PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

ACHIEVING A PROACTIVE POLICY FOR PATIENT FLOW MANAGEMENT IN A COMPLEX HOSPITAL NETWORK THROUGH REINFORCEMENT LEARNING

MATÍAS DE GEYTER MESSINA

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
JORGE RAFAEL VERA ANDREO
HOMERO LARRAÍN IZQUIERDO

Santiago de Chile, March 2024

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

ACHIEVING A PROACTIVE POLICY FOR PATIENT FLOW MANAGEMENT IN A COMPLEX HOSPITAL NETWORK THROUGH REINFORCEMENT LEARNING

MATÍAS DE GEYTER MESSINA

Members of the Committee:

JORGE RAFAEL VERA ANDREO

DocuSigned by:
JORGE RAFAEL VERA ANDREO
9BCDF27784E44D7...

HOMERO LARRAÍN IZQUIERDO

DocuSigned by:
HOMERO LARRAÍN IZQUIERDO
7E083725BE65439...

ALEJANDRO FRANCISCO MAC CAWLEY VERGARA

DocuSigned by:
3A98A756C1574D1...

JUAN CARLOS PATTILLO SILVA

DocuSigned by:
Juan Carlos Pattillo Silva
A7D2C9FBC6664B3...

RODRIGO FERNANDO CÁDIZ CÁDIZ

DocuSigned by:
RODRIGO FERNANDO CADIZ CADI
E0AFA235D0AD428...

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Santiago de Chile, March 2024

Grateful to my parents and
siblings, who accompanied me
throughout this process

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I want to express my deepest gratitude to all the individuals who contributed in any way to the completion of this thesis work.

First and foremost, I am grateful for the invaluable support of my family and loved ones. To my parents, Jorge and Paola, for their unconditional love, constant encouragement, and understanding during moments of intense dedication. To my siblings, Elisa, Agustín, Vicente, Clemente, and Joaquín, for their continuous encouragement and motivation. Also, I want to express my deep gratitude to Lucas for his generous contribution to improving this work.

I sincerely appreciate my thesis supervisors, Jorge and Homero, for their expert guidance, constant support, and valuable suggestions that guided this project from its inception to completion. Their dedication and commitment were essential in achieving the set objectives, as well as their support during the most challenging moments. I want to express my gratitude for their encouragement to present this work at the INFORMS Annual Meeting in 2023. This opportunity provided me with a deeper understanding of the academic world and the development of high-impact scientific research.

This work received financial support from the Agencia Nacional de Investigación y Desarrollo (ANID) through the National Master's Scholarship No. 22231602.

To all these individuals and institutions, my most sincere gratitude for their contribution to this academic achievement.

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Decision-making in healthcare is a complex problem due to uncertainty and dynamics. One very important issue is related to decisions on patient admissions. In a healthcare network composed of various hospitals, it should be convenient to consider the combined capacity of the whole system. Hence, besides admissions to a hospital, the transfer of patients to others is also possible. These decisions depend not only on the patient's medical condition but also on the current capacity of the various areas of the hospitals. In this work, we have formulated the problem as a Markov Decision Process, and we address it using a Reinforcement Learning approach where the expected future cost is estimated using an XGBoost model, combined with simulation, based on the Q-learning methodology. Based on the obtained policy via reinforcement learning and notions from the literature, general hospital managerial guidelines are proposed, which end up in a simplified and easy-to-implement policy with promising results, including a reduction in the average service time in a 13.38% and a Waiting List size reduction of a 65.86%. Compared to the base scenario, there is a reduction in the total costs of 71.44%, where the social costs (due to waiting times and patient bed opportunity costs) are reduced by 87.30%.

Keywords: **Healthcare Management - Approximate Dynamic Programming - Simulation - Markov Decision Process - Reinforcement Learning**.

RESUMEN

La toma de decisiones en el cuidado de la salud es un problema complejo debido a la incertidumbre y la realidad dinámica. Un tema muy importante está relacionado con las decisiones sobre admisiones de pacientes. En una red de atención médica compuesta por varios hospitales, debería ser conveniente considerar la capacidad combinada de todo el sistema. Por lo tanto, además de las admisiones a un hospital, también es posible transferir pacientes a otros. Estas decisiones dependen no solo de la condición médica del paciente, sino también de la capacidad actual de las diferentes áreas de los hospitales. En este trabajo, hemos formulado el problema como un Proceso de Decisión de Markov, y lo abordamos utilizando un enfoque de Aprendizaje por Refuerzo donde se estima el costo futuro esperado utilizando un modelo XGBoost, combinado con simulación, basado en la metodología de Q-learning. Basándonos en la política obtenida a través del aprendizaje por refuerzo y en nociones de la literatura, se proponen pautas generales de gestión hospitalaria, que resultan en una política simplificada y fácil de implementar con resultados prometedores, incluida una reducción en el tiempo de servicio promedio en un 13.38% y una reducción del tamaño de la lista de espera del 65.86%. En comparación con el escenario base, hay una reducción en los costos totales del 71.44%, donde los costos sociales (debidos a los tiempos de espera y los costos de oportunidad de las camas para pacientes) se reducen en un 87.30%.

Palabras Claves: **Gestión de la salud - Programación Dinámica Aproximada - Simulación - Proceso de Decisión de Markov - Aprendizaje por Refuerzo**.

1. INTRODUCTION

Healthcare plays an essential role in human development and quality of life. As stated, for example, by X. Zhang et al. (2017) and Wendt (2009), access to healthcare is a critical factor that significantly influences health outcomes, thus emphasizing the importance of providing full coverage to the population. This proposition is not new; the Millennium Development Goals (MDGs) in 2000 and, later, the Sustainable Development Goals (SDGs) in 2015, proposed by the United Nations (UN), also established the importance of healthcare for human well-being (United Nations, 2020).

The United Nations responded to this situation by proposing a target for 2030: achieving universal health coverage. This goal entails ensuring that all individuals have access to high-quality essential healthcare services.

Unfortunately, access to medical services is not always possible for everybody, especially in middle and low-income countries, resulting in hazardous living conditions for those in need. This situation led the UN to propose as a target for 2030 to achieve universal health coverage, where all individuals must have access to high-quality essential healthcare services. (United Nations, 2020). In this context, where resources are scarce, Operations Research (OR) can contribute to optimize the use of healthcare assets, resulting in better coverage and, ultimately, an improvement in human development and quality of life for society. A notable example in this regard is the research done by Lee et al. (2015), where they improved the Emergency Department quality of care and operational efficiency, while reducing avoidable readmissions, using a decision support system integrating machine learning, simulation, and optimization.

In this research, we will tackle one of the many problems healthcare services usually face. Specifically, it is relevant to understand and improve the usage of hospital beds in a Hospital Network by transferring patients between hospitals, and managing the movements of inpatients (i.e., patients who are admitted to a hospital) between hospital sections. Green (2005) mentions that this problem has vast opportunities for OR tools to improve efficiency and effectiveness in bed management decisions, which are also related to decisions about the admission of patients.

The lack of hospital beds in Chile's Public Healthcare System has been a problem for a long time. Although some initiatives have addressed this problem, bed availability is a primary concern, leading to the challenge of improving the system's management.

This situation is exemplified by the fact that Chile has 2.16 beds per 1,000 inhabitants, which is lower than the 2.8 beds per 1,000 inhabitants recommended by the World Health Organization (WHO) (Colegio Médico de Chile, 2019). Furthermore, the average of the Organization for Economic Co-operation and Development (OECD), which Chile is part of, is 4.8 beds per 1,000 inhabitants (Colegio Médico de Chile, 2019). This shows that Chile still has a long way to go in terms of providing adequate healthcare facilities to its citizens. Ultimately, this affects the patient, who ends up waiting, on average, 12 hours and up to 3 days to be attended, given that there are not enough beds to service the entire population. The lack of capacity and efficiency in the Chilean Public Healthcare System has resulted in a precarious situation. Statistical data from 2021 shows that almost 2.3 million people are waiting to be attended, and, eventually, it is suspected that approximately 40.000 people died waiting for medical procedures (Ministerio de Salud de Chile, 2023a).

To address this problem, the Chilean Ministry of Health created in 2009 the Centralized Bed Management Unit, which later changed its name to Centralized Case Management Unit (CCMU). This public entity coordinates and manages the beds in the Public System, in cooperation with the Private System when needed (Ministerio de Salud de Chile, 2023b). Given a medical request from a patient of the Public Healthcare System, the main decisions that the CCMU can take are to:

(i) Transfer: Move the patient between two hospitals from the Public Network.

(ii) Divert: Move the patient to a clinic in the Private System to provide a medical resolution.

(iii) Admission: Admit the patient to the hospital for medical care.

Although the CCMU has achieved certain milestones, its action policy is imperfect concerning efficiency and effectiveness in bed management decisions in a Hospital Network, requiring consideration of the inpatient management of each hospital in the network. It has been reported that its operations are not necessarily optimal and well-informed (Instituto de Sistemas Complejos de Ingeniería, 2020), resulting in situations where transfers and diversions are made only when a hospital's capacity does not allow the patient admission. This situation results in sub-optimal actions, thus ending in high treatment and operation costs for the Public System. Also, given that patient allocation is not necessarily managed correctly, the Hospital Network ends up with high congestion levels, which increases the waiting times to access medical treatment, resulting, for example, in wait times of over 200 days for patients in the Waiting List (Ministerio de Salud de Chile, 2023a).

Hence, given this situation, the present study proposes, as its primary objective, to enhance the policy of the CCMU, with a specific focus on reducing the sum of operational costs and the monetized social costs (i.e., patient waiting times and bed opportunity costs), thereby improving the overall experience for patients via a proactive methodology. Specifically, we propose a formulation via a Markov Decision Process (MDP), which can simultaneously handle the transfers, diversions, and admissions in a Hospital Network, as well as the inpatient management in each hospital that compose the network. Although designed for the Chilean reality, this model and related methodology can be applied to other situations where the healthcare system aims to operate similarly.

To achieve our main objective, this work is structured as follows: Section 2 reviews the literature and the state-of-the-art models and methodologies regarding the improvement of hospital bed usage in a Hospital Network, with a special interest in those research that develops proactive policies in terms of decisions and patient management. Section 3 describes the problem in detail, showing all the relevant features in this study, whereas Section 4 formulates the MDP model, given the previous section. Section 5 addresses the biggest challenges regarding the model resolution, describing the proposed solution approach. Section 6 shows the implementation and results of an example instance, based on real-case studies, which leads to further analysis of the obtained policy. Finally, Section 7 consolidates this study's major findings and discoveries while proposing extensions and future lines of research.

2. LITERATURE REVIEW

The availability of beds and the management of their usage in a Hospital Network are vital aspects of hospital operations and productivity. Using OR tools aims to avoid overcrowding and ensure a desirable patient flow and attention process in every hospital unit, ultimately increasing the healthcare system service quality and patient medical outcomes (Sun et al., 2013). The number of studies on this topic has steadily increased since the early 2000s, in both deterministic and stochastic approaches, although focusing on specific parts that compose the whole hospital capacity planning and optimization problem (Humphreys et al., 2022).

An important OR technique used to address this challenge is simulation. The advantages of this tool are that the system's complex interactions can be represented, including multiple decisions and uncertain events, such as those regarding patient arrivals and inpatient flow. This allows the decision-maker to test different policies and system provisions, which ultimately requires comparing them to determine the best decision. Also, this technique is pretty versatile; Banditori et al. (2013) developed a deterministic integer programming approach to optimize the master surgical scheduling problem in a hospital, testing the solution's robustness against the unpredictability of the duration of surgical procedures and the duration of hospital stays. On the other hand, Ahalt et al. (2016) proposed a discrete-event simulation approach to determine which crowding scores used at the Emergency Department are the best for predicting impending crowdedness. Another example is the study conducted by Devapriya et al. (2015), where they introduce a decision support tool for managing in terms of planning and budgeting the current and future bed capacity, also evaluating the benefits from possible improvements in bed provision.

Another vital aspect of OR tools applied in healthcare is the utilization of mathematical formulations. These formulations, particularly using queueing theory, aim to enhance bed management, reduce wait times, and address the challenge of increasing demand (Green, 2006). The study conducted by McManus et al. (2004) addressed the allocation problem of scarce resources in the Intensive Care Unit, based on real-world data, they constructed an accurate mathematical model of patient flow, which correctly predicted demand levels and helped to determine the appropriate supply of beds. Alternatively, the research done by de Bruin et al. (2007) studied the bottlenecks in the emergency care chain of cardiac in-patient flow in a university medical center, optimizing the bed allocation over it and retrieving useful data in terms of patient dynamics and Length of Stay (LOS), which ultimately gave insights for future decision-making.

Studies concentrating on deterministic and stochastic optimization are another form of OR that can be implemented in healthcare. For instance, Feng et al. (2015) developed and solved via heuristics a multi-objective stochastic model for medical resource allocation in Emergency Departments, minimizing the average length of stay of patients and medical resource waste costs. Similarly, Zhou et al. (2018) proposed a multi-objective stochastic programming model that aims to maximize hospital revenue and equity among different types of patients, allocating hospital ward resources, which are often pretty limited. Finally, Burdett et al. (2017) developed an analysis of hospital resources and capacity, coupled with different mixed integer linear programming models, which helped to determine the maximum number of patients that the system could attend in a certain time window, later validating these results in a real-case scenario.

It is important to note that these examples represent only a small portion of the extensive research that utilizes OR tools in healthcare. Even though the mentioned studies, as well as others not reported, achieve great results in describing the system or obtaining an optimized (and usually increased) bed provision and/or configuration via a mathematical formulation, a much more sophisticated and proactive focus, based on simulation and optimization simultaneously, is needed. The review done by Humphreys et al. (2022) manifests this issue, adding that previous research tends to focus on certain parts of the hospital, which ultimately neglects that a hospital is one body, where each component is intricately related to the others, needing an integral and detailed description to obtain significant and valuable results. This is essential to achieve a real-world solution that can be effectively implemented.

Given these research gaps proposed by the authors and the opportunity to increase efficiency with the available resources, instead of just increasing them, this study must develop a decision frame that pairs simulation and optimization. Also, it must be capable of optimizing different costs in a complex and detailed network, considering patient equity and the possibility of scaling up to real-world scenarios. To achieve these tasks, we will focus on specific research studies that have developed proactive policies regarding decisions and patient management, with real-world applications, ideally. In our analysis, we found that the bed management problem can be categorized into two scales. The first one, referred to as the macro-scale dimension, handles the admission, transfer, and diversion actions that the CCMU could take in a complex hospital network. On the other hand, the micro-scale dimension handles the inpatient management in each hospital.

Regarding the macro-scale dimension, the first research is titled *An approximate dynamic programming approach to the admission control of elective patients*, by J. Zhang, Dridi, and Moudni (2021). In this study, the authors tackle the problem associated with the elective demand admission problem, specifically regarding surgery scheduling for a complex and diverse waiting list regarding patient diagnoses and medical conditions. Given the problem, an MDP model is proposed, which is solved using Approximate Dynamic Programming, based on a novel reinforcement-learning-based approach that can handle the curses of dimensionality resulting from the large scale of realistically sized problems. The most important contribution of this research, in terms of our study, is that a dynamic and long-term proactive policy is obtained that can handle real-world problems, reduce operational costs and waiting times, and increase equity between patients. The biggest opportunities from this research are to extend the model to the downstream patient flow into other units following their surgery, as well as the necessity to incorporate a solution approach that can handle non-linear costs, which their ADP proposal neglects via a cost linearization.

Another relevant study in terms of the macro-scale problem is the one defined in *Identifying proactive ICU patient admission, transfer and diversion policies in a public-private hospital network*, by Marquinez, Sauré, Cataldo, and Ferrer (2021). In this research, an infinite-horizon MDP, with a cost-effective policy for transferring ICU patients between hospitals or diverting them to private clinics, is obtained for a real-case scenario. Given the courses of dimensionality, the MDP's solution approach is based on an equivalent linear programming model using column generation. This study excels in being the first dynamic bed allocation problem that handles multiple hospitals and multiple types of patients at the same time, as well as collaboration between public and private hospitals through the externalization of care services. Similarly to the previous research, the biggest opportunity is to extend the model to a complex and dynamic patient flow, considering the multiple units that interact inside a hospital. Also, the proposed Approximate Lineal Optimization solution method fails to represent and handle the non-linear costs of the real-life problem regarding wait-time and, subsequently, the health hazards that a patient must assume. Finally, it is possible to consider elective demand in a waiting list, just as J. Zhang et al. (2021) proposed in their model, to cover the complete reality and decisions of the Health System.

Regarding the micro-scale dimension, it is relevant to mention the study: *A proactive transfer policy for critical patient flow management*, by González, Ferrer, Cataldo, and Rojas (2018). This research proposes an MDP model to improve the inpatient flow between the Emergency Department and other hospital units. Similarly to Marquinez et al. (2021), the solution method, given the courses of dimensionality, is based on the Approximate Lineal Optimization method, retrieving an equivalent linear programming model using column generation. The main contribution is the possibility of obtaining a long-term action policy for the inpatient management flow in a complex hospital model that considers the interactions between critically ill patients and the different departments within the hospital. The biggest opportunity is to extend the model to a complex and dynamic network, considering multiple hospitals simultaneously, as well as the possibility of giving medical resolution to patients in the private system, something that Marquinez et al. (2021) propose. Furthermore, it is worth mentioning that the proposed linear approach does not necessarily adjust to reality's complexity in terms of costs.

In conclusion, the present study aims to gather and extend the opportunities these studies have as features, integrating them and representing the relationships between hospitals and their units in a complex network. To achieve this, an integral and detailed model, based on a Markov Decision Process (MDP), is formulated to obtain significant and valuable results.

3. PROBLEM DESCRIPTION

3.1. Macro-Scale Problem

In the context of the macro-scale dimension of the problem, we will consider a public hospital network similar to Marquinez et al. (2021). This network consists of H hospitals; each receives urgent patients at their Emergency Department (ED) and can transfer patients between them. Additionally, the Public Network can divert patient's requests to private hospitals. As suggested by the authors, the clinics in the Private System (PS) can accommodate a virtually infinite number of requests with various medical conditions. Figure 3.1 illustrates this dimension of the problem; in this case, it shows what Hospital 1 can do with the demand that has arrived at its ED.

