTX}}\right)$$

The probability of each patient surviving to this point in time post-transplant - P(EP) - can then be calculated:

$$\text{P(EP)} = \exp(\lambda_{PTX} \times \exp(\text{LP}_{PTX}) \times \text{EP}^{\nu_{PTX}})$$

To prioritise the waiting list the allocation policy must specify the weight that should be given to the risk-adjusted benefit (RAB_w). For each patient, the expected waiting list and post-transplant survival durations are calculated (WL and PTX respectively), the risk-adjusted benefit is then calculated as:

$$\text{Net Benefit} = PTX - WL$$

$$\text{Raw Benefit} = (1 - \text{RAB}_w) * \text{Net Benefit}$$

$$\text{Weighted Benefit} = \text{RAB}_w * \text{P(EP)} * \text{Net Benefit}$$

$$\text{Risk-adjusted Benefit} = \text{Raw Benefit} + \text{Weighted Benefit}$$

D.4.2 Conditional Survival

Conditional survival can be calculated as follows:

First the CON is calculated as outlined in:[138]

$$CS(t|s) = \frac{S(s+t)}{S(s)}$$

Where $CS(t|s)$ is the probability of living another t days given a patient has survived s days and S is the survival function.

The survival function can be derived from the Weibull distribution used to simulate survival durations and is given in[136] as:

$$S(t) = \exp(-\lambda t^{\nu})$$

To adjust this survival function for use in the simulation engine the waiting list scale and shape parameters are used (λ_{WL} and ν_{WL} respectively), and the equation is adjusted to take the patient's waiting list linear predictor (LP_{WL}) into account:

$$S(t) = \exp(-\lambda_{WL} \exp(\mathrm{LP}_{WL}) t^{\nu_{WL}})$$

This adjusted survival function can then be used to calculate the conditional survival - CS - for any patient, given their linear predictor of risk and the waiting list survival parameters. Each policy also specifies a WL-Ratio and a PTX-Ratio for the relative weight of waiting list and post-transplant survival, this can all be combined to give an adjusted survival score:

$$(\text{PTX-Ratio} \times \text{PTX-Survival}) - (\text{WL-Ratio} \times \text{WL-Survival} \times CS(t|s))$$

D.5 Flowcharts and Pseudocode

There are four main shapes used in the flowcharts in the following sections: rectangles, circles, diamonds and hexagons. There are also two types of arrows: filled and hollow.

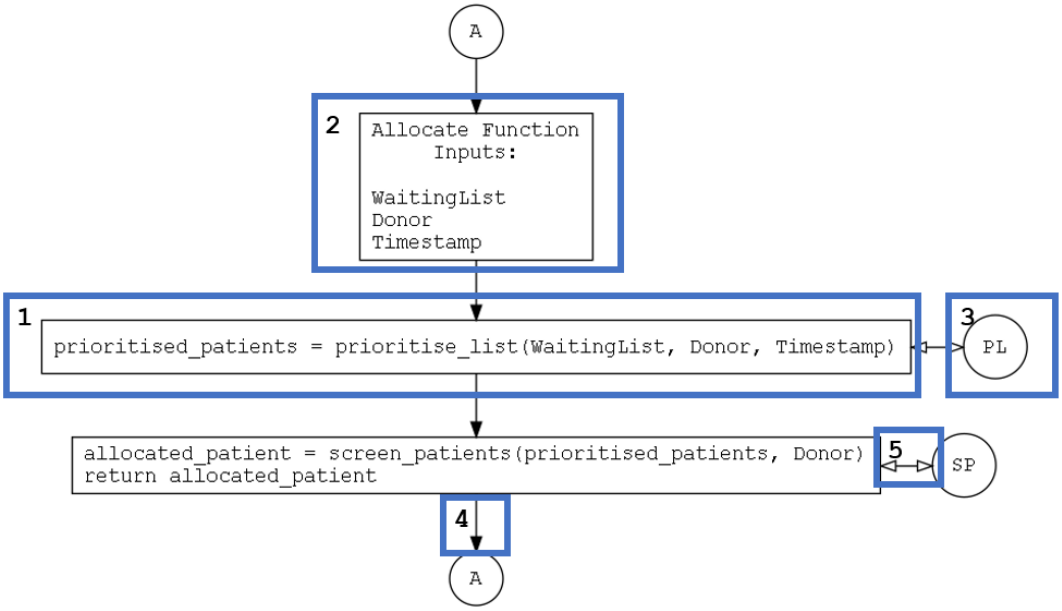
A rectangle defines a function, containing pseudocode explaining the sequence of steps to accomplish a specific task (see box 1). A rectangle is also used to define a function header, which gives the name of the function along with any required inputs (see box 2).

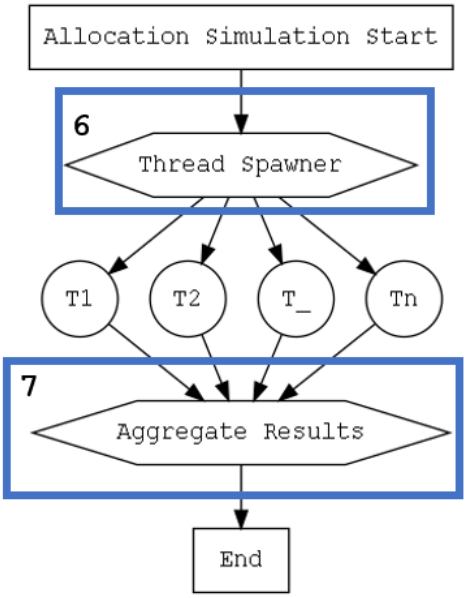
Circles are used as a convenient way to reference other functions (see box 3). An abbreviation of the function name is contained in the circle, with arrows showing the flow of execution. A solid filled arrow shows the primary flow of execution (see box 4), with hollow white arrows showing function calls (see box 5).

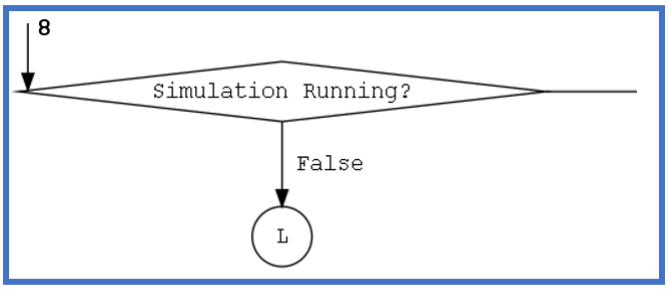
A single-headed hollow arrow shows a function or process that is initiated but doesn't return a value to the calling function. A double-headed hollow arrow indicates a function that returns a value (or the flow of execution) to the calling function (see box 5).

A hexagon indicates where the flow of execution branches into multiple parallel, independent threads of execution (see box 6), or where the results of multiple threads are aggregated (see box 7).

A diamond indicates a decision/branch in the flow of execution, depending on the condition contained in the diamond (see box 8).




Allocation Simulation Start
6
Thread Spawner
T1
T2
T_
Tn
7
Aggregate Results
End


8
Simulation Running?
False
L

D.5.1 Entry Point

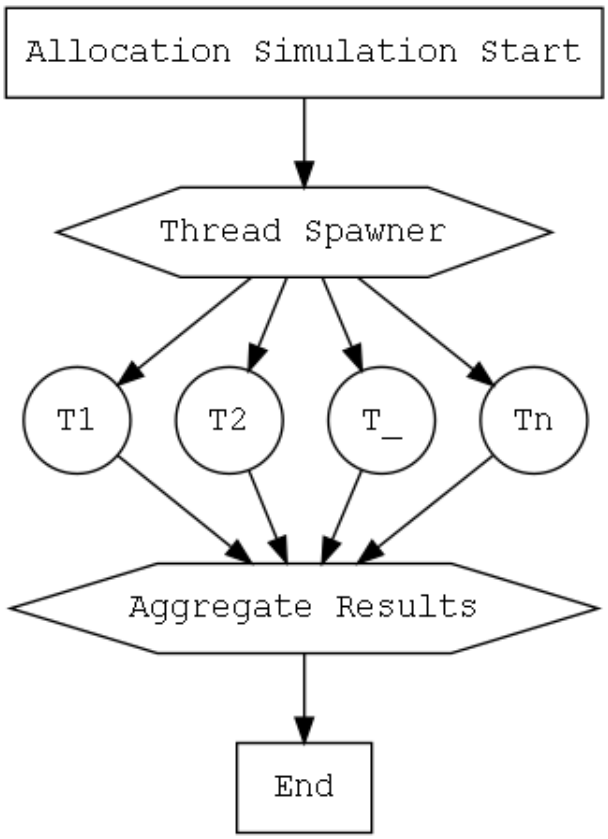


Figure D.5: At a high level, the simulation spawns a number of threads (T1 - T_n), each thread simulates 20 years of allocation in parallel with other threads. Once the required number of simulations in each thread have completed, the results are aggregated.

At the beginning of the simulation process, one or more simulation threads are spawned. Each thread is independent of the others and will perform a fixed number of simulation runs over a pre-specified simulation period in parallel with the other threads. In the case of the UK lung allocation simulations, 8 threads performed 5 simulation runs of 20 years each, resulting in a total of 40 simulation runs per simulated policy (and a speed increase of 8 times).

D.5.2 Simulation Thread

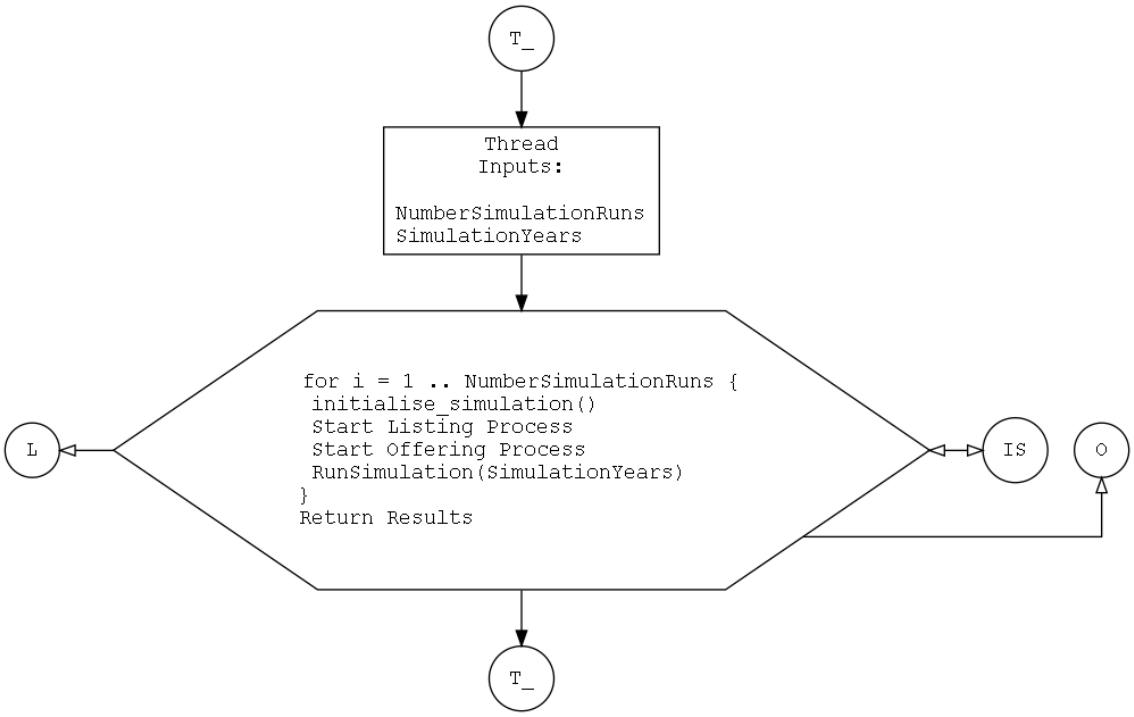


Figure D.6: Each simulation thread will run for `SimulationYears` and repeat `NumberSimulationRuns` times. For each simulation run, the simulation engine is initialised and then the patient listing and donor offering processes are started in parallel.

Each simulation thread takes two parameters as input: `NumberSimulationRuns` and `SimulationYears`. `NumberSimulationRuns` specifies how many unique simulation runs should be performed, and `SimulationYears` specifies the number of simulated years each simulation should run for. For simplicity one year is defined as 365 days.

Before the two main processes are started, the simulation engine must be initialised to ensure each simulation run is unique. The `initialise_simulation()` function handles the initialisation of the simulation engine and is abbreviated `IS` in the flow charts.

D.5.3 Simulation Initialisation

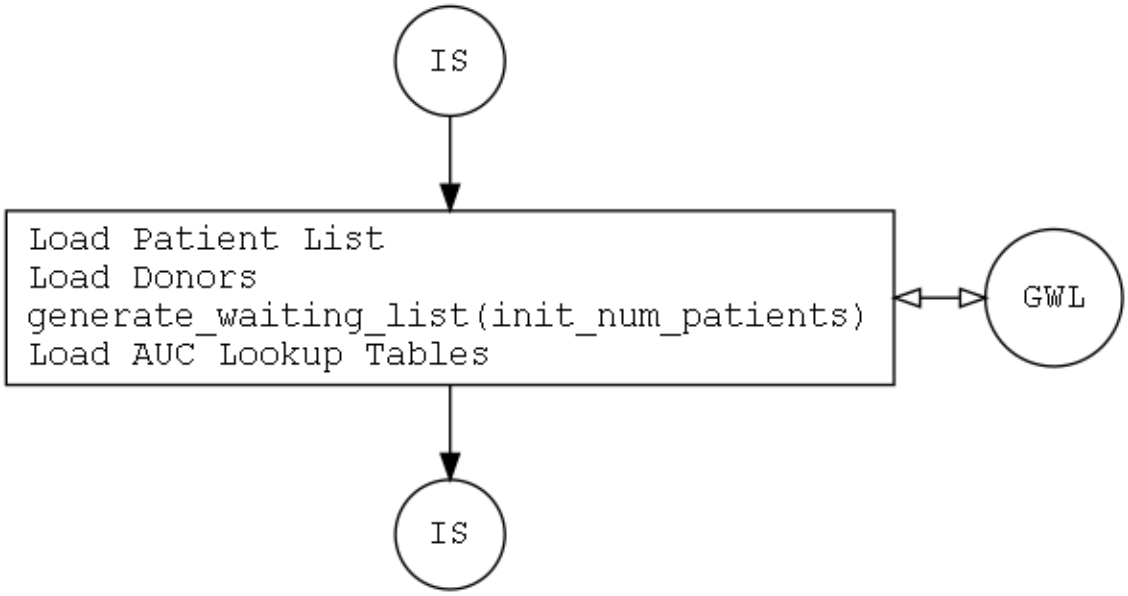


Figure D.7: The .csv files containing patient data and donor data are loaded into memory, along with the area-under-the-curve (AUC) lookup tables to quickly calculated expected survival durations. The waiting list is randomly initialised with `init_num_patients` patients.

To ensure each simulation is unique, the waiting list is initialised with `init_num_patients`, randomly selected from the loaded patient list. The patient list is a CSV file containing one patient per row, with each column corresponding to a variable associated with that patient. Donor data is also stored in a CSV file and is loaded into memory for later use. The `generate_waiting_list()` function handles the creation of the waiting list and is abbreviated `GWL` in the flow chart.

Allocation is based on expected survival duration, calculated using the AUC of the survival curve, integrated out to 20 years. In order to speed up computation, two AUC lookup tables are loaded into memory corresponding to waiting list survival and post-transplant survival. One column holds the LP value (rounded to 3 decimal places) and the other column contains the pre-computed AUC value (i.e., the expected survival for that LP value).

D.5.4 Generating Waiting List

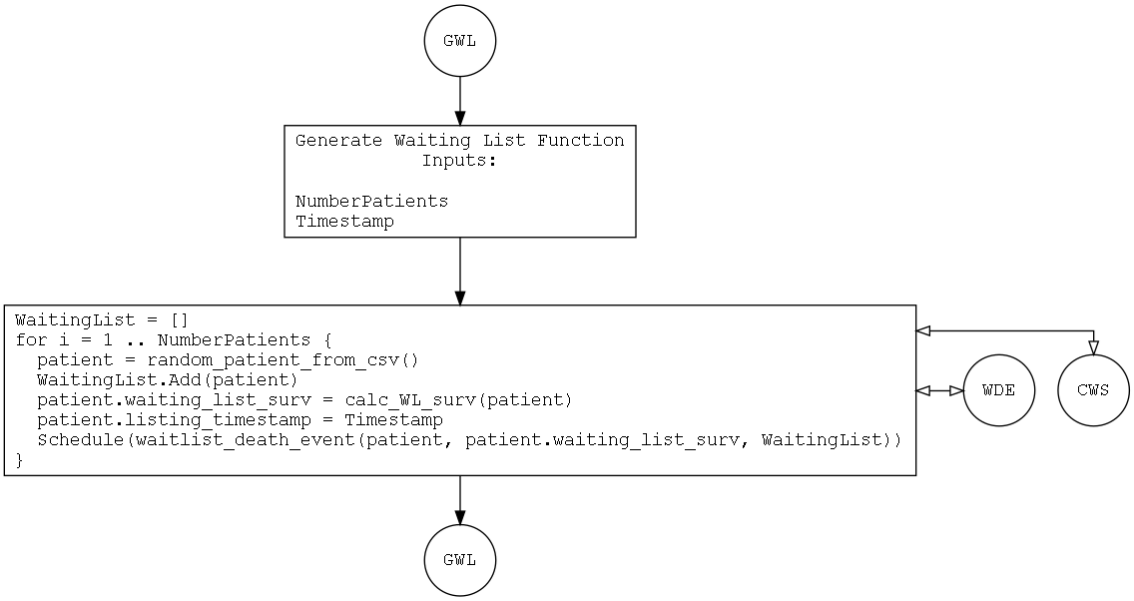


Figure D.8: A blank waiting list is populated with `NumberPatients` patients. For each patient a randomised waiting list survival duration is generated, and a waiting list death event is scheduled to occur once the survival duration has elapsed.

The waiting list is generated by looping `NumberPatients` times. First a random patient is selected using the `random_patient_from_csv()` function from the patients list (note: this list was loaded in the `initialise_simulation()` function). This patient is added to the waiting list and then a random waiting list survival duration is generated.

Next, the timestamp of the patient being listed is stored, this will be used later to calculate total waiting time. Finally, a waiting list death event is scheduled to occur after `waiting_list_surv` days.

The details of `calc_WL_surv()` (`CWS`) and `waitlist_death_event()` (`WDE`) will be given in later sections.

D.5.5 Patient Listing Process

The patient listing process runs in a loop as long as the simulation is still running. To accurately simulate the listing process four variables are used:

1. `WaitingListGaps` contains an ordered list of time gaps (in days) between subsequent listings observed from data
2. `WaitingListGapFreqs` contains a list where the element at each index corresponds to the frequency of the time gap at the same index in `WaitingListGaps`
3. `WaitingListCounts` is similar to `WaitingListGaps`, but contains an ordered list of the number of listings occurring on the same day
4. `WaitingListCountFreqs` is similar to `WaitingListGapFreqs`, but corresponds to the frequencies of listing occurrences stored in `WaitingListCounts`

Next, roulette selection is used to generate a time `gap` - which is the time gap in simulated days before another listing event occurs - and also a number of `new_patients` which will be added to the simulated waiting list.

The number of `new_patients` to add grows annually at a rate determined by `PopulationGrowthRate`, resulting in the number of listings increasing in line with population growth.

For each new patient added to the waiting list, the same sequence of events occur as specified in the `generate_waiting_list` (`GWL`) function.

The two key functions in the listing process are the `calc_WL_surv` (`CWS`) and `waitlist_death_event` (`WDE`) functions. These will be detailed in this section after the roulette selection and population growth processes are explained.

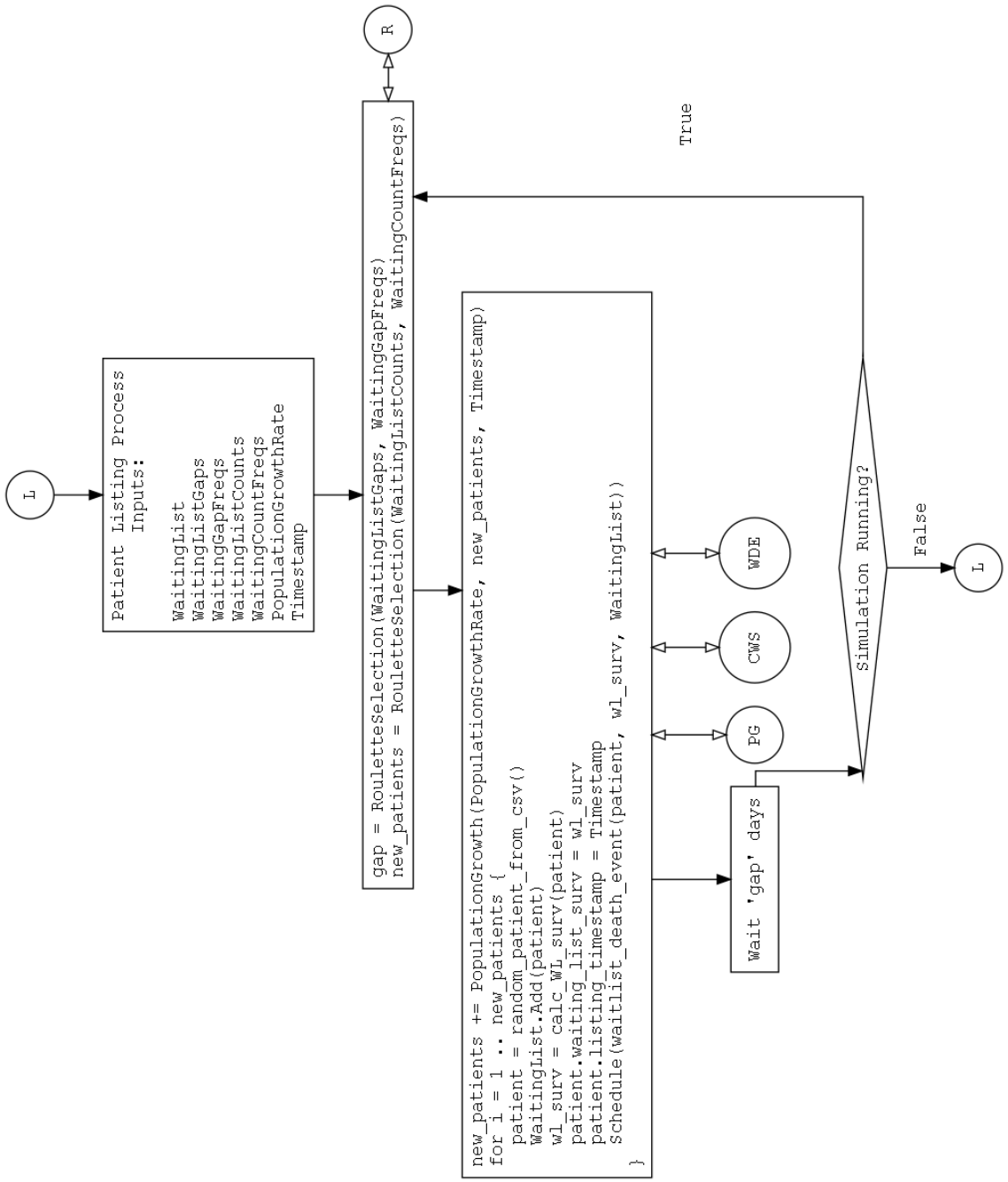


Figure D.9: Flowchart showing how the listing of new patients on the waiting list was simulated.

D.5.6 Roulette Selection

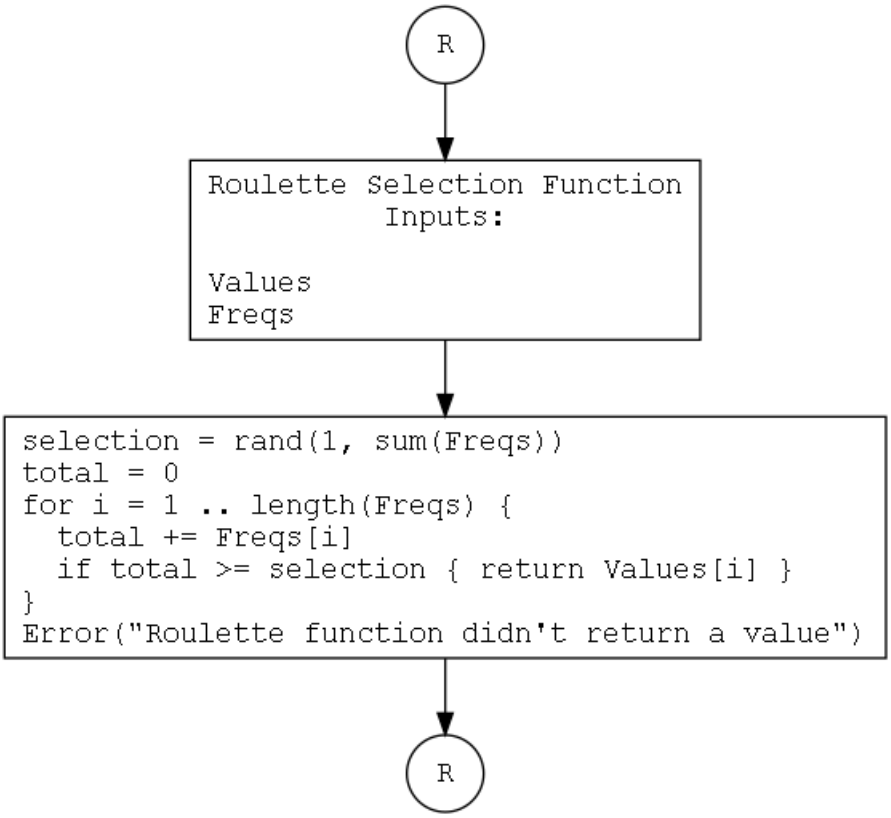


Figure D.10: The roulette selection function takes a list of `Values` and `Freqs` corresponding to the frequency of each value occurring. A randomised number is generated that will select a value from `Values` with the corresponding probability contained in `Freqs`.

