Operating Room Scheduling

A Patient Prioritization Approach

M.R. de Jong

Operating Room Scheduling

A Patient Prioritization Approach

by

M.R. de Jong

to obtain the degree of Bachelor of Science
at the Delft University of Technology,
to be defended publicly on Friday July 4, 2025 at 9:30 AM.

Student number:	5606063	
Project duration:	April 22, 2025 – July 4 2025	
Thesis committee:	Dr. ir. J.T. van Essen,	TU Delft, supervisor
	Dr. ir. G.F. Nane	TU Delft

This thesis is confidential and cannot be made public until July 4, 2025.

TU Delft

Contents

Laymen’s Summary

Hospitals often face challenges in deciding which patients should have surgery first, especially with limited resources. Instead of relying solely on doctors’ judgment, this study introduces an objective method using Disability-Adjusted Life Years (DALYs) to measure the health impact of surgical delays.

Using real data from Erasmus MC, the research creates models to schedule surgeries that reduce overall health loss. These models also consider doctors’ suggested maximum waiting times by adding penalties when these are exceeded, encouraging timely surgeries without strict deadlines.

Three models are tested: one minimizes total health loss, another focuses on fairness by helping the worst-off patients, and a third focuses on the waiting time. Comparing these shows the trade-offs between maximizing overall benefit and ensuring fairness.

This approach helps hospitals make fairer, more effective surgery schedules by combining health data with doctors’ insights, improving patient care.

Abstract

Efficient surgical scheduling is essential for maximizing patient outcomes and ensuring optimal use of hospital resources. This thesis proposes and evaluates optimization strategies that incorporate Maximum Waiting Times (MWTs) assigned by doctors—reflecting subjective urgency assessments—and medical urgency quantified objectively through Disability-Adjusted Life Years (DALYs), alongside operational constraints inherent in hospital scheduling. Using real-world surgical data from Erasmus MC, the study aims to identify approaches that effectively balance equity and efficiency in patient prioritization.

The scheduling problem is addressed through Integer Linear Programming (ILP). A core optimization model is developed to minimize the total DALY loss resulting from surgical delays while respecting MWTs. Two extensions of this model are introduced: one explicitly incorporating patient waiting time with a penalty on exceeding MWTs, and another designed to minimize the maximum DALY loss across patients to promote fairness. These three distinct models, each representing a different prioritization strategy, are empirically tested and compared. Additionally, a sensitivity analysis is conducted on the parameter γ, which penalizes excess waiting time beyond the MWT, to assess how varying emphasis on excess waiting time impacts prioritization outcomes and overall scheduling effectiveness.

The results highlight key trade-offs between system-level efficiency and individual patient fairness, offering actionable insights for improving surgical scheduling practices. The findings support the integration of objective health-outcome metrics alongside clinical judgment into operational decision-making, contributing to a more equitable and effective allocation of surgical resources.

Overall, this study contributes to the field of surgical scheduling by incorporating the objective measure of Disability-Adjusted Life Years (DALYs) into prioritization decisions and promoting ethically grounded, outcome-oriented scheduling policies. The development of three distinct optimization models—balancing medical urgency, fairness, operational constraints, and MWTs—scheds light on the complex trade-offs between minimizing total DALY loss, considering waiting times and subjective urgency, and ensuring equitable patient outcomes.

1 Introduction

In healthcare systems, particularly in hospitals operating under resource constraints, the scheduling of surgical procedures is a complex and high-stakes decision-making process. Hospitals must not only consider medical urgency but also balance competing priorities across patients whose conditions may vary greatly in severity. This makes the question of who gets treated first both ethically and operationally challenging. In scenarios where demand exceeds available capacity, such as a limited number of operating rooms or specialized surgical staff, the need for a fair, data-driven, and efficient method for allocating surgical resources becomes critical. The core of this challenge lies in developing prioritization strategies that are transparent, equitable, and grounded in measurable health outcomes.

One well-established approach to quantifying health outcomes is the use of Disability-Adjusted Life Years (DALYs), a metric that reflects the burden of disease by combining years of life lost due to premature death and years lived with disability. In 2015, the total disease burden in the Netherlands was nearly 5 million DALYs, with cancer (16.5%), cardiovascular diseases (16.3%), and mental disorders (14.0%) as the leading contributors (Hilderink et al. (2020)). These figures illustrate the immense health losses associated with these conditions and highlight the urgency of interventions that can reduce the disease burden—particularly through timely surgical procedures.

In collaboration with Erasmus MC, this thesis explores how DALYs can be integrated into a decision-making framework to prioritize patients awaiting surgery. By estimating the potential health loss caused by delayed treatment, the hospital can rank patients more objectively and ensure that those who stand to lose the most—measured in DALYs—are treated sooner. However, effective surgical scheduling must also account for operational limitations. These include surgeon availability, operating room capacity, and variability in surgical durations. Given that operating rooms are among the most costly and limited resources in a hospital, optimizing their use is essential.

To balance medical urgency and logistical constraints, an optimization model is proposed that aims to maximize operating room utilization while prioritizing patients based on their DALY impact. Although patients arrive over time, the current model operates in a static setting as a first step to determine the effect of using DALYs in surgical scheduling. In reality, surgical scheduling is a dynamic process. For this reason, a rolling horizon approach—where the schedule is continuously updated to accommodate new patients and shifting resource availability—is recommended for future research. Incorporating such a method would improve adaptability and responsiveness, making the model more suitable for real-world hospital environments.

The objective of this thesis is to develop, implement, and evaluate several optimization strategies that incorporate both the medical urgency of patients—quantified through DALYs—and the operational constraints involved in hospital scheduling. These strategies are tested using real-world data from Erasmus MC, with the aim of identifying an approach that maximizes both patient benefit and resource efficiency. In Chapter 2, we start with reviewing where the relevant literature to establish the context of surgical scheduling and to examine existing methodologies. In Chapter 3, the core optimization model is formulated, using DALYs to prioritize patients based on expected health loss from surgical delays. Chapter 4 extends this model with two variations: one that explicitly considers waiting time and another that focuses on minimizing the maximum DALY loss across patients. The dataset used to evaluate these models is described in Chapter 5, providing the empirical foundation for testing. In Chapter 6, a sensitivity analysis explores the impact of placing more or less emphasis on excess waiting time in the prioritization scheme. In Chapter 7 presents the results of the different models, comparing their performance and discussing the trade-offs involved, with the ultimate goal of informing a more equitable and efficient approach to surgical scheduling. The discussion and conclusions are presented in Chapter 8, which summarizes the key findings and insights from the preceding chapters.

2 Literature

A substantial body of research has focused on optimizing operating room (OR) efficiency. For instance, Cardoen et al. (2010) and Wang et al. (2021) present extensive reviews on operational research studies on operating room (OR) planning and scheduling. These studies explore a broad range of factors related to OR utilization, including patient characteristics, performance measures, decision structures, and the handling of uncertainty. The existing literature emphasizes operational performance; however, the focus of this thesis differs by prioritizing patients based on Disability-Adjusted Life Years (DALYs). This approach emphasizes patient-centered decision-making rather than purely operational performance, while still aiming to achieve effective OR utilization.

Van Oostrum (2009) provides a comprehensive analysis of surgical patient planning at Erasmus University Medical Center, where a cyclic and integrated planning method known as Master Surgery Scheduling (MSS) was developed. This approach iteratively schedules standard surgery types across ORs in fixed cycles, aiming to create predictable patient flows and balanced resource utilization across departments such as wards and intensive care units. The planning is supported by mathematical models such as bin-packing algorithms and portfolio techniques, which help reduce planned slack and improve OR utilization. Although the MSS concept was piloted and tested using real data from Erasmus MC, Van Oostrum (2009) notes that full-scale implementation requires organizational commitment and adaptation to hospital-specific constraints. Some aspects of the MSS framework were explored in collaboration with staff at Erasmus MC, indicating practical interest and feasibility, but widespread operational adoption is not explicitly confirmed.

Given the nature of elective procedures, this thesis specifically addresses the scheduling and prioritization of elective surgical patients—those whose conditions do not require immediate attention but for whom surgical delay can still lead to substantial health loss. In such settings, patient outcomes are particularly sensitive to the management of waiting times over extended periods, making a nuanced prioritization approach essential.

The Disability-Adjusted Life Year (DALY) is a standardized metric developed by the World Health Organization (WHO) to quantify the overall burden of disease. It combines the effects of both premature mortality and morbidity into a single value, enabling consistent comparison across diseases and interventions. One DALY represents one lost year of “healthy” life, and is calculated as the sum of Years of Life Lost due to early death and Years Lived with Disability due to illness or impairment (Murray & Lopez, 1996).

In this thesis, DALYs are employed as a metric for prioritizing elective surgical patients based on the health impact of surgical delay. While DALYs are conventionally reported per year or per month, this study adopts a daily resolution—expressing DALY loss per day of surgical delay—to facilitate fine-grained scheduling decisions. This approach builds on the work of Van Alphen et al. (2024), who developed a three-state cohort state-transition model (pre-operative, post-operative, deceased) to estimate health loss due to surgical delays. Their model incorporates patient-specific parameters including survival probabilities, quality of life before and after surgery, and timing of intervention. For each surgery type, expected DALY loss per month of delay was calculated and used to rank procedures by their relative urgency.

By converting these estimates into DALY loss per day, this thesis enhances the prioritization framework with a more detailed assessment of patient urgency. A higher DALY loss per day indicates a greater daily decline in patient health while waiting for surgery and thus warrants higher scheduling priority. This patient-centered approach shifts the focus away from operational metrics, while still supporting effective OR utilization.

Efficient scheduling and prioritisation of patients from a waitlist requires dynamic methods that adapt to uncertainties such as variability in clinical urgency. In this regard, rolling horizon approaches have demonstrated significant advantages. Addis et al. (2016) introduced a rolling horizon framework for operating room scheduling, focusing on minimising delays and rescheduling costs under uncertainty. Their method inspired the design of dynamic rescheduling policies that can be adjusted at each decision point.

Extending this concept, Kamran et al. (2019) proposed an adaptive operating room planning and scheduling model under a rolling horizon framework, incorporating stochastic surgery durations, emergency arrivals, and block scheduling constraints. Their model emphasizes minimising patient waiting times, tardiness, cancellations, and surgeon idle times through a stochastic mixed-integer programming formulation. While this thesis is motivated by such dynamic approaches, the current model operates in a static setting. This choice was made deliberately as a first step to investigate the implications of incorporating DALYs into surgical scheduling. In reality, surgical scheduling is inherently dynamic; thus, future work will aim to extend this static model into a fully dynamic, rolling horizon framework that more accurately reflects the complexity and variability of real-world hospital environments.

3

DALY Model

In this study, we develop a model to schedule surgical patients based on their priority. In this case, a patient's priority is given by their health loss, denoted by Disability-Adjusted Life Years (DALYs). We assume that a master surgery schedule is available which specifies the allocating of surgical specialties to operating rooms on specific days. Since this means that there are no shared resources between the different specialties, we can model this problem for each specialty separately.

Scheduling Framework

At each iteration, a schedule is generated for the upcoming D days. We define $\mathcal{D} = \{1, 2, \ldots, D\}$ as the set of all days in this planning horizon. At the beginning of each planning horizon, the model includes all patients currently on the waiting list who have not yet been scheduled in any of the previous iterations. Importantly, because the schedule for each scheduling window is fixed at the start of the period, any new patients who enter the waiting list during the planning horizon are not considered immediately. Instead, they are deferred to the next iteration, where they will be eligible for scheduling in the subsequent planning horizon. This process is illustrated in Figure 3.1.

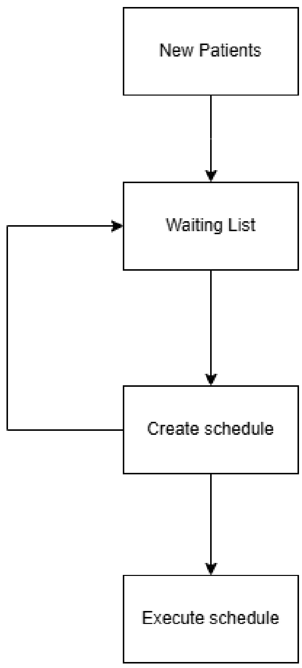


Figure 3.1: Flow of patients through the scheduling model.

Sets and Parameters

We define the following sets and parameters used in the model to represent patients, operating rooms, scheduling availability, and clinical priorities:
Let P represent the total number of patients on the current waiting list, and define the set of all patients as $\mathcal{P} = \{1, 2, \ldots, P\}$. Each element in this set corresponds to an individual patient requiring surgical care. Similarly, let R denote the number of available operating rooms in the facility. The set of all operating rooms is then given by $\mathcal{R} = \{1, 2, \ldots, R\}$.

Because the schedule is predefined with specific days allocated to each specialty and corresponding operating rooms, not all operating rooms are available on all days for the specialty under consideration.

To capture this, we define the set of valid operating room-day pairs for the specialty as:

$$\mathcal{Q} \subseteq \mathcal{D} \times \mathcal{R},$$

where each element $(d, r) \in \mathcal{Q}$ indicates that operating room $r \in \mathcal{R}$ is available on day $d \in \mathcal{D}$ for scheduling surgeries of that specific specialty. All scheduling decisions must respect this constraint; hence, surgeries can only be assigned to OR-days within $\mathcal{Q}$.