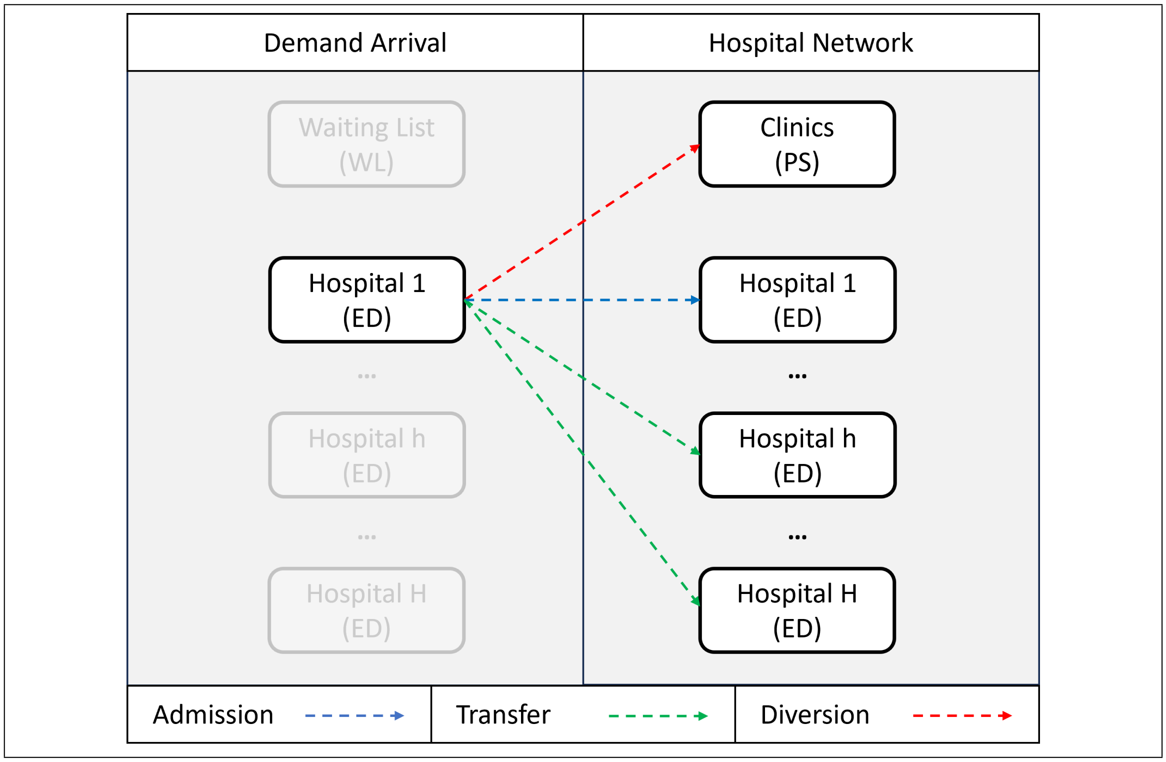

Figure 3.1. Diagram of the Macro-Scale Dimension of the Problem for the Emergency Demand

Additionally, the network features a Dynamic Elective Patient Waiting List (WL), which includes requests from the population that must be addressed within a specific time frame. Although this last part apparently simplifies the problem presented by J. Zhang et al. (2021), our proposal can effectively model the emergency patients that arrive at the hospitals, as well as the elective demand that the Public System must service. The Public Network can also divert patients from the Waiting List to the clinics in the Private System if necessary, satisfying all their health requirements. Figure 3.2 illustrates this dimension of the problem; in this case, it shows what the Decision Maker can do with the demand that has arrived at the Network WL.

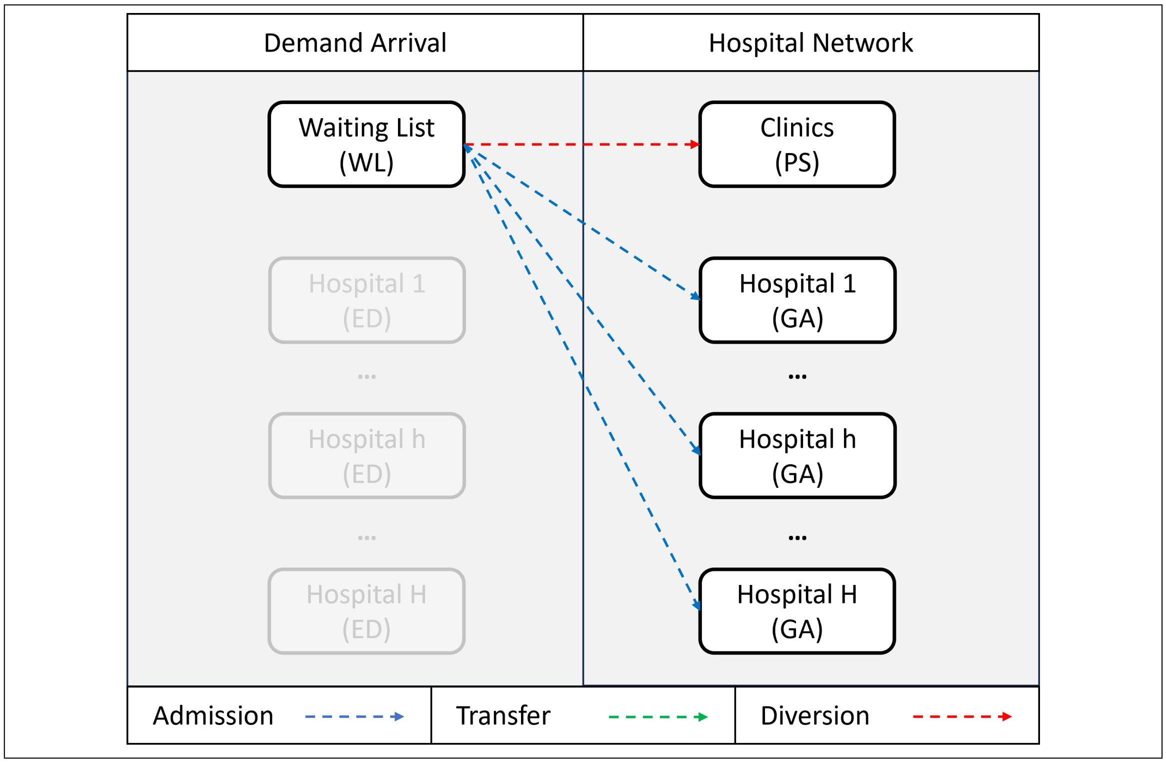

Figure 3.2. Diagram of the Macro-Scale Dimension of the Problem for the Elective Demand

3.2. Micro-Scale Problem

In the context of the small-scale dimension of the problem, which deals with inpatient flows within a hospital's units, we propose an extension from the problem studied by González et al. (2018). In this scenario, emergency patients arrive at the hospital's ED and can be admitted to the inner units; if not, they are transferred to another hospital's ED or diverted to the PS. Patients are assigned to specific units based on their medical needs, as follows: For surgery requirements, the Operating Room (OR); for high complexity conditions, the Intensive Care Unit (ICU); those with medium complexity to the Step Down Unit (SDU), and the ones with low complexity are accommodated in the Ward. Additionally, the elective demand from the WL first passes through the General Admission (GA) process before being transferred to the inner units (OR, ICU, SDU, WARD). Inpatients may move between the inner units until their medical needs are fully addressed; at this point, they are discharged from the WARD.

Compared to González et al. (2018), in our study, we will consider all kinds of possible transitions, which implies a greater complexity in the stochasticity and dynamism of the problem. Furthermore, we consider resource consumption in the OR, although we assume that all the requests for this resource can be accommodated in the hospital's surgical schedule.

Figure 3.3 provides a representation of the micro-scale dimension of the problem. At the top, it shows the type of actions taken, and at the bottom, it outlines the various types of possible transitions that a patient could take. It is important to observe that we outline some uncommon transitions to facilitate the representation of the case study. Also, it is important to mention that although inpatient transfers between hospitals are possible, for the sake of simplicity, we will not consider them in this study.

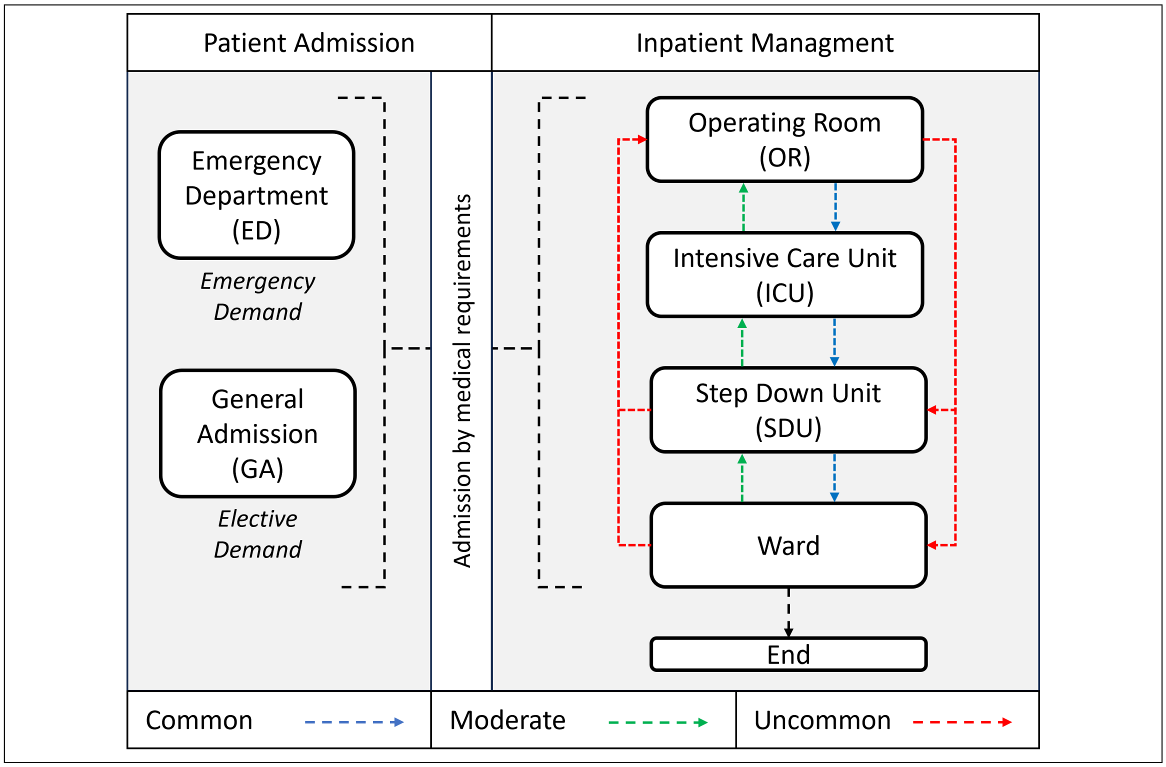

Figure 3.3. Diagram of the Micro-Scale Dimension of the Problem

Note that every hospital represented in Figures 3.1 is modeled assuming the proposal from Figure 3.3. This generates the connection between both scales, where the decisions from the Macro-Scale (diversions, admissions, and transfers) will affect each hospital's resource utilization, thus requiring optimal inpatient management between each hospital's units.

3.3. Decision Making Period and Horizon

To model this problem, a reasonable time frame between actions must be defined. However, one may think that a shorter time window means better modeling, which is not necessarily needed to achieve good results. As stated by González et al. (2018) and Marquinez et al. (2021), time discretization must be according to the decisions being taken. In our study, the decision-making timeline is represented in Figure 3.4.

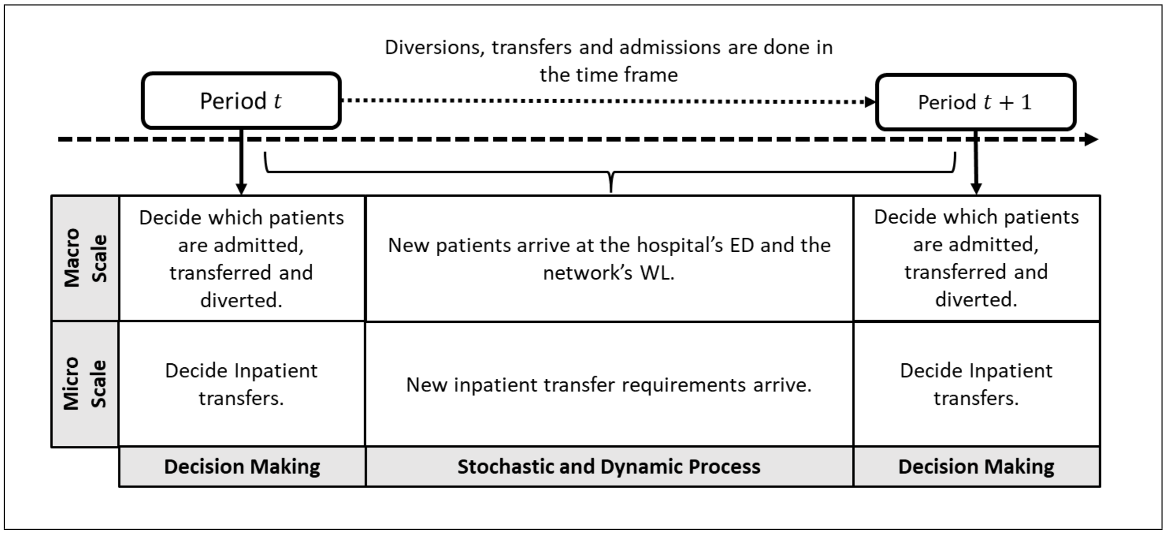
Diversions, transfers and admissions are done in the time frame
Period t
Period $t+1$
Macro Scale
Micro Scale
Decide which patients are admitted, transferred and diverted.
New patients arrive at the hospital's ED and the network's WL.
Decide which patients are admitted, transferred and diverted.
Decide Inpatient transfers.
New inpatient transfer requirements arrive.
Decide Inpatient transfers.
Decision Making
Stochastic and Dynamic Process
Decision Making

Figure 3.4. Diagram of the Decision Making Timeline

In our study, fractions of the day are considered. Following the literature (González et al., 2018; Martinez et al., 2019; J. Zhang et al., 2021), it is valid to consider two windows of 12 hours. We assume that the preparation times for beds, as well as the times for transfers, diversions, and admissions of patients, occur within the given time frame, provided that they are shorter than the time window itself. However, it is possible to further discretize time without affecting the modeling and solution strategy. Also, the model is constructed with an infinite horizon, aiming to derive a stationary operational policy, given the sequential decision process behind the modeling.

3.4. Patients

3.4.1. Aggregation by Diagnosis-Related-Groups

Although each patient has its characteristics, such as age, sex, comorbidities, diagnosis, procedures, and other aspects, developing a model or strategy to consider each particular case will make it unmanageable by any Operations Research methodology. To address this management problem, Fetter et al. (1980) and Fetter (1991) developed a classification scheme based on the relation between the demographic, diagnostic, and therapeutic characteristics of patients and the hospital outputs they utilize. This clustering, known as the Diagnosis-Related-Groups (DRG), leads to a reasonable amount of patient types to manage.

This classification has proven to be useful in different aspects; Zapata (2018) mentions that, in Chile, the DRG classification system has helped clinical and managerial staff to improve patient flow, making them work together. One key aspect of these improvements is the different metrics that we can retrieve to make comparisons of performance between different hospitals in a network and, even more relevant, between two different years of operations in the same system. This proposal is done by the Case Mix Report proposed by Fetter et al. (1980). For the sake of our study, let us assume that we have a DRG classification with K different groups. Also, given a reference system, which can be the non-optimized scenario (i.e., the hospital network outcomes under the current non-improved management policy), let us define P_k as the proportion of DRG k of the reference system cases, and A_k the average stay for DRG k in the reference system. Similarly, let us define p_k and a_k as the metrics for DRG k in the case study system, which can be the optimized scenario (i.e., the hospital network outcomes under the improved management policy), for the proportion and the average Length of Stay (LOS), respectively.

Given these values, Fetter et al. (1980) proposed the following metrics:

(i) Average Lenght of Stay (ALOS): For the reference system, as well as the case study system, we can obtain the average stay of the system as follows:

$$A = \sum_k A_k P_k$$

$$a = \sum_k a_k p_k$$

(ii) LOS Index: It is the relative measure of LOS. This index compares how the case study would attend to patients compared with the reference system's treatment duration standards. The formula is the following:

$$\text{LOS Index} = \frac{\sum_k a_k P_k}{\sum_k A_k P_k}$$

The numerator estimates the ALOS if we have attended the reference system Case Mix P with the case study treatment durations a. On the other hand, the denominator calculates the ALOS of the reference system. If we improve efficiency in treating the same patients, the desirable index would be lower than 1. Values higher than 1 represent a reduction in efficiency.

(iii) Case Mix Index: It is the relative measure of Case Mix Complexity (where the complexity is proportional to the patients' stay). This index compares how the reference system would have attended the case study patients compared with the reference system's Case Mix. The formula is the following:

$$\text{Case Mix Index} = \frac{\sum_k A_k p_k}{\sum_k A_k P_k}$$

The numerator estimates the ALOS if we have attended the case study Case Mix p with the reference system treatment durations A. On the other hand, the denominator calculates the ALOS of the reference system. If we increase the Case Mix complexity, this index will be higher than 1. Values lower than 1 represent a less complex case study.

(iv) Differences: Although we can directly compare a and A, we can not separate the effects due to changes in LOS (directly related to efficiency) and Case Mix. To separate them, we can use the proposition by Kitagawa (1955), which is mentioned by Fetter et al. (1980) in the proposal of the Case Mix Report.

$$\underbrace{a-A}_{\Delta\ \text{ALOS}} = \underbrace{\sum_k P_k(a_k - A_k) +}_{\text{Difference due to LOS}} \underbrace{\sum_k A_k(p_k - P_k)}_{\text{Difference due to Case Mix}} \underbrace{+\sum_k (a_k - A_k)(p_k - P_k)}_{\text{Difference due to the interaction}}$$

This formula was reviewed by Gupta (1978), resulting in a simplified version:

$$\underbrace{a-A}_{\Delta\ \text{ALOS}} = \underbrace{\sum_k \frac{(P_k + p_k)(a_k - A_k)}{2}}_{\text{Difference due to LOS}} + \underbrace{\sum_k \frac{(A_k + a_k)(p_k - P_k)}{2}}_{\text{Difference due to Case Mix}}$$

Separating these terms lets us better understand the reasons behind changes in ALOS between the systems.