The roulette selection function is used to randomly select from a range of `Values` with probability frequencies `Freqs`.

The first step is to generate a uniform random number between 1 and the sum of all values contained in the `Freqs` array (inclusive). Next, the frequencies are iterated over and added to a running `total`. If the running `total` is greater than or equal to the random `selection` then the corresponding value from the `Values` array is returned.

Finally, the function should have returned before the loop terminates, if not an error is raised. This is a sense check to ensure the function always returns a value.

D.5.7 Population Growth

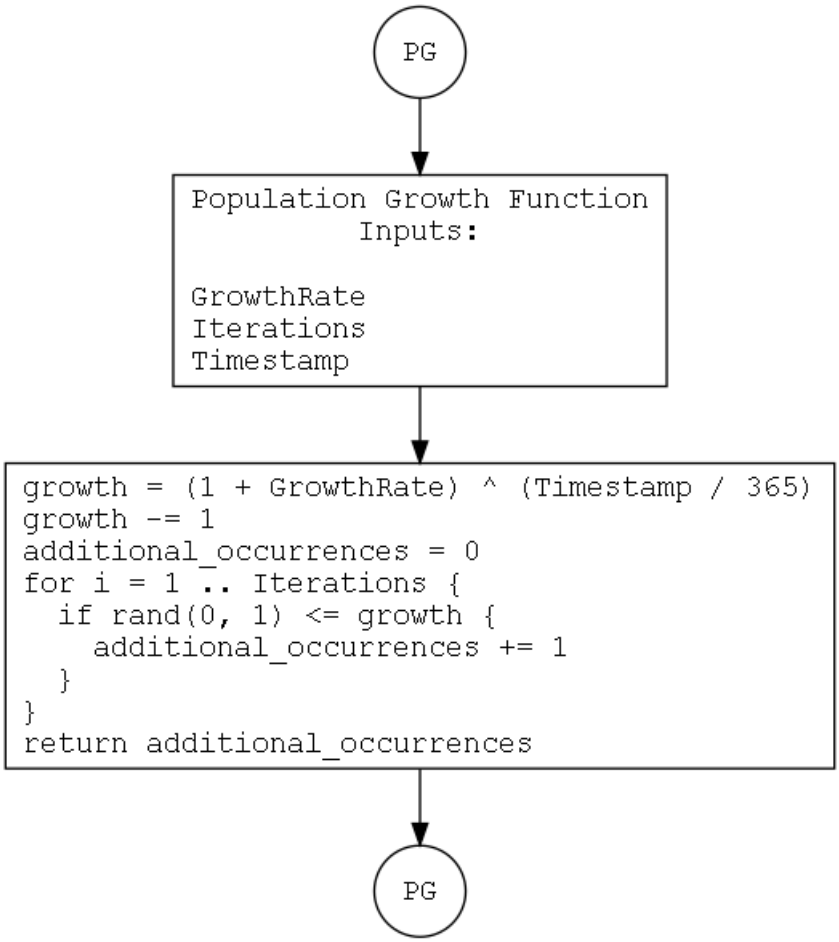


Figure D.11: Flowchart showing how population growth was simulated.

The population growth function (`PG`) is a generic function that can simulate a % increase in the number of occurrences of events annually as the size of the population grows. This function is used for three purposes:

1. Simulate the increase in the number of listings as a result of population growth
2. Simulate the increase in the number of donors offered for allocation as a result of population growth
3. Simulate an increased number of donors being available for allocation as a result of increased utilisation / donation rates

As the simulation progresses over multiple years, it is assumed that the population of the country will continue to grow. This will result in an increase in the number of patients added to the waiting list, and also the number of donors that are available. In the UK dataset the number of listings and donor offers grow on average 3% per year.

The additional population growth is calculated as:

$$\texttt{growth} = (1 + \texttt{PopulationGrowthRate})^{\frac{\texttt{Timestamp}}{365}}$$

This will result in `growth` having a value ≥ 1, the next step is to subtract 1 to set the range to $[0, 1]$. To calculate the number of `additional_patients`, a loop executes `new_patients` times. For each `new_patient` that would be added, there is a probability determined by `growth` that an additional patient will also be listed. A random value is generated in the range $[0, 1]$ and if it is $\leq$ `growth` then the number of `additional_patients` are incremented.

It is important to note that this technique of simulating additional listings only works up to an additional growth of 100%. This is due to adding *one* additional listing if `rand(0, 1)` $\leq$ `growth`. If the population growth is $\geq 100\%$ then `rand(0, 1)` $\leq$ `growth` will always evaluate to **true**, therefore for every new listed patient there will always be one additional patient, resulting in a maximum increase of 100%.

D.5.8 Calculating Waiting List Survival

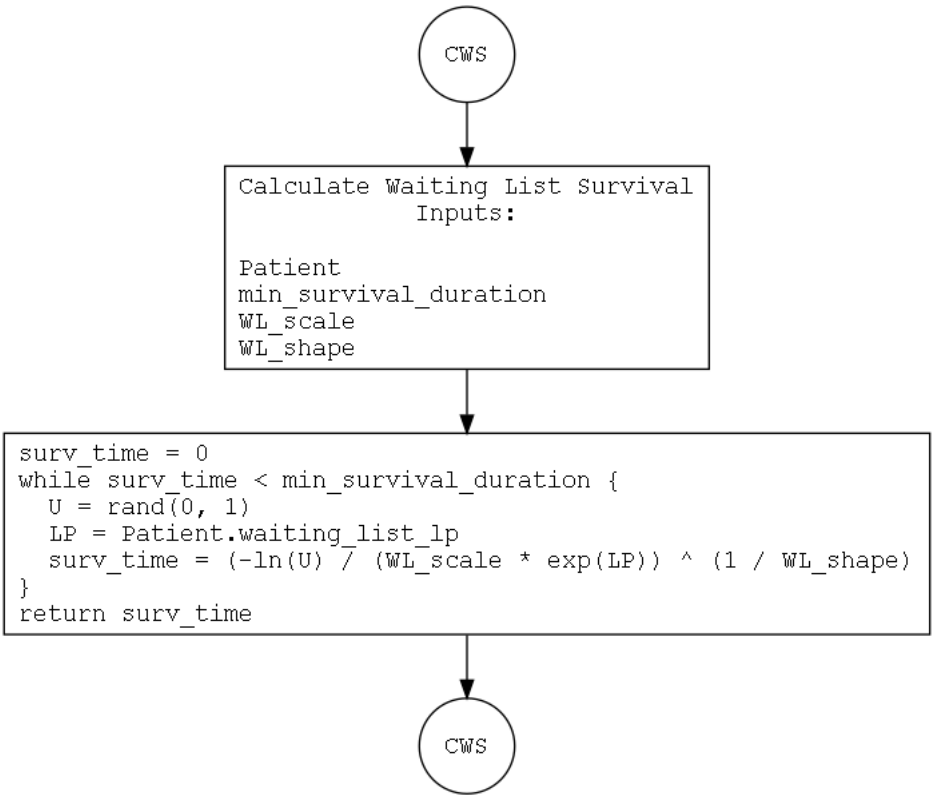


Figure D.12: Flowchart of the process used to generate random but realistic waiting list survival durations.

The methods for generating realistic survival times are outlined in more detail in section 4.2.1. It is unrealistic that a patient is listed with a very short life expectancy on the waiting list. To account for this a `min_survival_duration` is specified, in the case of UK lung allocation simulations, a minimum threshold of 14 days was used. The main loop repeats until a survival duration $\geq$ `min_survival_duration` is generated.

The survival duration is randomised so that a patient with the same characteristics will survive a different duration on the waiting list with each simulation. The survival durations are generated from a probability distribution determined by `WL_scale` and `WL_shape`, corresponding to the scale (λ) and shape (ν) parameters of the Weibull distribution fit to the waiting list survival data.

The survival duration is adjusted to the patient's individual risk profile. This is accomplished using the linear predictor (`LP`) value calculated for that patient, which is the dot-product of the variables associated with the patient (x) and the coefficients from the Cox model (β). The `LP` value will shift the probability distribution to lower values as `LP` increases and vice versa.

D.5.9 Waiting List Death Event

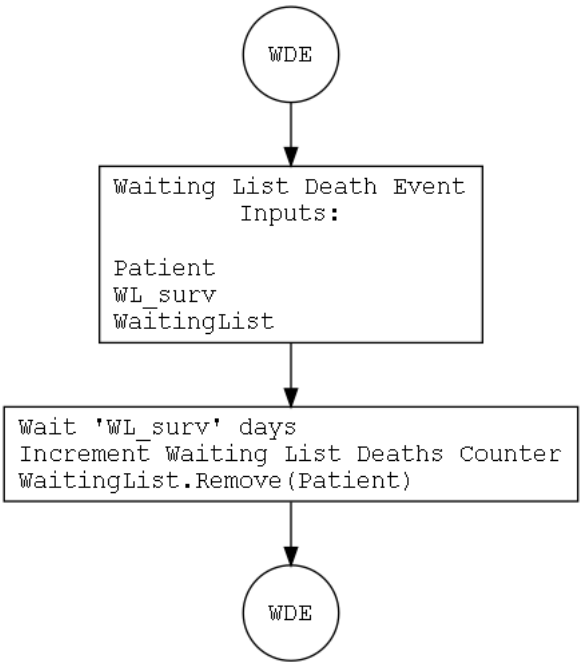


Figure D.13: Flowchart showing the steps taken when a 'waiting list death event' occurs.

Once a randomised waiting list survival duration has been generated for a patient, a waiting list death event is scheduled to occur once the survival duration (`WL_surv`) has elapsed. If a patient is transplanted this event is cancelled and will not occur, however if a waiting list death event does occur, the number of waiting list deaths is incremented and the patient is removed from the `WaitingList`.

D.5.10 Donor Offering Process

The donor offering process follows a similar pattern to patient offering with a few key differences:

1. A donor offering event triggers an 'Allocation' event where the allocation policy is simulated
2. The baseline donor offering rate can be increased to simulate different scenarios using the `DonorIncreaseRate` variable

The time gap between donors being offered for allocation and the number of donors available for allocation are generated using the roulette selection function. Next, if `DonorIncreaseRate` is set to a value greater than 0, the `PopulationGrowth` function is used to randomly increase the number of donors available, with an average rate of `DonorIncreaseRate`. For example, a `DonorIncreaseRate` of 0.05 would result in 5% more donors being available for allocation compared to the baseline calculated from the transplant data set.

The next step is to simulate population growth, this is accomplished in the same way as with patient listings, however the key difference is that this compounds with the `DonorIncreaseRate`. For example, with a population growth rate of 3% and a 5% increase in the donor rate, after two years there would be an 11.4% increase in the donor rate:

$$1.05 \times 1.03^2 = 1.114$$

Once the number of donors to be offered has been determined, the main donor offering loop is executed `new_donors` times. Within the loop a new donor is selected randomly from the dataset, then the allocation policy of interest is simulated (`A` in the flow charts). There may be scenarios where there is no suitable recipient to match with the donor, in this case the loop continues with the next donor, or terminates if all donors have been simulated.

If a recipient has been identified from the allocation policy, the donor is assigned to the recipient (`allocated_patient.assigned_donor`), the recipient is removed from the waiting list, the recipient's waiting list death event is cancelled, and finally post-transplant survival metrics are calculated.

Once all donors have been allocated / discarded, the donor offering process waits `gap` days (as determined by the roulette selection earlier) before repeating. This process will run until the simulation stops running at the end of the simulation period.

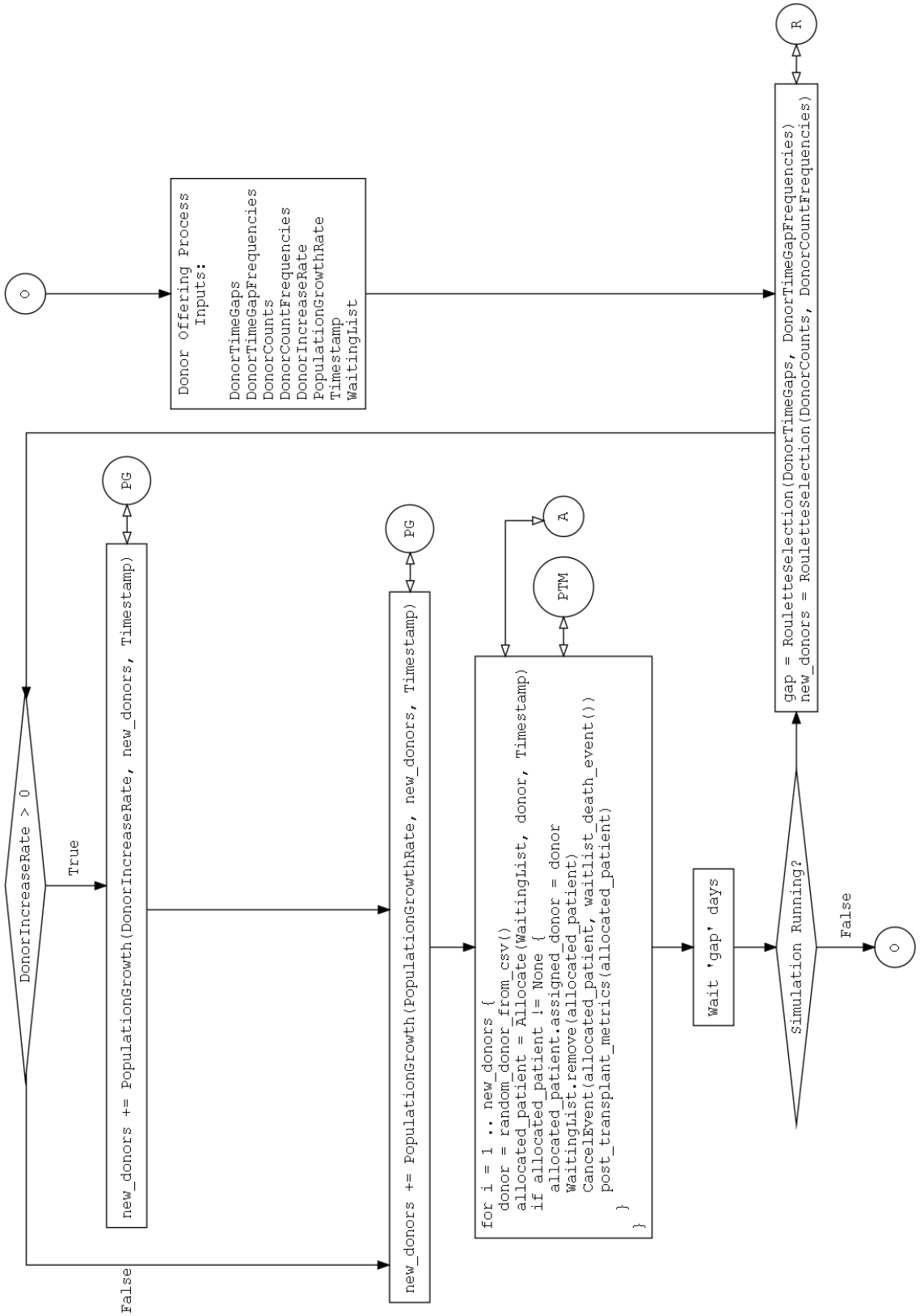


Figure D.14: Flowchart showing the donor offering process, which triggers the allocation policy and selection of a candidate for transplant.

D.5.11 Donor Allocation Function

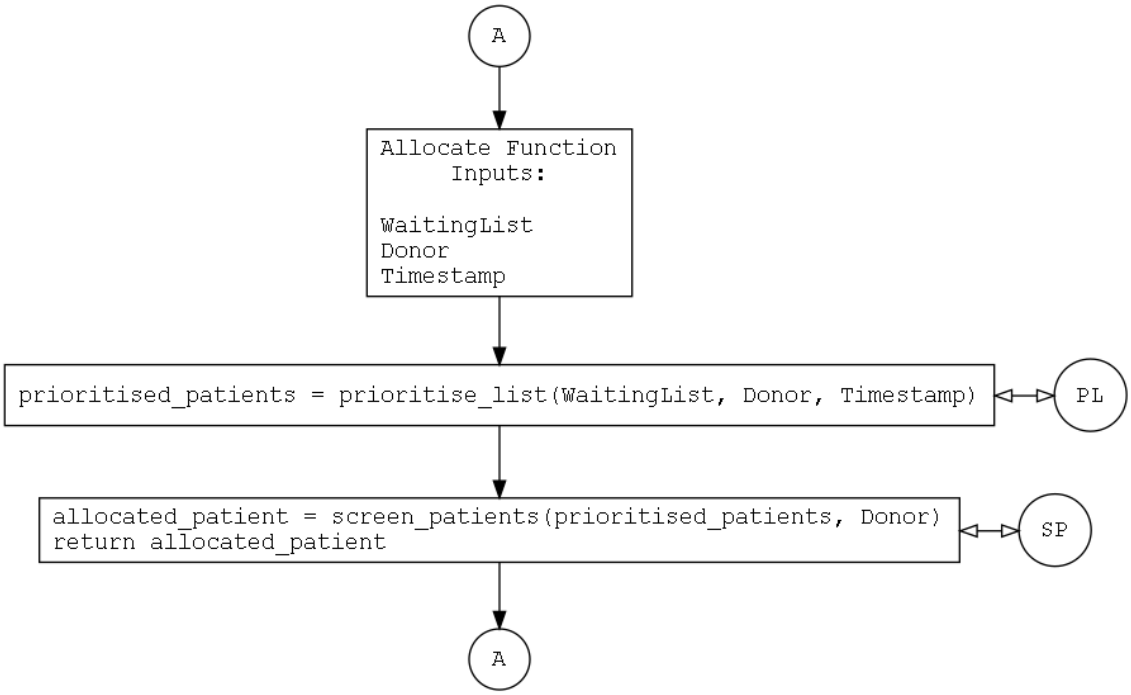


Figure D.15: Flowchart showing the two main steps taken when allocating a donor: prioritising the waiting list and screening incompatible patients.

The allocation process is split into two processes, the first is the `prioritise_list` function (`PL`), which ranks the patients on the waiting list from highest to lowest priority according to the allocation policy. Next, screening criteria are applied to ensure only compatible patients are matched with the donor (compatibility in this case is simulated using ABO and height matching criteria).

D.5.12 Prioritising Waiting List Function

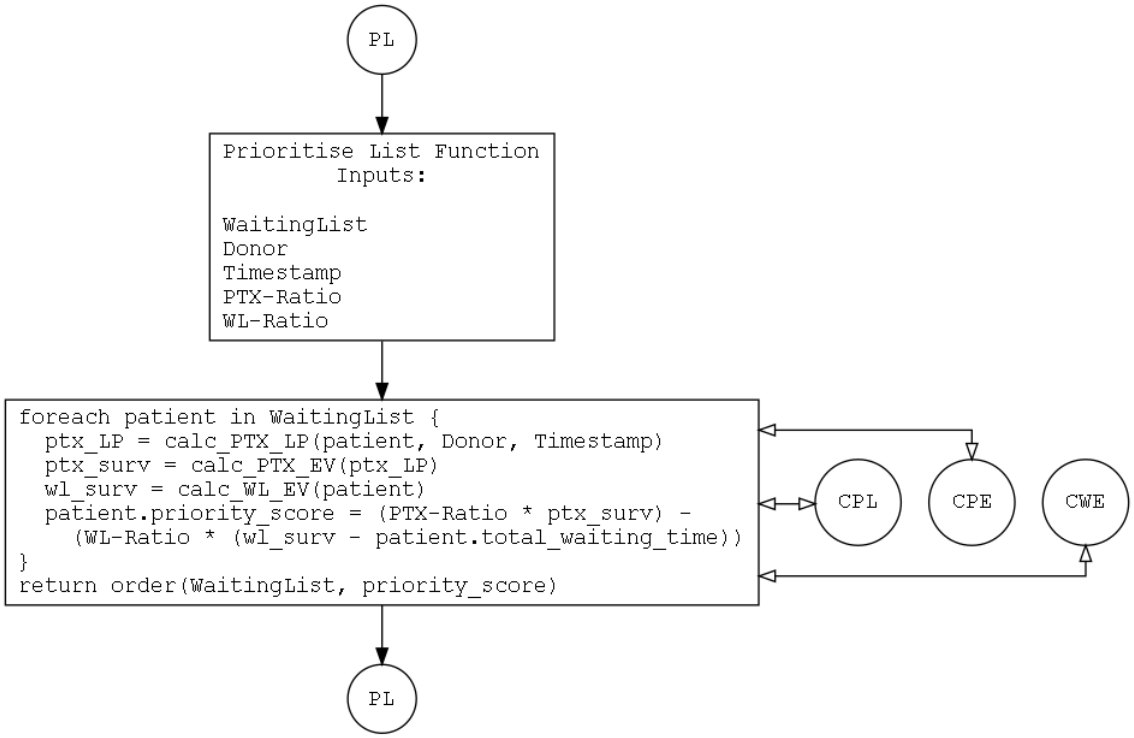


Figure D.16: Flowchart showing how the waiting list is prioritised according to the allocation scoring system in place.

The waiting list is prioritised by first calculating a `priority_score` for each patient. First the linear predictor (`LP`) for post-transplant survival is calculated using the combination of patient and donor characteristics using the `calc_PTX_LP()` (`CPL`) function.

The `ptx_LP` value is then mapped to an expected survival duration using the `calc_PTX_EV()` (`CPE`) function. The expected waiting list survival duration is also calculated using the `calc_WL_EV()` (`CWE`) function. A `priority_score` can then be calculated using the expected waiting list and post-transplant survival durations. The priority score depends on the relative importance of post-transplant survival (`PTX-Ratio`) and waiting list survival (`WL-Ratio`). These priorities are determined by the allocation policy.

If adjusted priority scores are being used, each patient's `total_waiting_time` is subtracted from their expected waiting list survival duration, meaning that if there were two patients with identical characteristics, the patient with the longer waiting time will be higher priority. This subtraction is not performed for unadjusted priority scores.

Finally the waiting list is returned to the allocation function with all patients ordered from highest to lowest priority score.

This function calls three other functions: `calc_PTX_LP` (`CPL`), `calc_PTX_EV` (`CPE`) and `calc_WL_EV` (`CWE`), which are detailed in the following sections.

D.5.13 Calculating Post-transplant Linear Predictor

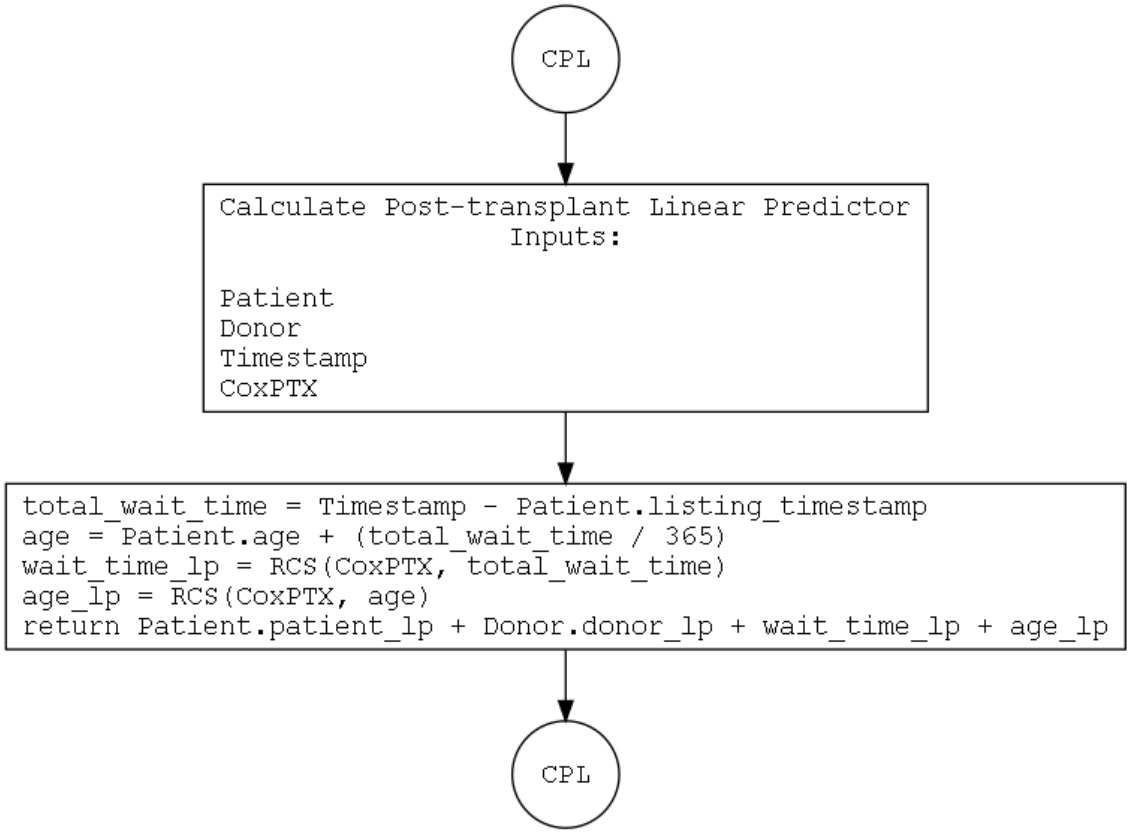


Figure D.17: Flowchart showing the process where donor, recipient, and other dynamic variables are combined to calculate the linear predict for post-transplant survival.