For each patient $p \in \mathcal{P}$, several parameters are defined. The parameter $DALY_p$ represents the Disability-Adjusted Life Years per day associated with patient $p \in \mathcal{P}$, providing a measure of the health benefit gained from performing the surgery. The parameter WT_p denotes the time that patient $p \in \mathcal{P}$ has already spent on the waiting list, while MWT_p indicates the maximum acceptable waiting time for that patient. This threshold is assigned by the treating physician based on clinical judgment. Finally, Dur_p represents the expected duration of the surgical procedure required by patient $p \in \mathcal{P}$. The parameter C_{dr} represents the total available time for scheduling surgeries in operating room $r \in \mathcal{R}$ on day $d \in \mathcal{D}$. In addition, we introduce a parameter $\gamma > 0$, which determines the weight of the penalty associated with exceeding the maximum acceptable waiting time for a patient. A higher value of γ reflects a stronger emphasis on minimizing violations of acceptable waiting times, thus penalizing tardiness more heavily in the objective function.

Decision Variables

We define the following decision variables in the model:

$$x_{pdr} = \begin{cases} 1, & \text{if patient } p \in \mathcal{P} \text{ is scheduled on day } d \in \mathcal{D} \text{ in operating room } r \in \mathcal{R}, \\ 0, & \text{otherwise.} \end{cases}$$

This binary variable indicates whether a specific patient is scheduled for surgery on a given day in a particular operating room.

$$y_{dr} = \begin{cases} 1, & \text{if a long surgery is scheduled on day } d \in \mathcal{D} \text{ in operating room } r \in \mathcal{R}, \\ 0, & \text{otherwise.} \end{cases}$$

This variable identifies whether a long surgery, i.e., a surgery with a duration greater than C_{dr} minutes, is scheduled in an operating room on a specific day.

$$s_p \geq 0, \text{excess waiting time for patient } p \in \mathcal{P}\text{, i.e., the number of days beyond } MWT_p.$$

This variable stands for any waiting time beyond the acceptable threshold for each patient, allowing for a violation of the maximum waiting time with a corresponding penalty.

3.1. DALY Model

$$\min \quad \sum_{p\in\mathcal{P}} DALY_p \left(\sum_{(d,r)\in\mathcal{Q}} \left((WT_p + d - 1)x_{pdr}\right) + (WT_p + D)\left(1 - \sum_{(d,r)\in\mathcal{Q}} x_{pdr}\right)\right) + \gamma \sum_{p\in\mathcal{P}} \frac{s_p}{MWT_p} \tag{3.1a}$$

$$\text{s.t.} \quad \sum_{(d,r)\in\mathcal{Q}} x_{pdr} \leq 1 \quad \forall p \in \mathcal{P} \tag{3.1b}$$

$$\sum_{\substack{p\in\mathcal{P} \\ Dur_p \leq C_{dr}}} Dur_p\, x_{pdr} \leq C_{dr}(1 - y_{dr}) \quad \forall (d,r) \in \mathcal{Q} \tag{3.1c}$$

$$\sum_{\substack{p\in\mathcal{P} \\ Dur_p > C_{dr}}} x_{pdr} \leq y_{dr} \quad \forall (d,r) \in \mathcal{Q} \tag{3.1d}$$

$$\sum_{(d,r)\in\mathcal{Q}} (WT_p + d - 1)x_{pdr} + (WT_p + D)\left(1 - \sum_{(d,r)\in\mathcal{Q}} x_{pdr}\right) \leq MWT_p + s_p \quad \forall p \in \mathcal{P} \tag{3.1e}$$

$$x_{pdr} \in \{0,1\} \quad \forall p \in \mathcal{P} \quad \forall (d,r) \in \mathcal{Q} \tag{3.1f}$$

$$y_{dr} \in \{0,1\} \quad \forall (d,r) \in \mathcal{Q} \tag{3.1g}$$

$$s_p \geq 0 \quad \forall p \in \mathcal{P} \tag{3.1h}$$

The objective function (3.1a) minimizes the total health loss across all patients. For each patient, the DALY parameter represents the health impact per day of delayed surgery. If a patient is scheduled during the current planning horizon on day $d \in \mathcal{D}$, the delay is given by their current waiting time WT_p plus the number of days until their scheduled procedure, $d - 1$. If a patient is not scheduled within the current planning horizon, they are assumed to wait at least until the next planning horizon, resulting in a total delay of $WT_p + D$. This delay is multiplied by the patient's DALY value to estimate the total health loss due to waiting. An additional penalty term, scaled by γ, is applied for any time spent on the waiting list beyond the maximum acceptable waiting time; this penalty is normalized by the maximum waiting time to prioritize patients with more urgent needs.

Constraints (3.1b) ensure that each patient $p \in \mathcal{P}$ can be scheduled at most once, meaning that every patient is either scheduled in the current planning horizon or later. Constraints (3.1c) address the operating room time for short surgeries: if a surgery has an expected duration of C_{dr} or less, the total scheduled surgery time in operating room $r \in \mathcal{R}$ on day $d \in \mathcal{D}$ must not exceed C_{dr}. Constraints (3.1d) handle long surgeries (i.e., those exceeding C_{dr}), requiring that these can only be scheduled alone in operating room $r \in \mathcal{R}$ on day $d \in \mathcal{D}$, ensuring that no other procedures are scheduled at the same time in this OR. Finally, Constraints (3.1e) determine the excess waiting time of patient $p \in P$ if the patient has to wait beyond their maximum waiting time (MWT_p) for their surgery. This excess waiting time s_p of patient $p \in P$ is penalized in the objective function.

3.2. Conclusion

This chapter presents a mathematical optimization model that utilizes DALYs to prioritize and schedule surgical patients. By incorporating individual patient characteristics such as current wait time, maximum acceptable waiting time, and procedure duration, the model aims to minimize the total health burden caused by surgical delays. The inclusion of DALYs in the objective function allows the model to focus scheduling decisions on patients for whom delays are most harmful in terms of health outcomes.

The model operates on a fixed planning horizon of D days and generates a new schedule at the start of each planning horizon. While new patients are not incorporated immediately upon arrival, they are included in subsequent planning horizons. This periodic re-optimization supports a degree of responsiveness and equity in scheduling, while remaining compatible with the operational constraints of specialty-specific operating room assignments and fixed daily capacities. The model also accounts for both standard and long-duration surgeries by enforcing capacity-based constraints for each operating room-day combination.

4 Alternative Models

To evaluate the performance and prioritization strategies of the DALY model presented in Chapter 3, we introduce two alternative models. These models are constructed with modifications to the original objective function, while keeping the underlying constraints and scheduling mechanics identical. The goal is to compare how different prioritization criteria influence surgical scheduling outcomes.

4.1. Waiting Time Model

The first alternative, referred to as the Waiting Time (WT) model, removes the DALY-based prioritization and instead bases scheduling decisions on how long patients have waited relative to their maximum acceptable waiting time. In this model, the DALY parameter in the objective function is replaced by the inverse of the patient's maximum waiting time, $\frac{1}{MWT_p}$. It is important to note that the maximum waiting time, MWT_p, is assigned by the treating physician based on clinical judgment and therefore represents a subjective threshold rather than a standardized or purely objective measure. This shift changes the focus away from clinical severity and toward time-based fairness, giving higher priority to patients with shorter maximum allowable waiting times regardless of the potential health benefit of their surgery. In this model, we introduce a separate penalty parameter $\bar{\gamma}$ for excess waiting time, which will be discussed in more detail in Chapter 6.

$$\min \quad \sum_{p\in\mathcal{P}} \frac{1}{MWT_p}\left(\sum_{(d,r)\in\mathcal{Q}} \big((WT_p + d - 1)x_{pdr}\big) + (WT_p + D)\left(1 - \sum_{(d,r)\in\mathcal{Q}} x_{pdr}\right)\right) + \bar{\gamma}\sum_{p\in\mathcal{P}} \frac{s_p}{MWT_p} \tag{4.1a}$$

s.t. (3.1b) - (3.1h)

4.2. MinMax DALY Model

The second alternative, the MinMax DALY model, shifts the objective entirely. Instead of focusing on minimizing the total DALY-adjusted waiting time across all patients, this model minimizes the worst individual experience in terms of DALY impact. The goal is to prevent any single patient—particularly those with the highest health impact from delays—from facing excessively long waiting times, thereby promoting a form of maximum individual fairness.

To implement this, a new variable $z \geq 0$ is introduced to represent the highest experienced DALYs among all patients. The objective function then becomes minimizing this maximum value, along with penalties for any patient who waits longer than their allowed maximum waiting time.

$$\min \quad z + \gamma \sum_{p \in \mathcal{P}} \frac{s_p}{MWT_p} \tag{4.2a}$$

$$\text{s.t.} \quad z \geq DALY_p \left(\sum_{(d,r) \in Q} \left((WT_p + d - 1) x_{pdr} \right) + (WT_p + D) \left(1 - \sum_{(d,r) \in Q} x_{pdr} \right) \right) \quad \forall p \in \mathcal{P} \tag{4.2b}$$

$$\text{(3.1b)} - \text{(3.1h)} \tag{4.2c}$$

$$z \geq 0 \tag{4.2d}$$

After solving this model, we obtain the optimal value z^*, representing the minimum achievable upper bound on the DALY burden experienced by any individual patient due to waiting. This bound is then incorporated as an additional constraint in the original DALY minimization model (Section 3.1) to ensure that no patient suffers a DALY impact exceeding z^*.

$$\min \quad \sum_{p \in \mathcal{P}} DALY_p \left(\sum_{(d,r) \in Q} \left((WT_p + d - 1) x_{pdr} \right) + (WT_p + D) \left(1 - \sum_{(d,r) \in Q} x_{pdr} \right) \right) + \gamma \sum_{p \in \mathcal{P}} \frac{s_p}{MWT_p} \tag{4.3a}$$

$$\text{s.t.} \quad z* \geq DALY_p \left(\sum_{(d,r) \in Q} \left((WT_p + d - 1) x_{pdr} \right) + (WT_p + D) \left(1 - \sum_{(d,r) \in Q} x_{pdr} \right) \right) \quad \forall p \in \mathcal{P} \tag{4.3b}$$

$$\text{(3.1b)} - \text{(3.1h)} \tag{4.3c}$$

By enforcing this upper bound, the revised DALY model not only minimizes the total health burden across all patients but also ensures that no individual patient experiences excessive health burden due to treatment delays. This approach helps to safeguard the most vulnerable patients by reducing extreme outcomes in terms of health loss.

In this chapter, we introduced two alternative models to the DALY-based scheduling approach developed earlier. While all three models operate under identical scheduling constraints, they differ fundamentally in their prioritization logic. The DALY model emphasizes efficiency by minimizing the total health burden accumulated across all patients. In contrast, the WT model prioritizes time-based fairness, focusing on how long patients have waited relative to their maximum acceptable waiting time, irrespective of health severity. Finally, the MinMax DALY model aims to protect the most vulnerable patients by minimizing the maximum individual harm, ensuring that no single patient bears a disproportionately high burden. These alternatives offer distinct perspectives on what constitutes an equitable and effective surgical scheduling policy. Their implications are compared in Chapter 6 in terms of waiting times and total DALYs, revealing the trade-offs between efficiency, fairness, and protection of the most critical cases.

5
Data

This chapter describes the dataset used in the development of the scheduling model. The data was obtained from the Erasmus University Medical Center (Erasmus MC) and includes information on patients who were on the surgical waiting list during the year 2019.

5.1. Data Selection

For this study, only patients who were on the waiting list at any point during 2019 were considered, provided that a Disability-Adjusted Life Year (DALY) value was available for their condition. As DALY estimates are currently limited to a specific set of diseases, this resulted in a filtered subset of the full dataset. DALY-per-day values were available for 72 different conditions, ranging from 0.000185 to 0.004299. Figure 5.1 shows the distribution of these values.

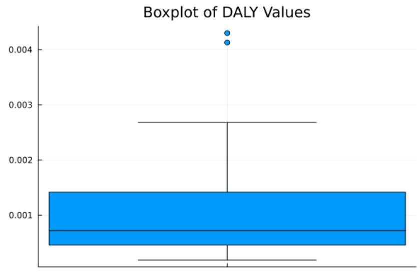


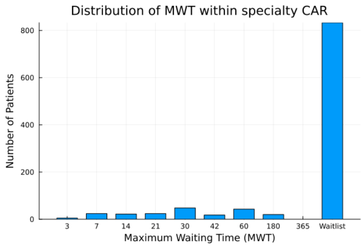
Figure 5.1: Distribution of DALY values

To ensure data quality and model relevance, only a selected set of specialties was included. The following five specialties were chosen based on the number of recorded surgeries and availability of DALY data:

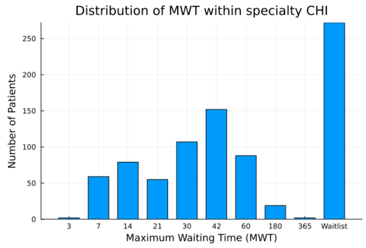
- **Cardiology (CAR)** – 1037 surgeries
- **General Surgery (CHI)** – 835 surgeries
- **Cardiothoracic Surgery (CTC)** – 401 surgeries
- **Gynecology (GYN)** – 208 surgeries
- **Ear, Nose, and Throat (KNO)** – 193 surgeries

Specialties with fewer than approximately 100 surgeries were excluded due to insufficient data.

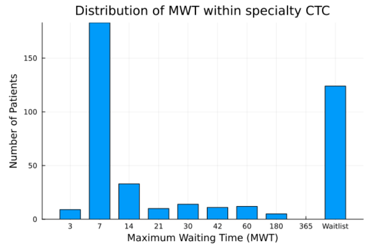
The dataset includes only non-urgent cases, defined as patients with a maximum allowed waiting time (MWT) of three days or more. The urgency categories range from 3 days to 1 year, as well as an open-ended category labeled “according to waiting list.” Figure 5.2 shows the distribution of these categories per specialty.