One big issue regarding the DRG classification system is the necessity of many personnel compiling and classifying each patient regularly to achieve as close to real-time classification as possible. This staffing problem is coupled with human error, resulting in misclassifications that endanger the DRG applications. Thankfully, new approaches that use Machine Learning have been developed to address these problems, such as *Machine Learning Approaches for Early DRG Classification and Resource Allocation* by Gartner et al. (2015) and *Early prediction of diagnostic-related groups and estimation of hospital cost by processing clinical notes* by Liu et al. (2021).

3.4.2. Length of Stay and Patient Flow Model

Given the DRG classification system, it is easier to incorporate desirable properties for each group. The first one is the LOS distribution, which tends to fit well with lognormal-based models in every hospital unit, as reported in the literature (Min & Yih, 2010; Samudra et al., 2016; X. Zhang et al., 2019; Zhu et al., 2018). We want to point out that choosing this distribution does not necessarily imply that it is the only one possible in real-life scenarios, existing other options that the proposed model can handle either way.

Regarding patient flow modeling, it can be assumed that the entity arrival follows a Poisson distribution (Astaraky & Patrick, 2015; Samudra et al., 2016; Truong, 2015;

Van Riet & Demeulemeester, 2015). In contrast, the inpatient requirements distributions, including the discharge, can be modeled with a Markov Chain, using the distinctions proposed by González et al. (2018). Similar to the LOS modeling, it is important to remark that there could be other alternatives to represent real-life phenomena regarding this last aspect. Also, it is important to mention that these requirements originate from medical staff decisions, where the model at the Micro-Scale dimension is a tool to handle, organize, and optimize inpatient flow based on these medical requests.

3.4.3. Waiting Time Costs

The mortality rate in Chilean hospitals rose in connection with the waiting time for patients with non-prioritized health conditions (Martinez et al., 2019). This situation is supported by the study conducted by Plunkett et al. (2011), where evidence indicates that delays in admission and transfers are linked to unfavorable outcomes in terms of mortality. To achieve patient equity and healthcare optimization, as well as a way to compare the economic costs regarding operations from transfers and diversions, a simple and effective way to assign economic value to waiting time effects is needed.

Although there are different approaches, we will focus on a strategy that could merge a statistical approach with the expert knowledge of medical staff. To achieve that, first of all, let us define T as the maximum waiting time (in periods) allowed in the system for any patient at any unit. Based on that, the statistical approach suggests that we could estimate the death (or irreparable damage) probability for any waiting time period $t \leq T$. There are numerous studies regarding this issue, such as the ones conducted by Martinez et al. (2019) and Plunkett et al. (2011). Although many of these insights are concentrated in the Emergency Department, they fail to explain this phenomenon in other hospital units. Also, there is not always enough historical information to completely understand each DRG probabilities at every unit, requiring other sources of information. Specifically, medical staff, based on their experiences, can help fill the missing information gaps. Thus, not only is the model more precise in terms of quantifying waiting time effects, but it also includes the clinician's opinions and expertise, which gives the model greater validity and appeal before healthcare professionals.

As an example of what can be obtained, Figure 3.5 represents three different evolutions over time. It is observed that, independent of the risk level, every curve has at the cumulative waiting time T a coefficient equal to 1. In that situation, the patient's integrity is completely in danger. For instance, the High-Risk curve resembles the increasing mortality rate for a patient waiting at the ED, as reported by Plunkett et al. (2011). The Medium-Risk curve could be appropriate for an inpatient waiting to be transferred from the Ward to the SDU, while the Low-Risk curve is suitable, for example, for an elective patient waiting at the WL.

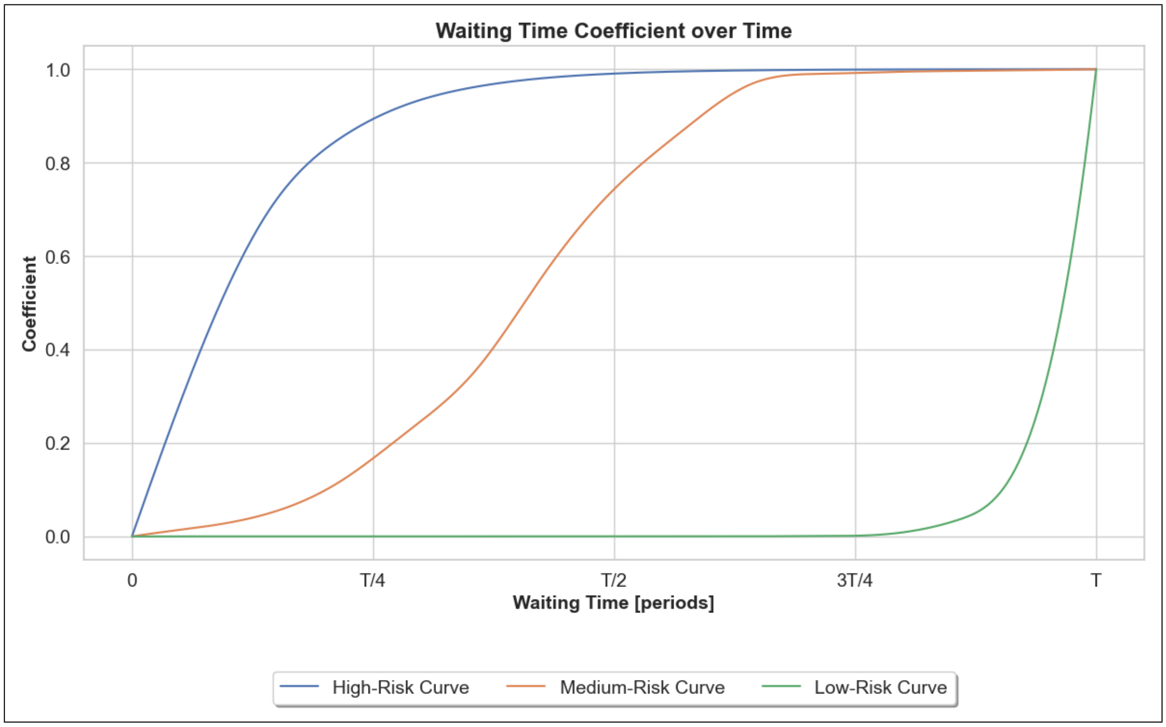

Figure 3.5. Representation of Different Waiting Time Coefficients over Time

Given these coefficients, we can multiply them by the statistical value of life, usually used in Chile in the Social Evaluation of Projects, to obtain a monetary value for the risks associated with the waiting time. It is important to mention that, for practical purposes, we seek to evaluate the economic appreciation that society assigns to a marginal decrease in fatality risk. Consequently, it should not be interpreted as the monetary value of a life, since the latter is invaluable.

3.5. Bed Opportunity Cost

It is also important to consider the detrimental effects in terms of operational efficiency and costs that the healthcare system must assume in case of inpatient overstay at certain hospital units. As stated by Madeira et al. (2021), there is evidence to determine that prolonged stays considerably affect the operational efficiency, as well as the costs of the hospital network. In this way, it is necessary to reduce these situations to optimize the use of public resources.

Batista et al. (2021) proposes that inpatient overstay costs, in terms of the detriment of patient service, require hospital managers to establish specific values, following their preferences and internal policies. On the other hand, authors such as Seidel, Whiting, and Edbrooke (2006) and Tan et al. (2012), were able to estimate a monetary value for the inpatient overstay, considering the misutilization of resources (staffing, equipment, etc.) that the hospital is assuming. Similarly to the waiting time costs, we consider that the Bed Opportunity Cost must consider both statistical and hospital manager preferences approach in order to have a robust estimation, but also a credible and attractive one to healthcare professionals.

4. MODEL FORMULATION

It is important to simplify the problem before creating a mathematical model. This involves identifying decision variables that focus on the Macro-Scale actions, such as the admission and diversion of patients, and the transfers between emergency departments to manage the emergency demand. Additionally, it is equally important to manage the Micro-Scale actions which involve the flow of inpatients to ensure that medical requests are promptly serviced.

Given the actions taken, an operational cost is incurred due to the Macro-Scale actions (transfers and diversions). Also, an opportunity cost is assumed at the end of each modeled period due to the waiting times that some patients are experiencing and the beds they occupy that could be attending the correct type of inpatient. These two costs represent the immediate costs. Additionally, the resulting state of the system (i.e., how the different types of patients are occupying the system's capacity at every hospital) at the end of a period affects the expected future costs that the system will have. Thus, the problem requires minimizing the sum of the immediate costs with the discounted expected future costs.

At the end of each period, the problem must satisfy certain constraints regarding bed capacity by each unit. This ensures that all inpatients who have been accommodated can receive proper attention and care.

Let us assume a set of all types of patients P that the system can treat. This set includes all the potential combinations of DRGs, medical requirements, and waiting periods. This set will be detailed in section 4.3. Specifically, let us consider the patients from type $p \in P$. If we delineate the issue concerning patients of type p, Figure 4.1 offers a visual depiction of this specific patient management challenge. Note that we only present a pair of hospitals for a more precise representation. Immediately, it can be noted that this situation resembles a network flow problem for each $p \in P$. The capacity constraints bind all these sub-problems into one big problem.

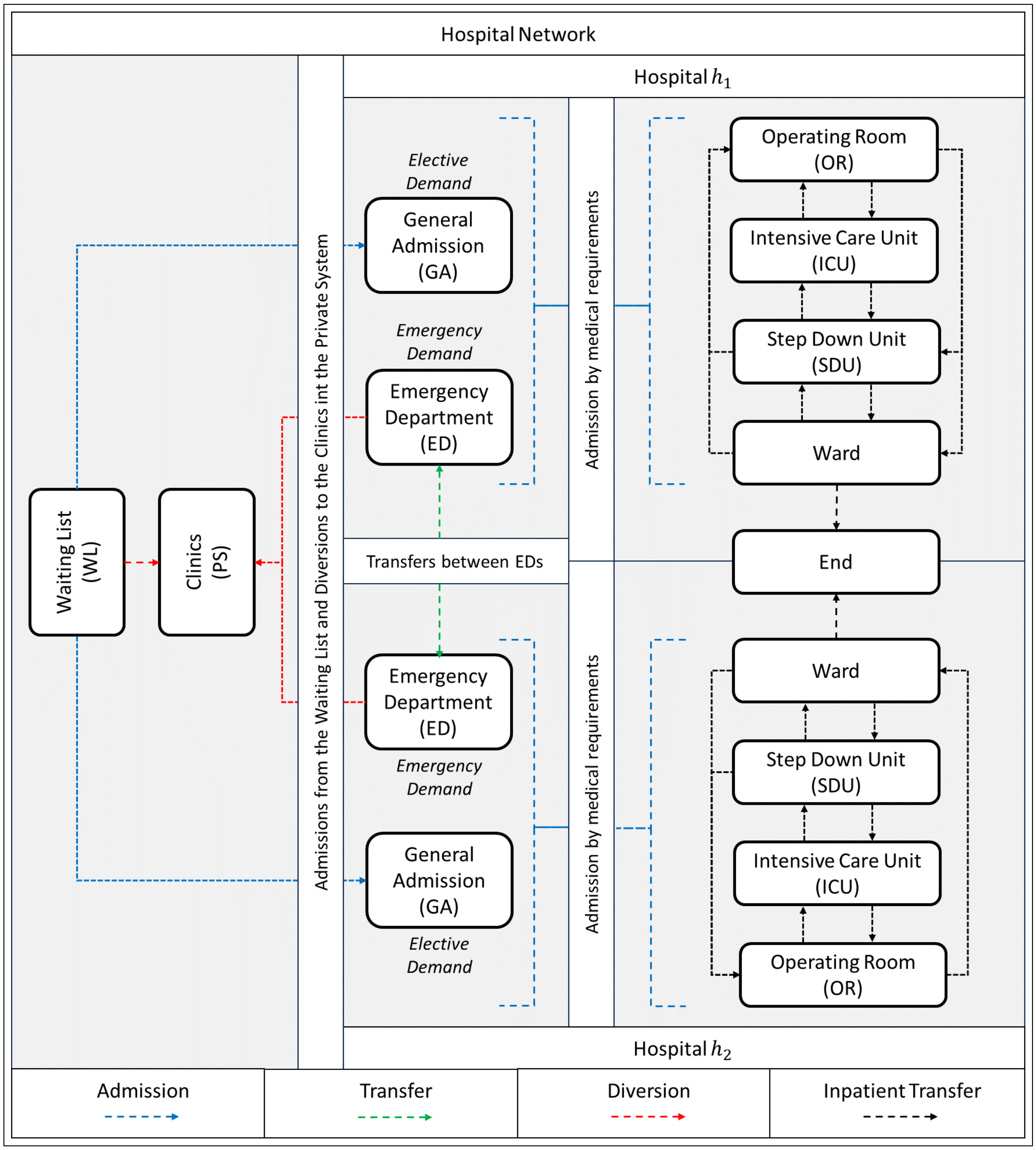

Figure 4.1. Diagram of the Problem on a 2 Hospital Network for Each Type of Possible Patient

Additionally, we can define a synthetic notation to facilitate the formulation of the different metrics and, subsequently, the mathematical model notation based on Gupta (1978).

For instance, given a two-index variable x_{ij} and coefficient α_{ij}, where $(i,j) \in I \times J$, we can define:

$$x_{*j}\alpha_{*j} = \sum_{i \in I} x_{ij}\alpha_{ij}$$
$$x_{i*}\alpha_{i*} = \sum_{j \in J} x_{ij}\alpha_{ij}$$
$$x_{**}\alpha_{**} = \sum_{i \in I}\sum_{j \in J} x_{ij}\alpha_{ij}$$

Obviously, this notation can be extended to an N dimensional variable.

Given the problem description, the mathematical formulation of a cost-minimization MDP will be presented in the next sections.

4.1. Assumptions

To ensure the convergence and feasibility of the model solution and to maintain a consistent formulation regarding the system dynamics, it is imperative to declare our assumptions:

(i) Patients are assumed to not die during the care process.

(ii) Patients have infinite waiting disposition.

(iii) There is the availability of logistic resources to perform all transfers and derivations that the decision-maker deems appropriate.

(iv) Time discretization allows for patient transfers and sanitation and bed preparation processes.

(v) The model has finite returns, i.e.:

$$|c(\vec{s_1}, \vec{a})| \;<\; \infty \quad \forall \vec{s_1} \in \mathbb{S}_1,\; \forall \vec{a} \in \mathbb{X}(\vec{s_1})$$

(vi) The process is stationary, meaning returns and probabilities are independent of time.

4.2. Bellman Equations

The general Bellman Equations for an MDP, where it is expected to minimize a cost function, are:

$$C(\vec{s_1}) = \min_{\vec{a} \in \mathbb{X}(\vec{s_1})} \left\{ c(\vec{s_1}, \vec{a}) + \lambda \sum_{\vec{u} \in \mathbb{S}_1} C(\vec{u}) p\left(\vec{u} | \vec{s_2}\right) \right\} \quad \forall \vec{s_1} \in \mathbb{S}_1$$

Where $\vec{s_1}$ represents the pre-decision state, $\vec{s_2}$ the post-decision state, $\vec{a}$ represents the actions, $c(\vec{s_1}, \vec{a})$ the Cost-to-Go, and $\vec{u}$ represents the possible future states. Each component will be detailed in the following sections.

4.3. Sets

The sets considered in the model are the following:

Hospitals: $h \in \{1, 2, \ldots, H\}$
Admission Units (N_1): $n \in \{\text{ED, GA}\}$
Service Units (N_2): $n \in \{\text{OR, ICU, SDU, WARD}\}$
Hospital Units (N): $N_1 \cup N_2$
Diagnosis Related Groups (DRG): $g \in \{1, 2, \ldots, G\}$
Modeled Periods for Patient Waiting Time ($\mathbb{T}$): $t \in \{1, \ldots, T\}$
Patient Types: $p \in P = \{N_2 \times DRG \times \mathbb{T}\}$, where p represents a patient with need of care $n(p)$, DRG $g(p)$ and waiting time $t(p)$.

4.4. Parameters

The parameters considered in the model are the following:

$\Phi_h^n \in \mathbb{Z}^+$: Capacity of unit $n \in N$ in hospital h.
$\bar{\omega} \in \mathbb{Z}^+$: Capacity of the WL.

ρ_p: Average cost of deriving a patient of type $p \in P$ to the PS from the WL at the beginning of the period.

$\theta_p^h \in \mathbb{R}^+$: Average cost of deriving to the PS an ED patient of type $p \in P$ from hospital h at the beginning of the period.

$\vartheta_p^{h_1 h_2} \in \mathbb{R}^+$: Average cost of transferring an ED patient of type $p \in P$ between hospitals h_1 and h_2 at the beginning of the period.

$\nu_p \in \mathbb{R}^+$: Average cost per waiting period for a patient of type $p \in P$ in the WL at the end of the period.

$\phi_p^{hn} \in \mathbb{R}^+$: Average cost per waiting period for a patient of type $p \in P$ in hospital h and unit $n \in N$ at the end of the period.

$\mu \in \mathbb{R}_0^+$: Weight factor for waiting costs.

4.5. States

As recommended by Powell (2011, 2012), the model states can be divided into a Pre-Decision State and a Post-Decision State, where the Post-Decision State is achieved after the optimal decision for the Pre-Decision State is taken. The benefits of using this modeling approach include its ability to simplify the handling of the various types of curses of dimensionality, which often arise in MDP formulations.

4.5.1. Start of a Period (Pre-Decision State)

$$\mathbb{S}_1 := \{\vec{s_1} : (\vec{\omega_1}, \vec{\alpha_1})\}$$

Where:

$\omega_1^p \in \mathbb{Z}_0^+$: Number of patients of type $p \in P$ on the WL at the start of the period.

$\alpha_1^{phn} \in \mathbb{Z}_0^+$: Number of patients of type $p \in P$ in hospital h and unit $n \in N$ at the start of the period.

4.5.2. End of a Period (Post-Decision State)

$$\mathbb{S}_2 := \{\vec{s_2} : (\vec{\omega_2}, \vec{\alpha_2})\}$$

Where:

$\omega_2^p \in \mathbb{Z}_0^+$: Number of patients of type $p \in P$ on the WL at the end of the period.
$\alpha_2^{phn} \in \mathbb{Z}_0^+$: Number of patients of type $p \in P$ in hospital h and unit $n \in N$ at the end of the period.