First, the total waiting time for the patient is calculated, then the additional *fractional* years of life since listing are added to the patient's age. Next, the individual linear predictors for waiting time and age are calculated using restricted cubic splines (`RCS`) and the corresponding coefficients from the post-transplant Cox model (`CoxPTX`).

The linear predictor for variables intrinsic to the patient are stored in `Patient.patient_lp`, and for variables intrinsic to the donor, `Donor.donor_lp`.

The final post-transplant linear predictor is then simply: `Patient.patient_lp + Donor.donor_lp + wait_time_lp + age_lp`.

D.5.14 Calculating Expected Survival Durations

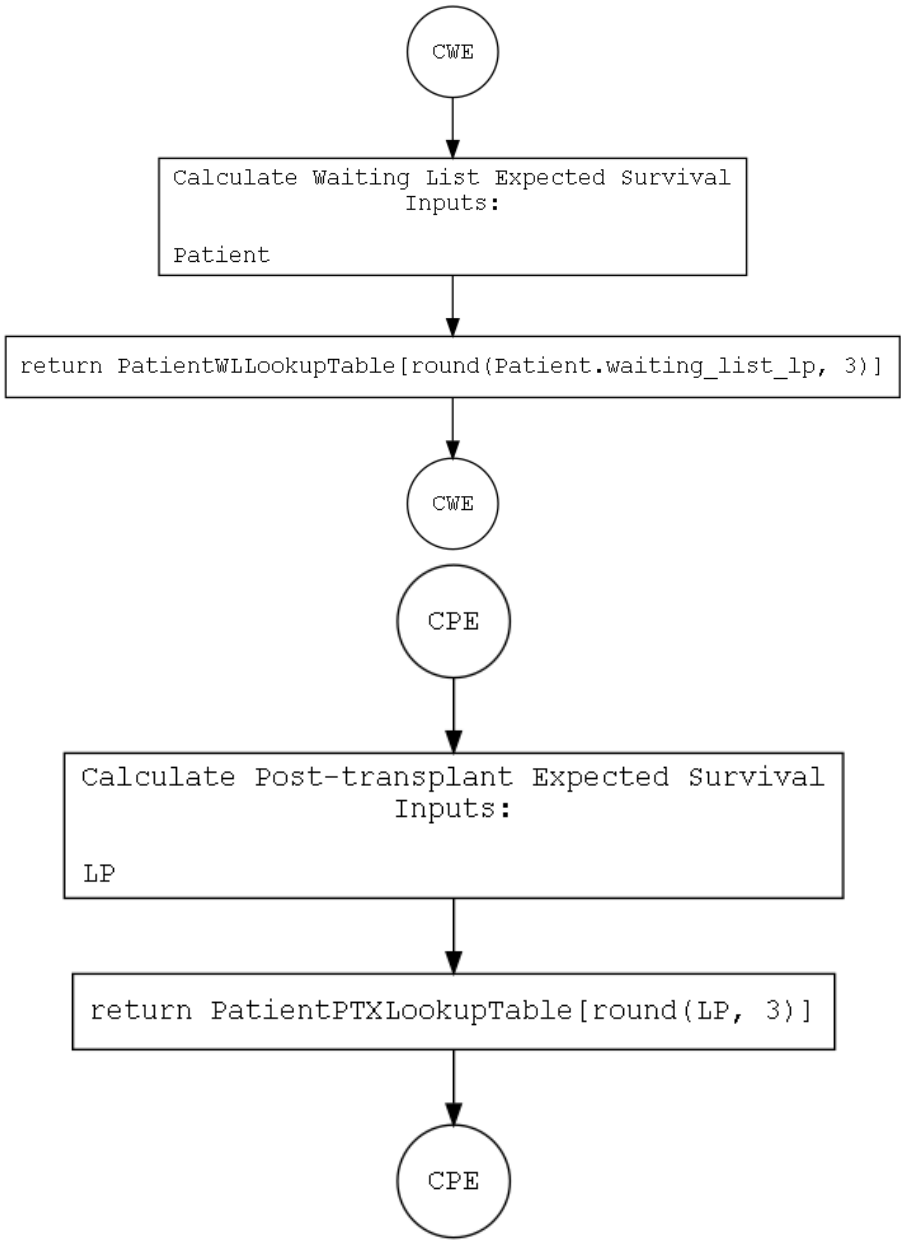


Figure D.18: Two flowcharts showing how the expected survival duration on the waiting list is calculated (`CWE`) and post-transplant (`CPE`)

There are two separate lookup tables for mapping linear predictors for waiting list and post-transplant survival to expected survival durations. Expected survival in the case of UK lung allocation simulations is the restricted mean up to 20 years.

The linear predictor value is rounded to three decimal places and then matched to a survival duration in the corresponding lookup table.

D.5.15 Screen Waiting List Function

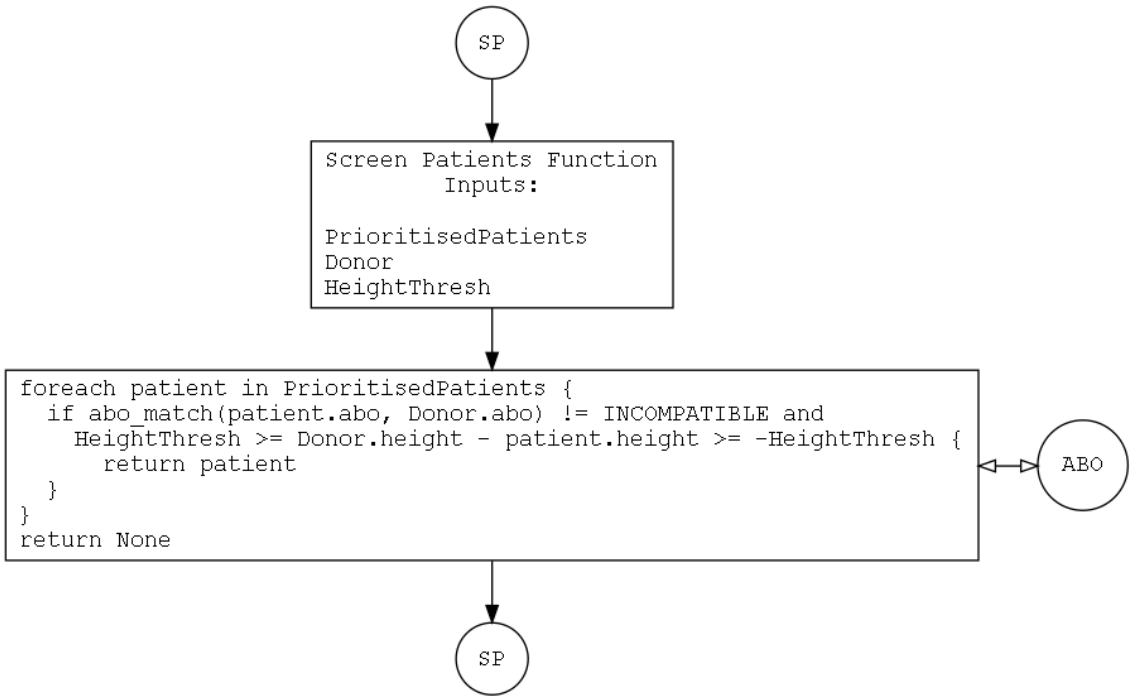


Figure D.19: Flowchart showing how the waiting list is screened to only allow suitable candidates to be matched to a donor.

The screening process iterates over every patient on the waiting list, starting at the highest priority patient and progressing towards lower priority patients. For each patient the blood group ABO compatibility must not be `INCOMPATIBLE` (and therefore, must be `COMPATIBLE` or `IDENTICAL`). The donor-recipient height difference must also be within the height difference threshold specified by `HeightThresh`.

The loop will continue until the highest priority patient meeting the screening criteria is identified, this patient will be allocated the donor lungs in the allocation function.

D.5.16 ABO Matching

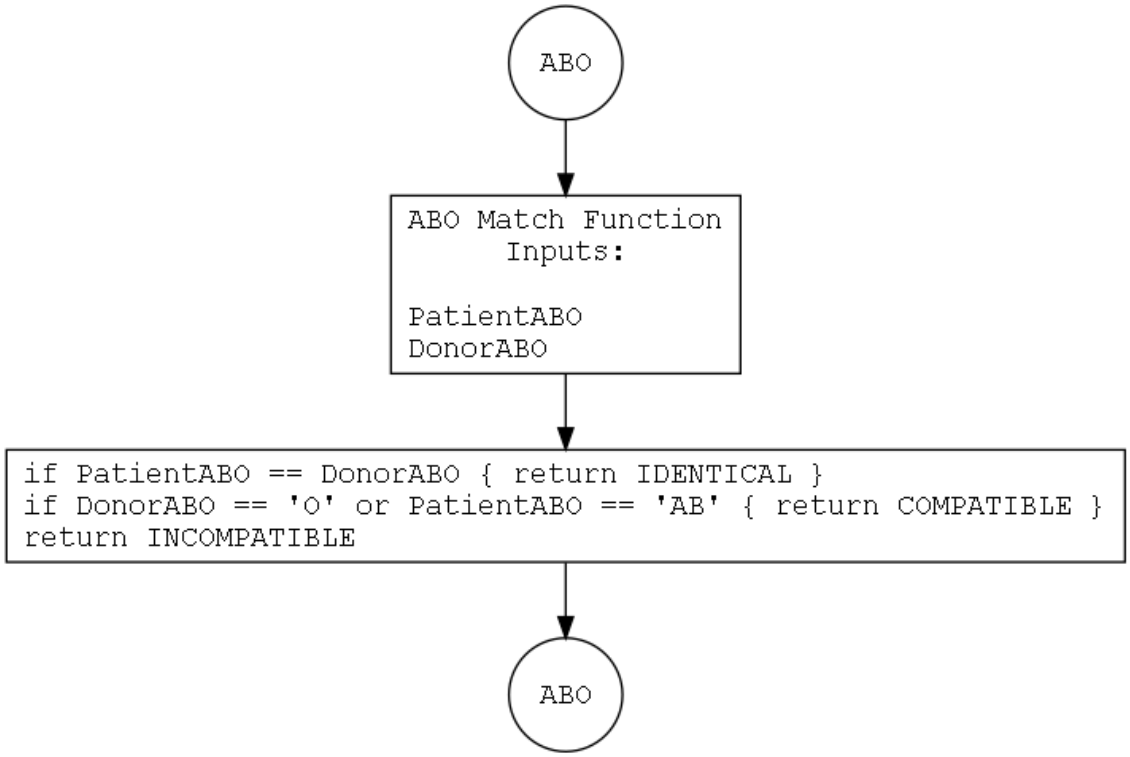


Figure D.20: Flowchart showing the process where donor and candidate blood types are determined to be 'identical', 'compatible', or 'incompatible'.

The ABO matching function first checks if the donor and recipient blood groups are the same, if so, it returns `IDENTICAL`. If the donor's blood group is 'O', they are a universal donor and therefore compatible with all other blood groups. Likewise, a patient with blood group 'AB' is compatible with any donor blood group. In both these cases the function returns `COMPATIBLE`. If none of the previous conditions hold, the ABO match is `INCOMPATIBLE`.

D.5.17 Post-transplant Metrics

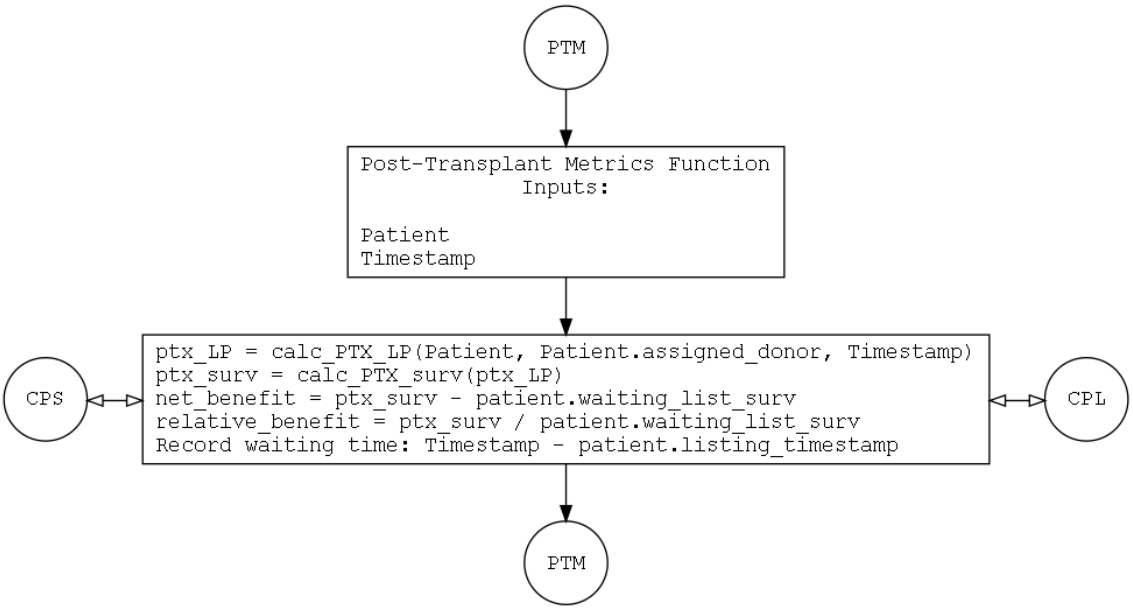


Figure D.21: Flowchart showing how post-transplant metrics (post-transplant survival duration, net benefit, relative benefit, and waiting time) were recorded.

To calculate the survival metrics post-transplant, the post-transplant linear predictor is calculated using the `calc_PTX_LP` function that has already been described.

Next, the post-transplant linear predictor is used to generate a random post-transplant survival duration using the `calc_PTX_surv` (`CPS`) function. Finally, net benefit can be calculated as the difference between post-transplant survival and waiting list survival, and relative benefit can be calculated as the ratio of the two survival durations. Waiting time is simply the the time difference between the current `Timestamp` in the simulation and the timestamp when the patient was first listed.

D.5.18 Randomised Post-transplant Survival Duration

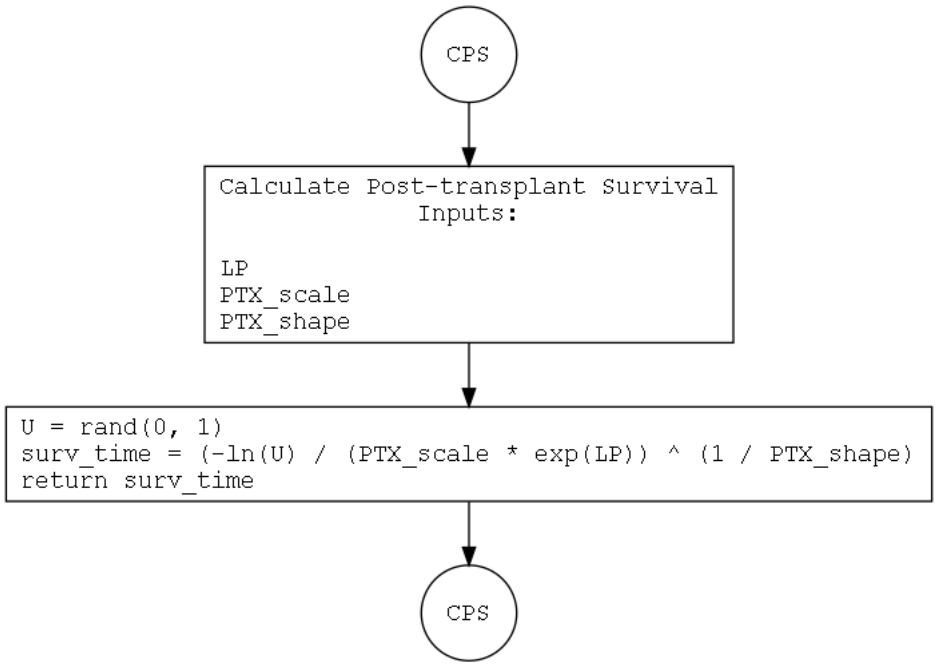


Figure D.22: Flowchart showing how random (but realistic) post-transplant survival durations were generated.

The randomised post-transplant survival function works in almost the exact same way as the waiting list survival function, however there is no minimum survival threshold, and the *scale* and *shape* parameters are from the post-transplant survival dataset rather than the waiting list survival dataset. The randomised survival time is used to calculate the post-transplant metrics.

Appendix E

Simulation Validation Results

E.1 Survival Times Validation Tables

Waiting List Survival Evaluation - Training Dataset(n = 3424)

Subset Size	Simulated Restricted Mean	Observed Restricted Mean	Difference
100	1724 (145.2)	1853 (368.7)	-129
200	1711 (104.3)	1806 (263.1)	-95
300	1716 (89.7)	1794 (226.6)	-77
400	1718 (73.5)	1805 (207.9)	-86
500	1713 (66.7)	1793 (182.6)	-79
600	1717 (61.4)	1792 (169.2)	-74
700	1716 (55.9)	1780 (164.1)	-64
800	1717 (53.3)	1785 (149.3)	-67
900	1717 (51.2)	1783 (152.6)	-65
1000	1714 (47.2)	1779 (142)	-65
1100	1715 (43.8)	1777 (129.4)	-62
1200	1717 (42.3)	1774 (133.1)	-56
1300	1715 (41.2)	1780 (134.1)	-65
1400	1717 (40.4)	1779 (129.6)	-62
1500	1717 (39.2)	1771 (126.6)	-53
1600	1714 (38.6)	1778 (117.8)	-64
1700	1714 (36.3)	1777 (118.3)	-62
1800	1716 (35.7)	1768 (110.6)	-51
1900	1716 (34.4)	1769 (113.7)	-53
2000	1718 (31.6)	1767 (110.5)	-48
2100	1716 (32)	1769 (107.3)	-53
2200	1717 (31.5)	1772 (103.1)	-54
2300	1715 (33.3)	1772 (102)	-56
2400	1716 (30.6)	1772 (96.8)	-56
2500	1716 (29.9)	1764 (96.4)	-48
2600	1717 (29.4)	1764 (95.9)	-46
2700	1717 (28.4)	1763 (94.4)	-45
2800	1716 (27.5)	1767 (90.7)	-50
2900	1716 (28)	1764 (87.7)	-48
3000	1716 (27.1)	1764 (88.5)	-48
3100	1717 (27)	1773 (86.3)	-56
3200	1717 (26.8)	1767 (86.1)	-49
3300	1717 (25.1)	1770 (84.5)	-53
3400	1715 (25.2)	1764 (83.9)	-49

Waiting List Survival Evaluation - Validation Dataset (n = 856)

Subset Size	Simulated Restricted Mean	Observed Restricted Mean	Difference
100	1246 (97.1)	1275 (192.6)	-29
200	1249 (68.5)	1244 (131.6)	5
300	1247 (56.9)	1245 (113.3)	2
400	1252 (49.1)	1243 (93.8)	8
500	1252 (42.4)	1251 (83.1)	1
600	1253 (39.4)	1247 (75.2)	6
700	1250 (35.6)	1249 (71.7)	1
800	1250 (33.6)	1245 (67.4)	5

Post-transplant - Training Dataset (n = 1705)

Subset Size	Simulated Restricted Mean	Observed Restricted Mean	Difference
100	2917 (255.3)	2814 (295.8)	103
200	2929 (183.5)	2812 (206.4)	116
300	2926 (150.2)	2803 (173.5)	122
400	2923 (129.4)	2810 (150.4)	113
500	2923 (117.6)	2811 (129)	111
600	2923 (103.7)	2806 (122.2)	116
700	2922 (101.8)	2808 (113.1)	113
800	2924 (91.2)	2801 (102.5)	122
900	2929 (86)	2810 (101.1)	118
1000	2923 (82.9)	2805 (89.4)	118
1100	2929 (72.2)	2814 (90)	114
1200	2922 (73.7)	2814 (85.4)	108
1300	2922 (69.9)	2806 (86.2)	115
1400	2921 (67.5)	2811 (80.2)	110
1500	2927 (64.4)	2805 (78)	122
1600	2920 (64)	2814 (73.8)	106
1700	2924 (61)	2810 (72.6)	113

Post-transplant - Validation Dataset (n = 426)

Subset Size	Simulated Restricted Mean	Observed Restricted Mean	Difference
100	2930 (263.9)	2924 (305.5)	6
200	2935 (186.1)	2902 (206.7)	32
300	2940 (151.9)	2905 (181.7)	34
400	2935 (129)	2902 (149.9)	32

E.2 Waiting List Survival Times Validation Survival Curves

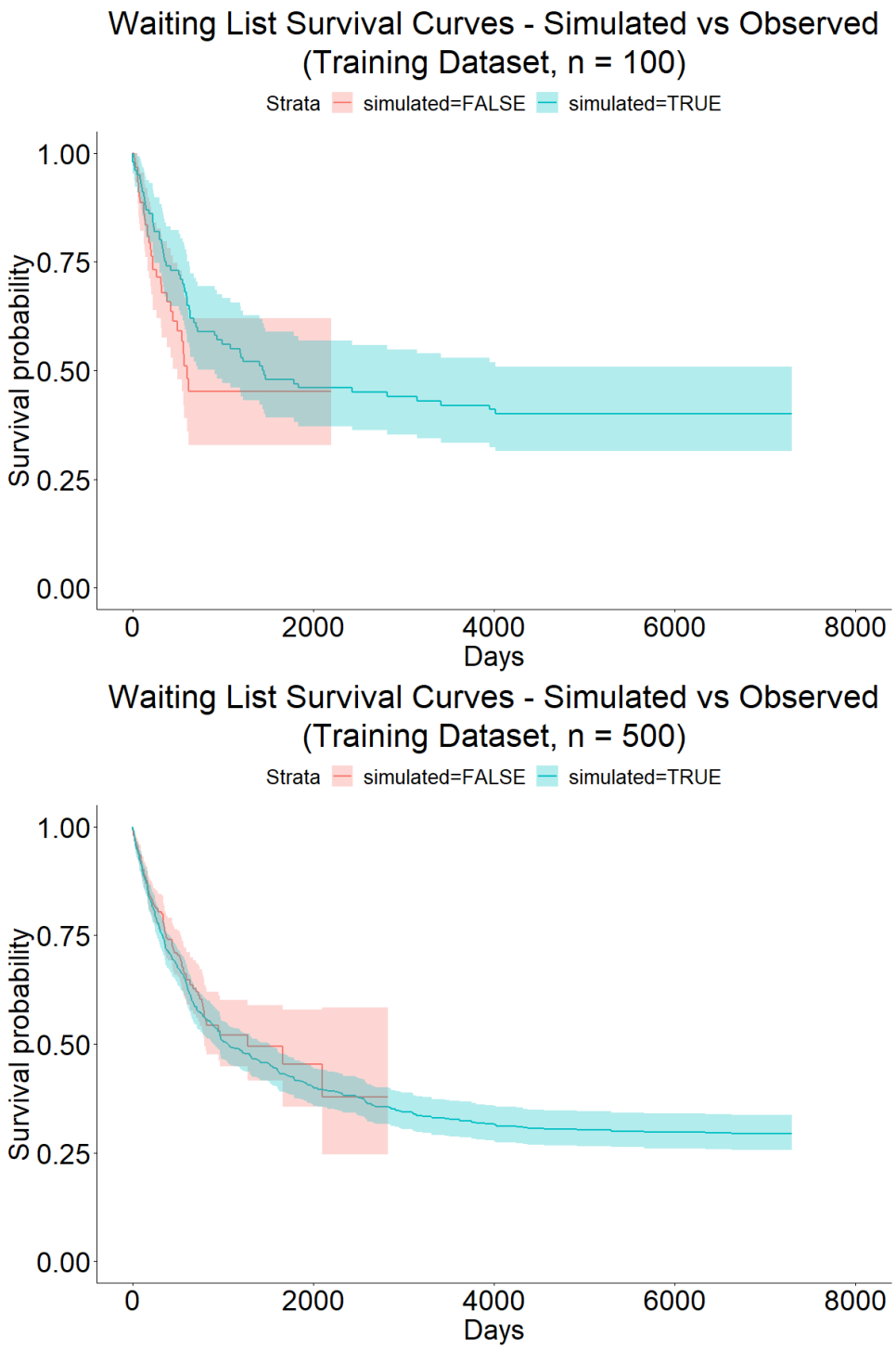


Figure E.1: Comparison of observed versus simulated waiting list survival curves, using a random subset of 100 and 500 lung transplant candidates to plot the survival curves.