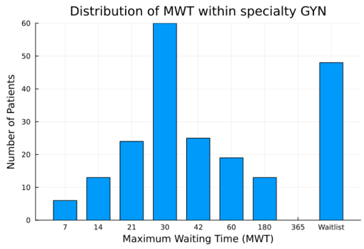
(a) Cardiology (CAR)



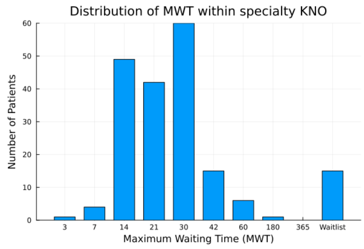
(b) General Surgery (CHI)



(c) Cardiothoracic Surgery (CTC)



(d) Gynecology (GYN)



(e) Ear, Nose, and Throat (KNO)

Figure 5.2: Distribution of maximum allowed waiting time (MWT) per specialty.

5.2. Time and Inflow Analysis

Each patient record in the dataset includes the date they were added to the waiting list. This time-stamp made it possible to compute the actual waiting time WT_p for each patient $p \in \mathcal{P}$, which was used as a key parameter in the scheduling model. This allowed the model to evaluate not just planned scheduling, but also how long patients had already been waiting at the start of the planning horizon.

The dataset also includes estimated durations for each surgery, which are used for planning purposes. Actual durations are not used because they are only known post-operation and could introduce bias. However, discrepancies between estimated and actual durations can lead to overtime or under-utilization in practice. While the current model assumes fixed durations, future extensions could consider stochastic variations or robust optimization to improve schedule feasibility. Figure 5.3 shows the distribution of estimated surgery durations across all included specialties.

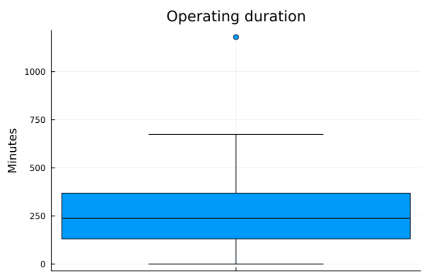
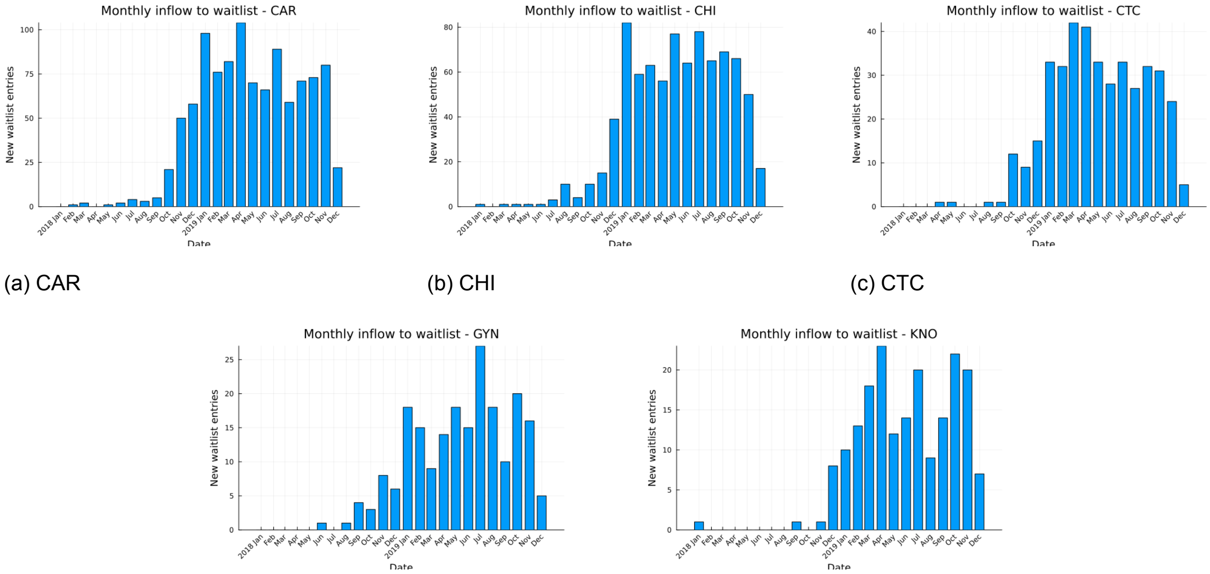


Figure 5.3: Estimated surgery duration distribution

To understand the dynamics of patient arrivals, Figure 5.4 presents the monthly inflow of patients to the waiting list for each specialty during 2018 and 2019. It shows that a significant number of patients were already on the waiting list prior to the start of the scheduling horizon in January 2019. This reflects the realistic situation in hospitals, where scheduling begins with an existing backlog rather than an empty list. As such, the model does not assume a clean slate but incorporates these initial waiting times, allowing for more accurate planning and prioritization based on how long patients have already waited.



(a) CAR (b) CHI (c) CTC

(d) GYN (e) KNO

Figure 5.4: Monthly inflow to the waiting list for each specialty (2018–2019).

5.3. Operating Room Allocation

Erasmus MC has an existing operating room schedule indicating when each specialty can perform surgeries. However, this schedule was not representative for this study due to the reduced dataset. Therefore, we constructed a new MSS that allocates operating room time to each specialty in such a way that approximately 80% of surgeries could be scheduled. This allows for more balanced and comparable results across specialties.

	Week 1					Week 2				
	Mon	Tue	Wed	Thu	Fri	Mon	Tue	Wed	Thu	Fri
OR 1	CAR	CAR	CAR	CAR	CAR	CAR	CAR	CAR	CAR	CAR
OR 2	CHI	CHI	CHI	CHI	CHI	CHI	CHI	CHI	CHI	CHI
OR 3	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	Other
OR 4	GYN	Other	Other	Other	KNO	GYN	Other	Other	Other	Other

6
Sensitivity Analysis

In this chapter, we analyze and compare the performance of the three scheduling models introduced in earlier chapters: the DALY model, the Waiting Time (WT) model, and the MinMax DALY model. We evaluate these models based on three criteria: the total Disability-Adjusted Life Years (DALYs), the spread of the DALYs and a fairness metric derived from waiting time excesses.

6.1. Evaluation Metrics

We investigate the total DALY, defined as the sum of $DALY_p$ multiplied by the waiting time across all scheduled patients. This quantity serves as an indicator for the total health gain achieved through the model's scheduling decisions. Since DALYs measure the health burden prevented by timely surgery, a lower total DALY indicates that the model has prioritized patients for whom surgery yields higher health benefits. Therefore, this metric reflects how a model reduces population-level health loss.

In addition to the total DALYs, we consider the spread of DALYs across patients to evaluate how the health burden is distributed.

To address disparities in how long patients wait beyond clinically acceptable limits, we introduce a third metric: the Normalized Excess Waiting Time (NEWT), defined as the sum of $\frac{s_p}{MWT_p}$ across all patients. Here:

- s_p denotes the Excess Waiting Time—the number of days a patient is scheduled beyond their Maximum Waiting Time (MWT), and
- $\frac{s_p}{MWT_p}$ normalizes the delay by the patient's individual Maximum Waiting Time, placing greater weight on delays experienced by patients with shorter permissible waiting times.

Together, the total DALY, the spread of the DALYs and NEWT allow us to evaluate both the effectiveness and the fairness of each model.

6.2. Analytical Comparison of Trade-offs in a Two-Patient Case

To better understand how the parameter γ influences scheduling priorities, we analyze a simplified case with two patients and compare the objective function under different patient assignments. We consider the objective function of patient $p \in \mathcal{P}$ in the DALY model:

$$\text{Obj}_p = DALY_p \cdot w_p + \gamma \cdot \frac{s_p}{MWT_p}$$

where w_p is the time the patient waits on the waiting list and $s_p = max(0, w_p - MWT_p)$ is the excess waiting time. Assume that two people arrive on the waiting list at the same time, and assume in an optimal solution, we have $w_1 < w_2$, so $w_2 = w_1 + c$.

We consider two patients:

- **Patient 1**: has a DALY value of $DALY_1$ and a maximum waiting time of MWT_1.

- **Patient 2**: has a DALY value of $DALY_2$ and a maximum waiting time of MWT_2.

We compare two scheduling orders for these patients:

Objective value	First Scheduled Patient	Second Scheduled Patient
C_1	Patient 1	Patient 2
C_2	Patient 2	Patient 1

Taking the difference $C_2 - C_1$ allows us to directly compare the overall impact of the two scheduling options, based on a scoring function that combines DALYs and penalties for exceeding the maximum waiting time.

Since we minimize the objective and we assumed that in the optimal solution patient 1 is scheduled first, i.e. $w_1 < w_2$, we have that $C_2 - C_1 > 0$. This comparison allows us to analyze how the trade-off between health benefit and waiting time penalties, influenced by γ, affects the preferred scheduling order in a simplified setting.

This gives the following:

$$\begin{aligned}
C_1 &= DALY_1 w_1 + \gamma \frac{max(0, w_1 - MWT_1)}{MWT_1} + DALY_2 w_1 + DALY_2 c + \gamma \frac{max(0, w_1 + c - MWT_2)}{MWT_2} \\
C_2 &= DALY_1 w_1 + DALY_1 c + \gamma \frac{max(0, w_1 + c - MWT_1)}{MWT_1} + DALY_2 w_1 + \gamma \frac{max(0, w_1 - MWT_2)}{MWT_2} \\
C_2 - C_1 &= \gamma \left(\frac{max(0, w_1 + c - MWT_1) - max(0, w_1 - MWT_1)}{MWT_1} \right) \\
&\quad - \gamma \left(\frac{max(0, w_1 + c - MWT_2) - max(0, w_1 - MWT_2)}{MWT_2} \right) + (DALY_1 - DALY_2)c > 0
\end{aligned}$$

So, scheduling patient 1 first is optimal as long as:

$$\begin{aligned}
DALY_1 - DALY_2 > \frac{\gamma}{c} \Bigg(&\frac{\max(0, w_1 + c - MWT_2) - \max(0, w_1 - MWT_2)}{MWT_2} \\
&- \frac{\max(0, w_1 + c - MWT_1) - \max(0, w_1 - MWT_1)}{MWT_1} \Bigg)
\end{aligned}$$

Scenario 1: Patient 1 Has Higher DALY

Assume $DALY_1 > DALY_2$, and define $\delta = DALY_1 - DALY_2 > 0$.
So scheduling patient 1 first is optimal as long as:

$$\delta > \frac{\gamma}{c} \left(\frac{max(0, w_1 + c - MWT_2) - max(0, w_1 - MWT_2)}{MWT_2} - \frac{max(0, w_1 + c - MWT_1) - max(0, w_1 - MWT_1)}{MWT_1} \right)$$

Now if $MWT_1 < MWT_2$, it will always be optimal to schedule Patient 1 first. So for the remainder, we assume that $MWT_1 > MWT_2$. We can further distinguish the following cases:

Both patients are scheduled before their MWTs for both orderings:

No penalties apply. Since $DALY_1 > DALY_2$, Patient 1 is always scheduled first.

Both patients are scheduled after their MWTs for both orderings:

$$\delta > \gamma \left(\frac{1}{MWT_2} - \frac{1}{MWT_1} \right) \Rightarrow \gamma < \frac{(DALY_1 - DALY_2) \cdot MWT_1 \cdot MWT_2}{MWT_1 - MWT_2}$$

Since patients are always scheduled after their MWTs in both orderings, penalties always apply. This inequality determines whether DALY or MWT dominates the scheduling decision. If γ is below the threshold, DALY decides the order. If γ is larger than this threshold, the patient with the shorter MWT is prioritized, even if their DALY is lower.

The threshold range is known to be $[1.3 \times 10^{-7}, 3.0]$ because MWT and DALY values come from a finite set of clinical scenarios, and since for this case is assumed that both patients are scheduled after their MWTs in both orderings, it was possible to compute the full range of possible threshold values.

Patient 1 scheduled before MWT in both orderings, and Patient 2 after their MWT in both orderings:

$$c\delta > \gamma \frac{c}{MWT_2} \Rightarrow \gamma < MWT_2(DALY_1 - DALY_2)$$

This inequality defines the threshold at which the scheduler shifts from prioritizing DALY reduction to avoiding MWT violations. Since only patient 2 exceeds their maximum waiting time, the inequality captures the trade-off. If γ is less than this threshold, then the additional health benefit from scheduling patient 1 first outweighs the penalty of delaying patient 2. The range of the right-hand side can be computed and is approximately $[1.3 \times 10^{-7}, 1.5]$.

Patient 1 scheduled after MWT only in second schedule, Patient 2 always after MWT:

$$\delta > \frac{\gamma}{c}\left(\frac{c}{MWT_2} - \frac{w_1 + c - MWT_1}{MWT_1}\right) \Rightarrow \gamma < \frac{DALY_1 - DALY_2}{\left(\frac{1}{MWT_2} - \frac{w_1+c-MWT_1}{cMWT_1}\right)}$$

This inequality captures the trade-off when Patient 2 always exceeds their MWT, and Patient 1 exceeds theirs only in one of the schedules. If γ is below this threshold, the DALY difference dominates and Patient 1 is scheduled first. If γ is above the threshold, then the health benefit is not enough to justify the added penalty, so patient 2 is scheduled first.

The threshold for γ is highly sensitive to the values of w_1, c, DALY and the MWT values. Because of this variability, no general conclusion can be drawn about the influence of γ in this scheduling configuration — it depends strongly on the specific timing parameters.