4.6. Actions

The admissions, diversions, and transfers in the model are represented as follows:

$$\mathbb{X}(\vec{s_1}) := \{\vec{a} : (\vec{\eta}, \vec{\kappa}, \vec{v}, \vec{w}, \vec{x}) \in \mathbb{F}(\vec{s_1})\} \quad \forall \vec{s_1} \in \mathbb{S}_1$$

Where:

$\eta_p \in \mathbb{Z}_0^+$: Number of patients of type $p \in P$ on the WL derived to the PS at the start of the period.
$\kappa_p^h \in \mathbb{Z}_0^+$: Number of patients of type $p \in P$ on the WL who are admitted to GA of hospital h at the start of the period.
$v_p^h \in \mathbb{Z}_0^+$: Number of patients of type $p \in P$ derived to the PS from the ED of hospital h at the start of the period.
$w_p^{h_1 h_2} \in \mathbb{Z}_0^+$: Number of patients of type $p \in P$ transferred from the ED between hospital h_1 and h_2 at the start of the period.
$x_p^{h n_1 n_2} \in \mathbb{Z}_0^+$: Number of patients of type $p \in P$ transferred within hospital h between units $n_1 \in N$ and $n_2 \in N_2$ at the start of the period.

Each action $\vec{a}$ belongs to $\mathbb{F}(\vec{s_1})$, meaning that it satisfies:

(i) Transition within a period: The action taken can ensure a correct transition between the pre-decision and post-decision states, respecting their respective nature (non-negativity). Note that the set of constraints resembles a network-flow problem, where the patient flow is being optimized.

- Waiting List:

(a) The number of patients of type $p \in P$ of the post-decision state is equal to the number of patients of type $p \in P$ minus the ones diverted to the PS and the admitted to the hospitals.

$$\omega_2^p = \omega_1^p - \eta_p - \kappa_p^* \quad \forall p \in P$$

- Hospital's GA:

(a) The number of patients of type $p \in P$ with a waiting time equal to 1 in each hospital h of the post-decision state is equal to the number of patients of type $p \in P$ with a waiting time equal to 1 in each hospital h plus the ones admitted from the WL with the same DRG g and medical requirement $n \in N_2$.

$$\alpha_2^{phn} = \alpha_1^{phn} + \sum_{u \in \mathbb{U}(p)} \kappa_h^u \qquad \forall p \in \{p \in P : t(p) = 1\}$$

$$\forall h \in \{1, \ldots, H\}$$

$$\forall n \in \{\text{GA}\}$$

Where $\mathbb{U}(p) = \{u \in P : n(u) = n(p) \wedge g(u) = g(p)\}$

(b) The number of patients of type $p \in P$ with a waiting time higher than 1 in each hospital h of the post-decision state is equal to the number of patients of type $p \in P$ with a waiting time higher than 1 in each hospital h minus the ones admitted to the service unit $n_2 \in N_2$.

$$\alpha_2^{phn_1} = \alpha_1^{phn_1} - x_p^{hn_1n_2} \qquad \begin{aligned} &\forall p \in \{p \in P : 1 < t(p) \wedge n(p) = n_2\} \\ &\forall h \in \{1, \ldots, H\} \\ &\forall n_1 \in \{\text{GA}\} \\ &\forall n_2 \in N_2 \end{aligned}$$

- Hospital's ED:

(a) The number of patients of type $p \in P$ in each hospital h of the post-decision state is equal to the number of patients of type $p \in P$ in each hospital h, minus the ones diverted to the PS, plus the ones transferred from the other hospital's ED, minus the ones transferred to the other hospital's ED and minus the ones admitted to the service unit $n_2 \in N_2$.

$$\alpha_2^{phn_1} = \alpha_1^{phn_1} - v_p^h + w_p^{*h} - w_p^{h*} - x_p^{hn_1n_2} \qquad \begin{aligned} &\forall p \in \{p \in P : n(p) = n_2\} \\ &\forall h \in \{1, \ldots, H\} \\ &\forall n_1 \in \{\text{ED}\} \\ &\forall n_2 \in N_2 \end{aligned}$$

- Hospital's Service Units:

(a) The number of patients of type $p \in P$ with a waiting time equal to 1 in each hospital h of the post-decision state is equal to the number of patients of type $p \in P$ with a waiting time equal to 1 in each hospital h plus the ones transferred from the different units of the hospital with the same DRG g and medical requirement $n \in N_2$.

$$\alpha_2^{phn} = \alpha_1^{phn} + \sum_{u \in \mathbb{U}(p)} x_u^{h*n} \qquad \forall p \in \{p \in P : t(p) = 1 \wedge n(p) = n\}$$
$$\forall h \in \{1, \dots, H\}$$
$$\forall n \in N_2$$

Where $\mathbb{U}(p) = \{u \in P : n(u) = n(p) \wedge g(u) = g(p)\}$

(b) The number of patients of type $p \in P$ with a waiting time higher than 1 in each hospital h of the post-decision state is equal to the number of patients of type $p \in P$ with a waiting time higher than 1 in each hospital h.

$$\alpha_2^{phn} = \alpha_1^{phn} \qquad \forall p \in \{p \in P : 1 < t(p) \wedge n(p) = n\}$$
$$\forall h \in \{1, \dots, H\}$$
$$\forall n \in N_2$$

(c) The number of patients of type $p \in P$ with a waiting time higher than 1 in each hospital h of the post-decision state is equal to the number of patients of type $p \in P$ with a waiting time higher than 1 in each hospital h minus the ones transferred to the other hospital's service unit.

$$\alpha_2^{phn_1} = \alpha_1^{phn_1} - x_p^{hn_1n_2} \qquad \forall p \in \{p \in P : 1 < t(p) \wedge n(p) \neq n_1\}$$
$$\forall h \in \{1, \dots, H\}$$
$$\forall n_2 \in N_2$$

(ii) Capacity constraints: The decision taken must ensure that, after reaching the post-decision state, each unit capacity must be respected.

- Waiting List:

$$\omega_2^* \leq \bar{\omega}$$

- Hospital's Units:

$$\begin{aligned} \alpha_2^{*hn} \leq \Phi_h^n \qquad & \forall h \in \{1, \ldots, H\} \\ & \forall n \in N \end{aligned}$$

It is important to remember that for all initial states $\vec{s_1} \in \mathbb{S}_1$ and all possible actions $\vec{a}$ in $\mathbb{A}(\vec{s_1})$, it holds that:

$$\vec{s_2} := f(\vec{s_1}, \vec{a})$$

Where the Post-Decision State $\vec{s_2}$ is achieved by the result of the implementation of the actions $\vec{a}$ over the Pre-Decision State state $\vec{s_1}$.

4.7. Period Transition

After reaching the post-decision state in period t, the system status must be updated following the process outlined in Figure 3.4, thereby transitioning to a new pre-decision state in period $t + 1$. Given the process's inherent stochasticity, the network's inpatient status and medical requirements (i.e., the service unit that must attend to the patient) could change. Also, new medical requests can arrive at the WL and the hospital's EDs. However, certain transitions are still deterministic. It follows that:

(i) Deterministic Part:

The following constraints ensure that patients in the system add one period to their waiting time status.

- Waiting List:

Patients of type $p \in P$ that end up waiting a period for medical attention in the WL.

$$\omega_1^{u(p)} = \omega_2^p \qquad \forall p \in \{p \in P : 1 < t(p) < T\}$$

$$\omega_1^p = \omega_2^p + \omega_2^{z(p)} \qquad \forall p \in \{p \in P : t(p) = T\}$$

Where $u(p) = \{u \in P : n(u) = n(p) \wedge g(u) = g(p) \wedge t(u) = t(p) + 1\}$
Where $z(p) = \{z \in P : n(z) = n(p) \wedge g(z) = g(p) \wedge t(z) = t(p) - 1\}$

- Hospital's GA:

Patients of type $p \in P$ that end up waiting a period for medical attention at each hospital's GA.

$$\alpha_1^{phn} = 0 \qquad \begin{aligned} &\forall p \in \{p \in P : t(p) = 1\} \\ &\forall h \in \{1, \dots, H\} \\ &\forall n \in \{\text{GA}\} \end{aligned}$$

$$\alpha_1^{u(p)hn} = \alpha_2^{phn} \qquad \begin{aligned} &\forall p \in \{p \in P : 1 < t(p) < T\} \\ &\forall h \in \{1, \dots, H\} \\ &\forall n \in \{\text{GA}\} \end{aligned}$$

$$\alpha_1^{phn} = \alpha_2^{phn} + \alpha_2^{z(p)hn} \qquad \begin{aligned} &\forall p \in \{p \in P : t(p) = T\} \\ &\forall h \in \{1, \dots, H\} \\ &\forall n \in \{\text{GA}\} \end{aligned}$$

Where $u(p) = \{u \in P : n(u) = n(p) \wedge g(u) = g(p) \wedge t(u) = t(p) + 1\}$
Where $z(p) = \{z \in P : n(z) = n(p) \wedge g(z) = g(p) \wedge t(z) = t(p) - 1\}$

- Hospital's ED:

Patients of type $p \in P$ that end up waiting a period for medical attention at each hospital's ED.

$$\alpha_1^{u(p)hn} = \alpha_2^{phn} \qquad \begin{aligned} &\forall p \in \{p \in P : 1 < t(p) < T\} \\ &\forall h \in \{1, \ldots, H\} \\ &\forall n \in \{\text{ED}\} \end{aligned}$$

$$\alpha_1^{phn} = \alpha_2^{phn} + \alpha_2^{z(p)hn} \qquad \begin{aligned} &\forall p \in \{p \in P : t(p) = T\} \\ &\forall h \in \{1, \ldots, H\} \\ &\forall n \in \{\text{ED}\} \end{aligned}$$

Where $u(p) = \{u \in P : n(u) = n(p) \wedge g(u) = g(p) \wedge t(u) = t(p) + 1\}$
Where $z(p) = \{z \in P : n(z) = n(p) \wedge g(z) = g(p) \wedge t(z) = t(p) - 1\}$

- Hospital's Service Units:

Patients of type $p \in P$ that end up waiting a period for medical attention at each hospital's service unit $n \in N_2$, waiting to be transferred to a service unit

$n(p) \neq n$.

$$\alpha_1^{phn} = 0 \qquad \forall p \in \{p \in P : t(p) = 1 \wedge n(p) \neq n\}$$
$$\forall h \in \{1, \ldots, H\}$$
$$\forall n \in N_2$$

$$\alpha_1^{u(p)hn} = \alpha_2^{phn} \qquad \forall p \in \{p \in P : 1 < t(p) < T \wedge n(p) \neq n\}$$
$$\forall h \in \{1, \ldots, H\}$$
$$\forall n \in N_2$$

$$\alpha_1^{phn} = \alpha_2^{phn} + \alpha_2^{z(p)hn} \qquad \forall p \in \{p \in P : t(p) = T \wedge n(p) \neq n\}$$
$$\forall h \in \{1, \ldots, H\}$$
$$\forall n \in N_2$$

Where $u(p) = \{u \in P : n(u) = n(p) \wedge g(u) = g(p) \wedge t(u) = t(p) + 1\}$
Where $z(p) = \{z \in P : n(z) = n(p) \wedge g(z) = g(p) \wedge t(z) = t(p) - 1\}$

(ii) Stochastic Part:

Let us define a new set of patients: $Q = \{q \in Q : N_2 \times DRG\}$. This set represents all the possible combinations of medical requirements and DRGs the system can attend. These types of patients are the ones that arrive at the WL and the hospital's EDs with $t(q) = 1$ (i.e., waiting time equal to 1).

- Waiting List:

Let W^q be the random variable representing arrivals of patients of type q to the WL. Thus:

$$W^q = \omega_1^p \qquad \forall (p,q) \in \{P \times Q : n(p) = n(q) \land g(p) = g(q) \land t(p) = 1\}$$

Each arrival event is independent of the rest, with probability $\mathbb{P}\left(W^q = \omega_1^p\right)$. The previous expression represents the probability of arrivals of patients from type q being ω_1^p, which feeds the new pre-decision state for the next period. For simplicity, the compound probability from all these variables is equal to $\mathbb{P}\left(\vec{W}^q = \vec{\omega_1^p}\right)$

- <u>Hospital's ED:</u>

Let A_h^q be the random variable representing arrivals of patients of type q to the hospital's h ED. Thus:

$$\begin{aligned} A_h^q = \alpha_1^{phn} \qquad & \forall (p,q) \in \{P \times Q : n(p) = n(q) \land g(p) = g(q) \land t(p) = 1\} \\ & \forall h \in \{1, \dots, H\} \\ & \forall n \in \{\text{ED}\} \end{aligned}$$

Each arrival event is independent of the rest, with probability $\mathbb{P}\left(A_h^q = \alpha_1^{phn}\right)$. The previous expression represents the probability of the number of arrivals of patients from type q in hospital h being α_1^{phn}, which feeds the new pre-decision state for the next period. For simplicity, the compound probability from all these variables is equal to $\mathbb{P}\left(\vec{A} = \vec{\alpha_1}\right)$

- <u>Hospital's Unit:</u>

Let C_p^{hnk} be the variable representing the number of patients of type $p \in P$ from hospital h located in unit $n \in N_2$, and transitioning to care need $k \in N_2$ at the end of the period. It is known that a patient served in unit n can either maintain their care needs ($n = k$) or change them ($n \neq k$).

By construction, it holds that:

$$\alpha_2^{phn} = C_p^{hn*} \qquad \forall p \in \{p \in P : n(p) = n\}$$
$$\forall h \in \{1, \ldots, H\}$$
$$\forall n \in N_2$$

We aim to study the probability of transitioning to:

(a) Patients that remain being treated at service unit n:

$$\alpha_1^{u(p)hn} = C_p^{hnn} \qquad \forall p \in \{p \in P : n(p) = n \wedge 1 < t(p) < T\}$$
$$\forall h \in \{1, \ldots, H\}$$
$$\forall n \in N_2$$

$$\alpha_1^{phn} = C_p^{hnn} + C_{z(p)}^{hnn} \qquad \forall p \in \{p \in P : n(p) = n \wedge t(p) = T\}$$
$$\forall h \in \{1, \ldots, H\}$$
$$\forall n \in N_2$$

Where $u(p) = \{u \in P : n(u) = n(p) \wedge g(u) = g(p) \wedge t(u) = t(p) + 1\}$

Where $z(p) = \{z \in P : n(z) = n(p) \wedge g(z) = g(p) \wedge t(z) = t(p) - 1\}$

(b) Patients that update their treatment requirements to service unit $k \neq n$:

$$\alpha_1^{phn} = \sum_{u \in \mathbb{U}(p)} C_u^{hnk} \qquad \forall p \in \{p \in P : n(p) \neq n \wedge n(p) = k \wedge t(p) = 1\}$$
$$\forall h \in \{1, \ldots, H\}$$
$$\forall k \in N_2$$

Where $\mathbb{U}(p) = \{u \in P : n(u) = n \wedge g(u) = g(p)\}$

Without loss of generality, for any $t(p)$, we want to study the random variable:

$$\mathbb{P}\left(C_p^{hn1} = c_1, \ldots, C_p^{hnN_2} = c_{N_2} \mid \alpha_2^{phn} = C_p^{hn*}\right)$$

It is important to remark that $0 \leq c_k \leq \alpha \quad \forall k \in N_2$. It can be shown that this expression is equivalent to:

$$\mathbb{P}\left(C_p^{hn2} = c_2, \ldots, C_p^{hnN_2} = c_{N_2} \mid C_p^{hn1} = c_1,\ \alpha_2^{phn} - c_1 = \sum_{k=2}^{N_2} C_p^{hnk}\right) \mathbb{P}\left(C_p^{hn1} = c_1\right)$$

Given that $C_p^{hn1} = c_1$, it is important to remark that $0 \leq c_k \leq \alpha_2^{phn} - c_1 \quad \forall k \in N_2 \setminus \{1\}$. Now, it can be shown that this expression is equivalent to:

$$\mathbb{P}\left(C_p^{hn3} = c_3, \ldots, C_p^{hnN_2} = c_{N_2} \mid C_p^{hn2} = c_2,\ C_p^{hn1} = c_1,\ \alpha_2^{phn} - c_1 - c_2 = \sum_{k=3}^{N_2} C_p^{hnk}\right)$$
$$\cdot\, \mathbb{P}\left(C_p^{hn2} = c_2 \mid C_p^{hn1} = c_1\right) \cdot \mathbb{P}\left(C_p^{hn1} = c_1\right)$$

Now, given that $C_p^{hn1} = c_1$ and $C_p^{hn2} = c_2$, it is important to remark that $0 \leq c_k \leq \alpha_2^{phn} - c_1 - c_2 \quad \forall k \in N_2 \setminus \{1, 2\}$. Following this logic, we arrive at the following expression:

$$\prod_{k \in N_2} \mathbb{P}\left(C_p^{hnk} = c_k \mid C_p^{hnk-1} = c_{k-1}, \ldots, C_p^{hn1} = c_1\right)$$

Let $\delta_p^{hnk} \geq 0$ be the probability that a patient of type $p \in P$ from hospital h located in unit n, transitions to care needs of type $k \in N_2$ at the end of the period. It can be shown that:

$$\mathbb{P}\left(C_p^{hnk} = c_k \mid C_p^{hnk-1} = c_{k-1}, \ldots, C_p^{hn1} = c_1\right) =$$
$$\mathbb{P}\left(\text{Binomial}\left(n = \alpha_2^{phn} - \sum_{u=1}^{k-1} c_u,\ p = \frac{\delta_p^{hnk}}{\sum_{u=k}^{N_2} \delta_p^{hnu}}\right) = c_k\right)$$

Note that it holds:

$$\begin{aligned} \delta_p^{hn*} = 1 \qquad & \forall p \in \{p \in P : n(p) = n\} \\ & \forall h \in \{1, \ldots, H\} \\ & \forall n \in N_2 \end{aligned}$$

Thus, it holds:

$$\mathbb{P}\left(\vec{\mathbb{C}} = \vec{C} \mid \vec{\alpha_2}\right) = \prod_{h=1}^{H} \prod_{n \in N_2} \prod_{p \in P} \left\{ \prod_{k \in N_2} \mathbb{P}\left(C_p^{hnk} = c_k \mid C_p^{hnk-1} = c_{k-1}, \ldots, C_p^{hn1} = c_1\right) \right\}$$

Once we determine $\vec{C}$, it is easy to determine the new state $\vec{1}$. Additionally, in the case of $n =$ WARD, it must be considered that a possible transition is the end of patient care, which ultimately frees up hospital resources.