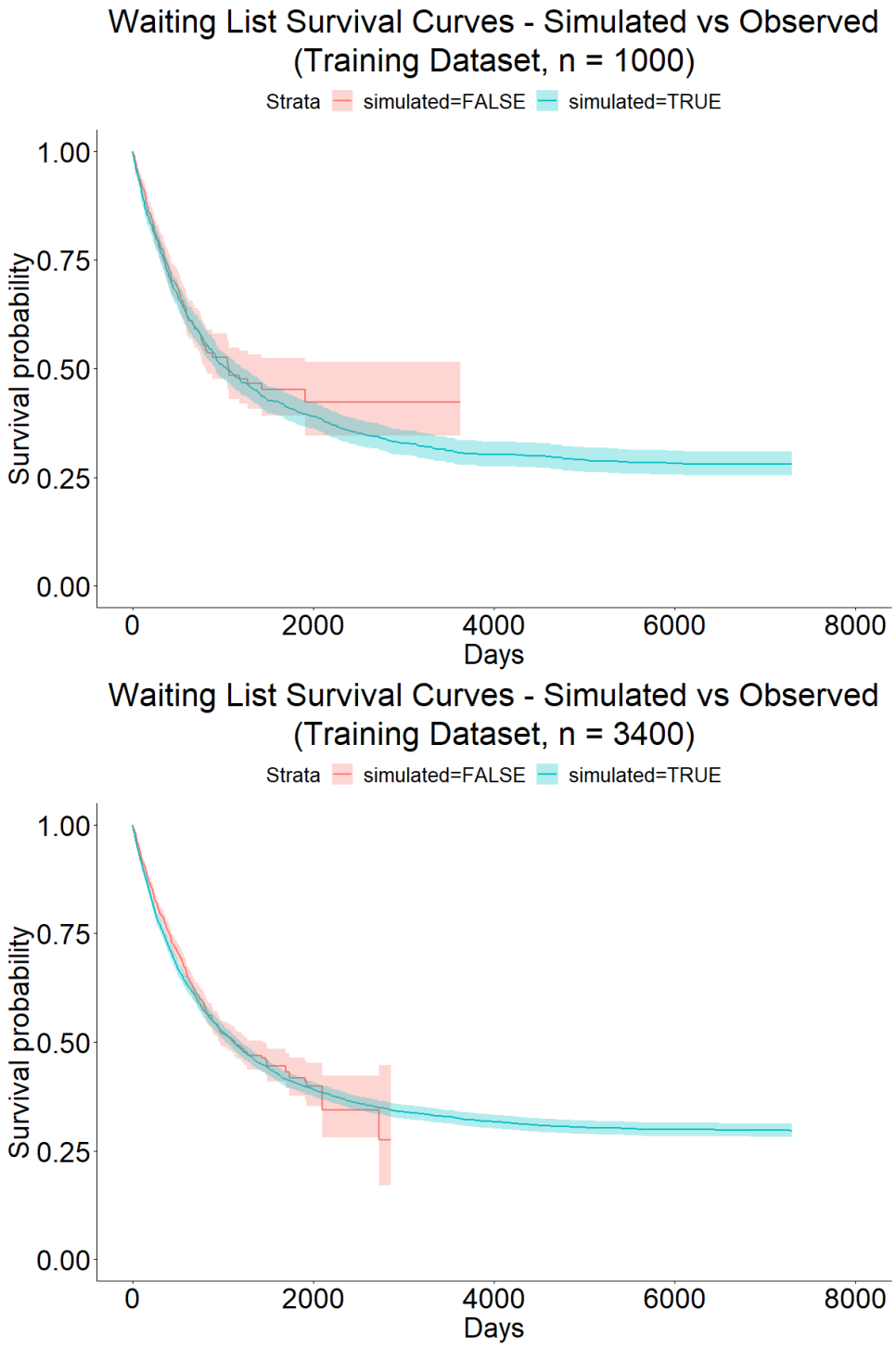


Figure E.2: Comparison of observed versus simulated waiting list survival curves, using a random subset of 1000 and 3400 lung transplant candidates to plot the survival curves.

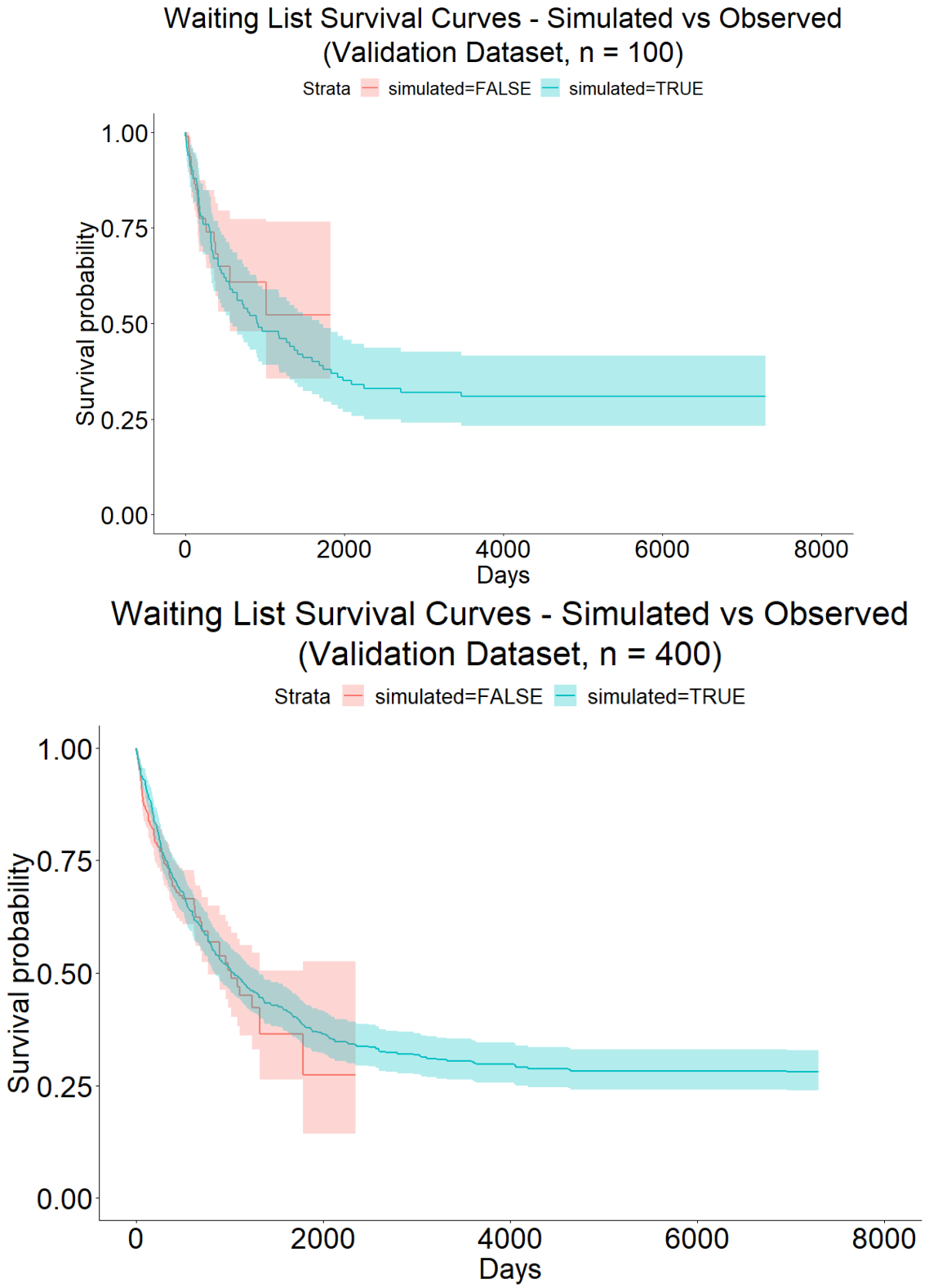


Figure E.3: A comparison of survival curves for simulated and observed waiting list survival durations, for a random selection of individuals from the validation dataset, with subset sizes 100 and 400.

Figure E.4: A comparison of survival curves for simulated and observed waiting list survival durations, for a random selection of individuals from the validation dataset, with subset size 800.

E.3 Comparison of Simulated and Observed Post-transplant Survival Curves

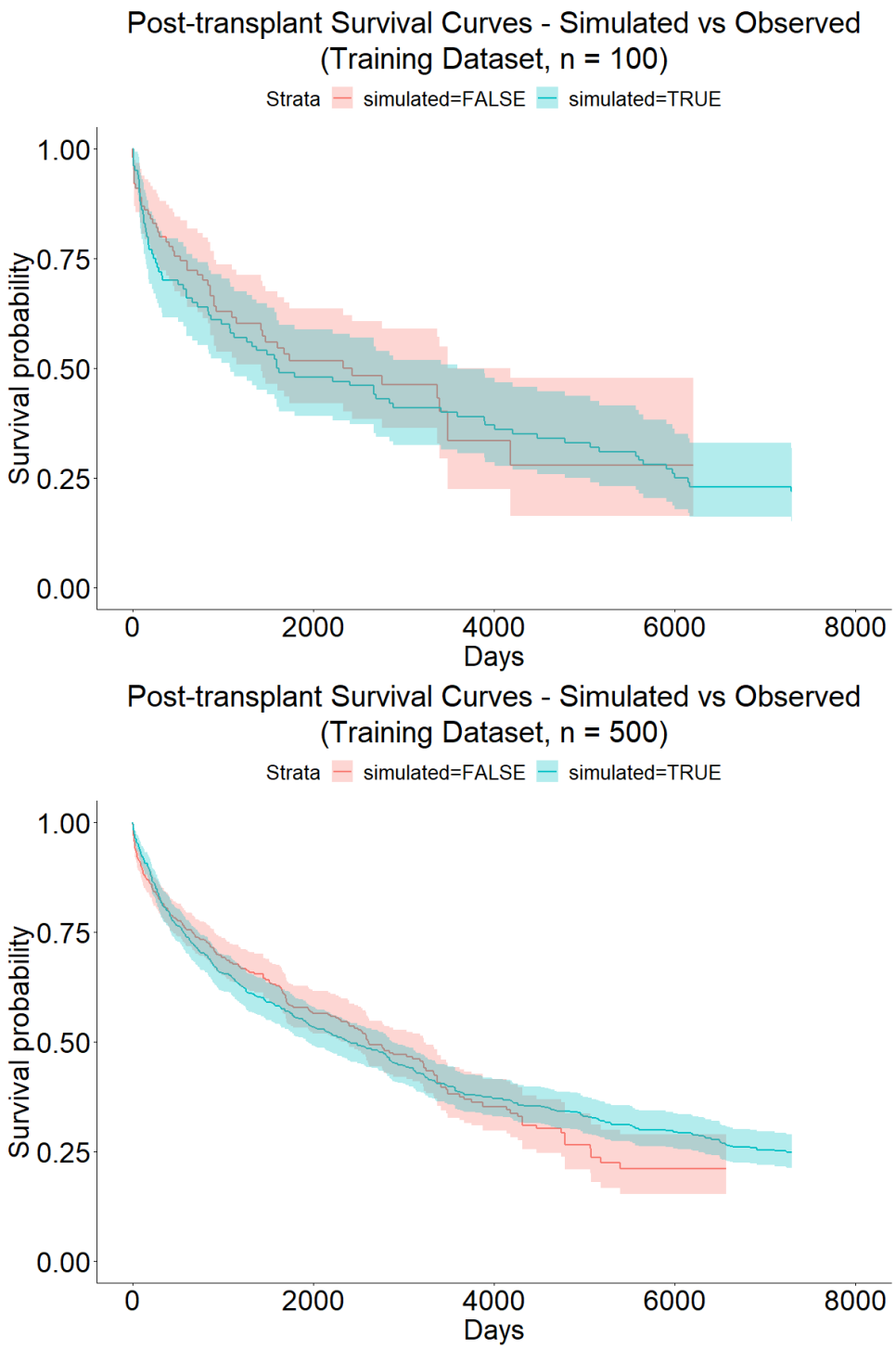


Figure E.5: Comparison of observed versus simulated post-transplant survival curves, using a random subset of 100 and 500 lung transplant recipients to plot the survival curves.

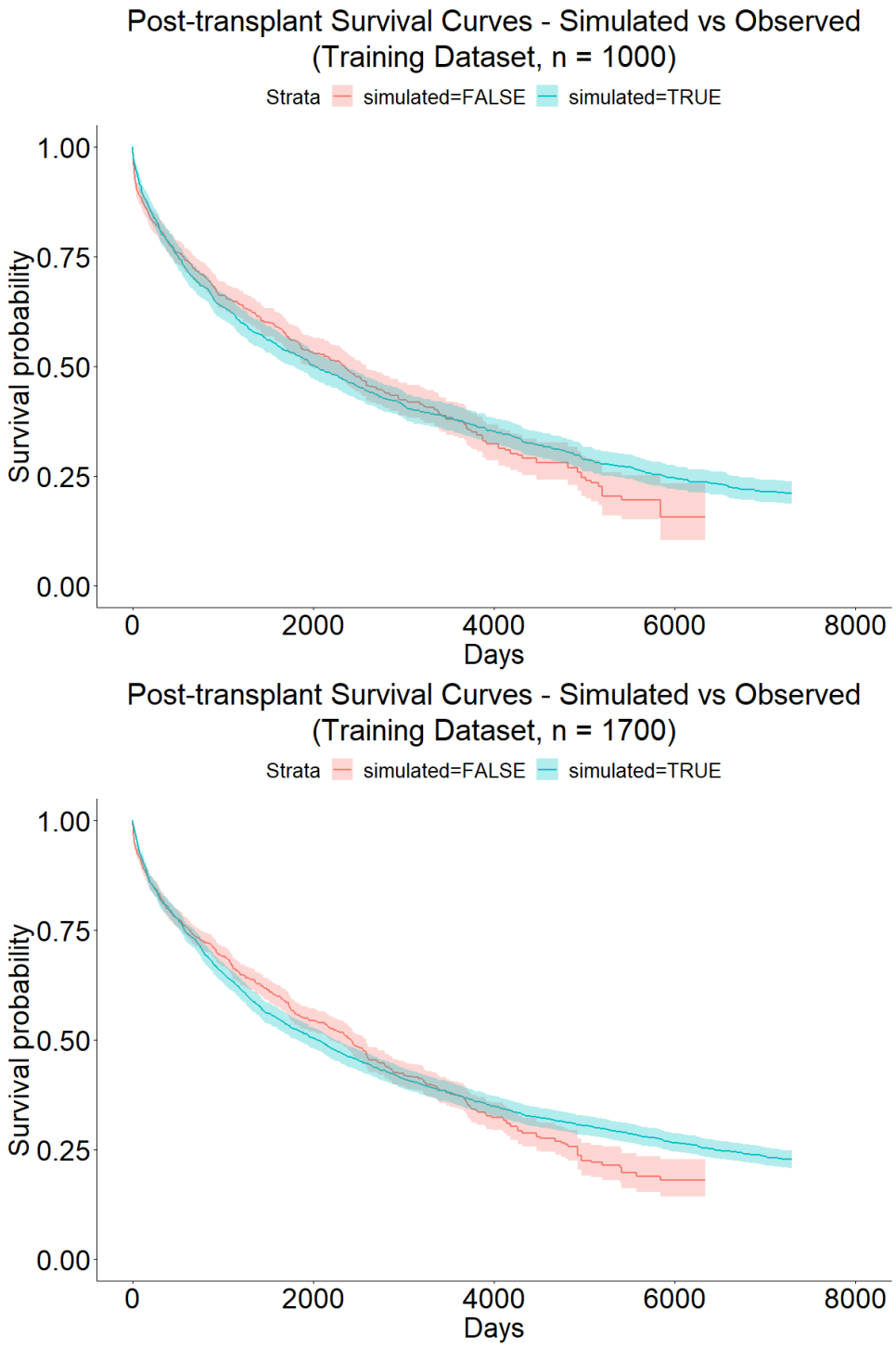


Figure E.6: Comparison of observed versus simulated post-transplant survival curves, using a random subset of 1000 and 1700 lung transplant recipients to plot the survival curves.

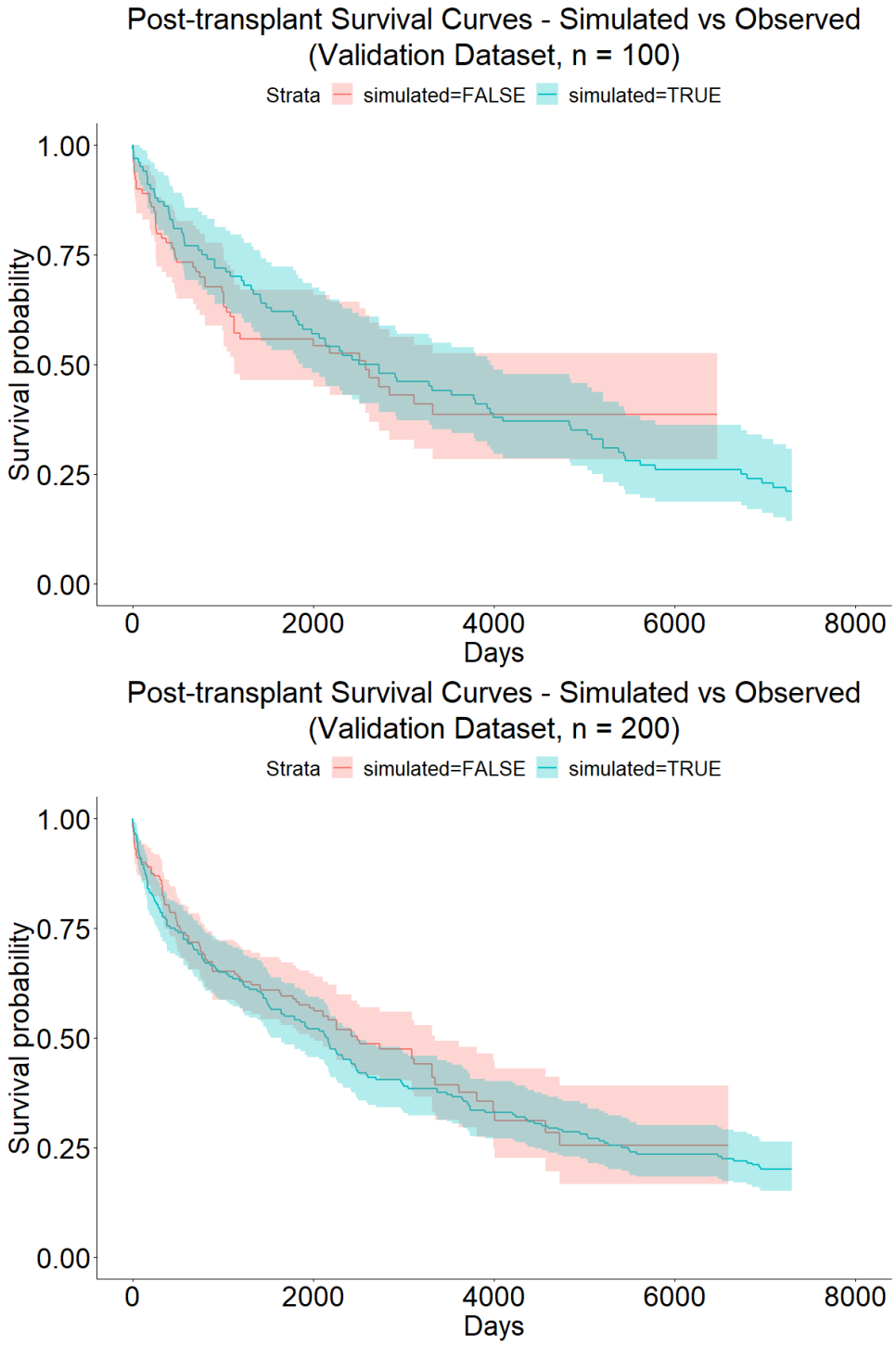


Figure E.7: Comparison of observed versus simulated post-transplant survival curves, using a random subset of 100 and 200 lung transplant recipients from the validation dataset to plot the survival curves.

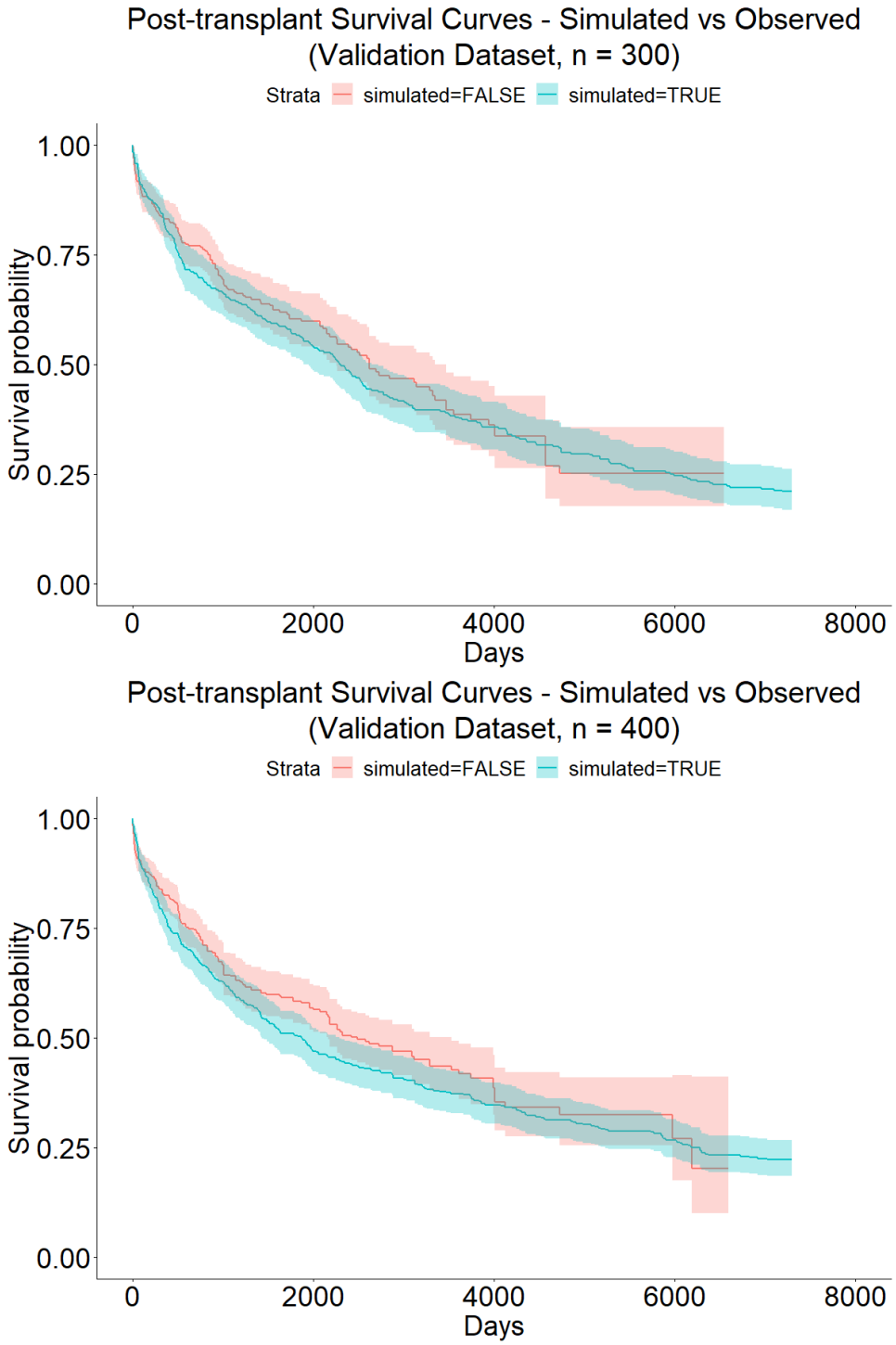
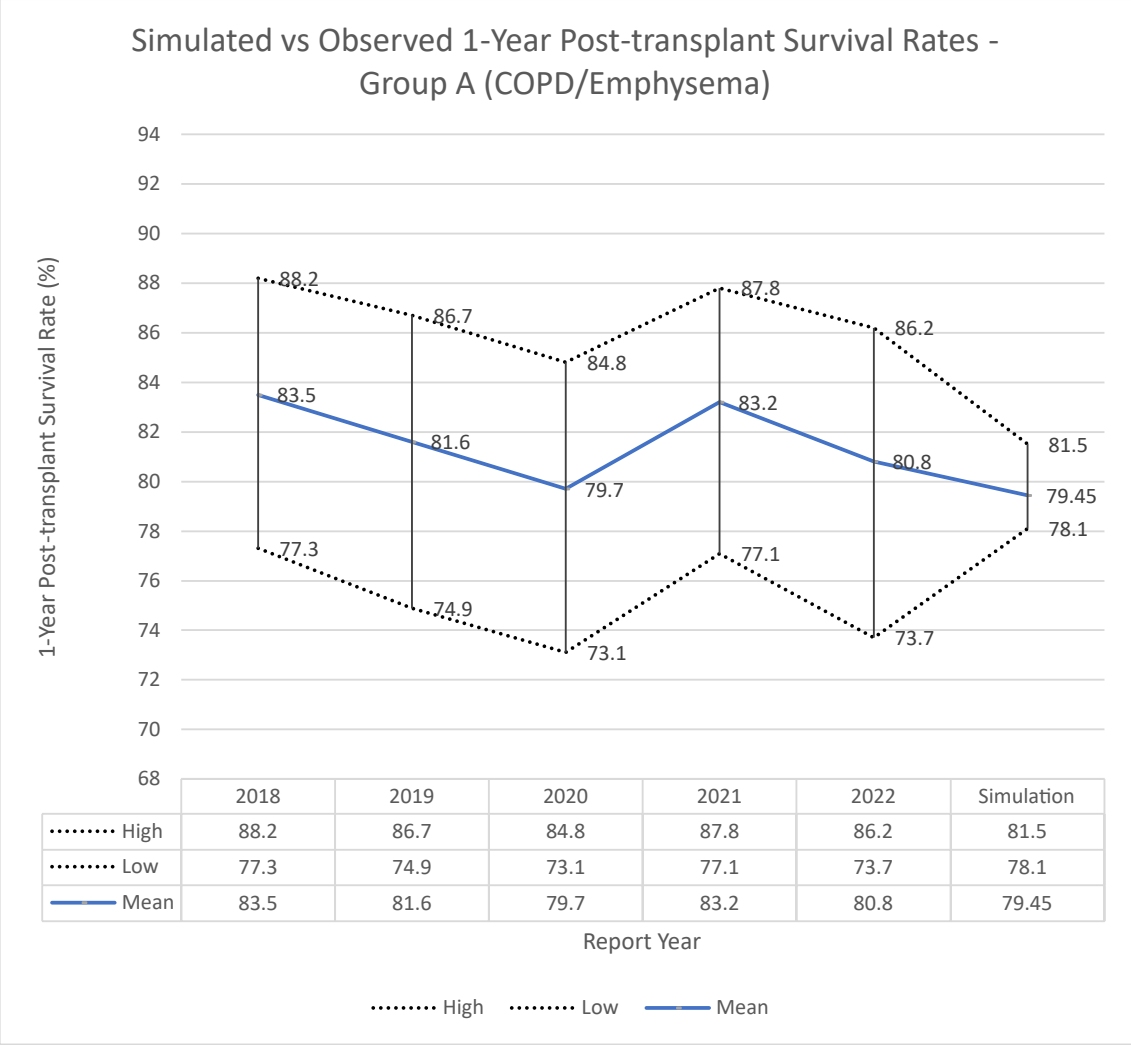


Figure E.8: Comparison of observed versus simulated post-transplant survival curves, using a random subset of 300 and 400 lung transplant recipients from the validation dataset to plot the survival curves.