Patient 1 is scheduled before MWT in both orderings, Patient 2 after MWT in first ordering:

$$\delta > \frac{\gamma}{c}\left(\frac{w_1 + c - MWT_2}{MWT_2}\right) \Rightarrow \gamma < \frac{c(DALY_1 - DALY_2)}{\frac{w_1+c-MWT_2}{MWT_2}}$$

This inequality captures the trade-off when patient 1 is scheduled before their MWT, while Patient 2 exceeds their MWT in the first ordering. If γ is below this threshold, the difference in health impact dominates the decision, so patient 1 is prioritized despite patient 2's delay. If γ is above the threshold, the MWT and excess waiting time become more important. Because this threshold depends strongly on w_1, c, DALY and the MWT values. Because of this variability, no general conclusion can be drawn about the influence of γ in this scheduling configuration — it depends strongly on the specific timing parameters.

Patient 1 is scheduled after MWT in the second ordering, Patient 2 after MWT in first ordering:

$$\delta > \frac{\gamma}{c}\left(\frac{w_1 + c - MWT_2}{MWT_2} - \frac{w_1 + c - MWT_1}{MWT_1}\right) \Rightarrow \gamma < \frac{c(DALY_1 - DALY_2)}{\frac{w_1+c-MWT_2}{MWT_2} - \frac{w_1+c-MWT_1}{MWT_1}}$$

The key factor that determines the scheduling decision is the difference in how much each patient's waiting time goes beyond their allowed limit relative to their maximum waiting times. The model then weighs whether the difference in health impacts between the patients justifies accepting that additional delay for one of them. So, if γ is below the threshold defined by the model, then Patient 1 is scheduled first. However, because this threshold depends heavily on the specific values of waiting times, maximum waiting times, and DALYs, it is not possible to make general statements.

Scenario 2: Patient 2 Has Higher DALY

Here we consider the case where the DALYs are reversed, i.e., $DALY_1 < DALY_2$, so $\delta = DALY_1 - DALY_2 < 0$. The cost difference becomes:

$$\delta > \frac{\gamma}{c}\left(\frac{max(0, w_1 + c - MWT_2) - max(0, w_1 - MWT_2)}{MWT_2} - \frac{max(0, w_1 + c - MWT_1) - max(0, w_1 - MWT_1)}{MWT_1}\right)$$

If $MWT_1 > MWT_2$, and $DALY_1 < DALY_2$, then Patient 1 has both a lower DALY and a longer allowed waiting time. In that case, Patient 1 will always be scheduled first. Therefore, we assume in the following analysis that:

$$MWT_1 < MWT_2$$

Both patients scheduled before their MWTs for both orderings:
No penalties apply. Since $DALY_1 < DALY_2$, Patient 2 is always scheduled first.

Both patients scheduled after their MWTs:

$$\delta > \gamma\left(\frac{1}{MWT_2} - \frac{1}{MWT_1}\right) \Rightarrow \gamma > \frac{(DALY_2 - DALY_1)MWT_1 MWT_2}{MWT_2 - MWT_1}$$

This condition mirrors the case where $DALY_1 > DALY_2$, but with the inequality direction flipped due to dividing by $\delta < 0$. The threshold compares the importance of reducing DALY impact versus penalizing MWT violations. If γ exceeds this threshold, penalties dominate the decision and the patient with the shorter MWT (Patient 1) is prioritized, even if their DALY is lower.

The expression is highly sensitive to large MWTs, as the DALY difference is scaled by $MWT_1 \cdot MWT_2$ and divided by their difference. The full range of this threshold can be computed, it lies approximately in the interval $[1.3 \times 10^{-7},\ 3.0]$.

Patient 1 always after MWT, Patient 2 always before:

$$\delta > -\frac{\gamma}{MWT_1} \Rightarrow \gamma > MWT_1(DALY_2 - DALY_1)$$

In this case, only Patient 1 incurs a penalty, while Patient 2 is still within their maximum waiting time. The right-hand side scales the DALY difference by MWT_1, making the threshold sensitive to high values of MWT_1.

If γ exceeds this threshold, scheduling Patient 1 first becomes more favorable. The full range of this threshold can be computed, it lies approximately in the interval $[1.3 \times 10^{-7},\ 1.5]$.

Patient 1 late in second schedule only, Patient 2 late in first schedule only:

$$\delta > \frac{\gamma}{c}\left(\frac{w_1 + c - MWT_2}{MWT_2} - \frac{w_1 + c - MWT_1}{MWT_1}\right) \Rightarrow \gamma > \frac{c(DALY_1 - DALY_2)}{(w_1 + c)\left(\frac{1}{MWT_2} - \frac{1}{MWT_1}\right)}$$

Here, Patient 1 is late only in the second schedule, while Patient 2 is late only in the first schedule. If γ exceeds the threshold, the model prioritizes Patient 1, accepting a worse health outcome from delaying Patient 2 in order to avoid violating Patient 1's stricter MWT. The threshold for γ is sensitive to the values of $w_1, c,$ DALY and the MWT values, hence it is not possible to make general conclusions.

Patient 1 always late, Patient 2 late only in the first schedule:

$$\delta < \frac{\gamma}{c}\left(\frac{w_1 + c - MWT_2}{MWT_2} - \frac{c}{MWT_1}\right) \Rightarrow \gamma > \frac{c(DALY_1 - DALY_2)}{\left(\frac{w_1 + c - MWT_2}{MWT_2} - \frac{c}{MWT_1}\right)}$$

Patient 1 is always late, while Patient 2 is late only in the first schedule. If γ exceeds the threshold, the model prefers scheduling Patient 1 first, even though it has a lower DALY.

As before, the threshold's sensitivity depends strongly on w_1, c, DALY and MWT.

Patient 1 is scheduled after MWT in the second ordering, Patient 2 after MWT in first ordering:

$$\delta > \frac{\gamma}{c}\left(\frac{w_1 + c - MWT_2}{MWT_2} - \frac{w_1 + c - MWT_1}{MWT_1}\right) \Rightarrow \gamma > \frac{c(DALY_1 - DALY_2)}{\frac{w_1 + c - MWT_2}{MWT_2} - \frac{w_1 + c - MWT_1}{MWT_1}}$$

In this scenario, Patient 1 is late in the second schedule and Patient 1 is late in the first schedule. If γ exceeds the threshold, the model chooses to schedule Patient 1 first. As before, the threshold's sensitivity depends strongly on w_1, c, DALY and MWT.

Key takeaways:
In the two-patient case, for the DALY model, the scheduling priority depends on the balance between DALY differences and waiting penalties. For low γ (no penalty), the higher-DALY patient goes first; for high γ (heavy waiting penalty), the patient with the more urgent MWT tends to go first. The threshold γ where this switches depends on the parameters DALY gap δ, time gap c, and MWTs. The transition is most sensitive in the γ range [0,3].

6.2.1. Waiting Time Model

We also analyze the WT model, with the following objective function value for patient $p \in P$ given by:

$$\text{Obj}_p = \frac{w_p}{MWT_p} + \gamma \cdot \frac{s_p}{MWT_p}$$

Again, assume without loss of generality that $w_1 < w_2$, and define $w_2 = w_1 + c$, with $c > 0$.
Then:

$$C_2 - C_1 = \gamma \left(\frac{max(0, w_1 + c - MWT_1) - max(0, w_1 - MWT_1)}{MWT_1} - \frac{max(0, w_1 + c - MWT_2) - max(0, w_1 - MWT_2)}{MWT_2} \right) + c \left(\frac{1}{MWT_1} - \frac{1}{MWT_2} \right)$$

Assume $MWT_1 < MWT_2$. Then:

- It follows that $\frac{1}{MWT_1} > \frac{1}{MWT_2}$
- Define the function:
$$f(x) = \frac{max(0, w_1 + c - x) - max(0, w_1 - x)}{x}$$
This function is decreasing in x for $x > 0$, so:
$$f(MWT_1) > f(MWT_2)$$

Hence, $C_2 - C_1 > 0$, and it is optimal to schedule Patient 1 first.

Conversely, if $MWT_1 > MWT_2$, then the inequalities reverse:

$$\frac{1}{MWT_1} < \frac{1}{MWT_2}, \quad f(MWT_1) < f(MWT_2)$$

So $C_2 - C_1 < 0$, and it is optimal to schedule Patient 2 first.

In this two-patient comparison under fixed MWTs, the order of scheduling is determined entirely by the values of MWT. γ scales both patients' excess waiting time terms equally and thus does not affect the scheduling order. So we expect that γ does not influence the WT model's scheduling decisions overall, provided that it scales all relevant terms equally and the decision problem is comparative.

6.3. Normalization of Trade-off Parameter to Balance DALY/day and Waiting Time

Since DALY per day and normalized waiting times (i.e. $\frac{1}{MWT}$) are not on the same scale—DALY/day values lie approximately in the range $[0.000185; 0.004299]$, while $\frac{1}{MWT}$ values range from $[0.00138; 0.333]$—directly combining them in the objective function without adjustment would distort the balance between minimizing health loss and ensuring timely access to care.

To enable meaningful comparison and fair weighting, we normalize the trade-off by adjusting the waiting time weight parameter γ relative to a baseline parameter $\bar{\gamma}$, while keeping γ fixed. Specifically, we scale $\bar{\gamma}$ as follows:

$$\bar{\gamma} = \gamma \frac{mean(DALY/day)}{mean\left(\frac{1}{MWT}\right)}$$

This scaling aligns the magnitudes of the two components, ensuring that the effect of the γ parameters reflects a genuine trade-off between health outcomes and waiting time fairness, rather than an artifact of incompatible units.

This normalization also facilitates a more interpretable sensitivity analysis when comparing how models behave under varying prioritization weights.

6.4. Exploring the Trade-Off Parameter for Cardiology

In the previous section, we identified ranges for the trade-off parameter γ, noting significant variation in outcomes between values of 0 and 3. To explore this variation, we examine $\gamma \in \{0, 0.01, 0.1, 1, 3, 5\}$.

At $\gamma = 0$, the models operate solely based on their respective objectives without accounting for waiting time violations—this reflects pure optimization for either DALY minimization or waiting time fairness, depending on the model. This serves as a reference scenario. In the following sections, we analyze intermediate values of γ to explore how the balance between health outcomes and timely access can be optimized in practice.

While it would be possible to fine-tune γ separately for each medical specialty to better reflect clinical priorities, this thesis does not pursue that direction. Determining such values should be informed by medical expertise and practical considerations in the healthcare domain. Therefore, the current analysis provides a methodological foundation, while the actual calibration of γ is left for future work involving clinical stakeholders.

Comparison of Model Performance at $\gamma = 0$

This subsection presents a detailed comparison of the DALY, WT, and MinMax models under the condition $\gamma = 0$, where only their respective objectives without accounting for waiting time violations are minimized. In this regime, neither waiting time nor maximum waiting time (MWT) are explicitly constrained, so any performance related to these aspects emerges indirectly from each model's inherent prioritization logic.

We evaluate the models using three key metrics:

- **Waiting Time** — the duration between a patient's entry into the system and the time they receive treatment;
- **Excess Waiting Time** — the time patients wait beyond their designated MWT;
- **Disability-Adjusted Life Years (DALYs)** — a measure of cumulative health loss resulting from delayed treatment.

By examining these metrics, we assess not only the overall efficiency of each model but also the fairness and responsiveness of their scheduling behavior. Each subsection below highlights how the models allocate treatment and the resulting implications for patient outcomes.

Waiting Time (Subfigures a–f): The DALY model (subfigures 6.4a, 6.4d) concentrates early treatment on a subset of patients, resulting in very short waiting times for some patients and extremely long delays—nearing 700 days—for other patients. This suggests the model prioritizes patients whose health deteriorates fastest over time, leaving others untreated for extended periods. The combined plots confirm this skewed distribution, with a long tail of high waiting times.

In contrast, the WT model (6.4b, 6.4e) produces a flatter waiting time distribution. Without directly optimizing for DALY loss, the model effectively treats all patients with a MWT similarly, leading to moderate waiting times across the board—except for those prioritized strictly by their position on the waiting list. This approach causes patients who could afford to wait longer to do so, potentially leading to deterioration in their health while they wait.

The MinMax model (6.4c, 6.4f) results in the most consistent range of waiting times, with all patients treated within approximately 400 days. Waiting times are more tightly concentrated, reflecting the model's aim to reduce extreme delays and enforce a uniform allocation policy. However, this uniformity comes at the cost of responsiveness to urgency: patients who previously required only short waiting times now face significantly longer delays, undermining timely access for those who need it most.

Excess waiting time (Subfigures g–l): The DALY model (6.4g, 6.4j) shows the highest levels of excess waiting, with some patients exceeding their MWT by up to 400 days. Since the model schedules based solely on minimizing health loss, patients with lower marginal DALY impact are often deprioritized, leading to substantial delays, especially for those with longer MWTs.

The WT model (6.4h, 6.4k) results in almost no excess waiting across all patient categories. Because it follows a strict first-come, first-served logic based solely on the waiting list, most patients are treated within their MWT. However, this approach overlooks DALY, meaning that patients who would benefit most from timely intervention are not prioritized. As a result, even though their MWTs are technically respected, outcomes may still be suboptimal for those with high time-sensitivity in treatment

The MinMax model (6.4i, 6.4l) lies between the two extremes. While delays are more constrained than in the DALY model, excess waiting still reaches up to 200 days in some cases. The model limits the worst violations of MWT and distributes delays more evenly across patient groups, but does not eliminate excess waiting time entirely. This reflects its goal of minimizing maximum deviation rather than prioritizing urgency or strictly adhering to waiting list order.

DALY Spread (Subfigures m–r): In the DALY model (6.4m, 6.4p), health losses are highly uneven but the average DALY is the lowest among all models. The model concentrates resources on patients with the highest marginal health gains, fully protecting some from DALY accumulation. Others, however—typically those deemed less urgent in terms of health impact—are left untreated for long periods and experience significant health loss. This creates a wide DALY range across the population, reflecting a strategy that prioritizes minimizing total DALY over ensuring equity. The trade-off is clear: better outcomes on average, but with a cost in fairness and protection for lower-priority patients.