Based on the previously developed probabilities, it can be concluded that the probability of transition from the post-decision state $\vec{s_2}$ to a pre-decision state $\vec{u} \in \mathbb{S}_1$ is:

$$p(\vec{u}|\vec{s_2}) = \mathbb{P}\left(\vec{W^q} = \vec{\omega_1^p}\right) \cdot \mathbb{P}\left(\vec{A} = \vec{\alpha_1}\right) \cdot \mathbb{P}\left(\vec{\mathbb{C}} = \vec{C} \mid \vec{\alpha_2}\right)$$

4.8. Cost-to-Go

Finally, the cost function is formulated. It can be divided between the pre-decision state (operational costs) and the post-decision state (opportunity costs). Also, the decision-maker could modify the value μ, which we will consider equal to 1, for the sake of this study.

$$c(\vec{s_1}, \vec{a}) = c_1(\vec{s_1}, \vec{a}) + \mu \cdot c_2(\vec{s_2}) \quad \forall \vec{s_1} \in \mathbb{S},\ \forall \vec{a} \in \mathbb{X}(\vec{s_1})$$

The cost at the beginning of the period considers the diversions and transfer costs, which are defined as:

$$c_1(\vec{s_1}, \vec{a}) = \eta_* \rho_* + v_*^* \theta_*^* + w_*^{**} \vartheta_*^{**}$$

The first term represents the sum of diversion costs from the WL, the second term represents the sum of diversion costs from the hospital's EDs, and the last term represents the sum of transfer costs between the hospital's EDs.

The cost at the end of the period considers the waiting times and bed opportunity costs, which are defined as:

$$c_2(\vec{s_2}) = \omega_2^* \nu_* + \alpha_2^{***} \phi_*^{**}$$

The first term represents the sum of waiting time costs from patients at the WL, while the second one considers the waiting time costs for patients at the different hospital's units, as well as the bed opportunity costs from the beds that are being 'misused' by these patients waiting.

5. SOLUTION APPROACH

The following chapter outlines the methodologies implemented to optimize the problem. We start with the Q-learning and Reinforcement-Learning approach, focusing on the simulation and optimization techniques implemented. Additionally, different Q-learning enhancements are described, as well as the mathematical formulation behind the Machine Learning model.

5.1. The $\mathbb{Q}$ Function

Given the complexity of the problem, which involves a vast number of possible system states resulting in a high degree of complexity in both the states and actions dimensions, as well as the numerous stochastic transitions, it is necessary to solve this problem using an approximate method (Powell, 2011, 2012). To achieve this, first of all, we must make some changes to the classic Bellman Equations. Specifically:

$$C(\vec{s_1}) = \min_{\vec{a} \in \mathbb{X}(\vec{s_1})} \left\{ c(\vec{s_1}, \vec{a}) + \lambda \sum_{\vec{u} \in \mathbb{S}_1} C(\vec{u}) p\left(\vec{u} | \vec{s_2}\right) \right\} \quad \forall \vec{s_1} \in \mathbb{S}_1$$

We define:

$$\sum_{\vec{u} \in \mathbb{S}_1} c(\vec{u}) p(\vec{u} | \vec{s_2}) = \mathbb{E}_{\vec{u}} \left[c(\vec{u}) \mid \vec{s_2} \right] = \mathbb{Q}\left(\vec{s_2}\right) \quad \forall \vec{s_2} \in \mathbb{S}_2$$

Where $\mathbb{Q}\left(\vec{s_2}\right)$ is a function that represents the expected value of the post-decision state $\vec{s_2}$. The Bellman Equations are reduced to:

$$C(\vec{s_1}) = \min_{\vec{a} \in \mathbb{X}(\vec{s_1})} \left\{ c(\vec{s_1}, \vec{a}) + \lambda \mathbb{Q}\left(\vec{s_2}\right) \right\} \quad \forall \vec{s_1} \in \mathbb{S}_1$$

Now, the main task is to determine a good approximation of $\mathbb{Q}$. There are several possible techniques, such as a Miopic Adjustment, Rolling Horizon, Lookahead, and Parameterization methods, among others. In this research, we will focus on the Q-learning approach, which is a straightforward method for agents to acquire optimal behavior in regulated Markovian environments (Watkins & Dayan, 1992). The Q-learning approach will be coupled with a Reinforcement-Learning methodology, which involves acquiring effective strategies for sequential decision challenges by optimizing a cumulative future reward signal (Sutton & Barto, 2018), which is expected to be represented by the function $\mathbb{Q}$.

5.2. General Overview of the Q-learning and Reinforcement-Learning Approach

The Q-learning and Reinforcement-Learning strategies can be summarized in the following figure:

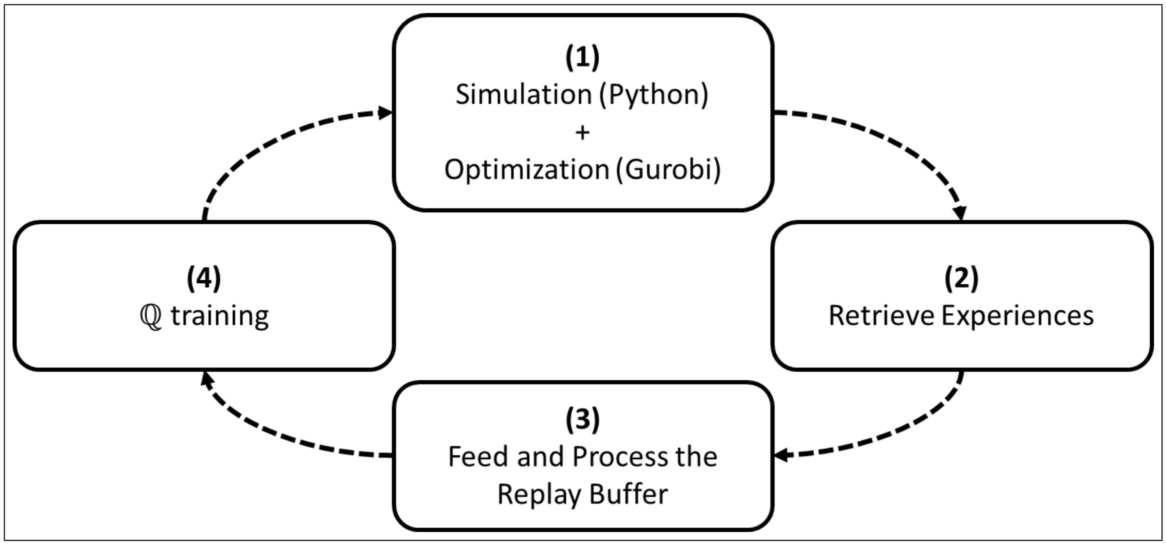

Figure 5.1. General Diagram of the Q-learning and Reinforcement-Learning Approach

Stages (1) and (2) consider a Simulation and Optimization approach, where the agent samples experiences that later can be used to learn a better policy. In this study, the simulation is programmed in Python 3.10.11, while the optimization problem is programmed in Gurobi 10.0.2. Steps (3) and (4) are related to the learning process based on the experiences. Each of these components, as well as their details, will be explained in the next section. Although one can iterate indefinitely, a stopping criteria is needed. In our case, we used two convergence criteria; the first one is based on the new experiences values, while the second one measured the $\mathbb{Q}$ values that were predicted and their changes over time. When both measures achieved a steady state, the training was ended.

It is worth noting that this approach involves offline training to minimize the computational resources required for real-time processing and problem-solving. Only the first stage is utilized during the actual implementation of the model in a real-world scenario. However, all relevant data must be collected for future retraining of the $\mathbb{Q}$ model.

5.3. Useful Methodologies

5.3.1. Stage 1: Simulation and Optimization

To correctly understand the solution approach, first of all, it is necessary to define certain elements. We can solve our patient management problem using a discrete-event simulation approach. This method allows the system to evolve dynamically between time frames, and we can make optimal decisions at each period to get the best results. By applying this approach, we can obtain specific elements at each time frame t:

- An initial pre-decision state $\vec{s_1} \in \mathbb{S}_1$
- An action $\vec{a} \in \mathbb{X}(\vec{s_1})$ taken by the current policy π .
- A cost $c(\vec{s_1}, \vec{a})$ incurred.
- An final post-decision state $\vec{s_2} \in \mathbb{S}_2$, given the pre-decision state and the action taken.

Figure 5.2 represents this framework, where an agent (i.e., the decision-maker) chooses an action, given an initial state $\vec{s_1}$. This decision is supported by an optimization model, optimized within the policy π step.

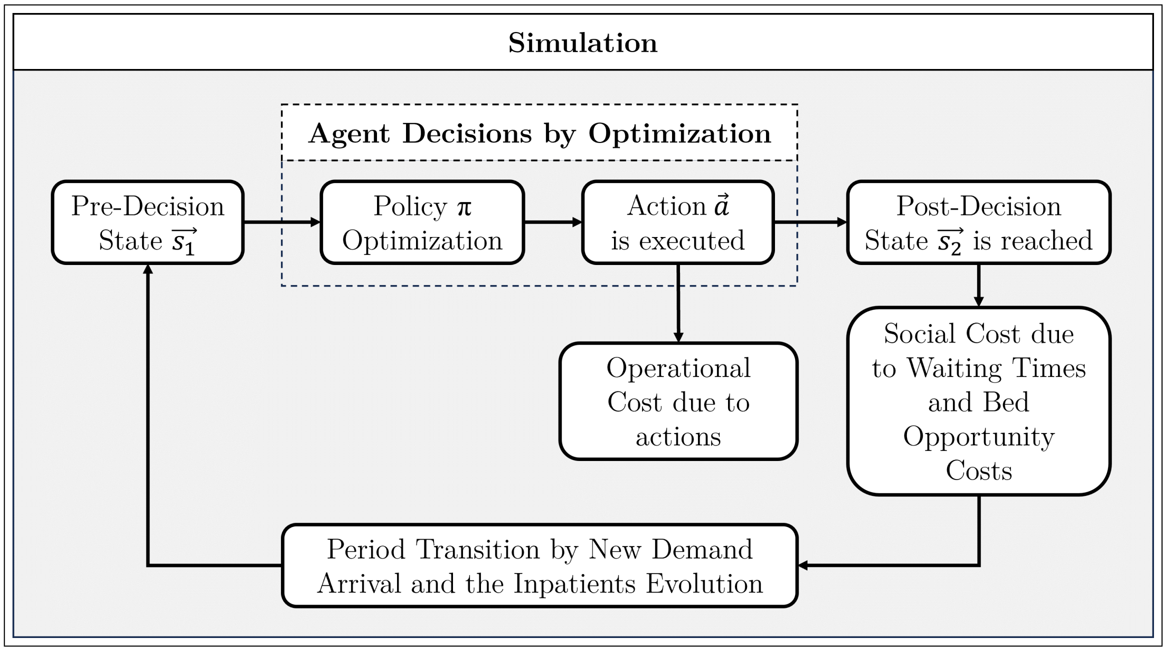
Simulation
Agent Decisions by Optimization
Pre-Decision State $\vec{s_1}$
Policy π Optimization
Action $\vec{a}$ is executed
Post-Decision State $\vec{s_2}$ is reached
Operational Cost due to actions
Social Cost due to Waiting Times and Bed Opportunity Costs
Period Transition by New Demand Arrival and the Inpatients Evolution

Figure 5.2. Diagram of the Simulation and Optimization Approach

From now on, and to keep the important elements, we will register at each time step t the tuple (S_t, C_t), where $S_t = \vec{s_2}$ and $C_t = c(\vec{s_1}, \vec{a})$.

5.3.2. Stages 2 and 3: Replay Buffer

Based on Lin (1992), a Replay Buffer is a set of previously observed transitions that are stored for a certain amount of time. These stored transitions are used multiple times; thus, the learning agent recalls its previous encounters and iteratively feeds these experiences to its learning algorithm, simulating the agent's re-experience of past occurrences. Eventually, as the Replay Buffer fills its capacity, the older experiences are changed by new ones. In our case, given a simulation run i with J time frames sampled, its Replay Buffer $\mathcal{R}_i$ is represented as:

$$\mathcal{R}_i = \left\{(S_1^i, C_1^i), \ldots, (S_t^i, C_t^i), \ldots, (S_J^i, C_J^i)\right\}$$

Now, given I simulations runs, the final Replay Buffer $\mathcal{R}$ is defined as:

$$\mathcal{R} = \bigcup_{i=1}^{I} \mathcal{R}_i$$

In our case, each simulation run sampled 1000 time frames in the stationary phase, with a Replay Buffer size of 120 simulation runs or, in other terms, 120,000 total experiences. We sampled 3,000 experiences between learning phases.

5.3.3. Stage 4: Machine Learning Approach and Enhancements

5.3.3.1. Justification

In this study, the function $\mathbb{Q}$ will be acquired through a machine learning approach, which later can be embedded in the optimization model. One of the reasons behind the selection of this approach is the possibility of modeling non-linear relations, which González et al. (2018) and Marquinez et al. (2021) propose as an improvement, compared to the Approximate Linear Optimization Approach used by both of the authors. In our research, we implemented a traditional solution approach using a multi-step Lookahead, similar to the Sample Average Approximation (SAA) method. However, we faced two major issues during this process. Firstly, we encountered endogeneity, which refers to the situation where decisions made in the present affect future probabilities. For instance, transferring a patient to a different hospital can impact the future possibilities regarding the availability of beds at that hospital. Secondly, we faced an explosive increase in the size of the problem due to scenario sampling, which led to intractable execution times for the case study.

There has been increasing interest in recent years to incorporate machine learning models in optimization models. This has resulted in numerous publications. A notable example is the research conducted by Maragno et al. (2023) titled *Mixed-Integer Optimization with Constraint Learning*, where the authors suggest a comprehensive process for data-informed decision-making, where the constraints and objectives are derived directly from the data through machine learning. The resulting models are then integrated into an optimization framework, testing it in a diet planning application and a cancer treatment optimization problem.

This has also resulted in the development of research software packages. In our case, we pay attention to the Gurobi Machine Learning package. As it is stated by Gurobi Optimization (2023), it consists in:

> A Python package to help use trained regression models in mathematical optimization models. The package supports a variety of regression models (linear, logistic, neural networks, decision trees,...) trained by different machine learning frameworks (scikit-learn, Keras and PyTorch).

5.3.3.2. Base Guidelines

$\mathbb{Q}$ must be considered the output from a regression model, where its prediction depends on the post-decision state S_t, but also on a set of parameters θ. Initially, based on Van Hasselt et al. (2016), the selected machine learning model for this research was a Deep Neural Network with ReLU activation, aiming to implement their Deep Reinforcement Learning approach. With that in mind, briefly, it is proposed that, given an iteration k, where the neural network parameters are θ_k, we can define:

$$\mathbb{Q}(S_t) \approx Y_t^i = C_{t+1}^i + \lambda \mathbb{Q}(S_{t+1}^i, \theta_k)$$

Where Y_t^i represents the new estimation of the expected value for the post-decision state S_t. The new parameters θ_{k+1} are obtained minimizing the loss function:

$$\sum_{i=1}^{I} \sum_{j=1}^{J-1} \left(Y_t^i - \mathbb{Q}(S_t^i, \theta_{k+1}) \right)^2$$

5.3.3.3. Challenges and Extensions

When optimizing the model, we encountered errors due to the linearization of ReLU expressions in the neural network embedded in the optimization model. This led to errors in the regressor's predictions, undermining $\mathbb{Q}$ value prediction.

To get over this problem, a new approach was proposed. Lookup Tables (Powell, 2011) and, later on, more sophisticated approaches, such as the Dynamic Lookup Tables (Ulmer, 2017), have proven useful in solving MDPs with the courses of dimensionality. Generally speaking, the idea behind this approach is to partition the post-decision vector space into subsets with similar $\mathbb{Q}$ values. In this case, a balance between broad and detailed divisions is desired in the post-decision space. Opting for a broad division may lead to a quick yet inadequate approximation, as it treats diverse states equally during evaluation. On the other hand, an overly detailed partitioning might lead to limited observations of states, consequently causing an inefficient approximation process by the overfitting produced by the scarce observations by partitioning.

5.3.3.4. Lookup Tables by Decision Trees

Although there are different methods to achieve good partitioning, decision tree classifier regressors and their variants caught our attention. Decision trees are a way of dividing observations into distinct leaves through a series of feature splits (Maragno et al., 2023). As stated by the authors, a split node makes a partition by the inequality $Ax \leq b$; on the other hand, a terminal node (i.e., leaf node), gives us a prediction p. Figure 5.3 represents an example of a generic decision tree.

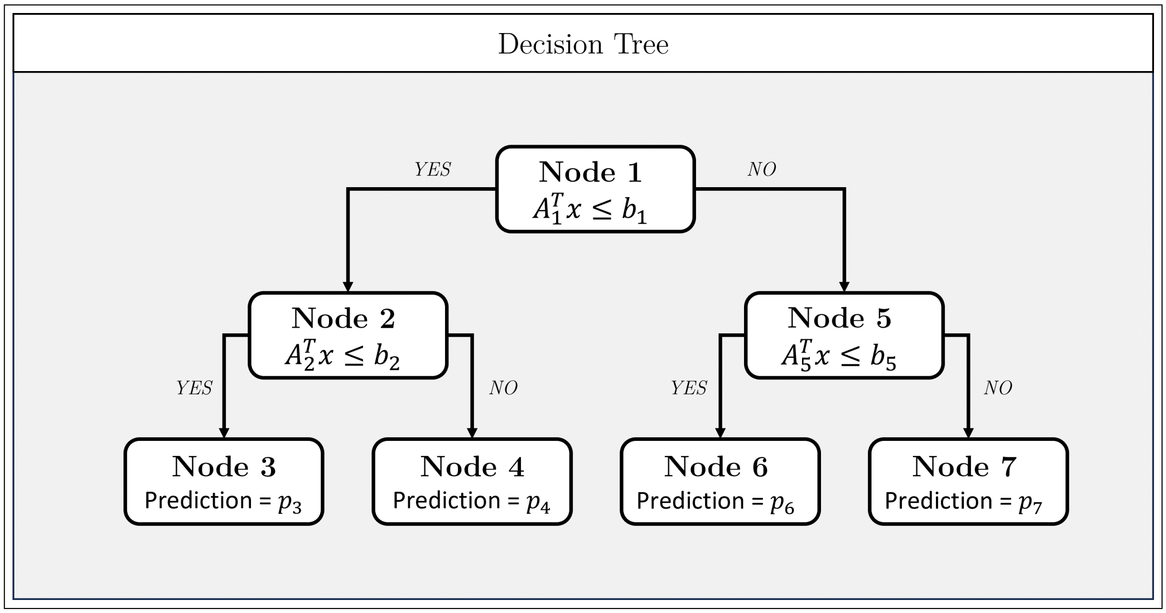

Figure 5.3. Diagram of a Generic Decision Tree

Every leaf can be characterized as a polyhedron, specifically a collection of linear constraints that all leaf members must adhere to (Maragno et al., 2023), thus resulting in the partitions that a Lookup Table aims to achieve. For instance, the partition achieved by Node 6 is represented as $\mathcal{P}_6 := \{x : A_1^T x > b_1, A_5^T x \leq b_5\}$.