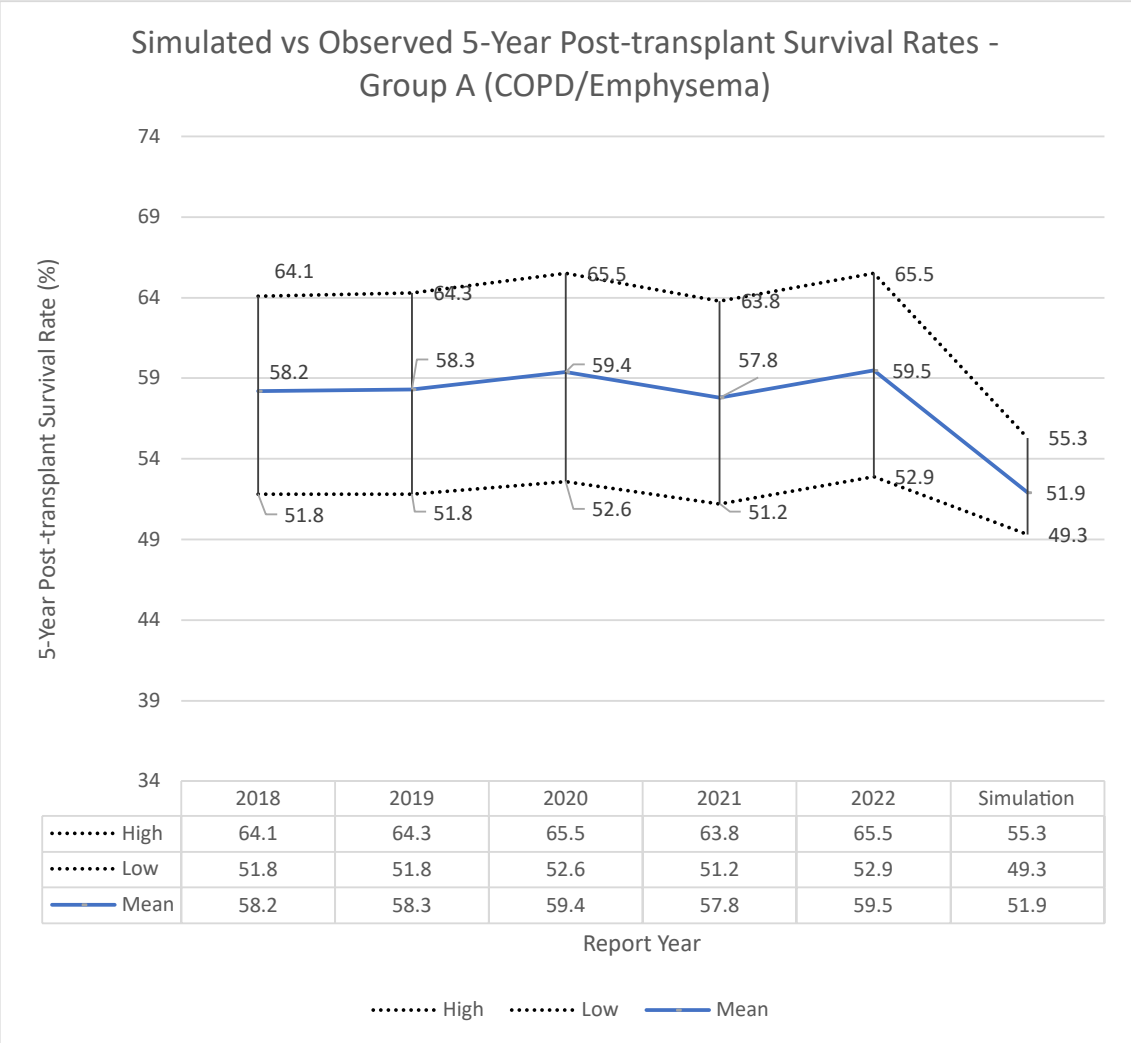
E.4 Simulated versus Observed 1- and 5-Year Post-transplant Survival Rates by Diagnosis Group

The following plots show the range of 1- and 5-year post-transplant survival rates from the NHS-BT organ specific reports from 2018 to 2022.[21,24–26,146] For each year, the high, low and mean from the report are shown, and the high, low and mean of the simulated NHS-BT and WL:PTX policies is shown on the right side of each plot.



	2018	2019	2020	2021	2022	Simulation
High	88.2	86.7	84.8	87.8	86.2	81.5
Low	77.3	74.9	73.1	77.1	73.7	78.1
Mean	83.5	81.6	79.7	83.2	80.8	79.45

Figure E.9: Simulated versus observed 1-year post-transplant survival rates for group A recipients (COPD/Emphysema)



	2018	2019	2020	2021	2022	Simulation
High	64.1	64.3	65.5	63.8	65.5	55.3
Low	51.8	51.8	52.6	51.2	52.9	49.3
Mean	58.2	58.3	59.4	57.8	59.5	51.9

Figure E.10: Simulated versus observed 5-year post-transplant survival rates for group A recipients (COPD/Emphysema)

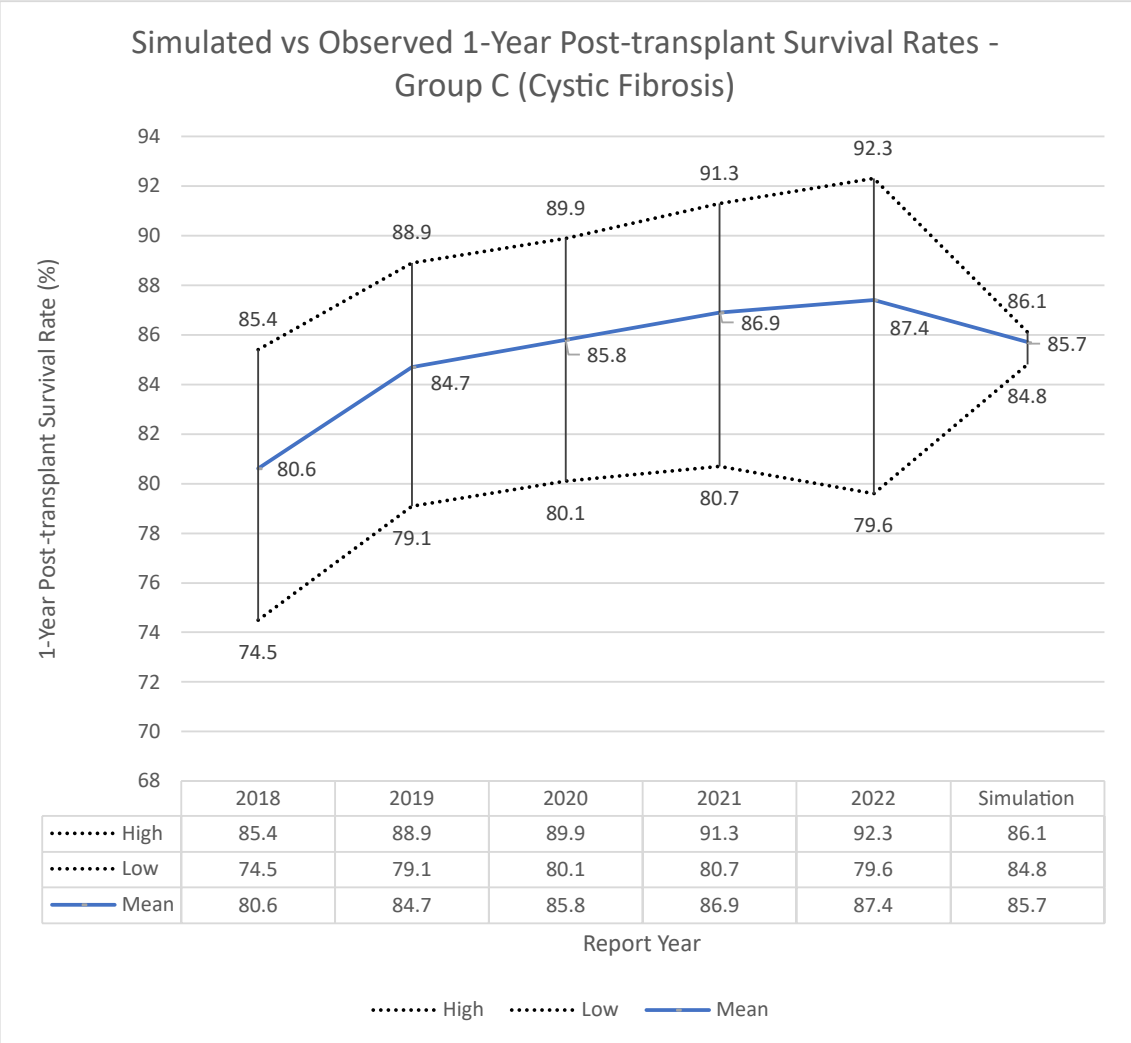


Figure E.11: Simulated versus observed 1-year post-transplant survival rates for group C recipients (cystic fibrosis)

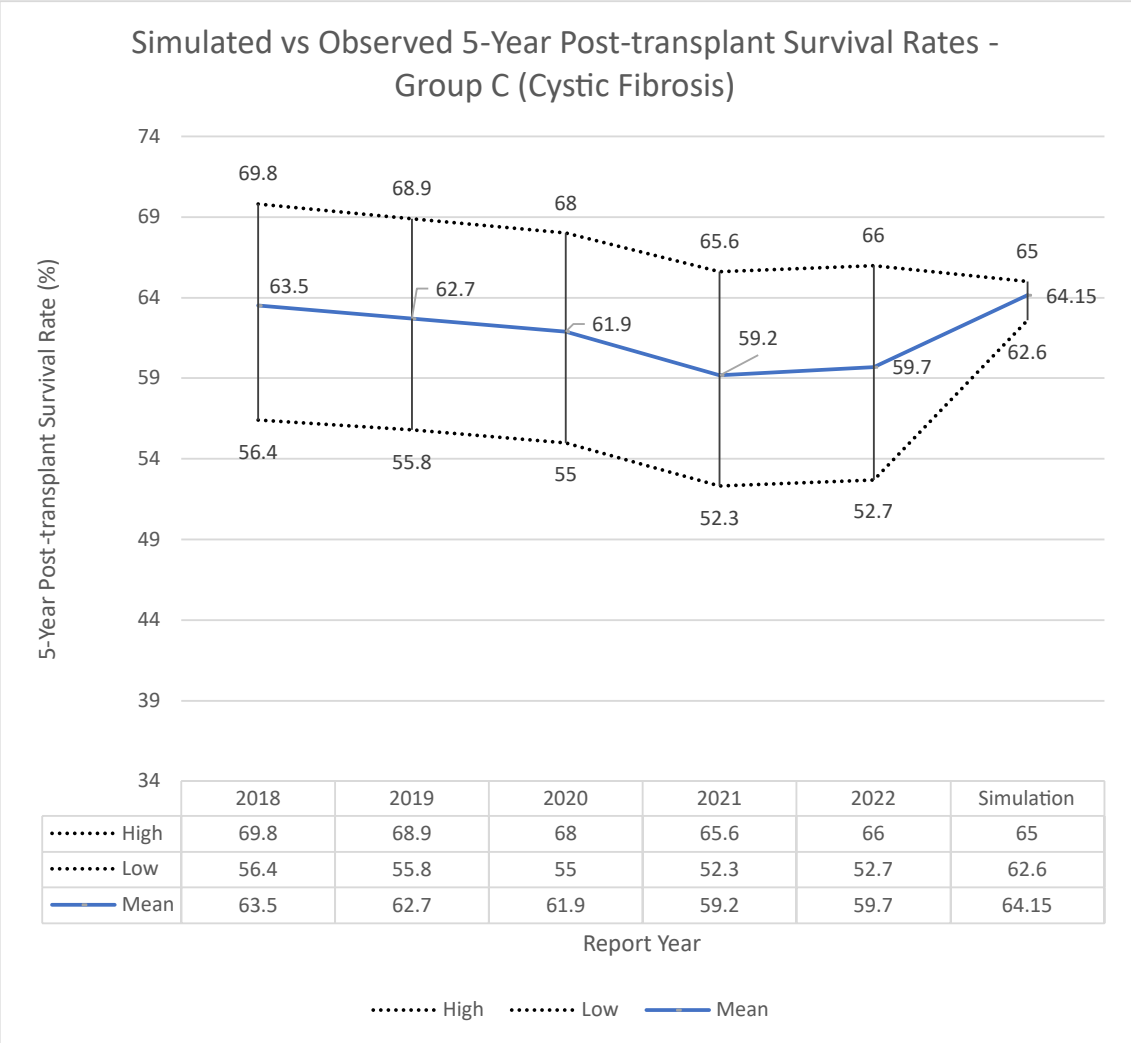


Figure E.12: Simulated versus observed 1-year post-transplant survival rates for group C recipients (cystic fibrosis)

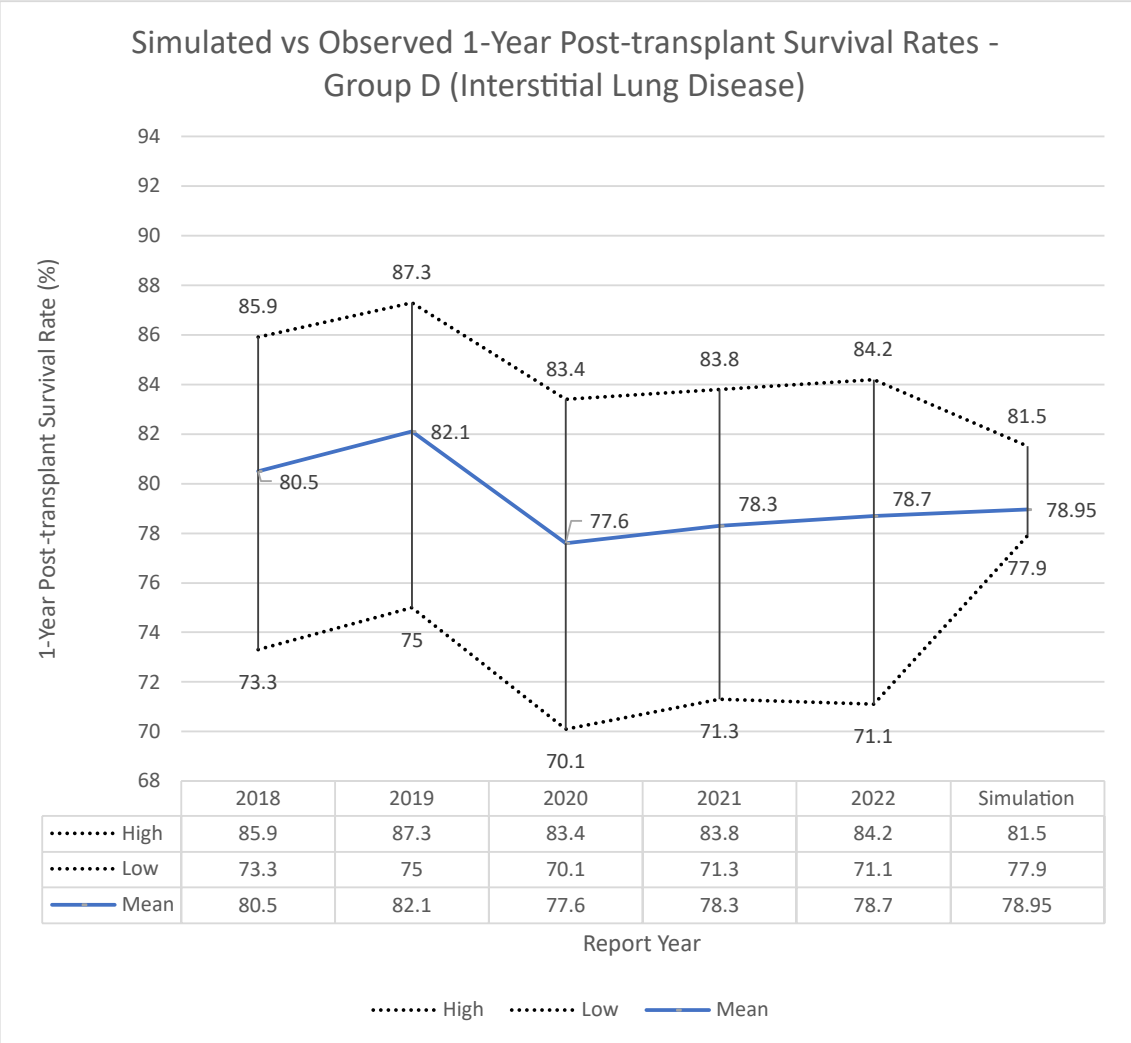


Figure E.13: Simulated versus observed 1-year post-transplant survival rates for group D recipients (ILD/IPF)

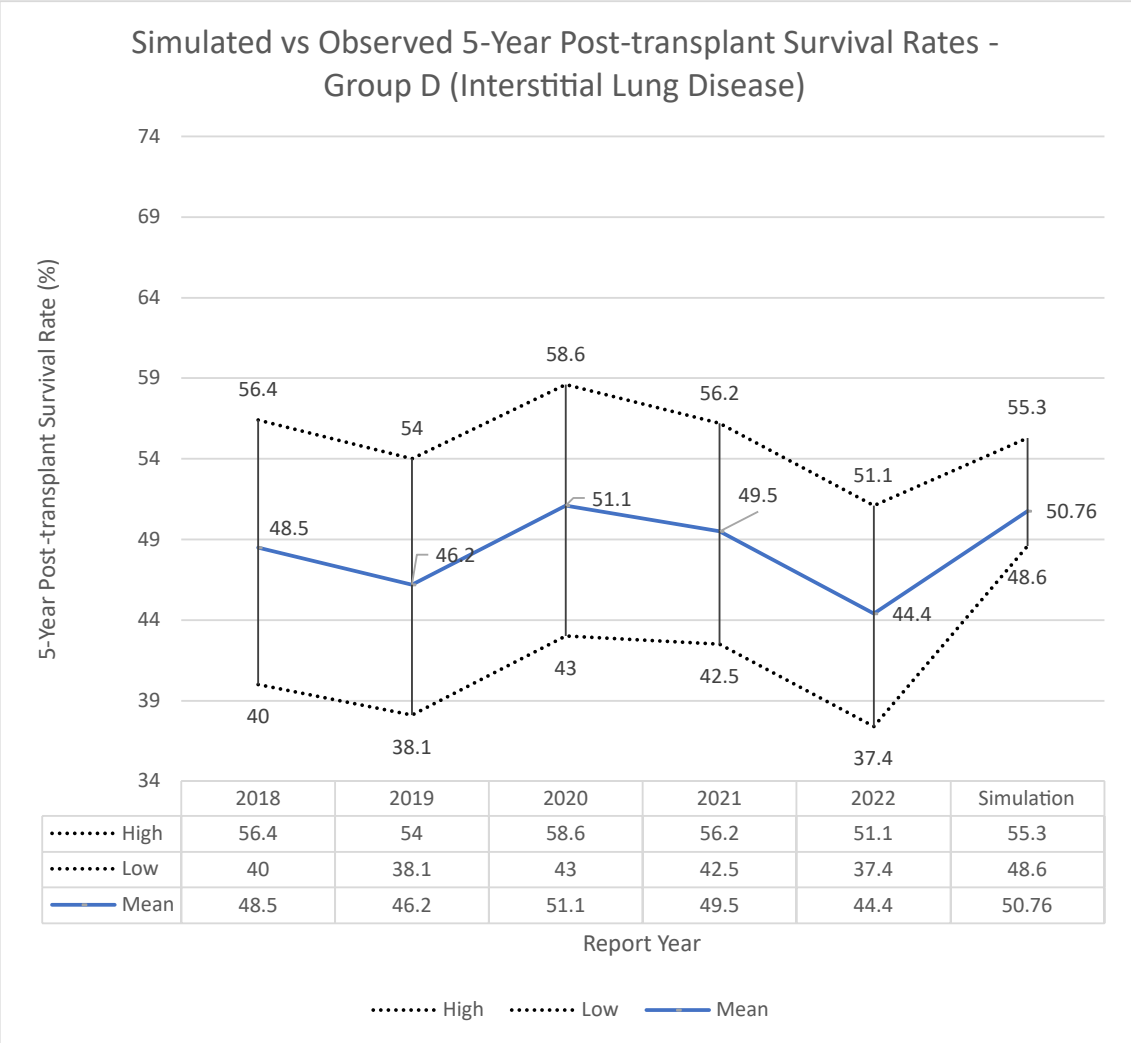


Figure E.14: Simulated versus observed 5-year post-transplant survival rates for group A recipients (ILD/IPF)

Appendix F

Full List of Waiting List and Post-transplant Dataset Variables

F.1 Waiting List Variables

Table F.1: The full list of variables available in the waiting list dataset. The descriptions were taken verbatim from the NHS-BT dataset, which are not public so a citation is not possible.

Variable	**Type**	**Description**
ARECIP_ID	Numeric	Anonymous ODT recipient number
centre	Factor	Anonymous unit that registered the patient for transplant
rwtime	Numeric	Non-urgent waiting time (days)
uwtime	Numeric	Urgent waiting time (days)
suwtime	Numeric	Super-urgent waiting time (days)
adate_on	Date	Registration date
init_urgency	Factor	Initial urgency
urgency	Factor	Highest urgency
final_urg	Factor	Final urgency
reg_outcome	Factor	Outcome of registration
SEX	Factor	Sex
bld_grp	Factor	Blood group
ethnic	Factor	Ethnicity
reg_age	Numeric	Age at registration (years)
reg_height	Numeric	Height at registration (cm)
reg_weight	Numeric	Weight at registration (kg)
dis_grp	Factor	Disease group
CMV	Factor	CMV status
HCV	Factor	HCV status
HBV	Factor	HBV status
HIV	Factor	HIV status
PREV_HEART_SURGERY	Factor	Previous open heart surgery operations at registration
PREV_THORACOTOMY	Factor	Previous thoracotomy at registration
PREV_SUDDEN_DEATH	Factor	Previous sudden death episode at registration
ANTIARRHYTHMICS	Factor	Antiarrhythmics (excluding digoxin) at registration
HYPERTENSION	Factor	Hypertension requring treatment in the last 5 years at registration
VASCULAR_DISEASE	Factor	Peripheral vascular disease with intervention performed or planned at registration

AICD	Factor	AICD at registration
CEREBROVASCULAR	Factor	Cerebrovascular disease at registration
CHOLESTEROL	Numeric	Cholesterol (mmol/l) at registration
DIABETES	Factor	Diabetes at registration
PREV_MALIGNANCY	Factor	Previous malignancy at registration
PREDNISOLONE	Numeric	Daily dose of prednisolone (mg) at registration
SMOKER	Factor	Smoker more than 5 a day within 6 months at registration
BILIRUBIN	Numeric	Bilirubin (μmol/l) at registration
CREATININE	Numeric	Creatinine (μmol/l) at registration
HOME_OXYGEN	Factor	Home oxygen
NYHA_CLASS	Factor	NYHA class at registration
PA_SYSTOLIC	Numeric	PA systolic (mm Hg) at registration
PA_MEAN	Numeric	PA mean (mm Hg) at registration
PCW	Numeric	PCW or LAP (mm Hg) at registration
CARDIAC_OUTPUT	Numeric	Cardiac output (l/min) at registration
EJECTION_FRACTION	Numeric	Ejection fraction (%) at registration
FEV1	Numeric	FEV1 (litres) at registration
FVC	Numeric	FVC (litres) at registration
VO2_MAX	Numeric	VO2 max (ml/kg/min) at registration
SIX_MIN_WALK	Numeric	6 minute walk test (m) at registration
IN_HOSPITAL	Factor	In hospital at registration
VENTILATED	Factor	Patient on ventilator whilst in hospital at registration
INOTRPOES	Factor	Patient on intopes whilst in hospital at registration
IABP	Factor	IABP whilst in hospital at registration
VAD	Factor	VAD whilst in hospital at registration
TAH	Factor	TAH whilst in hospital at registration
ECMO	Factor	ECMO whilst in hospital at registration

F.2 Waiting List Model Coefficients

```
groupB          groupC          groupD          groupOther
0.900855834     0.561848315     1.216314249     0.755787054

aboAB           aboB             aboO
0.115171359     -0.045771387     -0.009117243

rcs(rage)rage    rcs(rage)rage'    rcs(rage)rage''    rcs(rage)rage'''
-0.023685532     0.051664519       0.002615299        -1.502032419

rcs(rbmi)rbmi    rcs(rbmi)rbmi'    rcs(rbmi)rbmi''    rcs(rbmi)rbmi'''
-0.061741592     -0.402200019      1.527240158        -2.130369608

rcs(fev1)fev1    rcs(fev1)fev1'    rcs(fev1)fev1''    rcs(fev1)fev1'''
-1.331021371     20.346197113      -42.661130379      23.811978316

rcs(fvc)fvc     rcs(fvc)fvc'    rcs(fvc)fvc''    rcs(fvc)fvc'''
-1.018608855    6.504393563     -30.244115542    40.592330411

sexFemale       cmvPositive
-0.471347964    0.130035875

reg_diabetesYes - insulin dependent
0.386083312
reg_diabetesYes - not insulin dependent
0.042953679

reg_home_oxygenYes
0.556359018

smokerYes
0.442204969

rcs(reg_creatinine)reg_creatinine    rcs(reg_creatinine)reg_creatinine'
0.001997659                          -0.003724292
rcs(reg_creatinine)reg_creatinine''    rcs(reg_creatinine)reg_creatinine'''
0.164129370                            -0.407194067
```

F.3 Post-transplant Variables

Table F.2: The full list of variables available in the post-transplant dataset. The descriptions were taken verbatim from the NHS-BT dataset, which are not public so a citation is not possible.