The WT model (6.4n, 6.4q) produces a flatter spread but exhibits a peak in DALYs for individuals scheduled according to the waiting list, resulting in significantly worse health outcomes for the less severely ill.

Finally, the MinMax model (6.4o, 6.4r) achieves a more uniform distribution of DALY across the patient population than either alternative. By actively preventing extreme delays, the model reduces the risk of severe health loss, especially for patients with longer MWTs. However, this evenness comes at a cost: patients with shorter MWTs experience slightly higher DALYs than under the WT or DALY models, since the model does not aggressively prioritize them. Overall, the MinMax approach favors consistency and fairness in health outcomes over minimizing DALY for only the most urgent cases.

Table 6.1: Sensitivity analysis of DALY and NEWT for scheduled and not scheduled patients across models and γ values for CAR.

γ	Model	**Total DALY**			**DALY**		**NEWT**			**Scheduled %**
		Scheduled	Not Scheduled	Combined	Mean	Max	Scheduled	Not Scheduled	Combined	
0	DALY	46.12	91.10	137.22	0.1323	1.0146	200.05	327.31	527.36	79.17%
	WT	54.59	88.30	142.89	0.1378	1.0146	17.23	0.00	17.23	79.27%
	MinMax	106.17	45.54	151.71	0.1463	0.5113	419.35	246.87	666.22	74.45%
0.01	DALY	48.12	88.65	136.76	0.1319	1.0146	54.67	10.87	65.53	79.17%
	WT	52.95	90.33	143.28	0.1382	1.0146	16.92	0.00	16.92	79.17%
	MinMax	84.46	60.21	144.68	0.1395	0.5113	164.68	44.95	209.63	77.72%
0.1	DALY	49.80	89.03	138.82	0.1339	1.0146	35.84	0.00	35.84	78.78%
	WT	53.81	89.66	143.47	0.1384	1.0146	16.83	0.00	16.83	79.17%
	MinMax	93.08	54.08	147.16	0.1419	0.5128	72.09	4.84	76.93	76.28%
1	DALY	48.76	89.61	138.37	0.1334	1.0146	21.85	0.00	21.85	79.07%
	WT	54.76	89.48	144.25	0.1391	1.0146	16.83	0.00	16.83	78.98%
	MinMax	93.36	55.57	148.94	0.1436	0.5144	30.40	1.69	32.09	75.60%
3	DALY	51.29	88.40	139.69	0.1347	1.0146	17.24	0.00	17.24	78.69%
	WT	55.67	88.57	144.24	0.1391	1.0146	16.83	0.00	16.83	78.98%
	MinMax	97.61	53.71	151.32	0.1459	0.5144	18.06	0.02	18.08	75.02%
5	DALY	51.45	88.09	139.54	0.1346	1.0146	16.85	0.00	16.85	78.78%
	WT	55.78	88.16	143.94	0.1388	1.0146	16.83	0.00	16.83	79.07%
	MinMax	100.69	52.90	153.59	0.1481	0.5144	17.48	0.00	17.48	74.16%

6.5. Selection of the best trade-off parameter value

To determine the optimal value of γ, a sensitivity analysis was conducted across different models (DALY, WT, MinMax) using varying values of γ. Table 6.1 summarizes the effects of changing γ on total, average and maximum DALY loss, normalised excess waiting time (NEWT), and the percentage of scheduled patients.

At low values of γ (e.g., $\gamma = 0$), the DALY model minimizes total DALY loss, but this comes at the cost of extremely high waiting times, indicating highly unequal access to care. In contrast, higher values of γ (e.g., $\gamma = 3$ or $\gamma = 5$) lead to a more balanced trade-off between DALY loss and waiting time. For instance, with $\gamma = 5$, DALY increases only slightly, while the average waiting time drops drastically to levels comparable with the WT model.

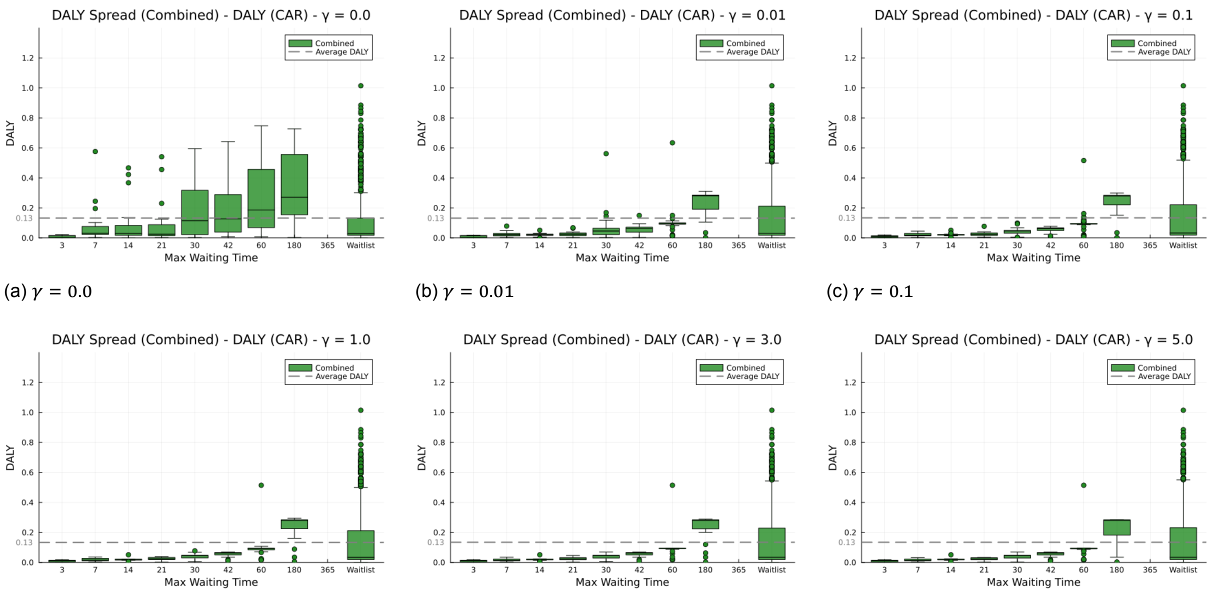
Interestingly, in the WT model, we expected the scheduling solution to remain unchanged as γ varies, since γ scales the excess waiting time terms uniformly for all patients and should not affect the relative ordering of schedules. However, contrary to this expectation, we observed that varying γ actually did change the scheduling decisions. This unexpected sensitivity suggests that the WT model may contain a structural issue or an unintended interaction in its objective function or constraints, causing γ to influence the scheduling order when theoretically it should not. This behavior indicates a potential flaw or limitation in the WT model formulation that requires further investigation to ensure its reliability and interpretability.

Figures 6.1 and 6.2 further illustrate how γ influences the distribution of Disability-Adjusted Life Years (DALYs) across different maximum waiting times (MWTs) in both the DALY and MinMax models. A small increase in γ, such as from 0.0 to 0.01, already results in a noticeable reduction in the spread of DALYs, particularly among patients with shorter MWTs. This trend becomes more pronounced at $\gamma = 3$, where patients with lower MWTs consistently experience significantly lower DALYs, highlighting the effect of prioritization.

While the DALY model shows a wide variation in outcomes as γ increases, the MinMax model—which is specifically designed to minimize the maximum DALY—exhibits a flatter and more uniform DALY distribution. While the MinMax model is designed to minimize the worst-case outcome, variations in γ influence the distribution of DALY values. As γ increases, there is a visible compression of DALY outcomes and a reduction in extreme cases, particularly among patients with shorter waiting times. This suggests that γ plays a meaningful role in controlling outcome dispersion, offering a way to balance efficiency with equity in scheduling decisions.

Figure 6.3 presents the excess waiting times for the different models at $\gamma = 3$. As expected, the WT model achieves the lowest excess waiting time and serves as a lower bound. However, both the DALY and MinMax models at this γ value perform comparably, with only slightly higher values.

Based on the sensitivity analysis, we select $\gamma = 3$ as a balanced trade-off between minimizing health burden and maintaining fairness in waiting times. The DALY model achieves the lowest health loss, but can result in disproportionately long delays for certain patients. The MinMax DALY model, particularly at $\gamma = 3$, offers a substantial reduction in mean excess waiting time (NEWT) while keeping the total DALY loss acceptably low. This makes it a suitable compromise in scenarios where both outcome efficiency and fairness are valued.



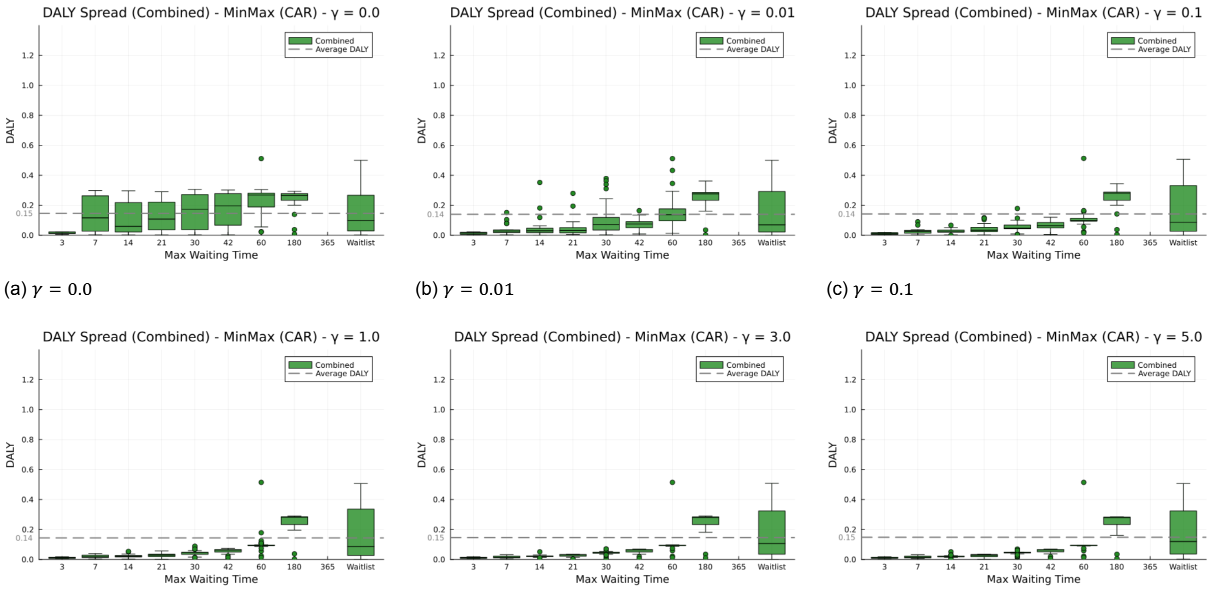
(a) $\gamma = 0.0$ (b) $\gamma = 0.01$ (c) $\gamma = 0.1$

(d) $\gamma = 1.0$ (e) $\gamma = 3.0$ (f) $\gamma = 5.0$

Figure 6.1: DALY model DALY spread for varying γ values.

6.6. Conclusion

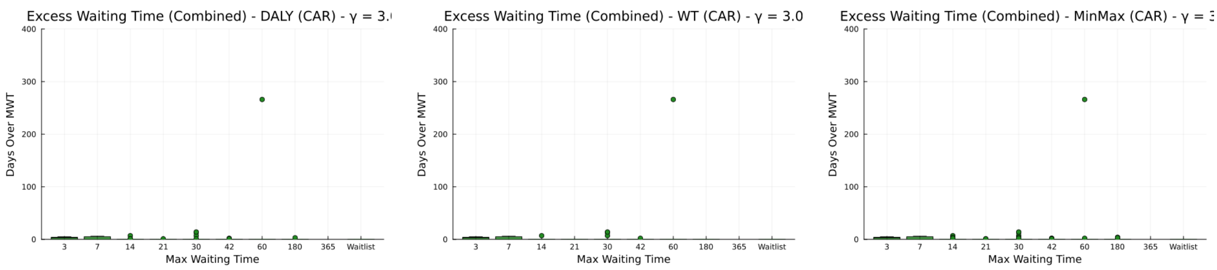
This sensitivity analysis highlights the fundamental trade-offs between minimizing health loss and promoting fairness in access to surgery. The DALY model is most effective in reducing total health burden, but without penalties, it allows some patients to wait excessively long. In contrast, the WT model prior-



(a) $\gamma = 0.0$ (b) $\gamma = 0.01$ (c) $\gamma = 0.1$

(d) $\gamma = 1.0$ (e) $\gamma = 3.0$ (f) $\gamma = 5.0$

Figure 6.2: MinMax model DALY spread for varying γ values.

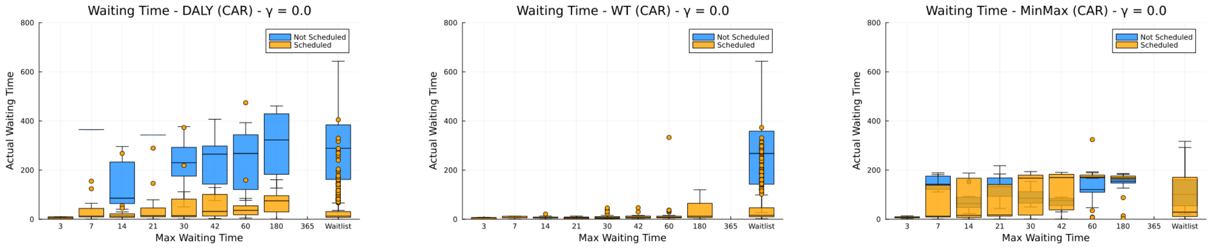


(a) CAR – DALY Model (b) CAR – WT Model (c) CAR – MinMax Model

Figure 6.3: Comparison of excess waiting times for CAR under different models with $\gamma = 3$

itizes timely access, resulting in lower waiting times but higher overall DALY loss. The MinMax DALY model offers a balanced approach, especially valuable in settings where fairness toward the most disadvantaged patients is a key consideration.