Regarding the mathematical formulation of this regressor, we will focus on the proposals by Bertsimas and Dunn (2017) and Maragno et al. (2023). Given a decision tree with $\mathcal{L}$ leaf nodes and $\mathcal{S}$ split nodes, let us define l_i as the binary variable equal to 1 if x belongs to partition $\mathcal{P}_i$ and p_i its associated prediction.

Let us call $\mathcal{S}^i \subset \mathcal{S}$ the splits that observations in leaf i must 'visit'. Given a leaf node i and a split node $j \in \mathcal{S}^i$, if i follows the left split from j, we can ensure that $A_j^T x \leq b_j$. Otherwise, we can ensure that $A_j^T x > b_j$, which is equivalent to $-A_j^T x < -b_j$. We can remove the strict inequalities using a sufficiently small ϵ parameter, giving us the expression $-A_j^T x \geq -b_j - \epsilon$.

Additionally, let us define $\overline{A}_j$, which is equal to A_j if leaf i follows the left split of a node j (i.e., $A_j^T x \leq b_j$) and $-A_j$, otherwise. Similarly, let us define $\overline{b}_j$, which equals b_j if leaf i follows the left split of a node j and $-b_j - \epsilon$, otherwise. We can represent the decision tree and its prediction y for x with the following equations:

$$\overline{A}_j^T x - M(1 - l_i) \leq \overline{b}_j \qquad \forall\, i \in \mathcal{L}, \quad \forall\, j \in \mathcal{S}^i$$

$$\sum_{i \in \mathcal{L}} l_i = 1$$

$$y = \sum_{i \in \mathcal{L}} p_i l_i$$

Where M is a sufficiently large number. For a better understanding of the previous equations, the generic regression tree presented in Figure 5.3 can be mathematically formulated as follows:

$$\begin{aligned}
A_1^\top \boldsymbol{x} - M\,(1 - l_3) &\leq b_1, \\
A_2^\top \boldsymbol{x} - M\,(1 - l_3) &\leq b_2, \\
A_1^\top \boldsymbol{x} - M\,(1 - l_4) &\leq b_1, \\
-A_2^\top \boldsymbol{x} - M\,(1 - l_4) &\leq -b_2 - \epsilon, \\
-A_1^\top \boldsymbol{x} - M\,(1 - l_6) &\leq -b_1 - \epsilon, \\
A_5^\top \boldsymbol{x} - M\,(1 - l_6) &\leq b_5, \\
-A_1^\top \boldsymbol{x} - M\,(1 - l_7) &\leq -b_1 - \epsilon, \\
-A_5^\top \boldsymbol{x} - M\,(1 - l_7) &\leq -b_5 - \epsilon, \\
l_3 + l_4 + l_6 + l_7 &= 1, \\
y - (p_3 l_3 + p_4 l_4 + p_6 l_6 + p_7 l_7) &= 0
\end{aligned}$$

Random Forest arises as a typical option among the tree-based models, which ensemble a collection of tree regressors (i.e., a forest). This regression method consists of a collection of tree predictors, where each tree relies on the values of a random feature vector that is independently sampled and follows the same distribution for all trees within the forest (Breiman, 2001). After the model is trained, predictions for unseen samples are made, averaging the predicted value for each regression tree as follows:

$$y = \frac{1}{N}\sum_{i=1}^{N} p_i$$

Recently, new tree-based models have been developed, among them, the XGBoost model stands out for its good results. Briefly, it is a tree-boosting technique, widely adopted and highly effective in the field of machine learning, being remarkable for its scalability in all kinds of scenarios (Chen & Guestrin, 2016). Given these properties, as well as its versatility, this model is selected. In terms of the model hyperparameters, they are optimized using Optuna, an automated hyperparameter optimization framework (Akiba et al., 2019). In contrast to Random Forest, a calibrated and trained XGBoost model output y is computed as:

$$y = \sum_{i=1}^{N} \beta_i y_i$$

Where y_i is the predicted value of the i-th tree regression model and β_i is the weight associated with the prediction. Given the trained model with N trees, the approximated

MDP model that we are solving is:

$$
\begin{aligned}
C(\vec{s_1}) &= \min_{\vec{a}\in\mathbb{X}(\vec{s_1})} \{c(\vec{s_1},\vec{a}) + \lambda\mathbb{Q}\left(\vec{s_2}\right)\} && \forall \vec{s_1}\in\mathbb{S}_1 \\
\mathbb{Q}\left(\vec{s_2}\right) &= \sum_{n=1}^{N} \beta_n y_n && (5.1)\\
y_n &= \sum_{i\in\mathcal{L}^n} p_i^n l_i^n && \forall n\in\{1,\cdots,N\} \quad (5.2)\\
(\overline{A_j^n})^T\vec{s_2} - M^n(1-l_i^n) &\le \overline{b_j^n} && \forall\, i\in\mathcal{L}^n, \quad \forall\, j\in\mathcal{S}_n^i, \\
& && \forall n\in\{1,\cdots,N\} \quad (5.3)\\
\sum_{i\in\mathcal{L}^n} l_i^n &= 1 && \forall n\in\{1,\cdots,N\} \quad (5.4)\\
y_n &= \sum_{i\in\mathcal{L}^n} p_i^n l_i^n && \forall n\in\{1,\cdots,N\} \quad (5.5)
\end{aligned}
$$

Constraint (6.1) calculates the $\mathbb{Q}\left(\vec{s_2}\right)$, given the decision tree models. Constraints (6.2), (6.3), (6.4) and (6.5) calculates the predicted value for each tree $n\in\{1,\cdots,N\}$ from the XGBoost model.

The two most significant benefits of the Lookup table represented via an XGBoost model for the $\mathbb{Q}$ function approximation, compared to the neural network approach, are: firstly, the proposed model has a mathematical formulation that is compatible with the optimization problem, preserving the core structure; secondly, the errors between the regressor and the mathematical formulation associated with it are, on average, lower than 10^{-4}%, which is acceptable, compared to the higher values that we obtained with the neural network approach.

5.3.3.5. Multi-step Learning

Classic Q-learning considers for each learning step the cost for the next period exclusively, resulting in the approximation:

$$Y_t^i = C_{t+1}^i + \lambda\mathbb{Q}(S_{t+1}^i,\theta_k)$$

Sutton (1988) proposed a multi-step learning approach, which can lead to a faster convergence if correctly calibrated. For instance, let us consider n future periods, the truncated n-step return is defined as:

$$Y_t^i(n) = \sum_{u=0}^{n-1} C_{t+1+u}^i + \lambda \mathbb{Q}(S_{t+n}^i, \theta_k)$$

Also, the loss function is changed, resulting in:

$$\sum_{i=1}^{I} \sum_{j=1}^{J-n} \left(Y_t^i(n) - \mathbb{Q}(S_t^i, \theta_{k+1})\right)^2$$

5.3.3.6. Decoupling the Q-learning Process

Although the classic Q-learning approach had achieved significant results, Van Hasselt et al. (2016) mentions that it can suffer from important biases. This is produced because, at each training iteration, we consider:

$$\mathbb{Q}(S_t^i) \approx C_t^i + \lambda \mathbb{Q}(S_{t+1}^i)$$

Thus, the $\mathbb{Q}$ function constantly feeds back on itself, resulting in subestimated values, given that we are minimizing. We propose an approach that decouples these values to avoid this undesired situation, reducing the Q-learning process's biases. To achieve this, we will learn multiple value functions; one will be the Target Model, with parameters θ_k, as well as the Critic Models, with parameters θ_k^r. The Target Model will be embedded in the optimization problem, estimating the $\mathbb{Q}$ function, while the Critic Models will be used to reduce biases.

Given an iteration k with a Replay Buffer $\mathcal{R}$, let us consider R critic models. Also, let us consider a truncated n-step return for the critic models, which is defined as:

$$Y_{C_t}^i = \sum_{u=0}^{n-1} C_{t+1+u}^i + \lambda \max_{r=1,\ldots,R} \left\{\mathbb{Q}(S_{t+n}^i, \theta_k^r)\right\}$$

The $\max$ operator makes it more likely to select overestimated values, which aims to counter the subestimations that the classic approach does (Van Hasselt et al., 2016). On the other hand, for the Target Model, let us consider the classic approach, where:

$$Y_{T_t}^{\,i} = C_{t+1}^i + \lambda \mathbb{Q}(S_{t+1}^i, \theta_k)$$

As a simulation run i becomes older, the experiences contained in the replay buffer $\mathcal{R}_i \subset \mathcal{R}$ become less ideal for good policy learning, as old experiences tend to be worse, thus needing to give the associated n-step returns less importance as they get older. This implies that the Y_t^i values to train the new value functions must be obtained using a weighted average, where the weight, defined as α_i for the term $Y_{C_t}^{\,i}$, decays over time, resulting in the expression:

$$Y_t^i = (1 - \alpha_i) Y_{T_t}^{\,i} + (\alpha_i) Y_{C_t}^{\,i}$$

Which can be extended as:

$$Y_t^i = (1 - \alpha_i) \left[C_{t+1}^i + \lambda \mathbb{Q}(S_{t+1}^i, \theta_k) \right] + (\alpha_i) \left[\sum_{u=0}^{n-1} C_{t+1+u}^i + \lambda \max_{r=1,\ldots,R} \left\{ \mathbb{Q}(S_{t+n}^i, \theta_k^r) \right\} \right]$$

The weight α_i can have different functional forms. It is important that, as the sample gets older, this weight tends to zero. For instance, and after some calibration experiments, given a simulation run i sampled u iterations ago, a good value for α_i is defined as:

$$\alpha_i = 0.99^u$$

After obtaining the Y set of values from $\mathcal{R}$, with their respective S states, they are split into the training set and the test set. The training set is divided into a random split for training data and validation data, independently for each Critic Model. The training data will be used to learn from the experiences and fit the model, while the validation data will be used to evaluate the model learning and avoid overfitting. On the other hand, the test set will be used as an unbiased evaluation of each Critic Model fit, where the one with the lowest error will pass its parameters to the target model. In our study, the metric implemented is the Mean Squared Error. Also, it is important to remark that all these splits must ensure that the partitions are correctly balanced for an optimal learning process.

The whole training process is presented in Algorithm 1.

Algorithm 1 Q-learning

```
Require:
    $\mathcal{R}$: Replay Buffer
    $\theta_k$: Target Model Hyper-parameters
    $\theta_k^r$: Critic Models Hyper-parameters

Ensure:
    $\theta_{k+1}$: New Target Model Hyper-parameters
    $\theta_{k+1}^r$: New Critic Models Hyper-parameters

    for $(i, t) \in \{(i, t) : i \in I, t \in J, t \leq J - n\}$ do        ▷ Get new $\mathbb{Q}$ values
        ${Y_C}_t^i = \sum_{u=0}^{n-1} C_{t+1+u}^i + \lambda \max_{r=1,\dots,R} \left\{\mathbb{Q}(S_{t+n}^i, \theta_k^r)\right\}$
        ${Y_T}_t^i = C_{t+1}^i + \lambda \mathbb{Q}(S_{t+1}^i, \theta_k)$
        $Y_t^i = (1 - \alpha_i){Y_T}_t^i + (\alpha_i){Y_C}_t^i$
    end for

    $S_{test},\ Y_{test},\ S_{training},\ Y_{training} \leftarrow S,\ Y$        ▷ Get the test and training split

    for $r \in R$ do        ▷ Train each Critic Model
        $S_{train},\ Y_{train},\ S_{validation},\ Y_{validation} \leftarrow S_{training},\ Y_{training}$
        $\theta_{k+1}^r \leftarrow model\_fit\,(S_{train},\ Y_{train},\ S_{validation},\ Y_{validation})$
    end for

    $\theta_{k+1} \leftarrow argmin\ _{r=1,\dots,R} \left\{MSE\left(\theta_{k+1}^r, S_{test},\ Y_{test}\right)\right\}$ ▷ Evaluate each Critic Model and obtain the Target Model
```

5.3.3.7. ϵ-greedy Exploration

In the basic reinforcement learning methodology, the agent optimizes the objective function to decide which decisions must be taken:

$$C(\vec{s_1}) = \min_{\vec{a} \in \mathbb{X}(\vec{s_1})} \{c(\vec{s_1}, \vec{a}) + \lambda \mathbb{Q}\left(\vec{s_2}\right)\} \quad \forall \vec{s_1} \in \mathbb{S}_1$$

The limitation of this approach is that the agent exclusively exploits the problem domain, with a high probability of ending in a sub-optimal policy. To improve the policy, the agent can act ϵ-greedily in terms of the optimization problem. This means assuming that:

$$\mathbb{Q}\left(\vec{s_2}\right) = 0 \quad \forall \vec{s_2} \in \mathbb{S}_2$$

This ends up in a situation where the agent decides with the lowest immediate value (thus, the greedy action). To balance exploration and exploration, the greedy action is taken with a probability ϵ, whereas the proactive action (where $\mathbb{Q}\left(\vec{s_2}\right) \neq 0$) is taken with probability $(1 - \epsilon)$. The idea behind this strategy is to introduce exploration by randomly selecting unnecessarily optimal actions, which can lead the agent to correct its current estimates, as well as the possibility of discovering new policies that it would not have considered before. In the case of this study, we considered a constant value of $\epsilon = 0.01$ during the training process.

6. EXPERIMENTS AND RESULTS

In this section, we present the main results for the solution approach proposed for obtaining an approximate optimal patient diversion, admission, and transfer policy. It is important to note that, although trained and tested on a specific instance, the methodology should adapt to any scenario without losing generality, thanks to the flexibility of the machine learning approach. Additionally, we will focus on the behavior and improvements of the obtained policy regarding costs, occupation, and waiting times, reviewing the actions being taken in general terms. The latter is because it is not feasible to determine the optimal action in every possible state.

6.1. Test Instance

The following test instance was generated based on the different research studies presented in Chapter 2: Literature Review and Chapter 4: Model Formulation.

6.1.1. Network Capacity

As mentioned earlier, the system has a limited number of beds available per unit. This resource is crucial for attending to patient requests, and as the system gets crowded, the number of available beds becomes scarce. The number of beds per unit is based on the research of Marquinez et al. (2021) and González et al. (2018). For our test instance, we have three hospitals with identical capacity per unit, which is not an accurate representation of real-life scenarios where each hospital has a different bed capacity per unit. However, this approach will help us interpret and comprehend the results while avoiding the phenomenon where a high-capacity hospital absorbs all the demand like a sink. The amount of beds by hospital and unit is presented in Table 6.1.

Table 6.1. Number of Beds by Hospital and Unit

	Hospital 1	Hospital 2	Hospital 3
ED	15	15	15
GA	15	15	15
OR	6	6	6
ICU	60	60	60
SDU/WARD	165	165	165

For simplicity, we will merge the SDU and WARD units and focus on analyzing the more complex units present at a hospital.

6.1.2. Demand

In terms of demand, we will consider 8 different DRGs, although in reality, there are more than 300 groups in the original classification system proposed by Fetter et al. (1980). The first 4 DRG are exclusively urgent requirements, meaning that patients with a DRG from these groups only arrive at the hospital's ED. On the other hand, the remaining DRGs are exclusively elective, meaning that patients with a DRG from these groups only arrive at the Network WL. Even though, in reality, this is not necessarily true, we made this on purpose to better analyze and interpret the results obtained. Currently, the healthcare system is experiencing high demand which is leading to reactive referrals across all hospitals and an increase in waiting lists. To address this issue, it is important to implement proactive diversions as suggested by Marquinez et al. (2021). This will help to decongest the system and maintain an optimal occupancy level, resulting in better service for patients.

6.1.2.1. Demand Composition

The emergency demand (DRG 1 to DRG 4), represents 49.69% of the total demand, while the elective demand (DRG 5 to DRG 8) sums up to the 50.31% remaining. This proportion aims to balance the demand between emergency requests at EDs and requests of the WL, which can be postponed in the face of emergencies but must be satisfied. The average arrivals by DRG per period are presented in table 6.2

Table 6.2. Average Arrivals per Period by DRG

DRG	Mean Arrival Rate [patients/period]	Percentage of Total Demand [%]
1	4.11	13.16%
2	3.92	12.54%
3	4.18	13.36%
4	3.32	10.63%
5	5.10	16.33%
6	3.93	12.57%
7	4.33	13.86%
8	2.36	7.55%
GENERAL	**31.25**	**100%**

The average arrivals by DRG per period in the network are presented in table 6.3. Additionally, table 6.4 shows the average relative arrivals by DRG per period in the Network.