Variable	Type	Description
adonor_id	Numeric	Anonymous ODT donor number
arecip_id	Numeric	Anonymous ODT recipient number
atx_id	Numeric	Anonymous ODT transplant number
rwtime	Numeric	Routine wait time (days)
uwtime	Numeric	Urgent wait time (days)
suwtime	Numeric	Super-urgent wait time (days)
rcod	Factor	Recipient cause of death
gstatus	Factor	Most recent outcome of graft
cof	Factor	Cause of failure
vtl_stat	Factor	Vital status of recipient
tsurv	Numeric	Transplant survival time (days)
tcens	Factor	Transplant censoring indicator
gsurv	Numeric	Graft survival time (days)
gcens	Factor	Graft censoring indicator
psurv	Numeric	Patient survival time (days)
pcens	Factor	Patient censoring indicator
centre	Factor	Anonymous unit that received and transplanted the organ
retunit	Factor	Anonymous unit that retrieved the organ
dhosp	Numeric	Anonymous donation centre
it	Numeric	Total ischaemic time hours from cross clamp of donor to reperfusion time (hours)
nhs_grp	Factor	Recipient NHS group
rethnic	Factor	Recipient ethnicity
rsex	Factor	Recipient sex
rage	Numeric	Recipient age
rpaed	Factor	Recipient paediatric (¡16) or adult
rbg	Factor	Recipient blood group

rheight	Numeric	Height of recipient (cm)
rweight	Numeric	Weight of recipient (kg)
rbmi	Numeric	Recipient BMI
dtype	Factor	Donor type
dage	Numeric	Donor age
dpaed	Factor	Donor paediatric (¡16) or adult
dage_grp2	Numeric	Donor age in 10 year increments
dsex	Factor	Donor sex
dheight	Numeric	Donor height (cm)
dweight	Numeric	Donor weight (kg)
dethnic	Factor	Donor ethnicity
dbmi	Numeric	Donor BMI
re_tx	Factor	Whether patient was re-transplanted
abomatch	Factor	Donor vs recipient blood group
pcd	Factor	Primary cause of cardiothoracic disease
reg_primary_disease	Factor	Primary disease at registration
reg_pa_systolic	Numeric	PA systolic (mm Hg) at registration
reg_pa_mean	Numeric	PA mean (mm Hg) at registration
reg_cardiac_output	Numeric	Cardiac output (l/min) at registration
reg_ejection_fraction	Numeric	Ejection fraction (%) at registration
reg_fev1	Numeric	FEV1 (litres) at registration
reg_fvc	Numeric	FVC (litres) at registration
reg_six_min_walk	Numeric	6 minute walk test (m) at registration
reg_ecmo	Factor	ECMO whilst in hospital at registration
urgent	Factor	Urgency status of transplant
ecmo	Factor	ECMO whilst in hospital at transplant
reg_diabetes	Factor	Diabetes at registration
reg_prev_malignancy	Factor	Previous malignancy at registration
reg_prednisolone	Numeric	Daily dose of prednisolone (mg) at registration
reg_smoker	Factor	Smoker more than 5 a day within 6 months at registration

reg_bilirubin	Numeric	Bilirubin (μmol/l) at registration
reg_creatinine	Numeric	Creatinine (μmol/l) at registration
reg_home_oxygen	Factor	Home oxygen
reg_in_hospital	Factor	In hospital at registration
reg_ventilated	Factor	Patient on ventilator whilst in hopsital at registration
cmv_tx	Factor	Recipient CMV status at transplant
tx_yr	Numeric	Year of transplant
tx_type	Factor	Type of transplant
org_txd	Factor	Type of cardiothoracic transplant
multi_tx	Factor	Multi-organ transplant indicator
organ	Factor	Organ transplanted
dcmv	Factor	Donor CMV test result
dpast_diabetes	Factor	Donor diabetes
dpast_smoker	Factor	Donor smoker
dpast_smoker_amount	Numeric	Donor smoker amount (number smoked per day)
rvent	Factor	Patient on ventilator whilst in hospital at transplant
rcreat	Numeric	Creatinine (μmol/l) at transplant
days_tx_hdu	Numeric	Number of days spent in HDU after transplant
days_tx_itu	Numeric	Number of days spent in ITU after transplant
days_tx_hosp	Numeric	Number of days spent in hospital after transplant
theatre	Factor	Complications - return to theatre
haemofiltration	Factor	Complications - haemofiltration/haemodialysis
other_post_op	Factor	Complications - other mechanical assistance post-op
lost_fup	Factor	Lost to follow up
gno_lng	Numeric	Number of lung transplants

F.4 Post-transplant Model Coefficients

aboAB	aboB	aboO
-0.000178869	0.150476675	-0.136174834

groupB	groupC	groupD	groupOther
-0.166802279	-0.418684819	-0.116178487	-0.207907918

height_delta
-0.008938175

rcs(rage)rage	rcs(rage)rage'	rcs(rage)rage''	rcs(rage)rage'''
-0.053912139	0.103493240	-0.274494375	0.099134897

rbmi	reg_creatinine	reg_home_oxygenYes
0.025338374	0.004484454	0.133490421

organLeft Lung	organRight Lung
0.409946505	0.076364869

ddiabYes
0.271796358

reg_prev_malignancyYes	dcmvPositive
0.344856145	0.221994142

rcs(dbmi)dbmi	rcs(dbmi)dbmi'	rcs(dbmi)dbmi''	rcs(dbmi)dbmi'''
-0.108806341	0.593904281	-2.159848254	2.623263404

dtypeDCD	dage
0.183683630	0.004406212

Appendix G

The Analytic Hierarchy Process (AHP) 💻

The AHP was first proposed by Thomas L. Saaty in the 1970's.[36] The AHP captures the three principles of problem solving: *decomposition*, *comparative judgements* and *synthesis of priorities*.[151]

Decomposition is achieved by starting with a goal, then breaking the goal down into multiple criteria which can be further broken down into sub-criteria. This process is continued until all factors relating to the overall goal have been included.

As a simple example, take the goal of selecting a house to purchase, as in section 2.2.3. Instead of just considering price and size, the goal can be decomposed into many sub-goals:

1. Purchase House
    (a) Location
        i. Proximity to local amenities
        ii. Proximity to schools
        iii. Ofsted ratings of local schools
        iv. Crime rates
        v. Parking facilities
        vi. Internet/TV/Phone service availability
        vii. Public transport links
    (b) Cost
        i. Price of house
        ii. Council tax band
        iii. Service Fees / Ground Rent / Other Charges
        iv. Energy efficiency and cost of utilities
        v. Condition of house, cost of repairs/improvements

(c) Size

i. Floor space

ii. Number of floors

iii. Size of driveway

iv. Size of garden

v. Built-in storage

Some of these sub-goals could be broken down further if desired, but the concept of decomposition has been illustrated in this case with 3 tiers of goals: the main goal, sub-goals and sub-sub-goals.

Comparative judgements are achieved using a pairwise comparison process. This process is repeated for every level in the hierarchy. Within each level, every pair of criteria are compared and decision makers are asked: "When making a decision (with respect to the goal) which of the two criteria is more important, and by how much?"

A verbal scale corresponding to a numeric scale is shown in table G.1.[152] Intermediate values can be chosen, for example '4' would correspond to 'moderate to strong importance'.

Table G.1: Saaty's verbal scale and corresponding numeric values.

Value	Definition
1	Equal Importance
3	Moderate importance of one over the other
5	Strong importance of one over the other
7	Very strong importance of one over the other
9	Extreme importance of one over the other

In the example given above, first the sub-goals would be compared: location vs cost, location vs size and cost vs size. Then within each sub-goal, the sub-sub-goals are compared pairwise - this leads to the concepts of *local weights*, *global weights* and *synthesis of priorities*.

For example, 'floor space' could have a local weight of 0.75 if it was very important compared to other goals in the same level (number of floors, size of driveway etc.) However, if the size of the property had a very low weight compared to the cost and the location (say for example, 0.05), then floor space has a global weight of $0.75 \times 0.05 = 0.0375$. This is illustrated in figure G.1.

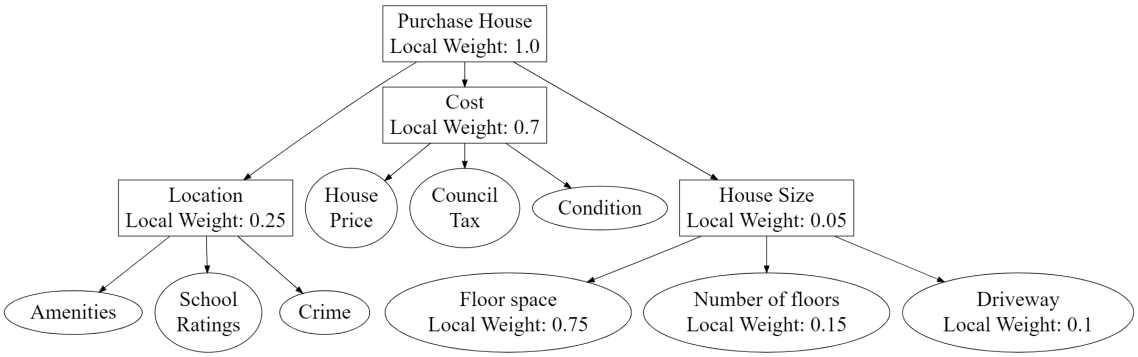


Figure G.1: An example of *decomposition* along with *local weights* that can be *synthesised* into global weights. Note that the root node of the graph always has a local weight of 1.

This process of calculating overall weights by combining local weights and global weights is the process of *synthesising priorities*.

G.1 AHP Debate

After the initial publication of the AHP there was debate over the validity of the AHP. The four main points of contention were:[174]

1. A lack of an axiomatic foundation for the AHP
2. Ambiguity in the questions asked during pairwise comparisons
3. The verbal 1 - 9 scale
4. Rank reversal

Each of these points will be discussed in the sections to follow.

Axiomatic Foundation Saaty published 'Axiomatic foundation of the analytic hierarchy process' in 1986[151] where four axioms are outlined: the reciprocal property, ρ-homogeneity, the principle of hierarchic composition and expectations.

The reciprocal property is described using the analogy of comparing the weights of two stones:[151] "If one stone is judged to be five times heavier than another, then the other is automatically one fifth as heavy as the first because it participated in making the first judgment." This property can be described mathematically for matrix A as: $A_{ij} = 1/A_{ji}$ for all i, j.

The axiom of ρ-homogeneity states that all elements in the same level of the hierarchy should be comparable on a scale bounded by some positive real number (ρ). For example using the scale given in table G.1, $\rho = 9$, therefore the entries in the matrix should range from 1/9 to 9. The importance of this property can be illustrated with the following question which violates ρ-homogeneity:

"On a scale of 1 - 9, how large is the Milky Way galaxy compared to a single marble?"

The sizes of the two are not comparable on a 1 - 9 scale and human judgement is not accurate when comparing criteria of massively different magnitudes. This is why in order to accurately capture the judgements of a decision maker the criteria in the same level of a hierarchy must be comparable. This axiom ensures this is the case.

The axiom of hierarchic composition allows the final weights of criteria to be calculated by multiplying the local weight of a criterion by the global weight of the parent criterion, all the way from the lowest nodes up to the original goal. This is ensured if: lower levels of the hierarchy are dependent on higher levels, elements in the same level of the hierarchy are not dependent on any other element at the same level and higher levels of the hierarchy are not dependent on lower levels of the hierarchy.

The final axiom of expectations states that if an individual has a reason for believing that alternatives should be ranked in a certain order, then they should expect the rankings generated by the AHP to represent their expectations if the hierarchy captures all the relevant criteria and alternatives.

Dyer claimed that the axiomatic foundation is flawed,[175] Harker and Vargas in[176] argued that the criticisms of the axioms are due to a lack of understanding of the underlying theory of AHP.

AHP Question Ambiguity The criticism of the questions asked during pairwise comparison of criteria were due to a misunderstanding of how the questions should be asked. Dyer[175] originally posed the question in the form:

"How much better is A_i than A_j on a criterion?"

Harker and Vargas[176] responded by showing a correct example:

"With respect to cost, which of the two cars (A or B) is preferred, and by how much?"

Watson and Freeling[177] also argued that the questions are meaningless due to there being no frame of reference. Harker and Vargas[174] argued that the frame of reference for answering the pairwise questions is irrelevant, since the entire exercise is to capture a specific individual's subjective perceptions of what they deem "more important" or "strongly more important". They also argued in their 1990 response[176] that the criterion ('cost' in the example question given previously) becomes the point of reference.

Scale of Preference Experimentation has shown that the 1 - 9 scale given by Saaty was the most effective scale to use when comparing perceived relative distances of various cities.[174] In[176] Harker and Vargas gave an example of a bounded scale (1 - 9) and an unbounded scale and calculated the difference in the weights calculated in each case, the difference was on the order of magnitude of 0.001.

Rank Reversal The most controversial criticism is the occurrence of rank reversals as a result of using the AHP. An example of rank reversal is given by Dyer and Wendell

in[178] using four alternatives and four criteria. When only $A_1 ... A_3$ are ranked the resulting ranking is: $A_3 > A_2 > A_1$. However, when A_4 is introduced then the resulting ranking is: $A_1 > (A_3, A_4) > A_2$. This however occurred in a hierarchy with feedback loops between the criteria and the alternatives and was "due to a misuse of the theory rather than a faulty axiomatic base".[174]

Belton and Gear[179] also gave an example of rank reversal starting with three alternatives $B > A > C$, and upon introducing D which was a duplicate of B the ranking becomes $A > (B, D) > C$. Harker and Vargas[174] refute this point by stating that axiom 4 (expectations) requires that all criteria and all alternatives must be specified before applying the AHP. If two alternatives are not distinguishable then new criteria must be added in order to distinguish them. Adding new criteria will result in a change in the structure of the overall hierarchy and thus a change in the weights and potentially a change in ranking.

Dyer[175] showed that rank reversal still occurs if an alternative is within 10% of another, so not an exact duplicate. Both[176] and[180] argue that rank reversals are a feature of the AHP and not a weakness of it.

Calculated Weights A more recent criticism in[181] shows that the weights derived from the AHP may not adequately represent the intensity of preference of the criteria. Given (for example) four alternatives: x_1, x_2, x_3, x_4, if x_1 is preferred to x_2 and x_3 is preferred to x_4 it is reasonable to assume that the weights of the alternatives maintain the relation: $x_1 > x_2$ and $x_3 > x_4$.

What also needs to be considered is the *intensity* of preference. It should be expected that if x_1 is preferred to x_2 more intensely than x_3 is preferred to x_4 then $x_1/x_2 > x_3/x_4$, however in this paper it is shown to not always be the case.

G.2 Applications of the AHP

Despite this debate, the AHP has been widely researched and utilised across numerous fields for various purposes. Vaidya[182] reviewed 150 application papers and outlined the various methods and sectors in which the AHP has been used. The methods the AHP was utilised for were: selection of a single best alternative, evaluation of several alternatives, cost-benefit analysis, allocation of resources, planning and development, prioritising and ranking alternatives, decision making and forecasting.

The sectors in which the AHP has been utilised were: medicine, personal, social, manufacturing, political, engineering, education, industry, government, management, banking, sports and finance.

The AHP has been utilised for multiple purposes in the context of health care.[183] It has been used for medical diagnosis, selection of treatment, selection of technologies, human resource planning, evaluation of hospital performance and organ transplantation. It was noted that there was some resistance from physicians about using a formal tool to

assist in the decision making process, however it was believed that this resistance would be overcome as more successful applications of the AHP were demonstrated.

G.2.1 AHP in Transplantation

The AHP has been proposed for use in the allocation of organs for transplantation. Two notable examples in the literature propose a system for livers and a system for kidneys. A framework is also proposed for transplantation in general.

Cook[184] proposed a system of allocating livers utilising the AHP. Surgeons, anaesthesiologists, procurements coordinators, transplantation coordinators, financial officers, ethicists and other experts were interviewed to determine which criteria should be included in the AHP model. Judgements were informed by both the transplantation literature and the personal experiences of surgeons and anaesthesiologists. After the initial pairwise comparisons were completed some criteria had very low weights assigned to them so were removed from the hierarchy. Screening criteria were defined, such as blood type compatibility (incompatible patients should be removed), donor size range and various risk factors. The decision was made to use ischaemic times rather than distance since the mode of transportation makes a large difference in total ischaemic time. The final consistency index was calculated as 0.04 showing that the group judgements were consistent. Due to the change in criteria and the weights assigned to them, the new rankings were different to the rankings of the old system. Cook concluded that the system in place at the time was never formally evaluated and that the AHP may provide the necessary equitable evaluation to justify the allocation of livers for transplantation.

Taherkhani[185] proposed a kidney allocation system to be used in Iran using the fuzzy delphi method (FDM) and fuzzy AHP. 'Fuzzy' numbers have a range of possible values as opposed to 'crisp' numbers which have a single specific value. Using the FDM, the criteria identified for allocation were:

1. Blood type compatibility
2. HLA matching
3. Panel reactive antibodies
4. Age difference
5. Recipient age
6. Location
7. Transplant status (first time vs repeat transplant)
8. Waiting time
9. Medical urgency

10. Predicted survival

11. Prior living donor

Three of these criteria were rejected for use in the new system. Location was rejected due to the way logistics are handled in Iran: it is believed that better outcomes are achieved by transporting the patient to the city in which they are to be transplanted, rather than transporting the kidney. Prior donation was rejected due to there being a market for selling kidneys. Transplantation status was rejected due to re-transplanted patients being treat similar to first-time patients. The remaining criteria were grouped into two clusters: equity and utility. The weights for the criteria are in table G.2.

Table G.2: Global and local weights of criteria used for prioritising kidney transplant patients, published by Taherkhani et al.[185]

Criteria	Sub-Criteria	Local Weight	Global Weight
Equity (0.33 Global Weight)	Medical Urgency	0.54	0.1782
	PRA	0.14	0.0462
	Recipient Age	0.27	0.0891
	Waiting Time	0.05	0.0165
Utility (0.67 Global Weight)	HLA Matching	0.35	0.2345
	Blood Type Compatibility	0.16	0.1072
	Age Difference	0.11	0.0737
	Predicted Survival	0.38	0.2546

The model was evaluated using the kidney transplantation dataset in Tehran. The dataset covered the period from October 2017 to December 2017 and included 484 candidates and 124 donors. The model was run and the chosen candidate was recorded for each donor. Data containing the existing system's choices were also available allowing comparison between the proposed and existing system. There was no outcome data available so the Estimated post transplant survival (EPTS) score was used. Lower EPTS scores indicate more years of graft function than higher scores.

Four measures of utility and three measures of equity were used for comparing the effectiveness of the proposed system. The results are in table G.3.

The proposed model improved all utility measures except identical blood type allocations, however this is due to the current system being based on matching using identical blood type. It also improved all equity measures except for average waiting time of urgent patients. There was a slight increase in the average waiting time for urgent patients (0.9 years to 1.1 years). The average waiting time for all patients was reduced however from 1.7 years to 1.25 years. Sensitivity analysis was performed and for top-level criteria (utility and equity), weights could vary by up to 50% without significantly changing the rankings, and sub-criteria weights could vary up to 30%.

Table G.3: AHP kidney allocation model performance compared to existing model, published by Taherkhani et al.[185]

Category	Measure	Developed Model	Existing Model
Utility	Number of recipients with EPTS $< 20\%$	124/228 (54%)	83/230 (36%)
	Average EPTS score of recipients	24.61%	41.37%
	Average donor-recipient age difference	5.3 years	8.1 years
	Number of identical blood type allocations	243/248 (98%)	248/248 (100%)
Equity	Average waiting time	1.25 years	1.7 years
	Average waiting time of urgent patients	1.1 years	0.9 years
	Number of paediatric allocations	20/22 (91%)	18/22 (82%)

The authors conclude that future research should compare the results for other organs using ordinary AHP, fuzzy AHP and intuitionistic fuzzy AHP (IF-AHP).

Overall this paper demonstrates that the AHP can be used to aid allocation of organs and improve allocation according to measures of utility and equity.

A general framework for applying the AHP to transplantation has also been proposed and applied to liver allocation.[186] The framework that was developed used the literature to identify criteria relevant to liver allocation, and then incorporated medical opinions as well as subjective and objective criteria in order to rank candidates. The authors proposed that the use of the AHP made the model easy to implement, use and update.

Four main factors were identified as top-level criteria in the AHP model:

1. Urgency - risk of death based on medical criteria
2. Efficiency - risk of transplant failure, post-transplant life expectancy and well-being.
3. Benefit - Combination of urgency and utility - who will benefit most?
4. Equity - The belief that all patients have equal rights to organs.

Unlike in the traditional use of the AHP, the alternatives (candidates) did not undergo pairwise comparison because the number of required comparisons would be intractable. Instead the technique of absolute measurement was used, this was accomplished by splitting measurements into groups and performing pairwise comparison on the groups. For example, Model for End-Stage Liver Disease (MELD) was broken down into five groups of ranges and each group was assigned a weight.

In order to combine the judgements of multiple experts, the geometric mean was used.

A sensitivity analysis was performed on the four top-level criteria and it was shown that adjusting the weights by $\pm$ 30% had no significant impact on the rankings, demonstrating

the stability of the developed model. The reliability of the model was also demonstrated with its agreement with the decisions made by OPTN.

The authors propose that future research should look into the Analytic Network Process (ANP) and believe that it may improve the accuracy of decision making.

G.2.2 AHP in Lung Transplantation

OPTN acknowledged the potential benefit of utilising MCDM methods in their concept paper.[187] They propose a 'Continuous distribution framework', which is defined as a system which prioritises candidates based on a combination of points for multiple factors related to transplantation.

The goal of continuous distribution is to address the problem of 'hard boundaries' which prevent optimal distribution and to increase transparency of the system. Some examples of hard boundaries are: distance from the donor, age and blood type compatibility. With the current system an 11 year old candidate will always receive an organ offer before a 12 year old candidate which is a similar distance from the donor, even if the 12 year old has a higher medical priority.

Hard boundaries will be removed by breaking down criteria into smaller sub-categories and if possible using continuous functions for criteria. This will prevent candidates from receiving a different priority just because they are slightly on one side of an arbitrary hard boundary. An important part of developing the new system is to assign weights to the criteria, the AHP is being used to accomplish this.[188]

A points based system could increase transparency by showing exactly how much each attribute contributed towards the overall ranking. The viability of a points-based lung allocation system was tested using a revealed preference analysis.[189] Discrete choice modelling was used to create an approximate composite scoring system. This composite system was compared to previous match runs and was found to have a Spearman correlation of 0.933 and a Kendall-Tau correlation of 0.803 for adults. For paediatric candidates the Spearman correlation was 0.911 and the Kendall-Tau correlation was 0.792. The correlation was strong between the current system and the composite scoring system while overcoming the problem of 'edge cases' near the hard boundaries.

The revealed preference analysis however doesn't reveal the *intended* weights of the criteria. Plans are being made for weights determined by the AHP to be compared against the weights in[189] to stimulate discussion on the appropriate weights for the new system.

Appendix H

Worked Example of AHP Calculations

To illustrate the AHP in action, four criteria labelled **A**, **B**, **C** and **D** will be used. **A** corresponds to the top row and left-most column of the comparison matrix, and **D** corresponds to the bottom row and right-most column.

For example, if criterion **A** was deemed 5 times more important than **C**, the entry at row 1 column 3 would be '5'. If **B** was judged as 3 times *less* important than **C** then the entry at row 2 column 3 would be '1/3':

$$\begin{pmatrix} & A & B & C & D \\ A & . & . & 5 & . \\ B & . & . & 1/3 & . \\ C & . & . & . & . \\ D & . & . & . & . \end{pmatrix}$$

The reciprocal entries are also contained in the matrix. Logically, if **A** is 5 times more important than **C**, then **C** is 1/5 as important as **A** and so on:

$$\begin{pmatrix} & A & B & C & D \\ A & . & . & 5 & . \\ B & . & . & 1/3 & . \\ C & 1/5 & 3 & . & . \\ D & . & . & . & . \end{pmatrix}$$

A criterion can not be more or less important than itself, so the diagonal elements are all 1:

$$\begin{pmatrix} & A & B & C & D \\ A & 1 & . & 5 & . \\ B & . & 1 & 1/3 & . \\ C & 1/5 & 3 & 1 & . \\ D & . & . & . & 1 \end{pmatrix}$$

Due to the matrix being a reciprocal matrix, only $\frac{n(n-1)}{2}$ comparisons need to be made to completely populate the matrix.

The AHP is not limited to a single decision maker. When performing pairwise comparisons group judgements can be taken and combined using the geometric mean.[151] This will allow the combination of multiple expert's opinions to be used for calculating the relative weight of importance of each criterion.