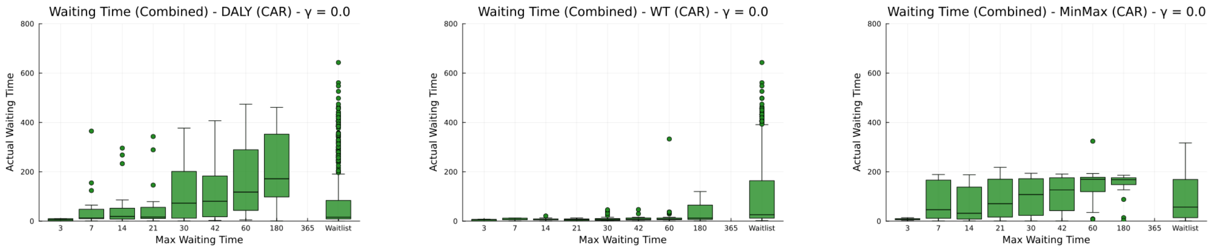
To balance health outcomes and fairness in waiting times, we select $\gamma = 3$ for further analysis. This value reflects a compromise: DALY loss remains acceptably low while the mean excess waiting time (NEWT) is significantly reduced, promoting a more equitable outcome without excessively sacrificing efficiency.



(a) DALY: Waiting Time Separated

(b) WT: Waiting Time Separated

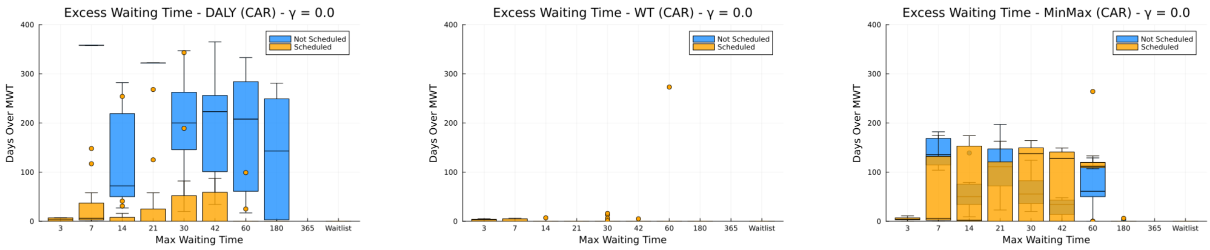
(c) MinMax: Waiting Time Separated



(d) DALY: Waiting Time Combined

(e) WT: Waiting Time Combined

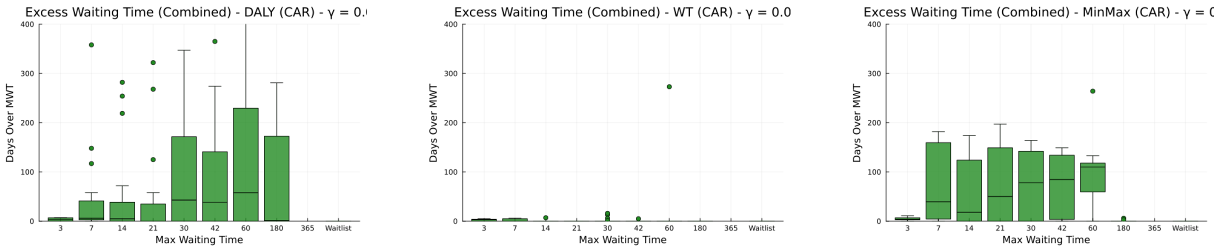
(f) MinMax: Waiting Time Combined



(g) DALY: Excess Waiting Separated

(h) WT: Excess Waiting Separated

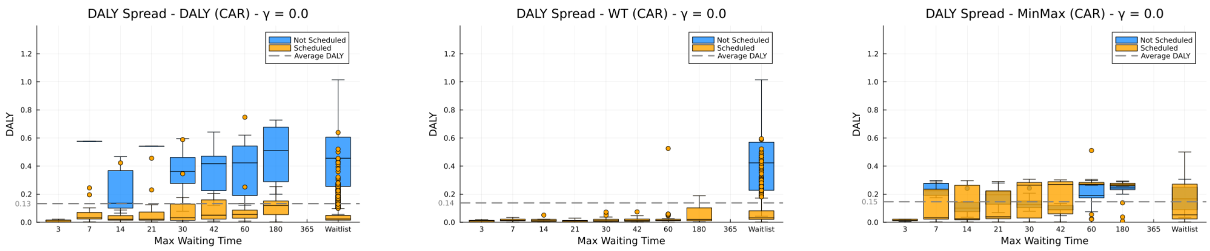
(i) MinMax: Excess Waiting Separated



(j) DALY: Excess Waiting Combined

(k) WT: Excess Waiting Combined

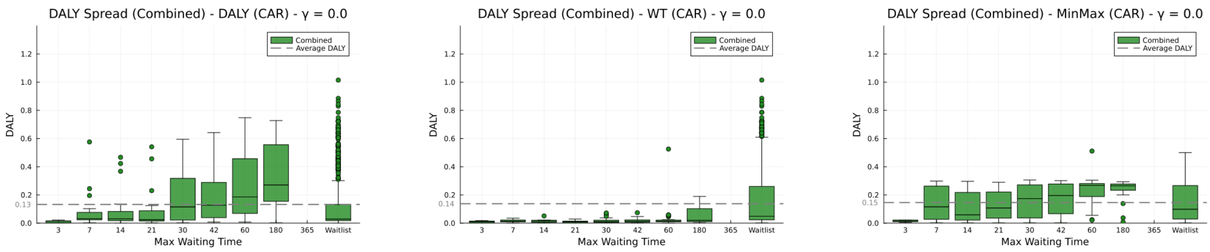
(l) MinMax: Excess Waiting Combined



(m) DALY: DALY Spread Separated

(n) WT: DALY Spread Separated

(o) MinMax: DALY Spread Separated



(p) DALY: DALY Spread Combined

(q) WT: DALY Spread Combined

(r) MinMax: DALY Spread Combined

Figure 6.4: Comparison of scheduled and not scheduled (top) and combined (bottom) results for the DALY, WT, and MinMax models at $\gamma = 0$, across different performance metrics.

7

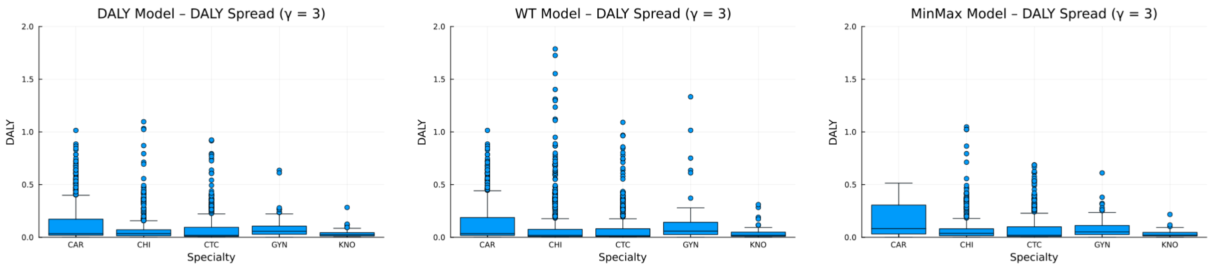
Results

This section presents the comparative performance of the DALY, WT, and MinMax models under a fixed 14-day scheduling horizon repeated throughout the year 2019, with a penalty parameter $\gamma = 3$. The models are evaluated across five medical specialties: Cardiology (CAR), General Surgery (CHI), Gynecology (GYN), Ear, Nose, and Throat (KNO), and Cardiothoracic Surgery (CTC). Figure 7.1 and Table 7.1 report key performance metrics, including total, mean, and maximum DALYs; Normalized Excess Waiting Time (NEWT); and the percentage of patients scheduled.

Each model is designed to prioritize a different scheduling objective:

- The DALY model aims to minimize the overall disease burden across patients.
- The WT model targets minimizing waiting times beyond each patient's clinically acceptable maximum.
- The MinMax model seeks to reduce the worst-case individual outcome by minimizing the highest DALY among all patients.

While these models are conceptually aligned with distinct goals, the results reveal that their real-world trade-offs are nuanced.



(a) DALY Model (b) WT Model (c) MinMax Model

Figure 7.1: DALY spread for $\gamma = 3$ for all models

7.1. Model comparisons

7.1.1. DALY Model

The DALY model has the lowest total DALY for CHI and CTC, as seen in Table 7.1. The mean DALY values are generally low for all specialties, indicating good average patient outcomes. However, it consistently performs poorly in terms of maximum DALY, particularly in specialties like CAR and CHI, where individual patients may face excessive delays.

This limitation is also visible in Figure 7.2, where the DALY model exhibits a wider spread in DALY values, with clear outliers. This supports the model's known trade-off: by optimizing population-level health outcomes, it may allow individual patients to experience disproportionately high delays—raising concerns about fairness.

Interestingly, the DALY model does not always achieve the lowest total DALY. For example, in the GYN specialty, the MinMax model slightly outperforms it in combined total DALY. This may appear counterintuitive, since the DALY model is designed to minimize exactly that metric. However, GYN includes a high proportion of urgent cases with short maximum waiting times (MWTs), as shown in Figure 5.2. In such cases, small scheduling delays can lead to steep increases in DALY. The MinMax model avoids these high-penalty cases by minimizing the worst individual DALY, which may indirectly reduce total DALY in high-urgency contexts.

7.1.2. WT Model

The WT model consistently achieves the lowest NEWT values across all specialties, confirming its effectiveness in minimizing the number of patients exceeding their MWT (see Table 7.1 and Figure 7.3). However, this emphasis on strict timeliness comes at a cost. The model frequently produces the highest total and maximum DALYs, as it prioritizes reducing delays over minimizing overall health loss. In doing so, it may schedule patients with less severe health needs earlier—potentially delaying those with more critical conditions and thereby worsening population-level health outcomes.

7.1.3. MinMax Model

The MinMax model strikes a balanced compromise across objectives. It consistently achieves the lowest maximum DALYs, demonstrating its strength in avoiding extreme outcomes for individual patients. Although its total and mean DALY values are not always the lowest, they remain competitive—often only slightly higher than those in the DALY model. Its NEWT values, while not as low as those in the WT model, are still moderate. This suggests that the MinMax approach offers a reasonable trade-off between fairness, timeliness, and health efficiency.

Table 7.1

Specialty	Model	**Total DALY**			**DALY**		**NEWT**			**Scheduled %**
		Scheduled	Not Scheduled	Combined	Mean	Max	Scheduled	Not Scheduled	Combined	
CAR	DALY	51.29	88.40	139.69	0.1347	1.0146	17.24	0.00	17.24	78.69%
	WT	55.67	88.57	144.24	0.1391	1.0146	16.83	0.00	16.83	78.98%
	MinMax	97.61	53.71	151.32	0.1459	0.5144	18.06	0.02	18.08	75.02%
CHI	DALY	32.09	28.56	60.65	0.0726	1.0970	125.59	37.80	163.39	77.13%
	WT	24.00	62.37	86.36	0.1034	1.7873	34.02	3.09	37.11	75.45%
	MinMax	33.32	29.95	63.28	0.0758	1.0477	141.04	32.79	173.83	75.81%
GYN	DALY	7.37	9.64	17.01	0.0818	0.6357	63.28	39.11	102.40	64.42%
	WT	9.14	12.00	21.14	0.1016	1.3334	60.19	29.70	89.89	63.94%
	MinMax	8.02	8.54	16.55	0.0796	0.6110	57.96	54.80	112.76	65.87%
KNO	DALY	5.13	1.13	6.26	0.0324	0.2836	56.41	22.95	79.36	87.05%
	WT	4.69	1.58	6.27	0.0325	0.3105	46.09	12.37	58.46	87.56%
	MinMax	4.97	0.92	5.89	0.0305	0.2170	46.72	17.12	63.85	88.60%
CTC	DALY	18.53	17.49	36.02	0.0898	0.9237	198.92	0.29	199.21	80.05%
	WT	13.26	24.54	37.80	0.0943	1.0911	145.13	0.29	145.41	79.80%
	MinMax	17.61	17.99	35.60	0.0888	0.6867	214.66	0.29	214.95	79.55%

7.2. Ethical and Clinical Considerations

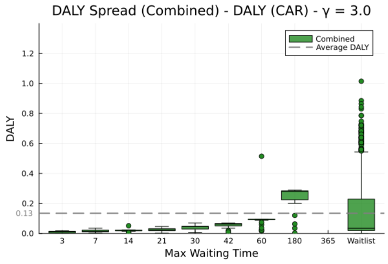
The choice of scheduling model reflects not only a technical preference but also an ethical position with real consequences for patient care. Each model implicitly supports different trade-offs—between efficiency, timeliness, and fairness—that should not be made by algorithm designers alone.