Table 6.3. Average Arrivals per Period in the Network

	Mean Arrival Rate [patients/period]			
DRG	**WL**	**HOSPITAL 1**	**HOSPITAL 2**	**HOSPITAL 3**
1	0.00	2.01	1.08	1.02
2	0.00	1.43	1.71	0.78
3	0.00	1.51	1.38	1.29
4	0.00	1.18	1.18	0.96
5	5.10	0.00	0.00	0.00
6	3.93	0.00	0.00	0.00
7	4.33	0.00	0.00	0.00
8	2.36	0.00	0.00	0.00
GENERAL	**15.72**	**6.12**	**5.35**	**4.06**

Table 6.4. Average Relative Arrivals by DRG per Period in the Network

	Relative Mean Arrival Rate [%]			
DRG	**WL**	**HOSPITAL 1**	**HOSPITAL 2**	**HOSPITAL 3**
1	0.00%	48.81%	26.31%	24.88%
2	0.00%	36.42%	43.64%	19.94%
3	0.00%	36.08%	32.94%	30.97%
4	0.00%	35.53%	35.65%	28.82%
5	100.00%	0.00%	0.00%	0.00%
6	100.00%	0.00%	0.00%	0.00%
7	100.00%	0.00%	0.00%	0.00%
8	100.00%	0.00%	0.00%	0.00%

From table 6.3, it is determined that there is an imbalance in demand. Hospital 1 receives, on average, a higher demand, whereas Hospital 3 receives the least. Considering that all of them have the same amount of resources (beds), there is an opportunity to transfer the emergency demand to increase the network efficiency. It has been observed from the data in table 6.4 that the demand for DRG is not uniformly distributed among hospitals. This presents an opportunity to optimize the efficiency of the network by rearranging the patients through transfers.

The average arrivals per period by DRG and medical requirement are presented in table 6.5. Additionally, table 6.6 shows the average relative arrivals by DRG and medical requirement per period.

Table 6.5. Average Arrivals by DRG and Requirement per Period

	Mean Arrival Rate [patients/period]		
DRG	**OR**	**ICU**	**SDU & WARD**
1	1.61	1.56	0.95
2	1.56	1.53	0.83
3	1.63	1.60	0.95
4	1.40	1.30	0.62
5	1.64	1.79	1.67
6	1.30	1.42	1.21
7	1.48	1.48	1.37
8	0.84	0.80	0.72
GENERAL	**11.45**	**11.48**	**8.33**

Table 6.6. Average Relative Arrivals by DRG and Requirement per Period

	Relative Mean Arrival Rate [%]		
DRG	**OR**	**ICU**	**SDU & WARD**
1	39.16%	37.86%	22.98%
2	39.67%	39.14%	21.19%
3	38.93%	38.32%	22.76%
4	42.06	39.25%	18.70%
5	32.14%	35.15%	32.72%
6	33.08%	36.02%	30.902%
7	34.24%	34.08%	31.68%
8	35.57%	33.78%	30.64%

From both tables, it is clear that medical requirements with lower complexity have, on average, lower demand in the case of emergency demand. Additionally, the proportions among elective DRGs have a balanced arrival rate among the units.

6.1.2.2. Demand Average LOS

It is essential to understand the length of stay for each DRG. The following table displays the average LOS by DRG based on their medical necessity.

Table 6.7. Average LOS by DRG and Medical Requirement

	Mean LOS [days]		
DRG	**OR**	**ICU**	**SDU & WARD**
1	0.59	3.44	9.26
2	0.61	5.06	10.82
3	0.57	1.64	4.46
4	0.58	2.72	8.11
5	0.50	2.02	5.57
6	0.50	1.48	3.72
7	0.55	3.27	8.99
8	0.54	2.99	8.36

Based on the data presented in Table 6.7, it can be concluded that the OR Average LOS is roughly the same for each DRG. However, when looking at the ICU, as well as the SDU & Ward, there are noticeable differences. For instance, DRG 2 and DRG 3 have significantly different LOS in these units, although they share the same type of demand (emergency).

6.1.2.3. Inpatient Evolution

The evolution of a patient's medical condition during hospitalization can be modeled using a Markov Chain, as proposed by González et al. (2018). When a patient completes their medical treatment in a particular unit $i \in N_2$, they transition to another unit $j \in N_2 : i \neq j$. Generally, a patient's condition tends to improve, but there is a possibility of deterioration, as suggested by González et al. (2018) in their research extensions. Furthermore, in the lowest complexity unit, in this case, the SDU & WARD, there is a chance that the patient may require further treatment in that unit. Although each DRG has a different transition matrix, we will illustrate the average transition matrix for clarity purposes.

Table 6.8. Average Transition Matrix for Medical Requirements

	OR	**ICU**	**SDU & WARD**	**Discharge**
OR	0.00	0.92	0.08	0.00
ICU	0.08	0.00	0.92	0.00
SDU & WARD	0.01	0.05	0.00	0.94

6.1.2.4. Costs by DRG

As previously mentioned, the primary goal of this research is to minimize the total operational and social costs (such as patient wait times and bed opportunity costs). Since the references used in this study have different currencies, a monetary unit has been defined to combine these costs. The costs were adjusted manually to ensure the scales between different expenses were reasonable. However, in real-world scenarios, it can be challenging to compile these costs as estimating bed opportunity costs and the statistical value of a life can be debatable.

In terms of the hazards associated with patient integrity due to waiting times, the methodology used is the one presented in subsection 3.4.3. It is important to note that, for this scenario, the statistical value of a life is considered to be 6.2660. Compared to other costs, this is the highest in the instance, highlighting the significance we place on the fairness and well-being of patients.

Tables 6.9, 6.10 and 6.11 present the bed opportunity cost for each unit and DRG by hopsital. This cost must be interpreted as the detrimental effects in terms of operational efficiency and costs that the healthcare system must assume in case of inpatient overstay at certain hospital units.

Table 6.9. Bed Opportunity Cost in Hospital 1 by Unit and DRG

	DRG 1	DRG 2	DRG 3	DRG 4	DRG 5	DRG 6	DRG 7	DRG 8
ED	0.0886	0.1605	0.0916	0.1407	0.0498	0.1521	0.0843	0.1112
GA	0.0664	0.1204	0.0687	0.1055	0.0374	0.1141	0.0632	0.0834
OR	0.0443	0.0802	0.0458	0.0704	0.0249	0.0760	0.0422	0.0556
ICU	0.0221	0.0401	0.0229	0.0352	0.0125	0.0380	0.0211	0.0278
SDU/WARD	0.0111	0.0201	0.0115	0.0176	0.0062	0.0190	0.0105	0.0139

Table 6.10. Bed Opportunity Cost in Hospital 2 by Unit and DRG

	DRG 1	DRG 2	DRG 3	DRG 4	DRG 5	DRG 6	DRG 7	DRG 8
ED	0.0975	0.0822	0.1726	0.1133	0.0763	0.0499	0.1256	0.0793
GA	0.0732	0.0616	0.1294	0.0850	0.0572	0.0374	0.0942	0.0594
OR	0.0488	0.0411	0.0863	0.0567	0.0382	0.0250	0.0628	0.0396
ICU	0.0244	0.0205	0.0431	0.0283	0.0191	0.0125	0.0314	0.0198
SDU/WARD	0.0122	0.0103	0.0216	0.0142	0.0095	0.0062	0.0157	0.0099

Table 6.11. Bed Opportunity Cost in Hospital 3 by Unit and DRG

	DRG 1	DRG 2	DRG 3	DRG 4	DRG 5	DRG 6	DRG 7	DRG 8
ED	0.1624	0.0928	0.0862	0.1246	0.1475	0.0510	0.0500	0.0834
GA	0.1218	0.0696	0.0646	0.0934	0.1106	0.0382	0.0375	0.0626
OR	0.0812	0.0464	0.0431	0.0623	0.0738	0.0255	0.0250	0.0417
ICU	0.0406	0.0232	0.0215	0.0311	0.0369	0.0127	0.0125	0.0209
SDU/WARD	0.0203	0.0116	0.0108	0.0156	0.0184	0.0064	0.0063	0.0104

From these tables, it is observed that the emergency DRGs (1 to 4) have higher costs than elective ones (5 to 8). Furthermore, the opportunity cost for the admission units (ED and GA) is higher than the service units (OR, ICU and SDU & WARD), as a result of not having these beds available for care means not being able to attend to future medical requests. It is essential to note that this cost structure may not be applicable in all cases. As previously mentioned, the opportunity cost of a bed should take into account both the statistical data and the preferences of the hospital manager, which can vary from case to case.

The table below shows the affinity (i.e., lowest bed opportunity cost) for each DRG by the hospitals in the network. A lower value (1) indicates a higher affinity, while a higher value (3) indicates a lower affinity.

Table 6.12. Hospital Affinity with the DRGs

	DRG 1	**DRG 2**	**DRG 3**	**DRG 4**	**DRG 5**	**DRG 6**	**DRG 7**	**DRG 8**
Hospital 1	1	3	2	3	1	3	2	3
Hospital 2	2	1	3	1	2	1	3	1
Hospital 3	3	2	1	2	3	2	1	2

The diversion costs must consider the average cost of resolution of a patient of each DRG with a specific medical need at the service units of a hospital. This costs were estimated as follows:

(i) In order to determine the average number of bed days consumed per care unit, we conducted a simulation of the hospital journey for a significant number of patients for each DRG. This estimation takes into consideration the possibility that patients may experience a decline in their condition, resulting in multiple visits to care units.

(ii) Using the average opportunity cost values in the network, we can calculate the amount of money spent on resources for an average patient of each type in the public healthcare system.

(iii) According to Martinez et al. (2019), the costs of medical treatment in private clinics are usually higher than in public clinics, ranging from 2 to 3 times more expensive. For our calculations, we assumed the highest possible value (i.e., 3). By multiplying the costs calculated in the previous step by 3, we obtained the costs of diversion in the network, which are presented in Table 6.14.

Table 6.14. Diversion Cost in the System by DRG and Medical Need

	DRG 1	**DRG 2**	**DRG 3**	**DRG 4**	**DRG 5**	**DRG 6**	**DRG 7**	**DRG 8**
OR	2.0054	2.4156	1.0084	1.8616	0.9125	0.6199	1.4876	1.2911
ICU	1.8641	2.2980	0.8395	1.7132	0.7977	0.5148	1.3445	1.1921
SDU/WARD	1.1074	1.2037	0.5002	1.0620	0.4910	0.2959	0.7964	0.7223

Finally, the transfer costs are displayed in the table below. For the purpose of simplicity, we assume that the costs are the same for any pair of hospitals in the network.

Table 6.15. Transfer Cost in the Network by DRG and Medical Need

	DRG 1	**DRG 2**	**DRG 3**	**DRG 4**	**DRG 5**	**DRG 6**	**DRG 7**	**DRG 8**
OR	0.0111	0.0133	0.0056	0.0103	0.0050	0.0034	0.0082	0.0071
ICU	0.0049	0.0061	0.0022	0.0045	0.0021	0.0014	0.0036	0.0032
SDU/WARD	0.0029	0.0031	0.0013	0.0028	0.0013	0.0008	0.0021	0.0019

In the previous model, it was shown that the costs of a problem depend on the DRG, patient medical needs, and waiting times. The costs presented in the tables assume that the patient waiting time is only one period. However, as the waiting time increases, the patient's health deteriorates, leading to higher costs for the operator due to technical requirements that ensure the integrity of a more complex patient. For instance, it is reasonable that transporting a patient with a waiting time of two weeks would cost more than a patient with a waiting time of a couple of hours.

To address this issue, we introduce a formula that can calculate the cost, $C(t)$, given any waiting time, t greater than 1 period, and any cost, C, from the tables: $C(t) = C(1+r)^{t-1} \quad (r > 0)$. For our research, we have utilized an appropriate value of $r = 0.01$.

6.2. Policies to be Compared

The following section presents the policies that were initially studied.

6.2.1. Myopic Policy: π_M

The Myopic Policy π_M assumes that every possible state's expected future cost is 0. This reactive policy (i.e., only optimizes the present value), can be formulated as follows:

$$C(\vec{s_1}) = \min_{\vec{a} \in \mathbb{X}(\vec{s_1})} \left\{ c(\vec{s_1}, \vec{a}) + \lambda \cancelto{0}{\mathbb{Q}(\vec{s_2})} \right\} \quad \forall \vec{s_1} \in \mathbb{S}_1$$

Which ends up solving the following problem:

$$C(\vec{s_1}) = \min_{\vec{a} \in \mathbb{X}(\vec{s_1})} \{ c(\vec{s_1}, \vec{a}) \} \quad \forall \vec{s_1} \in \mathbb{S}_1$$

6.2.2. Q-learning Policy: $\pi_{\mathbb{Q}}$

The Q-learning Policy $\pi_{\mathbb{Q}}$ assumes that we can estimate the expected future cost using a function $\mathbb{Q}$. In our study, we use an XGBoost model to learn this function via reinforcement learning. As demonstrated in section 5.3.3.4, given an XGBoost model with N trees, we can mathematically formulate this policy as follows:

$$C(\vec{s_1}) = \min_{\vec{a} \in \mathbb{X}(\vec{s_1})} \{c(\vec{s_1}, \vec{a}) + \lambda \mathbb{Q}\left(\vec{s_2}\right)\} \quad \forall \vec{s_1} \in \mathbb{S}_1$$

$$\mathbb{Q}\left(\vec{s_2}\right) = \sum_{n=1}^{N} \beta_n y_n \tag{6.1}$$

$$y_n = \sum_{i \in \mathcal{L}^n} p_i^n l_i^n \qquad \forall n \in \{1, \cdots, N\} \tag{6.2}$$

$$(\overline{A_j^n})^T \vec{s_2} - M^n(1 - l_i^n) \leq \overline{b_j^n} \qquad \forall\, i \in \mathcal{L}^n, \quad \forall\, j \in \mathcal{S}_n^i, \quad \forall n \in \{1, \cdots, N\} \tag{6.3}$$

$$\sum_{i \in \mathcal{L}^n} l_i^n = 1 \qquad \forall n \in \{1, \cdots, N\} \tag{6.4}$$

$$y_n = \sum_{i \in \mathcal{L}^n} p_i^n l_i^n \qquad \forall n \in \{1, \cdots, N\} \tag{6.5}$$

Constraint (6.1) calculates the $\mathbb{Q}\left(\vec{s_2}\right)$, given the decision tree models. Constraints (6.2), (6.3), (6.4) and (6.5) calculates the predicted value for each tree $n \in \{1, \cdots, N\}$ from the XGBoost model. It is worth mentioning that to obtain a highly proactive policy, we used $\lambda = 0.99$.

6.3. Q-learning and Reinforcement-Learning Process and Validation

The training process of the $\mathbb{Q}$, as well as its validation, are detailed in this section. A list of the computational resources used for these experiments is provided:

(i) **CPU:** 13th Gen Intel(R) Core(TM) i9-13980HX 2.20 GHz

(ii) **GPU:** NVIDIA GeForce RTX 4070 Laptop GPU GDDR6 (8GB - 140W)

(iii) **RAM:** 64 GB DDR5 - 4800 MHz

6.3.1. Training Process and Convergence

Multiple experiments were performed to evaluate the efficiency of the proposed Q-learning approach. Initially, each experiment began with 120 samples, each containing 1,000 experiences. These samples were fed into the Replay Buffer. After every training step, three new samples were collected, replacing the older ones. Each experiment was given a run-time of one day which was sufficient to obtain enough samples to update the Replay Buffer completely. Figure 6.1 presents the results of the training process for different trials.

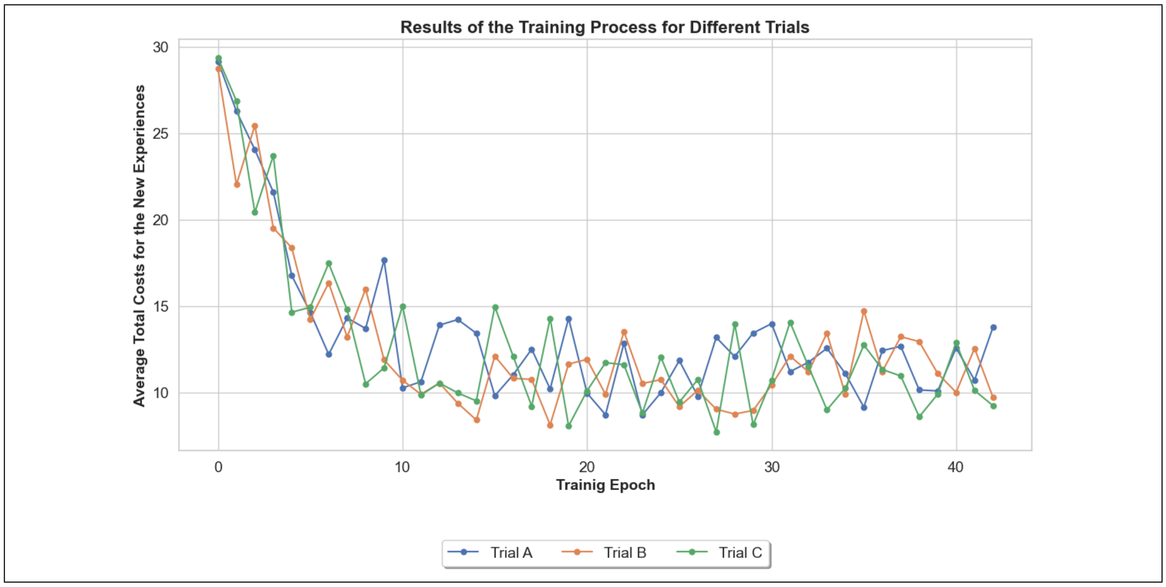

Figure 6.1. Results of the Training Process for Different Trials

Based on the figure, it can be concluded that the solution approach can quickly converge around a good solution (a good policy π, due to the $\mathbb{Q}$ function trained). Although asymptotic results are not yet evident, we can provide evidence that the methodology converges to a good Total Cost value. Further iterations will likely achieve asymptotic results, based on (Van Hasselt et al., 2016). Finally, the results demonstrate that different attempts can reach similar results of good quality, which indicates that the methodology is consistent in obtaining good results.

The XGBoost model was trained considering a maximum amount of 3,000 trees with a depth of 3 (the tree regressor can achieve 4 levels). We used a learning rate of 0.025, coupled with L1 and L2 regularization, with those parameters being calibrated using Optuna, which was previously mentioned in section 5.3.3.4. The best model ended with an optimal number of 2,348 trees from all the trials.

6.3.2. Validation of the Results

Validation of the trained model requires understanding its behavior, ensuring it accurately represents the relationships between variables and produces sensible output. To achieve this, we tested the Q-learning model on 30,000 post-decision states. These states were sampled by running the $\pi_{\mathbb{Q}}$ policy with the trained model in multiple independent scenarios. The post-decision states were then averaged, giving us a mean post-decision state $\bar{\vec{s}}_2$.