Once the matrix is populated each entry corresponds to the decision maker's estimate of the *ratio* of importance of the two criteria:

$$\begin{pmatrix} c_1/c_1 & c_1/c_2 & ... & c_1/c_n \\ c_2/c_1 & c_2/c_2 & ... & c_2/c_n \\ ... & ... & & ... \\ c_n/c_1 & c_n/c_2 & ... & c_n/c_n \end{pmatrix}$$

If the specific weights of each of the criteria $c_1...c_n$ were known then the following would hold:

$$\begin{pmatrix} c_1/c_1 & c_1/c_2 & ... & c_1/c_n \\ c_2/c_1 & c_2/c_2 & ... & c_2/c_n \\ ... & ... & & ... \\ c_n/c_1 & c_n/c_2 & ... & c_n/c_n \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ ... \\ c_n \end{pmatrix} = \begin{pmatrix} c_1 + c_1 + ... + c_1 \\ c_2 + c_2 + ... + c_2 \\ ... \\ c_n + c_n + ... + c_n \end{pmatrix} = n \begin{pmatrix} c_1 \\ c_2 \\ ... \\ c_n \end{pmatrix}$$

Where n is the number of criteria. Finding the vector $w = (c_1...c_n)^T$ such that $Aw = nw$ is the eigenvector problem.[36] By finding the corresponding eigenvector to matrix A, the elements of the eigenvector can be normalised to sum to one. Each entry then corresponds to the weight of that criterion.

Saaty in[36] defines consistency as: $A_{ij} \times A_{jk} = A_{ik}$ for all i, j. In a perfectly consistent and reciprocal matrix all eigenvalues of that matrix will be zero except a single eigenvalue that equals n. That is to say, the transitive property holds for all pairwise comparisons of the criteria. For example, if A is 2 times as important as B, and B is 3 times as important as C, then A should be 6 times more important than C; any deviation from from a value of '6' when comparing A and C results in an inconsistency. In the case of the AHP, entries in the matrix are estimates and thus prone to inconsistency.

Saaty places a limit on the amount of inconsistency that should be allowed in the

matrix by calculating the consistency index (CI):

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

Where λ_{max} is the largest eigenvalue of the matrix and n is the number of criteria. The CI can then be used to calculate the consistency ratio (CR):

$$CR = \frac{CI}{RI}$$

Where RI is the average CI of a large number of randomly generated reciprocal matrices. The CR should be below 0.1 for the pairwise comparisons to be considered consistent.

Appendix I

Lung Allocation Goals Survey - Patient Involvement

Patient Involvement - Lung Allocation Policy Goals and Values

The way in which available donor lungs are offered to transplant centres for patients waiting for lung transplantation is called the **Allocation Policy**. The performance of the current NHS lung allocation policy is currently being reviewed and we welcome your input into this process by completing this survey.

The purpose of the survey is to gain a greater understanding of what is important to patients waiting for lung transplantation, patients who have received a lung transplant, and also the perspectives of their families.

There are several different aims that an allocation policy could be designed to achieve:

- Reducing the number of patients dying on the waiting list before they are transplanted
- Increasing the total additional days of life gained as a result of having a lung transplant (also called the net transplant benefit or the extra time provided by lung transplant compared to days of life without a transplant)
- Increasing the total amount of life lived after transplant (also called the post-transplant survival duration)
- Reducing the time a patient must wait for a transplant

No single allocation policy can maximally achieve all of these aims and therefore compromises are inevitable.

By completing this survey, you will help us understand which of the above aims are most important to you. You will be presented with pairs of aims, and you will need to decide if one aim is more important than the other, and if so, how much more important. If the aims are similar in importance, select "No Preference".

The degree to which one aim is more important than the other can be indicated using the following scale:

1. Moderate
2. Strong
3. Very Strong
4. Extreme

This survey requires you to compare six pairs of allocation policy aims and will take approximately 10 minutes to complete.

--

Worked example:
To demonstrate how the comparison scale is used, here is an example in the context of purchasing a new car. There will be only three aims:

1 - Minimise cost
2 - Maximise number of seats
3 - Selecting favourite colour

First aims 1 and 2 are compared, so a decision must be made comparing the importance of minimising cost and maximising seating capacity.

There is no point selecting a car with a large seating capacity if it is unaffordable, but equally there is no point buying a cheap car that has insufficient number of seats . Given that any car that is unaffordable completely excludes it from consideration, it is decided that minimising cost is moderately more important than number of seats, so 'Minimise cost' is selected as being more important, and it is given an importance of 'Moderate'.

Next, goals 1 and 3 are compared. The colour of the car is much less important than the cost of the car, so 'Minimise cost' is selected as being more important, with an importance of 'Extreme'.

Finally, goals 2 and 3 are compared. Again, the colour is not as important as the number of seats in the car, so 'Maximise number of seats' is chosen as being more important. However the relative difference in importance is less than when comparing colour to the cost of the

1. What is your experience/relation to lung transplantation? *

- ◯ Patient on the active waiting list
- ◯ Family member of patient on active waiting list
- ◯ Lung transplant recipient
- ◯ Family member of lung transplant recipient
- ◯ Family member of a patient who was on the waiting list, but passed away before receiving a transplant
- ◯ Other

2. How would you rank the following goals in order of importance? (Highest priority first)

- Reducing the number of patients dying on the waiting list
- Reducing the time a patient must wait for a transplant
- Increasing the total amount of life lived after transplant
- Increasing the additional days of life gained as a result of having a lung transplant

3. Which is of higher importance?

- Reducing the number of patients dying on the waiting list
- Increasing the additional days of life gained as a result of having a lung transplant

*

- ◯ No Preference
- ◯ Reducing the number of patients dying on the waiting list
- ◯ Increasing the additional days of life gained from transplant

4. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

5. Which is of higher importance?

- Reducing the number of patients dying on the waiting list
- Increasing the total amount of life lived after transplant

*

- ◯ No Preference
- ◯ Reducing the number of patients dying on the waiting list
- ◯ Increasing the total amount of life lived after transplant

6. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

7. Which is of higher importance?

- Reducing the number of patients dying on the waiting list
- Reducing the time a patient must wait for a transplant

*

- ◯ No Preference
- ◯ Reducing the number of patients dying on the waiting list
- ◯ Reducing the time a patient must wait for a transplant

8. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

9. Which is of higher importance?

- Increasing the additional days of life gained as a result of having a lung transplant
- Increasing the total amount of life lived after transplant

*

- ◯ No Preference
- ◯ Increasing the additional days of life gained from transplant
- ◯ Increasing the total amount of life lived after transplant

10. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

11. Which is of higher importance?

- Increasing the additional days of life gained as a result of having a lung transplant
- Reducing the time a patient must wait for a transplant

*

- ◯ No Preference
- ◯ Increasing the additional days of life gained from transplant
- ◯ Reducing the time a patient must wait for a transplant

12. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

13. Which is of higher importance?

- Increasing the total amount of life lived after transplant
- Reducing the time a patient must wait for a transplant

*

- ◯ No Preference
- ◯ Increasing the total amount of life lived after transplant
- ◯ Reducing the time a patient must wait for a transplant

14. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

Appendix J

Lung Allocation Goals Survey - Clinicians

NHS-BT - Lung Allocation Policy Goals and Values

The way in which available donor lungs are offered to transplant centres for patients waiting for lung transplantation is called the **Allocation Policy**. The performance of the current NHS lung allocation policy is currently being reviewed and we welcome your input into this process by completing this survey.

The purpose of the survey is to gain a greater understanding of what is important to clinicians, cardiothoracic surgeons and other key stakeholders with respect to lung transplantation.

There are several different aims that an allocation policy could be designed to achieve:

- Reducing the number of patients dying on the waiting list before they are transplanted
- Increasing the total additional days of life gained as a result of having a lung transplant (also called the net transplant benefit or the extra time provided by lung transplant compared to days of life without a transplant)
- Increasing the total amount of life lived after transplant (also called the post-transplant survival duration)
- Reducing the time a patient must wait for a transplant

No single allocation policy can maximally achieve all of these aims and therefore compromises are inevitable.

By completing this survey, you will help us understand which of the above aims are most important to you. You will be presented with pairs of aims, and you will need to decide if one aim is more important than the other, and if so, how much more important. If the aims are similar in importance, select "No Preference".

The degree to which one aim is more important than the other can be indicated using the following scale:

1. Moderate
2. Strong
3. Very Strong
4. Extreme

This survey requires you to compare six pairs of allocation policy aims and will take approximately 10 minutes to complete.

--
Worked example:
To demonstrate how the comparison scale is used, here is an example in the context of purchasing a new car. There will be only three aims:

1 - Minimise cost
2 - Maximise number of seats
3 - Selecting favourite colour

First aims 1 and 2 are compared, so a decision must be made comparing the importance of minimising cost and maximising seating capacity.

There is no point selecting a car with a large seating capacity if it is unaffordable, but equally there is no point buying a cheap car that has insufficient number of seats . Given that any car that is unaffordable completely excludes it from consideration, it is decided that minimising cost is moderately more important than number of seats, so 'Minimise cost' is selected as being more important, and it is given an importance of 'Moderate'.

Next, goals 1 and 3 are compared. The colour of the car is much less important than the cost of the car, so 'Minimise cost' is selected as being more important, with an importance of 'Extreme'.

Finally, goals 2 and 3 are compared. Again, the colour is not as important as the number of seats in the car, so 'Maximise number of seats' is chosen as being more important. However the relative difference in importance is less than when comparing colour to the cost of the car, so an importance of 'Very Strong' is selected.
---

1. What is your experience/relation to lung transplantation? *

- ◯ Transplant Surgeon
- ◯ Transplant Physician
- ◯ Transplant Recipient Co-ordinator
- ◯ Specialist Nurse in Organ Donation (SNOD)
- ◯ Transplant Nurse / Nurse Practitioner
- ◯ Governance/Administration/Policymaker
- ◯ Ethicist
- ◯ Other

2. How would you rank the following goals in order of importance? (Highest priority first)

- Reducing the time a patient must wait for a transplant
- Increasing the total amount of life lived after transplant
- Increasing the additional days of life gained as a result of having a lung transplant
- Reducing the number of patients dying on the waiting list

3. Which is of higher importance?

- Reducing the number of patients dying on the waiting list
- Increasing the additional days of life gained as a result of having a lung transplant

*

- ◯ No Preference
- ◯ Reducing the number of patients dying on the waiting list
- ◯ Increasing the additional days of life gained from transplant

4. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

5. Which is of higher importance?

- Reducing the number of patients dying on the waiting list
- Increasing the total amount of life lived after transplant

*

- ◯ No Preference
- ◯ Reducing the number of patients dying on the waiting list
- ◯ Increasing the total amount of life lived after transplant

6. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

7. Which is of higher importance?

- Reducing the number of patients dying on the waiting list
- Reducing the time a patient must wait for a transplant

*

- ◯ No Preference
- ◯ Reducing the number of patients dying on the waiting list
- ◯ Reducing the time a patient must wait for a transplant

8. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

9. Which is of higher importance?

    - Increasing the additional days of life gained as a result of having a lung transplant
    - Increasing the total amount of life lived after transplant

    *

    ◯ No Preference

    ◯ Increasing the additional days of life gained from transplant

    ◯ Increasing the total amount of life lived after transplant

10. How much more important is your selected option? *

    ◯ Moderate

    ◯ Strong

    ◯ Very Strong

    ◯ Extreme

11. Which is of higher importance?

    - Increasing the additional days of life gained as a result of having a lung transplant
    - Reducing the time a patient must wait for a transplant

    *

    ◯ No Preference

    ◯ Increasing the additional days of life gained from transplant

    ◯ Reducing the time a patient must wait for a transplant

12. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

13. Which is of higher importance?

- Increasing the total amount of life lived after transplant
- Reducing the time a patient must wait for a transplant

*

- ◯ No Preference
- ◯ Increasing the total amount of life lived after transplant
- ◯ Reducing the time a patient must wait for a transplant

14. How much more important is your selected option? *

- ◯ Moderate
- ◯ Strong
- ◯ Very Strong
- ◯ Extreme

Appendix K

Description of Potential Future Methods

K.1 Decision Theory: Maximin, Maximax, and Minimax Regret

K.1.1 Maximin Strategy

The maximin strategy[164, 165] aims to maximise the minimum 'payoff' out of a set of choices. In the context of this work, the payoff would be the amount of net benefit a patient receives, and the set of choices would be offering the left lung first, the right lung first, or the lung pair.

For example, given the following choices for scenario '4' where the total net benefit for left-lung-first > right-lung first:

Scenario 4: Left-first ≥ Right-first

Table K.1: Allocation choices where allocating left-lung first results in total higher net benefit than right-lung-first.

Left Lung Allocated To	Candidate A Net Benefit	Candidate B Net Benefit	Total Net Benefit
Patient A	800 days	1500 days	2300 days
Patient B	900 days	1100 days	2000 days

If the left lung is allocated to Candidate A then the **minimum** net benefit received is 800 days. Alternatively, if the left lung is allocated to Candidate B, then the **minimum** net benefit received is 900 days. The **maxi**mum value from the possible **min**imum values is therefore 900 days, so the maximin strategy would choose to allocate right-then-left, giving Candidate B the left lung and Candidate A the left lung, even though the total overall net benefit is lower.

The example in table K.1 is just one possible permutation of values. In the example

scenario, Candidate A always has a lower net benefit than Candidate B, however there may be cases where one patient gains more net benefit than the other patient loses when comparing offering orders, the maximin strategy will still choose the ordering that maximises the minimum net benefit received in these cases.

K.1.2 Maximax Strategy

The maximax strategy[165] aims to maximise the maximum payoff from a set of options. Using the example in table K.1 the maximum net benefit received from allocating the left lung to Candidate A is 1500 days, and the maximum net benefit from allocating the left lung to Candidate B is 1100 days. In this case, the maximum net benefit is maximised by allocating the left lung to Candidate A.

K.1.3 Minimax Regret Strategy

The minimax regret strategy[165] aims to minimise the maximum regret from a set of options. Continuing to use table K.1 as an example, if the left lung is allocated to Candidate A, then Candidate A experiences 100 days 'regret' as they received 800 days of net benefit compared to the 900 days of net benefit if Candidate B had received the left lung. Conversely, in this case Candidate B would experience 0 days regret due to receiving the highest possible net benefit available to them.

The first step would be to calculate a regret table:

Scenario 4: Left-first $\geq$ Right-first

Table K.2: Regret experience by each candidate depending on ordering of offers.

Left Lung Allocated To	Candidate A Regret	Candidate B Regret
Candidate A	100 days	0 days
Candidate B	0 days	400 days

Next, the maximum regret for each choice would be calculated: 100 days in the case of Candidate A receiving the left lung, and 400 days if Candidate B receives the left lung. So in this case Candidate A would be allocated the left lung as this results in the lowest possible regret experienced.

References

[1] Sue Madden. Lung utilisation in dbd and dcd donors. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/31374/241-lung-utilisation-metrics_lung-summit-feb2023.pdf.

[2] NHS-BT. Organ and tissue donation and transplantation activity report 2022/23. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/30198/activity-report-2022-2023-final.pdf.

[3] The Cardiothoracic Advisory Group. Pol230/15 - donor lung distribution and allocation. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/30087/pol230.pdf.

[4] The Secretary of State for Health. Nhs blood and transplant (england) directions 2005. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/1864/nhsbt_directions_2005.pdf.

[5] Aurélie Moreau, Emilie Varey, Ignacio Anegon, and Maria-Cristina Cuturi. Effector mechanisms of rejection. *Cold Spring Harbor perspectives in medicine*, 3(11), 2013.

[6] C Yu Nelson, Marcus T Haug III, Saeed U Khan, Marlene Goormastic, L Kathleen Hague, Atul C Mehta, and Janet R Maurer. Does the donor-recipient abo blood group compatibility status predict subsequent lung transplantation outcomes? *The Journal of heart and lung transplantation*, 18(8):764–768, 1999.

[7] D Demos, G Divine, Gaetano Paone, J Borgi, J Morgan, Lisa Allenspach, Lisa Stagner, and Hassan Nemeh. Abo compatibility in lung transplantation. *The Journal of Heart and Lung Transplantation*, 35(4):S368, 2016.

[8] Mark L Barr, Robert C Bourge, Jonathan B Orens, Kenneth R McCurry, W Steves Ring, Tempie E Hulbert-Shearon, and Robert M Merion. Thoracic organ transplantation in the united states, 1994–2003. *American journal of transplantation*, 5(4p2):934–949, 2005.

[9] NHSBT. Pol231/5 – lung candidate selection criteria. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/26636/pol231.pdf.

[10] ISHLT. Adult lung transplantation focus theme. https://ishltregistries.org/downloadables/slides/2022/Adult_Lung_Transplantation_Focus_Theme.pptx.

[11] NHS. Chronic obstructive pulmonary disease (copd). https://www.nhs.uk/conditions/chronic-obstructive-pulmonary-disease-copd/.

[12] NHS. Pulmonary hypertension. https://www.nhs.uk/conditions/pulmonary-hypertension/.

[13] DJ Fox and RS Khattar. Pulmonary arterial hypertension: classification, diagnosis and contemporary management. *Postgraduate medical journal*, 82(973):717–722, 2006.

[14] NHS. Cystic fibrosis. https://www.nhs.uk/conditions/cystic-fibrosis/.

[15] Naritaka Kimura, Muhammad S Khan, Marc Schecter, Raheel Rizwan, Roosevelt Bryant III, Erin Wells, Christopher Towe, Farhan Zafar, and David LS Morales. Changing demographics and outcomes of lung transplantation recipients with cystic fibrosis. *The Journal of Heart and Lung Transplantation*, 35(10):1237–1244, 2016.

[16] Michael M Rey, Michael P Bonk, and Denis Hadjiliadis. Cystic fibrosis: emerging understanding and therapies. *Annual review of medicine*, 70:197–210, 2019.

[17] NHS. Idiopathic pulmonary fibrosis. https://www.nhs.uk/conditions/idiopathic-pulmonary-fibrosis/.

[18] Newcastle upon Tyne Hospitals. Alpha-1 antitrypsin deficiency (aatd). https://www.newcastle-hospitals.nhs.uk/services/clinical-genetics-service/information-for-healthcare-professionals/care-of-genetic-conditions-in-primary-care/alpha-1-antitrypsin-deficiency-aatd/.

[19] M Valapour, CJ Lehr, MA Skeans, JM Smith, E Miller, R Goff, T Mupfudze, K Gauntt, and JJ Snyder. Optn/srtr 2020 annual data report: Lung. *American Journal of Transplantation*, 22:438–518, 2022.

[20] Maryam Valapour, Carli J Lehr, David P Schladt, Jodi M Smith, Rebecca Goff, Tatenda G Mupfudze, Kaitlin Swanner, Katrina Gauntt, and Jon J Snyder. Optn/srtr 2021 annual data report: Lung. *American Journal of Transplantation*, 23(2):S379–S442, 2023.

[21] NHS-BT. Annual report on cardiothoracic organ transplantation. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/19874/nhsbt-annual-report-on-cardiothoracic-organ-transplantation-201920.pdf.

[22] NHS Blood and Transplant. Organ specific reports. https://www.odt.nhs.uk/statistics-and-reports/organ-specific-reports/.

[23] NHS-BT. Annual report on cardiothoracic organ transplantation. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/5418/cardiothoracic-annual-report-2016-17.pdf.

[24] NHS-BT. Annual report on cardiothoracic organ transplantation. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/25266/nhsbt-annual-report-on-cardiothoracic-organ-transplantation-202021.pdf.

[25] NHS-BT. Annual report on cardiothoracic organ transplantation. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/27816/nhsbt-annual-report-on-cardiothoracic-organ-transplantation-202122.pdf.

[26] NHS-BT. Annual report on lung transplantation. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/30885/nhsbt-lung-transplantation-report-2223.pdf.

[27] Amartya Sen. *Collective choice and social welfare*. Harvard University Press, 2018.

[28] Steven M Goldman. An axiomatic treatment of issues of equity and efficiency in the allocation of transplant organs: Reflections from economics. 2002.

[29] Thomas M Egan. Ethical issues in thoracic organ distribution for transplant. *American Journal of Transplantation*, 3(4):366–372, 2003.

[30] Catharyn T Liverman, Sarah Domnitz, and James F Childress. *Opportunities for organ donor intervention research: Saving lives by improving the quality and quantity of organs for transplantation*. National Academies Press, 2018.

[31] John C Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 41(2):148–164, 1979.

[32] Michael N Katehakis and Arthur F Veinott Jr. The multi-armed bandit problem: decomposition and computation. *Mathematics of Operations Research*, 12(2):262–268, 1987.

[33] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multi-armed bandit problem. *Machine learning*, 47:235–256, 2002.

[34] Volodymyr Kuleshov and Doina Precup. Algorithms for multi-armed bandit problems. *arXiv preprint arXiv:1402.6028*, 2014.

[35] Long Tran-Thanh, Archie Chapman, Enrique Munoz De Cote, Alex Rogers, and Nicholas R Jennings. Epsilon–first policies for budget–limited multi-armed bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 24, pages 1211–1216, 2010.

[36] Thomas L Saaty. A scaling method for priorities in hierarchical structures. *Journal of mathematical psychology*, 15(3):234–281, 1977.

[37] The Cardiothoracic Advisory Group (CTAG). Donor lung distribution and allocation. `https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/20278/pol230.pdf`.

[38] J. Smits et al. Et thoracic allocation system (ethas). `https://www.eurotransplant.org/wp-content/uploads/2020/01/H6-EThAS.pdf`.

[39] Optn policies. `https://optn.transplant.hrsa.gov/media/1200/optn_policies.pdf`.

[40] Guidelines for organ exchange in the scandiatransplant area. `http://www.scandiatransplant.org/members/nttr/Guideline_SHLG_oct_2017.pdf`.

[41] Clinical guidelines for organ transplantation from deceased donors. `https://tsanz.com.au/storage/documents/TSANZ_Clinical_Guidelines_Version-14.pdf`.

[42] United Network for Organ Sharing et al. A guide to calculating the lung allocation score. 2020. `https://unos.org/wp-content/uploads/unos/lung-allocation-score.pdf`.

[43] Organ Procurement and Transplantation Network. Title 42. `https://www.ecfr.gov/current/title-42/part-121#p-121.8(a)(8)`.

[44] Nelson C Yu, Marcus T Haug, Saeed U Khan, Marlene Goormastic, L.Kathleen Hague, Atul C Mehta, and Janet R Maurer. Does the donor-recipient abo blood group compatibility status predict subsequent lung transplantation outcomes? *The Journal of Heart and Lung Transplantation*, 18(8):764–768, 1999.

[45] Stephanie H Chang, Luis Angel, Deane E Smith, Julius Carillo, Darya Rudym, Melissa Lesko, Kimberly Sureau, Robert A Montgomery, Nader Moazami, and Zachary N Kon. A simple prioritization change to lung transplant allocation may result in improved outcomes. *The Annals of Thoracic Surgery*, 111(2):427–435, 2021.

[46] Elsevier B.V. Scopus. `https://www.scopus.com`.

[47] Clarivate. Web of science. `https://www.webofscience.com`.

[48] National Institutes of Health. Pubmed. `https://pubmed.ncbi.nlm.nih.gov/`.

[49] Thomas M Egan and Robert M Kotloff. Pro/con debate: lung allocation should be based on medical urgency and transplant survival and not on waiting time. *Chest*, 128(1):407–415, 2005.

[50] A Titman, CA Rogers, RS Bonser, NR Banner, and LD Sharples. Disease-specific survival benefit of lung transplantation in adults: a national cohort study. *American Journal of Transplantation*, 9(7):1640–1649, 2009.

[51] Johan De Meester, Jacqueline MA Smits, Guido G Persijn, and Axel Haverich. Listing for lung transplantation: life expectancy and transplant effect, stratified by type of end-stage lung disease, the eurotransplant experience. *The Journal of heart and lung transplantation*, 20(5):518–524, 2001.

[52] Andrew C Chang, Kevin M Chan, Robert J Lonigro, Christine L Lau, Vibha N Lama, Kevin R Flaherty, Ros Florn, Allan Pickens, Susan Murray, Fernando J Martinez, et al. Surgical patient outcomes after the increased use of bilateral lung transplantation. *The Journal of thoracic and cardiovascular surgery*, 133(2):532–540, 2007.

[53] Thomas M Egan, S Murray, RT Bustami, TH Shearon, Keith P McCullough, LB Edwards, MA Coke, ER Garrity, SC Sweet, DA Heiney, et al. Development of the new lung allocation system in the united states. *American Journal of Transplantation*, 6(5p2):1212–1227, 2006.

[54] Wayne M Tsuang, Laurie D Snyder, and Marie M Budev. Perspectives on donor lung allocation from both sides of the atlantic: The united states. *Clinical transplantation*, 34(7):e13873, 2020.

[55] Timothy S Lancaster, Jacob R Miller, Deirdre J Epstein, Nicholas C DuPont, Stuart C Sweet, and Pirooz Eghtesady. Improved waitlist and transplant outcomes for pediatric lung transplantation after implementation of the lung allocation score. *The Journal of Heart and Lung Transplantation*, 36(5):520–528, 2017.

[56] Alexander Iribarne, Mark J Russo, Ryan R Davies, Kimberly N Hong, Annetine C Gelijns, Matthew D Bacchetta, Frank D'Ovidio, Selim Arcasoy, and Joshua R Sonett. Despite decreased wait-list times for lung transplantation, lung allocation scores continue to increase. *Chest*, 135(4):923–928, 2009.

[57] Benjamin D Kozower, Bryan F Meyers, Michael A Smith, Nilto C De Oliveira, Stephen D Cassivi, Tracey J Guthrie, Honkung Wang, Beverly J Ryan, K Robert Shen, Thomas M Daniel, et al. The impact of the lung allocation score on short-term transplantation outcomes: a multicenter study. *The Journal of thoracic and cardiovascular surgery*, 135(1):166–171, 2008.