Crucially, these trade-offs must be informed and validated by clinical stakeholders. Physicians, surgeons, and healthcare administrators are best positioned to judge what constitutes an acceptable delay, which outcomes are clinically tolerable, and how much variation in patient experience is justifiable. For example:

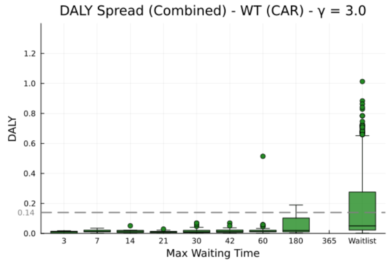
- A high maximum DALY might be clinically unacceptable in specialties with rapid disease progression.
- Timeliness, as prioritized by the WT model, may be more critical in emergency-driven disciplines.
- Equity in outcomes, as emphasized in the MinMax model, might align more closely with ethical standards in public healthcare systems.

From the empirical results, we observe that with $\gamma = 3$, the outcome distributions across the DALY and MinMax model are quite similar. However, the MinMax model results in fewer outliers, suggesting greater consistency across patients. That said, it still exhibits some excess waiting time, highlighting the complexity of balancing fairness with timeliness.

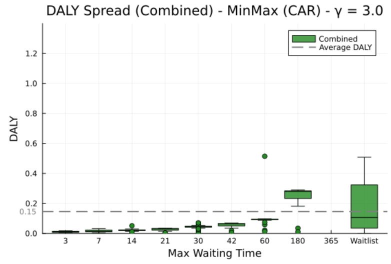
Therefore, model selection and interpretation of results should be done in close collaboration with clinicians to ensure that ethical priorities are grounded in medical realities and patient needs. This intersection of algorithmic design and clinical judgment is essential for responsible and effective implementation.



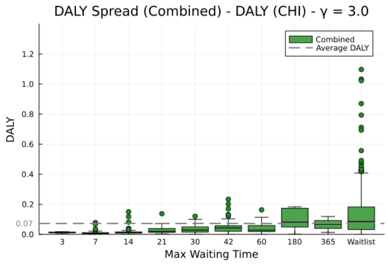
(a) CAR – DALY Model



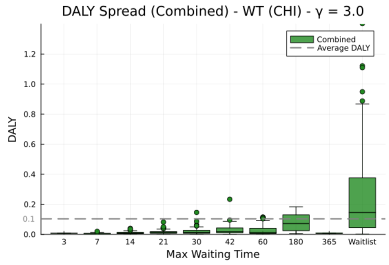
(b) CAR – WT Model



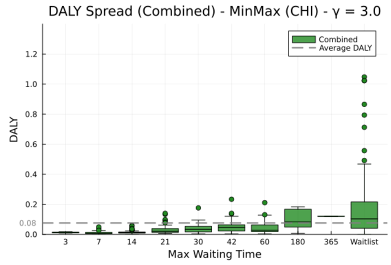
(c) CAR – MinMax Model



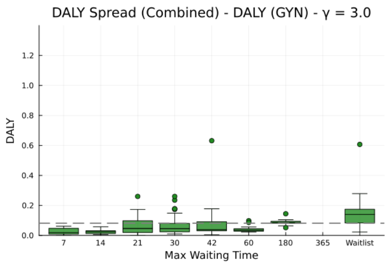
(d) CHI – DALY Model



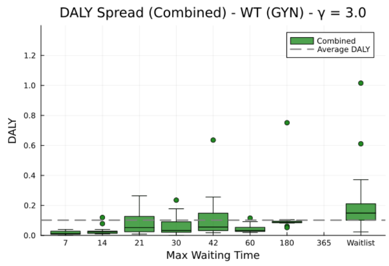
(e) CHI – WT Model



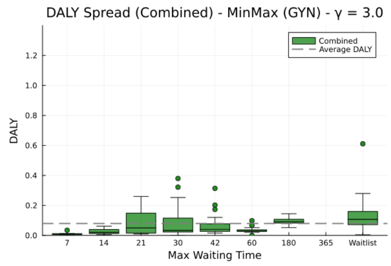
(f) CHI – MinMax Model



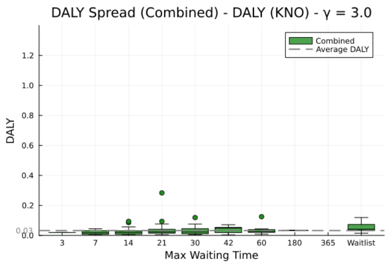
(g) GYN – DALY Model



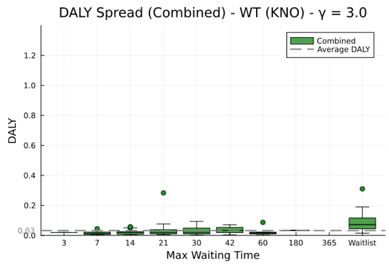
(h) GYN – WT Model



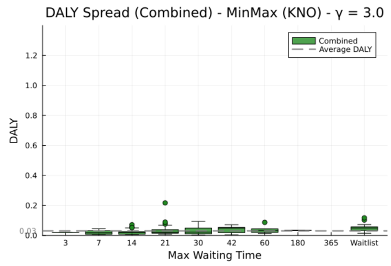
(i) GYN – MinMax Model



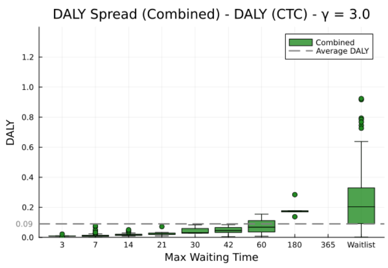
(j) KNO – DALY Model



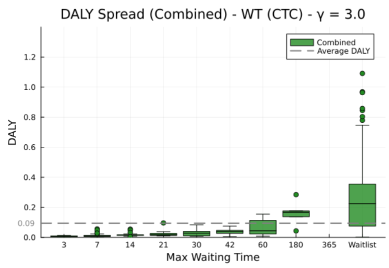
(k) KNO – WT Model



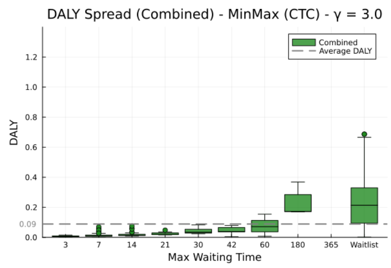
(l) KNO – MinMax Model



(m) CTC – DALY Model

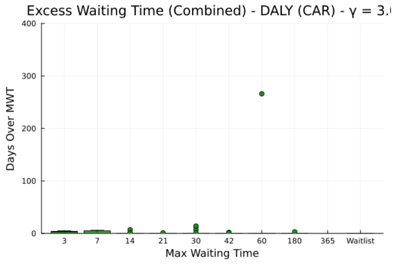


(n) CTC – WT Model

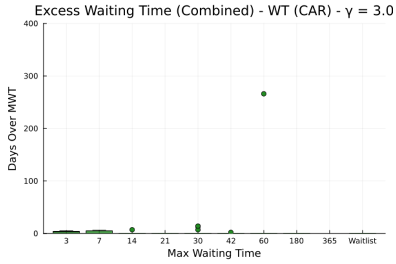


(o) CTC – MinMax Model

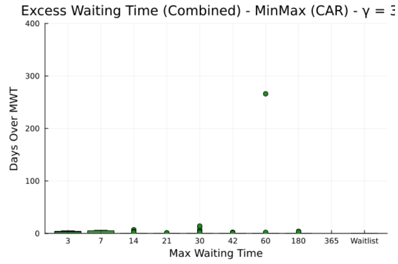
Figure 7.2: DALY spread comparison for $\gamma = 3.0$ across different models and specialties.



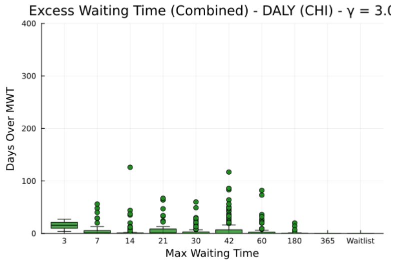
(a) CAR – DALY Model



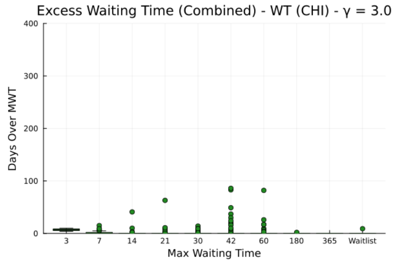
(b) CAR – WT Model



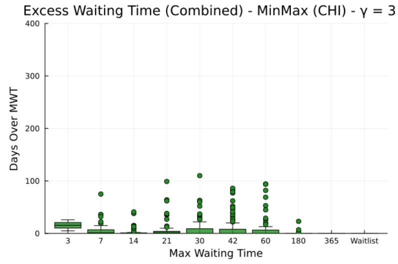
(c) CAR – MinMax Model



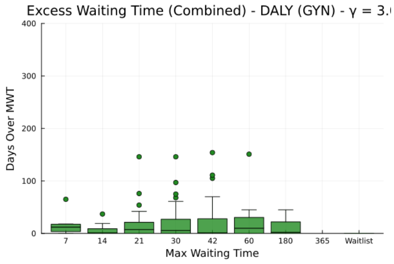
(d) CHI – DALY Model



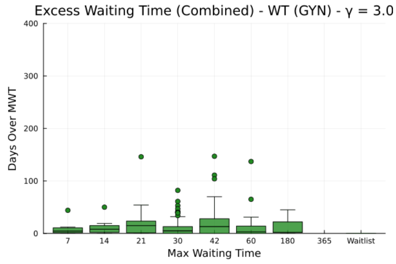
(e) CHI – WT Model



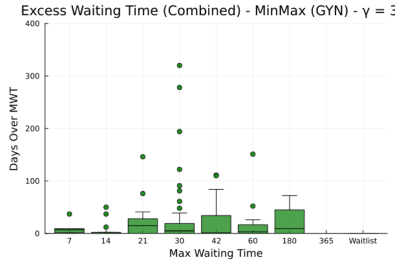
(f) CHI – MinMax Model



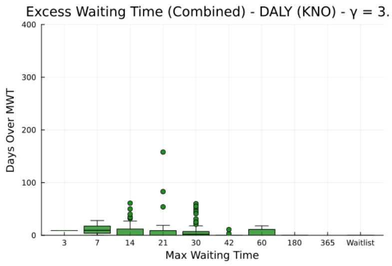
(g) GYN – DALY Model



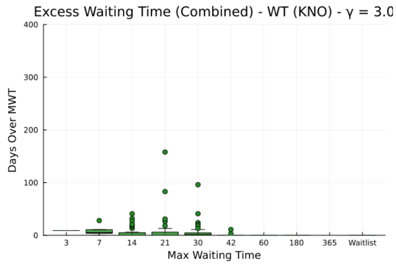
(h) GYN – WT Model



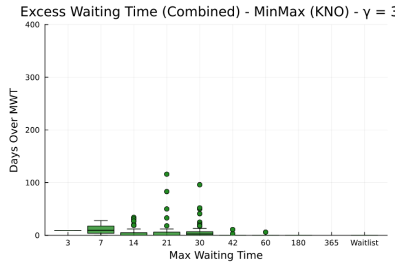
(i) GYN – MinMax Model



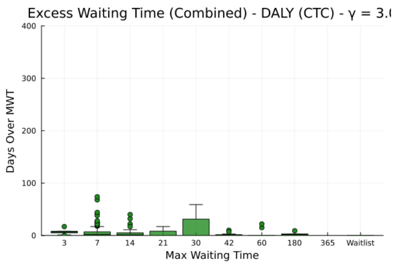
(j) KNO – DALY Model



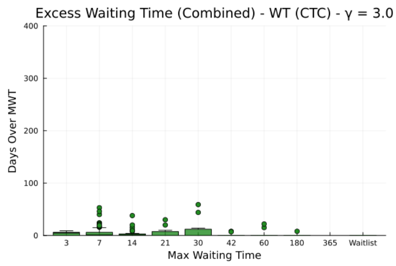
(k) KNO – WT Model



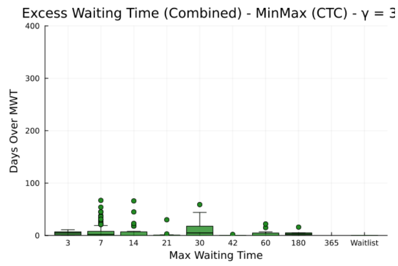
(l) KNO – MinMax Model



(m) CTC – DALY Model



(n) CTC – WT Model



(o) CTC – MinMax Model

Figure 7.3: Excess waiting time comparison for $\gamma = 3.0$ across different models and specialties.

8 Conclusion

This thesis introduced and evaluated three optimization models for surgical scheduling: the DALY-based model, the Waiting Time (WT) model, and the MinMax model. Each represents a distinct prioritization philosophy, aiming to balance competing objectives such as medical urgency, fairness, and operational efficiency in a resource-constrained hospital environment.

All models were formulated as Integer Linear Programs and implemented using real-world surgical data from Erasmus MC. The DALY model prioritizes overall health impact by minimizing total Disability-Adjusted Life Years (DALYs) lost due to surgical delays. The WT model integrates clinician-defined Maximum Waiting Times (MWTs), penalizing excess delays in accordance with subjective urgency assessments. The MinMax model, in contrast, prioritizes individual health outcomes by minimizing the highest single DALY loss across all patients—ensuring that no individual experiences a disproportionately severe health impact due to surgical delays.

A key component in the models is the parameter γ, which governs the trade-off between excess waiting time and health loss. Sensitivity analysis revealed that model outcomes are most responsive to changes of γ in the range $[0, 3]$. This informed our choice of $\gamma = 3$ as a balanced compromise between objective and subjective prioritization. Importantly, this parameter can be tailored to different specialties, reflecting variations in clinical urgency and ethical norms.

The comparative results suggest that each model serves a different purpose. The DALY model is efficient in reducing overall health burden, the WT model respects clinical guidelines on urgency, and the MinMax model offers the most equitable distribution of health outcomes. While context-dependent, the MinMax model appears best suited for general implementation due to its inherent focus on fairness and its robustness across specialties.