[58] Rajiv Lingaraju, Nancy P Blumenthal, Robert M Kotloff, Jason Christie, Vivek N Ahya, Jeffrey S Sager, Alberto Pochettino, and Denis Hadjiliadis. Effects of lung allocation score on waiting list rankings and transplant procedures. *The Journal of heart and lung transplantation*, 25(9):1167–1170, 2006.

[59] David M Vock, Michael T Durheim, Wayne M Tsuang, C Ashley Finlen Copeland, Anastasios A Tsiatis, Marie Davidian, Megan L Neely, David J Lederer, and Scott M Palmer. Survival benefit of lung transplantation in the modern era of lung allocation. *Annals of the American Thoracic Society*, 14(2):172–181, 2017.

[60] Justin M Schaffer, Steve K Singh, David L Joyce, Bruce A Reitz, Robert C Robbins, Roham T Zamanian, and Hari R Mallidi. Transplantation for idiopathic pulmonary arterial hypertension: improvement in the lung allocation score era. *Circulation*, 127(25):2503–2513, 2013.

[61] Hubert Chen, Stephen C Shiboski, Jeffrey A Golden, Michael K Gould, Steven R Hays, Charles W Hoopes, and Teresa De Marco. Impact of the lung allocation score on lung transplantation for pulmonary arterial hypertension. *American journal of respiratory and critical care medicine*, 180(5):468–474, 2009.

[62] Mardi Gomberg-Maitland, Cherylanne Glassner-Kolmin, Sydeaka Watson, Robert Frantz, Myung Park, Adaani Frost, Raymond L Benza, and Fernando Torres. Survival in pulmonary arterial hypertension patients awaiting lung transplantation. *The Journal of Heart and Lung Transplantation*, 32(12):1179–1186, 2013.

[63] Are Martin Holm, Franz Immer, and Christian Benden. Lung allocation for transplant: the european perspective. *Clinical transplantation*, 34(7):e13883, 2020.

[64] Jacqueline M Smits, George Nossent, Patrick Evrard, György Lang, Christiane Knoop, Johanna M Kwakkel-van Erp, Frank Langer, Rene Schramm, Ed van de Graaf, Robin Vos, et al. Lung allocation score: the eurotransplant model versus the revised us model–a cross-sectional study. *Transplant International*, 31(8):930–937, 2018.

[65] Wiebke Sommer, Christian Kühn, Igor Tudorache, Murat Avsar, Jens Gottlieb, Dietmar Boethig, Axel Haverich, and Gregor Warnecke. Extended criteria donor lungs and clinical outcome: results of an alternative allocation algorithm. *The Journal of Heart and Lung Transplantation*, 32(11):1065–1072, 2013.

[66] Martin Kosztowski, Sheng Zhou, Errol Bush, Robert S Higgins, Dorry L Segev, and Sommer E Gentry. Geographic disparities in lung transplant rates. *American Journal of Transplantation*, 19(5):1491–1497, 2019.

[67] Joshua J Mooney, Jay Bhattacharya, and Gundeep S Dhillon. Effect of broader geographic sharing of donor lungs on lung transplant waitlist outcomes. *The Journal of Heart and Lung Transplantation*, 38(2):136–144, 2019.

[68] National organ transplant act enacted 30 years ago. `https://unos.org/news/national-organ-transplant-act-enacted-30-years-ago/`.

[69] George Sigounas. Hrsa letter to optn. `https://optn.transplant.hrsa.gov/media/2397/hrsa_letter_to_optn_20171121.pdf`.

[70] Varun Puri, Ramsey R Hachem, Christian Corbin Frye, Margaret Shea Harrison, Tara R Semenkovich, John P Lynch, Gene Ridolfi, Casey Rowe, Bryan F Meyers, George Alexander Patterson, et al. Unintended consequences of changes to lung allocation policy. *American Journal of Transplantation*, 19(8):2164–2167, 2019.

[71] Carli J Lehr, Melissa Skeans, and Maryam Valapour. Validating thoracic simulated allocation model predictions for impact of broader geographic sharing of donor lungs on transplant waitlist outcomes. *The Journal of Heart and Lung Transplantation*, 39(5):433–440, 2020.

[72] Claire Drolen, Edward Cantu, Hilary J Goldberg, Joshua M Diamond, and Andrew Courtwright. Impact of the elimination of the donation service area on united states lung transplant practices and outcomes at high and low competition centers. *American Journal of Transplantation*, 20(12):3631–3638, 2020.

[73] Nathan Haywood, J Hunter Mehaffey, Sarah Kilbourne, Hannah Mannem, Max Weder, Christine Lau, Alexander S Krupnick, and Avinash Agarwal. Influence of broader geographic allograft sharing on outcomes and cost in smaller lung transplant centers. *The Journal of Thoracic and Cardiovascular Surgery*, 2020.

[74] Priya Ranganathan and CS Pramesh. Censoring in survival analysis: potential for bias. *Perspectives in clinical research*, 3(1):40, 2012.

[75] Fang Xiang, Susan Murray, and Xiaohong Liu. Analysis of transplant urgency and benefit via multiple imputation. *Statistics in medicine*, 33(26):4655–4670, 2014.

[76] Nabihah Tayob and Susan Murray. Statistical consequences of a successful lung allocation system–recovering information and reducing bias in models for urgency. *Statistics in medicine*, 36(15):2435–2451, 2017.

[77] Fang Xiang and Susan Murray. Restricted mean models for transplant benefit and urgency. *Statistics in medicine*, 31(6):561–576, 2012.

[78] Jon J Snyder, Nicholas Salkowski, Andrew Wey, Joshua Pyke, Ajay K Israni, and Bertram L Kasiske. Organ distribution without geographic boundaries: A possible framework for organ allocation. *American Journal of Transplantation*, 18(11):2635–2640, 2018.

[79] Darren E Stewart, Dallas W Wood, James B Alcorn, Erika D Lease, Michael Hayes, Brett Hauber, and Rebecca E Goff. A revealed preference analysis to develop composite scores approximating lung allocation policy in the us. *BMC Medical Informatics and Decision Making*, 21(1):1–11, 2021.

[80] Luke J Benvenuto, Joseph Costa, Davide Piloni, Meghan Aversa, Michaela R Anderson, Lori Shah, Hilary Y Robbins, Bryan Stanifer, Joshua R Sonett, Selim M Arcasoy, et al. Right single lung transplantation or double lung transplantation compared with left single lung transplantation in chronic obstructive pulmonary disease. *The Journal of Heart and Lung Transplantation*, 39(9):870–877, 2020.

[81] Carli J Lehr, Melissa Skeans, Elliott Dasenbrook, Aliza Fink, Gabriela Fernandez, Albert Faro, and Maryam Valapour. Effect of including important clinical variables on accuracy of the lung allocation score for cystic fibrosis and chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine*, 200(8):1013–1021, 2019.

[82] Darren Stewart. Moving toward continuous organ distribution. *Current Transplantation Reports*, pages 1–13, 2021.

[83] Martha Pavlakis. Continuous distribution in organ allocation: Stepping back from the edge. *Transplantation*, 105(12):2517–2519, 2021.

[84] David Thompson, Larry Waisanen, Robert Wolfe, Robert M Merion, Keith McCullough, and Ann Rodgers. Simulating the allocation of organs for transplantation. *Health care management science*, 7(4):331–338, 2004.

[85] Marie De Coster. *Simulation analysis: the impact of allocation policies on waitlist survival, waiting time and post-transplant survival*. PhD thesis, Ghent University.

[86] Julien Riou, Pierre-Yves Boëlle, Jason D. Christie, and Gabriel Thabut. High emergency organ allocation rule in lung transplantation: a simulation study. *ERJ Open Research*, 3(4), 2017.

[87] Stephanie H. Chang, Luis Angel, Deane E. Smith, Julius Carillo, Darya Rudym, Melissa Lesko, Kimberly Sureau, Robert A. Montgomery, Nader Moazami, and Zachary N. Kon. A simple prioritization change to lung transplant allocation may result in improved outcomes. *The Annals of Thoracic Surgery*, 111(2):427–435, 2021.

[88] RT Bustami, TE Hulbert-Shearon, S Murray, KP McCullough, AM Rodgers, RA Wolfe, T Egan, and RM Merion. Equity effects of implemention of a new lung allocation policy. *The Journal of Heart and Lung Transplantation*, 24(2):S121–S122, 2005.

[89] JM Tikkanen, A Hirji, R Zamel, H Zhao, LG Singer, S Keshavjee, and K Tinckam. Hladqb-matching is feasible and does not prolong wait times for lung transplant recipients. *The Journal of Heart and Lung Transplantation*, 35(4):S371, 2016.

[90] Jingjing Zou, David J. Lederer, and Daniel Rabinowitz. Efficiency in lung transplant allocation strategies. *The Annals of Applied Statistics*, 14(3):1088 – 1121, 2020.

[91] Jeffrey C. Munson, Jason D. Christie, and Scott D. Halpern. The societal impact of single versus bilateral lung transplantation for chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 184(11):1282–1288, 2011. PMID: 21868502.

[92] Erin M. Schnellinger, Edward Cantu, Douglas E. Schaubel, Stephen E. Kimmel, and Alisa J. Stephens-Shields. Clinical impact of a modified lung allocation score that mitigates selection bias. *The Journal of Heart and Lung Transplantation*, 2022.

[93] Anne L. Stephenson, Kathleen J. Ramos, Jenna Sykes, Xiayi Ma, Sanja Stanojevic, Bradley S. Quon, Bruce C. Marshall, Kristofer Petren, Joshua S. Ostrenga, Aliza K. Fink, Albert Faro, Alexander Elbert, Cecilia Chaparro, and Christopher H. Goss. Bridging the survival gap in cystic fibrosis: An investigation of lung transplant outcomes in canada and the united states. *The Journal of Heart and Lung Transplantation*, 40(3):201–209, 2021.

[94] Theodore P Papalexopoulos. *Multi-Objective Optimization for Public Policy*. PhD thesis, Massachusetts Institute of Technology, 2022.

[95] W. M. Tsuang, K. M. Chan, M. A. Skeans, J. Pyke, M. I. Hertz, A. J. Israni, L. Robbins-Callahan, G. Visner, X. Wang, T. C. Wozniak, and M. Valapour. Broader geographic sharing of pediatric donor lungs improves pediatric access to transplant. *American Journal of Transplantation*, 16(3):930–937, 2016.

[96] Joshua J. Mooney, Jay Bhattacharya, and Gundeep S. Dhillon. Effect of broader geographic sharing of donor lungs on lung transplant waitlist outcomes. *The Journal of Heart and Lung Transplantation*, 38(2):136–144, 2019.

[97] Carli J. Lehr, Melissa Skeans, and Maryam Valapour. Validating thoracic simulated allocation model predictions for impact of broader geographic sharing of donor lungs on transplant waitlist outcomes. *The Journal of Heart and Lung Transplantation*, 39(5):433–440, 2020.

[98] M. Skeans, J. Pyke, K. Audette, R. Lehman, K. Uccellini, K. M. Chan, E. D. Lease, R. Daly, and M. Valapour. Simulation of dsa-free lung allocation. *AMERICAN JOURNAL OF TRANSPLANTATION*, 19(3, SI):1016, APR 2019. American Transplant Congress (ATC), Boston, MA, JUN 01-05, 2019.

[99] M. Skeans, J. Pyke, C. Lehr, R. Lehman, K. Uccellini, K. M. Chan, E. D. Lease, R. Daly, and M. Valapour. Simulation vs. reality: 250nm as first unit of lung allocation. *AMERICAN JOURNAL OF TRANSPLANTATION*, 19(3, SI):1017–1018, APR 2019. American Transplant Congress (ATC), Boston, MA, JUN 01-05, 2019.

[100] Maryam Valapour, Carli J. Lehr, Andrew Wey, Melissa A. Skeans, Jonathan Miller, and Erika D. Lease. Expected effect of the lung composite allocation score system on us lung transplantation. *American Journal of Transplantation*, n/a(n/a).

[101] Carli J. Lehr, Andrew Wey, Melissa A. Skeans, Erika D. Lease, and Maryam Valapour. Impact of incorporating long-term survival for calculating transplant benefit in the us lung transplant allocation system. *The Journal of Heart and Lung Transplantation*, 41(7):866–873, 2022.

[102] Carli J. Lehr, Melissa A. Skeans, Erika D. Lease, and Maryam Valapour. Effects of broader geographic distribution of donor lungs on travel mode and estimated costs of organ procurement. *American Journal of Transplantation*, 21(12):4012–4022, 2021.

[103] Asil Oztekin. *Data mining-based survival analysis and simulation modeling for lung transplant*. Oklahoma State University, 2010.

[104] J. T. Magruder, A. S. Shah, T. C. Crawford, J. C. Grimm, B. Kim, J. B. Orens, E. L. Bush, R. S. Higgins, and C. A. Merlo. Simulated regionalization of heart and lung transplantation in the united states. *American Journal of Transplantation*, 17(2):485–495, 2017.

[105] Corrado Lanera, Honoria Ocagli, Marco Schiavon, Andrea Dell'Amore, Daniele Bottigliengo, Patrizia Bartolotta, Aslihan Senturk Acar, Giulia Lorenzoni, Paola Berchialla, Ileana Baldi, et al. The surplus transplant lung allocation system in italy: An evaluation of the allocation process via stochastic modeling. *International journal of environmental research and public health*, 18(13):7132, 2021.

[106] Elisabeth Mahase. Covid-19: Many icu staff in england report symptoms of ptsd, severe depression, or anxiety, study reports. *BMJ*, 372, 2021.

[107] Stephen R Knight, Antonia Ho, Riinu Pius, Iain Buchan, Gail Carson, Thomas M Drake, Jake Dunning, Cameron J Fairfield, Carrol Gamble, Christopher A Green, et al. Risk stratification of patients admitted to hospital with covid-19 using the isaric who clinical characterisation protocol: development and validation of the 4c mortality score. *bmj*, 370, 2020.

[108] Kennedy S, Freitas L, Hardy A, Downsland S, and Turner M. Clinical prioritisation assistance tool (cpat) for covid-19.

[109] LL Turstone. A law of comparative judgements. *Psychological Reviews*, 34(4):272–286, 1927.

[110] Arthur B Kahn. Topological sorting of large networks. *Communications of the ACM*, 5(11):558–562, 1962.

[111] Professor Cliff B. Jones. Adjusting khan's algorithm to be deterministic. Personal Correspondence.

[112] Oracle. What is java technology and why do i need it?

[113] Python Software Foundation. Python. `https://www.python.org/`.

[114] Standard C++ Foundation. Get started!

[115] Alan Beaulieu. *Learning SQL: master SQL fundamentals.* " O'Reilly Media, Inc.", 2009.

[116] World Wide Web Consortium (W3C). Html living standard.

[117] World Wide Web Consortium (W3C). Css snapshot 2023.

[118] Siemens. Ladder logic (lad) for s7-300 and s7-400 programming.

[119] The LaTeX Project. The latex project.

[120] Kenneth Rockwood, Xiaowei Song, Chris MacKnight, Howard Bergman, David B Hogan, Ian McDowell, and Arnold Mitnitski. A global clinical measure of fitness and frailty in elderly people. *Cmaj*, 173(5):489–495, 2005.

[121] Microsoft Corporation. Microsoft excel. `https://office.microsoft.com/excel`.

[122] Encode OSS Ltd. django rest framework. `https://www.django-rest-framework.org/`.

[123] National Health Service England. Epr.

[124] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[125] Source available.

[126] Dalhousie University. Clinical frailty scale.

[127] Government of the United Kingdom. Discrimination: your rights. `https://www.gov.uk/discrimination-your-rights`.

[128] Stewart Robinson. *Simulation: the practice of model development and use.* Bloomsbury Publishing, 2014.

[129] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2022.

[130] random.org. random.org. `https://www.random.org/`.

[131] Robert Tibshirani. The lasso method for variable selection in the cox model. *Statistics in medicine*, 16(4):385–395, 1997.

[132] Michael Alin Efroymson. Multiple regression analysis. *Mathematical methods for digital computers*, pages 191–203, 1960.

[133] David Collett. *Modelling survival data in medical research*. CRC press, 2015.

[134] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[135] NHS Blood and Cardiothoracic Advisory Group Transplant. Review of super-urgent and urgent lung allocation schemes. `https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/10474/super-urgent-lung-allocation-scheme-activity-web-version.pdf`.

[136] Ralf Bender, Thomas Augustin, and Maria Blettner. Generating survival times to simulate cox proportional hazards models. *Statistics in medicine*, 24(11):1713–1723, 2005.

[137] Christopher Jackson. flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, 70(8):1–33, 2016.

[138] Stefanie Hieke, Martina Kleber, Christine König, Monika Engelhardt, and Martin Schumacher. Conditional survival: a useful concept to provide information on how prognosis evolves over time. *Clinical Cancer Research*, 21(7):1530–1536, 2015.

[139] SRTR. Thoracic simulated allocation model. `https://www.srtr.org/media/1294/tsam-2015-user-guide.pdf`.

[140] Klaus Müller Stefan Scherfke, Ontje Lünsdorf and Tony Vignaux. Discrete event simulation for python. `https://simpy.readthedocs.io/en/latest/`.

[141] NHS-BT. Annual report on cardiothoracic organ transplantation. `https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/12636/nhsbt-cardiothoracic-organ-transplantation-annual-report-2017-2018.pdf`.

[142] The Alan Turing Institute. Fair: Framework for responsible adoption of artificial intelligence in the financial services industry. `https://www.turing.ac.uk/research/research-projects/fair-framework-responsible-adoption-artificial-intelligence-financial`.

[143] Thomas M Egan. Lung donor allocation systems. In *Lung Transplantation*, pages 125–134. CRC Press, 2016.

[144] Paul A Gagniuc. *Markov chains: from theory to implementation and experimentation.* John Wiley & Sons, 2017.

[145] Norman L Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous univariate distributions, volume 2*, volume 289. John wiley & sons, 1995.

[146] NHS-BT. Annual report on cardiothoracic organ transplantation. `https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/16795/nhsbt-annual-report-on-cardiothoracic-organ-transplantation-201819.pdf`.

[147] William F Parker, Nicole E Dussault, Renea Jablonski, Edward R Garrity, and Matthew M Churpek. Assessing the accuracy of the lung allocation score. *The Journal of Heart and Lung Transplantation*, 41(2):217–225, 2022.

[148] Sasiharan Sithamparanathan, Logan Thirugnanasothy, Stephen Clark, John H Dark, Andrew J Fisher, Kate F Gould, Asif Hasan, James L Lordan, Gerard Meachery, Gareth Parry, et al. Observational study of lung transplant recipients surviving 20 years. *Respiratory Medicine*, 117:103–108, 2016.

[149] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[150] Peter Emerson. The original borda count and partial voting. *Social Choice and Welfare*, 40:353–358, 2013.

[151] Thomas L Saaty. Axiomatic foundation of the analytic hierarchy process. *Management science*, 32(7):841–855, 1986.

[152] Thomas L Saaty. How to make a decision: the analytic hierarchy process. *European journal of operational research*, 48(1):9–26, 1990.

[153] Microsoft. Microsoft forms. `https://forms.office.com/`.

[154] Kelly Deuerling and Ryan Kerrigan. How do i use ternary diagrams? depicting ratios of three variables in the earth sciences.

[155] Jay Wright Forrester. World dynamics. *Cambridge Mass.*, 1971.

[156] Leo Freitas and Aaron Buhagiar. Specification-based csv support in vdm. 03 2023.

[157] Jay M Brahmbhatt, Travis Hee Wai, Christopher H Goss, Erika D Lease, Christian A Merlo, Siddhartha G Kapnadak, and Kathleen J Ramos. The lung allocation score and other available models lack predictive accuracy for post-lung transplant survival. *The Journal of Heart and Lung Transplantation*, 41(8):1063–1074, 2022.

[158] Carli J Lehr, Andrew Wey, Melissa A Skeans, Erika D Lease, and Maryam Valapour. Impact of incorporating long-term survival for calculating transplant benefit in the us lung transplant allocation system. *The Journal of Heart and Lung Transplantation*, 41(7):866–873, 2022.

[159] Emily M Rosenberger, Andrea F DiMartini, Annette J DeVito Dabbs, Christian A Bermudez, Joseph M Pilewski, Yoshiya Toyoda, and Mary Amanda Dew. Psychiatric predictors of long-term transplant-related outcomes in lung transplant recipients. *Transplantation*, 100(1):239, 2016.

[160] Annina Seiler, Josef Jenewein, Chantal Martin-Soelch, Lutz Goetzmann, Ilhan Inci, Walter Weder, Macé M Schuurmans, Christian Benden, Angela Brucher, and Richard Klaghofer. Post-transplant outcome-clusters of psychological distress and health-related quality of life in lung transplant recipients. *Swiss medical weekly*, 145:w14236, 2015.

[161] Anna Bertram, Jan Fuge, Hendrik Suhling, Igor Tudorache, Axel Haverich, Tobias Welte, and Jens Gottlieb. Adherence is associated with a favorable outcome after lung transplantation. *PLoS One*, 14(12):e0226167, 2019.

[162] Patrick Royston and Mahesh KB Parmar. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in medicine*, 21(15):2175–2197, 2002.

[163] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[164] Michael Maschler, Shmuel Zamir, and Eilon Solan. *Game theory*. Cambridge University Press, 2020.

[165] Kaplan Financial Limited. Maximax, maximin and minimax regret. `https://kfknowledgebank.kaplan.co.uk/maximax-maximin-and-minimax-regret-`.

[166] NHS Blood and Transplant. Organ donation and transplantation directorate - the eighteenth meeting of the nhsbt ctag(l) lungs advisory group. `https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/24614/ctag-lungs-agenda-september-2021.pdf`.

[167] NHS Blood and Transplant. Organ donation and transplantation directorate - the nineteenth meeting of the nhsbt ctag(l) lungs advisory group. `https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/28339/ctagl-22-m-02-minutes-to-be-ratified.pdf`.

[168] NHS Blood and Transplant. Organ donation and transplantation directorate - the twentieth meeting of the nhsbt ctag(l) lungs advisory group. `https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/28112/ctagl-a-22-02-ctag-lungs-agenda-final.pdf`.

[169] Gilbert Keith Chesterton. *The thing*. Aeterna Press, 1929.

[170] Krishnaiyan Thulasiraman and Madisetti NS Swamy. *Graphs: theory and algorithms*. John Wiley & Sons, 2011.

[171] SJ Willems and M Fiocco. Inverse probability censoring weights for routine outcome monitoring data. *Leiden, The Netherlands: Universiteit Leiden*, 2014.

[172] J.F.C. Kingman. *Poisson Processes*. Oxford Studies in Probability. Clarendon Press, 1992.

[173] Adam Lipowski and Dorota Lipowska. Roulette-wheel selection via stochastic acceptance. *Physica A: Statistical Mechanics and its Applications*, 391(6):2193–2196, 2012.

[174] Patrick T Harker and Luis G Vargas. The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Management science*, 33(11):1383–1403, 1987.

[175] James S Dyer. Remarks on the analytic hierarchy process. *Management science*, 36(3):249–258, 1990.

[176] Patrick T Harker and Luis G Vargas. Reply to "remarks on the analytic hierarchy process" by js dyer. *Management Science*, 36(3):269–273, 1990.

[177] SR Watson and ANS Freeling. Assessing attribute weights by ratios-comment, 1983.

[178] James S Dyer and Richard E Wendell. *A critique of the analytic hierarchy process*. Department of Management, College of Business Administration and Graduate ..., 1985.

[179] Valerie Belton and Tony Gear. On a short-coming of saaty's method of analytic hierarchies. *Omega*, 11(3):228–230, 1983.

[180] MV Mikhalevich. Remarks on the dyer-saaty controversy. *Cybernetics and Systems Analysis*, 30(1):75–79, 1994.

[181] Carlos A Bana e Costa and Jean-Claude Vansnick. A critical analysis of the eigenvalue method used to derive priorities in ahp. *European Journal of Operational Research*, 187(3):1422–1428, 2008.

[182] Omkarprasad S Vaidya and Sushil Kumar. Analytic hierarchy process: An overview of applications. *European Journal of operational research*, 169(1):1–29, 2006.

[183] Matthew J Liberatore and Robert L Nydick. The analytic hierarchy process in medical and health care decision making: A literature review. *European Journal of Operational Research*, 189(1):194–207, 2008.

[184] D Ryan Cook, Sandra Staschak, and William T Green. Equitable allocation of livers for orthotopic transplantation: an application of the analytic hierarchy process. *European journal of operational research*, 48(1):49–56, 1990.

[185] Nasrin Taherkhani, Mohammad Mehdi Sepehri, Shadi Shafaghi, and Toktam Khatibi. Identification and weighting of kidney allocation criteria: a novel multi-expert fuzzy method. *BMC medical informatics and decision making*, 19(1):182, 2019.

[186] Carol S Lin and Shannon L Harris. A unified framework for the prioritization of organ transplant patients: analytic hierarchy process, sensitivity and multifactor robustness study. *Journal of Multi-Criteria Decision Analysis*, 20(3-4):157–172, 2013.

[187] OPTN Thoracic Organ Transplantation Committee. Continuous distribution of lungs. `https://optn.transplant.hrsa.gov/media/3111/thoracic_publiccomment_201908.pdf`.

[188] OPTN. Continuous distribution. `https://optn.transplant.hrsa.gov/governance/policy-initiatives/continuous-distribution/`.

[189] Darren Stewart, Dallas Wood, James Alcorn, Michael Hayes, Brett Hauber, and Rebecca Goff. A revealed preference analysis to develop composite scores approximating lung allocation policy in the us. 2020.