In summary, this thesis highlights the potential of integrating objective health metrics with clinical judgment to support surgical scheduling decisions. By offering three models with complementary strengths, it provides a flexible framework adaptable to diverse clinical contexts. These tools can help healthcare systems navigate the complex trade-offs between efficiency, urgency, and fairness—ultimately contributing to more equitable and effective care delivery.

8.1. Discussion

The most important point to acknowledge is that there is a mistake somewhere in the model or implementation that has not yet been identified. This could stem from a programming issue, a modeling error, or a flaw in how patients are written off and tracked through the system. As a result, the reported numbers may vary across different tables, and some results should be interpreted with caution. Identifying and resolving this issue is a priority for future work, as it could impact the validity of the findings.

Beyond this, several assumptions in the current study also suggest directions for further research. The model used a fixed two-week scheduling horizon, meaning patients added during one period were only considered for scheduling in the next. While this provides clarity, it limits the model's ability to respond in real time to urgent cases or newly arriving patients. A rolling horizon approach—updating the schedule continuously as new information arrives—could offer greater flexibility and potentially improve both patient outcomes and resource utilization.

Another simplification was the use of a uniform 80% utilization rate across all specialties. While this standardization made comparisons easier, it does not reflect the real variation in utilization between different surgical departments. Future iterations should incorporate specialty-specific targets derived from empirical data to better capture actual operating conditions.

A further modeling choice that warrants refinement is the tuning of the γ parameter. In the current version, γ was tuned jointly across all specialties. However, each specialty presents different clinical and ethical constraints. For instance, in certain specialties with highly time-sensitive conditions—such as emergency trauma—patients must be scheduled with minimal delay. Therefore, tuning γ individually for each specialty would allow the model to better reflect the urgency and risk profiles of different types of procedures.

Operational uncertainty also remains a challenge. Surgery durations in practice often deviate from estimates, leading to over- or under-utilization of time blocks. Incorporating stochastic elements or mechanisms such as buffers and overtime could improve the model's realism and robustness, particularly in high-variability environments.

Ethical considerations are deeply embedded in this work. Balancing minimum waiting times (MWTs), which reflect clinical thresholds, with metrics like DALYs involves normative trade-offs that cannot be resolved through quantitative modeling alone. Developing prioritization frameworks that are ethically sound requires collaboration with clinicians, ethicists, and policymakers to ensure decisions align with both medical evidence and societal values.

Finally, not all conditions require surgical intervention—some can be managed through non-surgical treatments. Future research could explore how incorporating alternative treatment pathways might influence scheduling decisions and improve overall resource allocation.

In summary, this thesis contributes to the ongoing conversation about surgical resource allocation by proposing a data-driven, adaptable, and ethically aware framework. While limitations remain, including potential model errors and simplifying assumptions, the work lays a foundation for more responsive and equitable surgical scheduling systems.

Appendix

Table 1: Sensitivity analysis of DALY and NEWT for scheduled and not scheduled patients across models and γ values for CHI.

γ	Model	Total DALY			DALY		NEWT			Scheduled %
		Scheduled	Not Scheduled	Combined	Mean	Max	Scheduled	Not Scheduled	Combined	
0.01	DALY	27.53	19.35	46.88	0.0561	1.0477	271.93	149.85	421.78	83.11%
	WT	24.68	61.19	85.87	0.1028	1.8243	47.83	5.35	53.18	76.65%
	MinMax	31.95	14.96	46.91	0.0562	1.0403	332.53	246.56	579.09	82.99%
0.1	DALY	30.88	23.04	53.92	0.0646	1.0502	229.86	62.65	292.51	79.52%
	WT	24.30	61.76	86.06	0.1031	1.8859	42.48	5.65	48.13	75.93%
	MinMax	33.02	23.16	56.18	0.0673	1.0403	245.95	88.09	334.04	77.96%
1	DALY	31.57	27.68	59.25	0.0710	1.0847	154.34	53.86	208.21	77.84%
	WT	25.13	61.24	86.36	0.1034	1.8021	38.77	1.89	40.66	75.81%
	MinMax	32.29	27.59	59.88	0.0717	1.0477	142.24	46.16	188.41	76.53%
3	DALY	32.09	28.56	60.65	0.0726	1.0970	125.59	37.80	163.39	77.13%
	WT	24.00	62.37	86.36	0.1034	1.7873	34.02	3.09	37.11	75.45%
	MinMax	33.32	29.95	63.28	0.0758	1.0477	141.04	32.79	173.83	75.81%
5	DALY	31.66	29.49	61.15	0.0732	1.0748	109.57	30.75	140.32	77.13%
	WT	25.90	60.27	86.17	0.1032	1.7996	41.75	2.06	43.81	75.81%
	MinMax	34.46	29.93	64.40	0.0771	1.0477	124.44	54.43	178.87	75.69%

Table 2: Sensitivity analysis of DALY and NEWT for scheduled and not scheduled patients across models and γ values for GYN.

γ	Model	Total DALY			DALY		NEWT			Scheduled %
		Scheduled	Not Scheduled	Combined	Mean	Max	Scheduled	Not Scheduled	Combined	
0	DALY	3.82	6.75	10.57	0.0508	0.2573	21.79	171.31	193.10	75.48%
	WT	4.26	9.87	14.13	0.0679	1.3334	16.95	63.04	79.99	70.19%
	MinMax	7.60	5.70	13.30	0.0639	0.1486	103.35	160.44	263.80	66.83%
0.01	DALY	4.74	6.34	11.08	0.0533	0.2796	66.66	91.90	158.56	73.56%
	WT	6.00	9.84	15.84	0.0762	1.3334	34.33	48.47	82.80	66.35%
	MinMax	5.66	5.93	11.59	0.0557	0.2105	85.57	100.11	185.68	72.12%
0.1	DALY	5.77	7.20	12.97	0.0623	0.2796	58.27	66.60	124.87	68.27%
	WT	9.04	11.10	20.14	0.0968	1.3334	59.58	36.46	96.04	64.42%
	MinMax	5.99	7.29	13.28	0.0638	0.2796	76.63	61.11	137.74	67.79%
1	DALY	7.52	9.20	16.72	0.0804	0.6110	64.38	60.99	125.38	63.46%
	WT	7.64	10.71	18.35	0.0882	1.3334	45.04	32.01	77.05	64.90%
	MinMax	8.02	8.45	16.46	0.0792	0.6110	66.27	58.14	124.41	66.35%
3	DALY	7.37	9.64	17.01	0.0818	0.6357	63.28	39.11	102.40	64.42%
	WT	9.14	12.00	21.14	0.1016	1.3334	60.19	29.70	89.89	63.94%
	MinMax	8.02	8.54	16.55	0.0796	0.6110	57.96	54.80	112.76	65.87%
5	DALY	8.04	9.36	17.39	0.0836	0.6357	59.89	39.01	98.90	64.90%
	WT	9.14	12.00	21.14	0.1016	1.3334	60.19	29.70	89.89	63.94%
	MinMax	8.02	8.54	16.55	0.0796	0.6110	57.96	54.80	112.76	65.87%

Table 3: Sensitivity analysis of DALY and NEWT for scheduled and not scheduled patients across models and γ values for KNO.

γ	Model	Total DALY			DALY		NEWT			Scheduled %
		Scheduled	Not Scheduled	Combined	Mean	Max	Scheduled	Not Scheduled	Combined	
0	DALY	2.60	0.89	3.49	0.0181	0.1189	23.96	75.38	99.34	91.19%
	WT	2.86	1.29	4.15	0.0215	0.2489	28.21	30.29	58.50	91.71%
	MinMax	3.86	0.68	4.55	0.0236	0.1028	86.12	32.97	119.09	87.56%
0.01	DALY	2.72	0.75	3.47	0.0180	0.3105	23.56	52.13	75.69	93.26%
	WT	3.13	1.28	4.41	0.0228	0.1947	35.00	18.36	53.36	90.67%
	MinMax	3.10	0.61	3.71	0.0192	0.1028	44.37	38.53	82.90	90.67%
0.1	DALY	3.20	1.19	4.40	0.0228	0.3105	32.88	37.70	70.57	91.19%
	WT	4.72	1.95	6.68	0.0346	0.3105	50.06	8.99	59.05	86.53%
	MinMax	4.06	0.79	4.85	0.0251	0.1195	52.96	22.52	75.48	89.12%
1	DALY	4.87	1.27	6.14	0.0318	0.2836	52.68	26.20	78.89	85.49%
	WT	4.74	1.44	6.18	0.0320	0.2836	42.97	15.06	58.03	88.08%
	MinMax	4.83	1.18	6.01	0.0311	0.1949	65.52	15.15	80.67	87.56%
3	DALY	5.13	1.13	6.26	0.0324	0.2836	56.41	22.95	79.36	87.05%
	WT	4.69	1.58	6.27	0.0325	0.3105	46.09	12.37	58.46	87.56%
	MinMax	4.97	0.92	5.89	0.0305	0.2170	46.72	17.12	63.85	88.60%
5	DALY	4.82	1.08	5.90	0.0306	0.1727	48.06	18.30	66.36	88.08%
	WT	4.76	1.60	6.35	0.0329	0.3105	47.02	13.83	60.85	87.56%
	MinMax	4.82	1.08	5.90	0.0306	0.1727	48.34	18.30	66.65	88.08%

Table 4: Sensitivity analysis of DALY and NEWT for scheduled and not scheduled patients across models and γ values for CTC.

γ	Model	Total DALY			DALY		NEWT			Scheduled %
		Scheduled	Not Scheduled	Combined	Mean	Max	Scheduled	Not Scheduled	Combined	
0	DALY	16.73	13.08	29.81	0.0744	0.5324	945.57	1034.65	1980.22	79.55%
	WT	13.05	24.56	37.61	0.0938	1.0911	108.74	0.29	109.02	79.80%
	MinMax	24.14	6.79	30.93	0.0771	0.5023	1353.15	590.55	1943.69	80.30%
0.01	DALY	14.47	15.10	29.57	0.0737	0.5192	895.90	131.27	1027.16	80.55%
	WT	12.68	24.79	37.47	0.0934	1.1269	106.58	0.29	106.86	79.80%
	MinMax	14.69	14.35	29.04	0.0724	0.5023	851.48	208.52	1060.01	81.05%
0.1	DALY	17.04	16.13	33.16	0.0827	0.7581	378.83	48.71	427.55	80.55%
	WT	12.42	25.37	37.78	0.0942	1.1871	154.91	0.29	155.20	79.80%
	MinMax	17.43	15.96	33.39	0.0833	0.5098	462.44	47.29	509.72	80.30%
1	DALY	17.18	18.47	35.65	0.0889	1.1664	204.22	0.29	204.50	80.30%
	WT	13.07	25.01	38.08	0.0950	1.0911	152.02	0.29	152.31	79.30%
	MinMax	17.81	17.29	35.10	0.0875	0.5098	307.13	0.29	307.42	79.80%
3	DALY	18.53	17.49	36.02	0.0898	0.9237	198.92	0.29	199.21	80.05%
	WT	13.26	24.54	37.80	0.0943	1.0911	145.13	0.29	145.41	79.80%
	MinMax	17.61	17.99	35.60	0.0888	0.6867	214.66	0.29	214.95	79.55%
5	DALY	17.56	18.59	36.15	0.0902	1.0742	207.07	0.29	207.36	79.80%
	WT	12.56	25.31	37.87	0.0944	1.0911	154.52	0.29	154.81	79.55%
	MinMax	17.65	17.90	35.56	0.0887	0.6867	215.14	0.29	215.43	79.80%

Bibliography

Addis, B., Carello, G., Grosso, A., & Tànfani, E. (2016). Operating room scheduling and rescheduling: A rolling horizon approach. *Flexible Services and Manufacturing Journal*, *28*(2–3), 206–232. https://doi.org/10.1007/s10696-015-9213-7

Cardoen, B., Demeulemeester, E., & Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, *201*(3), 921–932. https://doi.org/10.1016/j.ejor.2009.04.011

Hilderink, H. B. M., Plasmans, M. H. D., Poos, M. J. J. C., Eysink, P. E. D., & Gijsen, R. (2020). Dutch dalys, current and future burden of disease in the netherlands. *Archives of Public Health*, *78*, 85. https://doi.org/10.1186/s13690-020-00461-8

Kamran, M. A., Karimi, B., Dellaert, N., & Demeulemeester, E. (2019). Adaptive operating rooms planning and scheduling: A rolling horizon approach. *Operations Research for Health Care*, *22*, 100200. https://doi.org/10.1016/j.orhc.2019.100200

Murray, C. J., & Lopez, A. D. (1996). *The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020* [Published on behalf of the World Health Organization and the World Bank]. Harvard School of Public Health.

van Alphen, A. M. I. A., Krijkamp, E. M., Gravesteijn, B. Y., Baatenburg de Jong, R. J., & Busschbach, J. J. (2024). Surgical prioritization based on decision model outcomes is not sensitive to differences between the health-related quality of life values estimates of physicians and citizens. *Quality of Life Research*, *33*(2), 529–539. https://doi.org/10.1007/s11136-023-03544-5

van Oostrum, J. M. (2009). *Applying mathematical models to surgical patient planning* [Doctoral dissertation, Erasmus University Rotterdam]. https://repub.eur.nl/pub/35213

Wang, L., Demeulemeester, E., Vansteenkiste, N., & Rademakers, F. (2021). Operating room planning and scheduling for outpatients and inpatients: A review and future research. *Operations Research for Health Care*, *31*, 100323. https://doi.org/10.1016/j.orhc.2021.100323