Next, we analyzed two factors of the average post-decision state. We focused on a particular patient type denoted as $p \in P$, with DRG 5, an OR medical requirement, and a waiting time of 1 period. For this patient type, we examined the variable ω_2^p, which reflects the number of patients of the same type as p on the WL. Additionally, we considered the component α_2^{p3OR}, representing the number of patients of type p being treated at the OR in Hospital 3. If we map in a range from 0 to 40 patients both variables, *ceteris paribus* the rest of components from $\bar{\vec{s}}_2$, and evaluate these cases, we obtain the following:

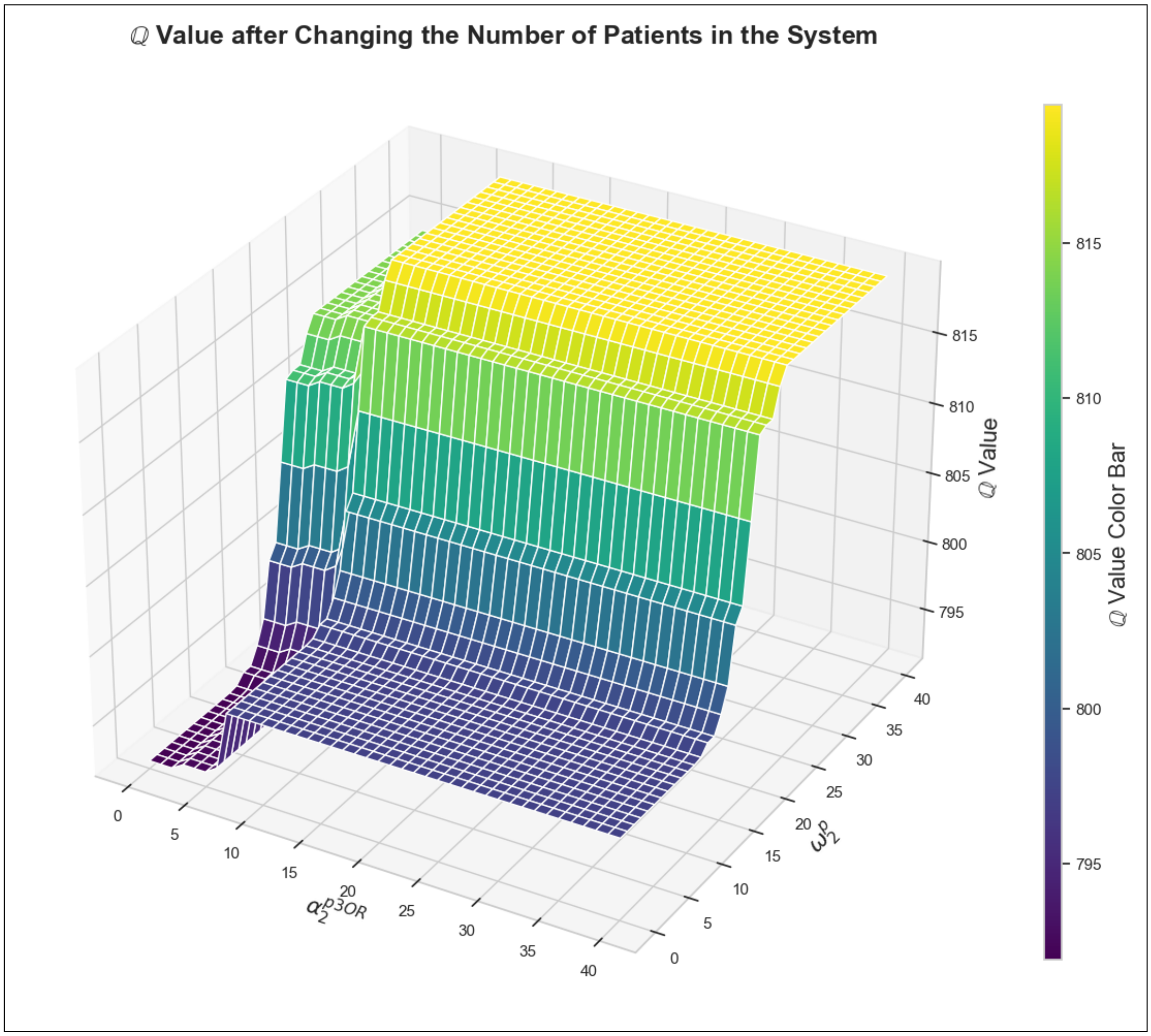

Figure 6.2. Profile 1 of the Predictions of ℚ from the Trained Model

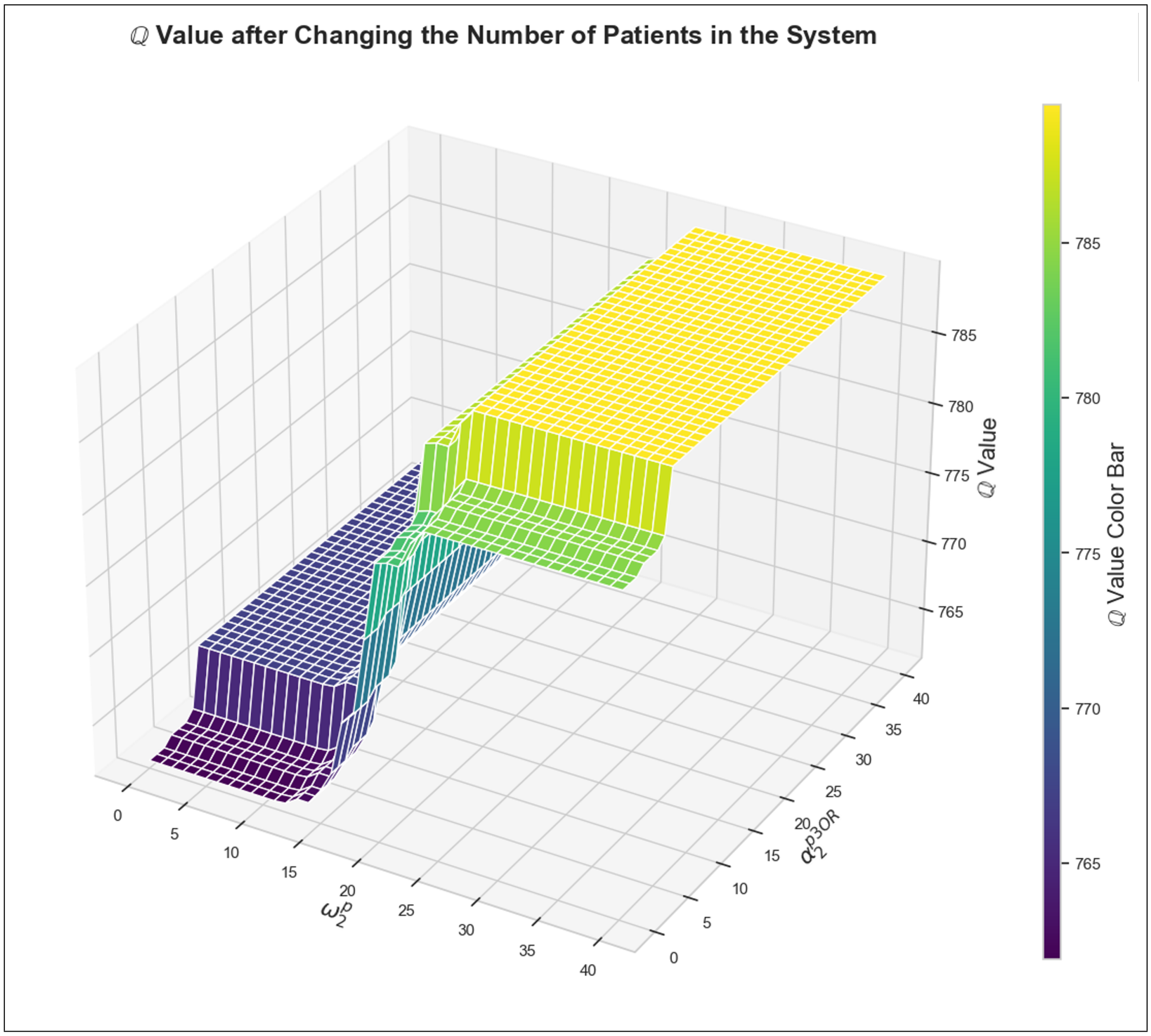

Figure 6.3. Profile 2 of the Predictions of $\mathbb{Q}$ from the Trained Model

Based on Figures 6.2 and 6.3, we can conclude that the model correctly represents the relationships between variables and the $\mathbb{Q}$ value. Specifically, as the number of patients increases, the $\mathbb{Q}$ value also increases. This supports the notion that a system experiencing congestion will have higher costs. Additionally, we observe that after 6 patients, the $\mathbb{Q}$ value for α_2^{p3OR} does not increase, which is consistent with the OR capacity that has 6 beds. It is worth mentioning that the model successfully incorporates a non-linear estimation of

$\mathbb{Q}$, which is a significant extension proposed by Marquinez et al. (2021) and González et al. (2018).

6.4. Main Results

After testing and validating the trained model for the $\pi_{\mathbb{Q}}$ policy, we can test it and compare it to the results obtained from policy π_M. Regarding optimization, the gap for training and testing instances was set at 0.01%, although the solver achieved a gap of 0%. The optimization model associated with each policy is briefly described in the table below:

Table 6.16. Summary of Optimization Models

Policy	**Average Execution Time [s]**	**Integer Variables**	**Continuous Variables**	**Constraints**
π_M	0.01	15,792	0	8,488
$\pi_{\mathbb{Q}}$	0.40	33,092 (17,300)	2,439	78,039

As the XGBoost model is embedded in the MDP, it increases the number of integer variables to 17,300, all of which are binary. Additionally, it introduces 2,439 continuous variables (one for each tree) and a variable that contains the final prediction from the forest. It is important to note that this formulation significantly increases the number of constraints, which ultimately explains the longer execution time.

6.4.1. Costs

To begin with, it is crucial to compare the costs associated with both policies. The simulation results for 50 independent runs throughout 1,300 periods are presented in Figures 6.4 and 6.5. The first 300 periods are dedicated to initializing the system, ensuring each run achieves a stationary result for the base scenario policy π_M. The next 200 periods are used for transitioning between policies, if necessary. Finally, the last 800 periods of each run are used to retrieve data, considering that a steady state regime is obtained.

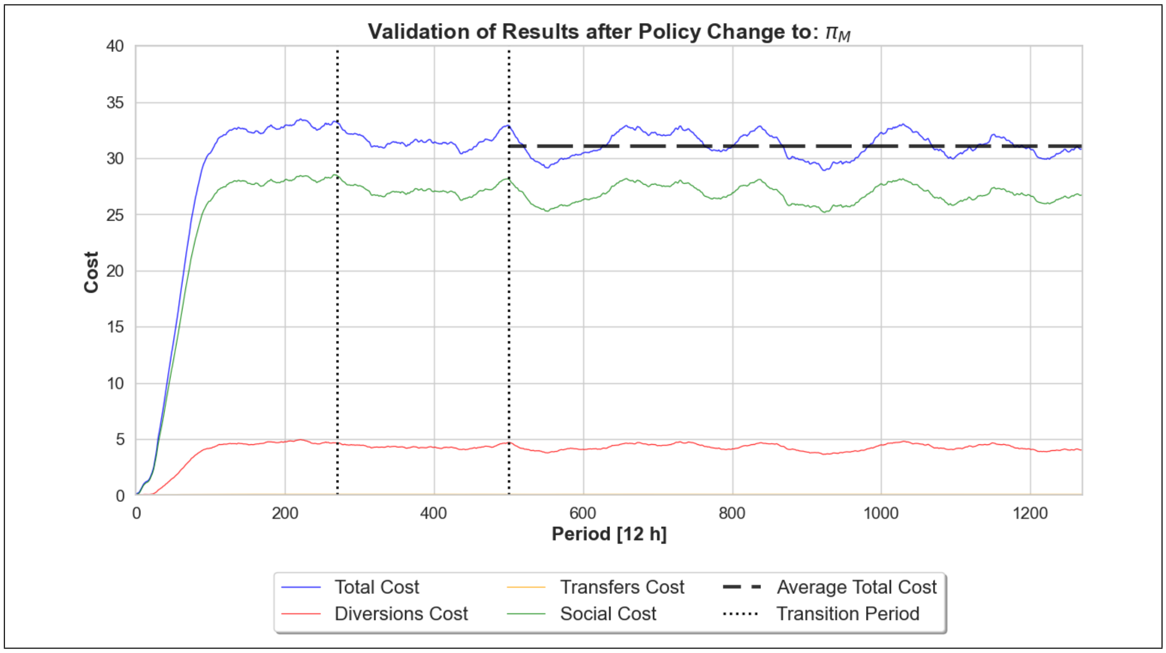

Figure 6.4. Results for Policy π_M

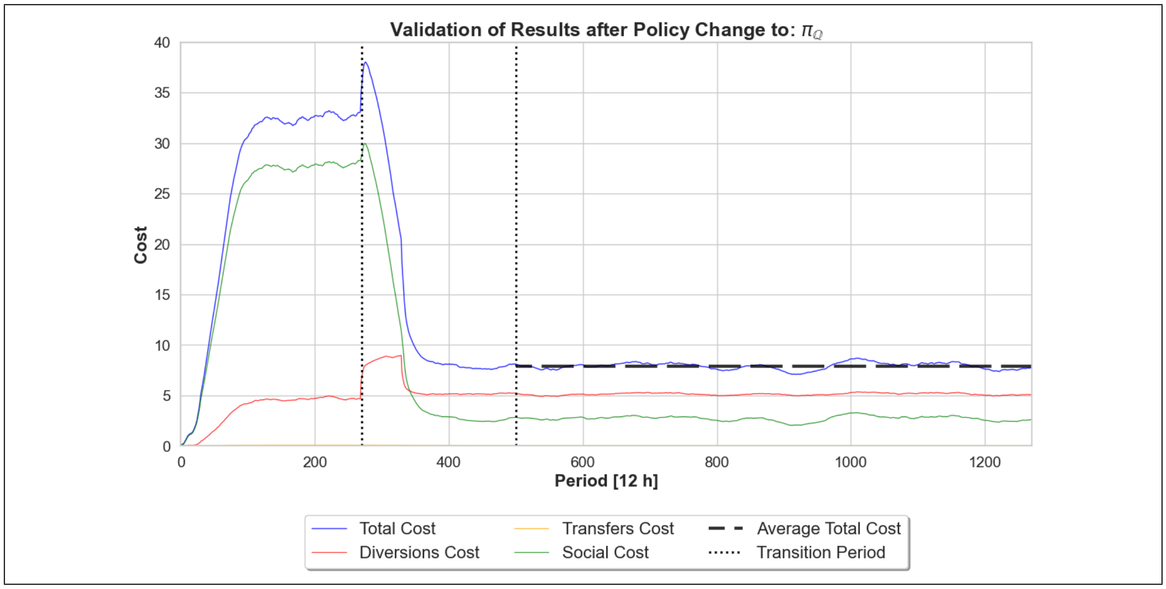

Figure 6.5. Results for Policy $\pi_{\mathbb{Q}}$

From both figures, it can be concluded that transfer costs are negligible compared to diversions and social costs. In terms of the base scenario with policy π_M; it is possible to conclude that the system is extremely congested, thus resulting in those high costs, where the diversions are being made because there is no more capacity.

Notably, the policy $\pi_{\mathbb{Q}}$ experiences a rough start during the initial phase of the transition. This phenomenon could be attributed to the system redistributing the demand to attain a superior state in the long run. Ultimately, although there is an increase in the diversions cost, the social cost is reduced at the point where the total cost is significantly lower, compared to the policy π_M. It can be concluded that the diversions improve the system occupancy rates.

The following table summarizes the results and compares the different cost items statistically. Each policy result is presented as a mean, with coefficients of variation in parentheses.

Table 6.17. Summary of Policies Results

	Total Cost	**Diversions Cost**	**Transfers Cost**	**Social Cost**
π_M	31.1043 (0.0504)	4.2535 (0.1012)	0.0806 (0.0422)	26.7708 (0.0438)
$\pi_{\mathbb{Q}}$	7.9148 (0.0606)	5.1354 (0.0334)	0.0571 (0.0455)	2.7225 (0.1314)
Difference	[23.0733, 23.42]	[-0.9504, -0.8222]	[0.0228, 0.0241]	[23.9768, 24.2429]

From the table, we can conclude that there is a significant statistical reduction (95% Confidence) in total costs, with an average reduction of 74.55% . This is achieved by increasing the diversions cost in 20.73%, which allows a reduction of the social cost of 89.83%. Notably, the dispersion of the diversion cost is significantly lower for $\pi_{\mathbb{Q}}$, indicating that this policy has a more consistent set of actions over time. However, the social cost has a higher dispersion, explained mainly by the lower magnitude of this cost for the $\pi_{\mathbb{Q}}$.

6.4.2. Policies Case-Mix Report

To better understand the results, a Case-Mix Report is proposed. We will consider as the reference system the base scenario with policy π_M, while the case study will consist of the system with policy $\pi_{\mathbb{Q}}$.

Table 6.18. Case-Mix Report

	A	**P**	**a**	**p**	**a-A**	**p-P**	**Δ LOS**	**Δ CM**
DRG 1	16.54	0.1497	16.19	0.1516	-0.35	0.0019	-0.05	0.03
DRG 2	19.92	0.1434	19.60	0.1443	-0.32	0.0009	-0.05	0.02
DRG 3	7.98	0.0722	7.62	0.1382	-0.36	0.0660	-0.04	0.51
DRG 4	14.04	0.1244	13.82	0.1263	-0.22	0.0019	-0.03	0.03
DRG 5	17.11	0.1641	11.79	0.1395	-5.32	-0.0246	-0.81	-0.36
DRG 6	14.28	0.1305	8.68	0.0984	-5.60	-0.0321	-0.64	-0.37
DRG 7	23.94	0.1394	17.42	0.1244	-6.52	-0.015	-0.86	-0.31
DRG 8	20.96	0.0763	16.42	0.0773	-4.54	0.0010	-0.35	0.02
Total							**-2.47**	**-0.44**

Based on the table provided, it can be concluded that we achieved an estimated average of 2.47 days saved per patient in terms of operational efficiency. In column $a - A$, it can be observed that all DRGs reduced their ALOS, which contributed to the Δ LOS sum. Regarding Case-Mix, column $p - P$ shows a tendency for the system to increase the proportion of emergency DRGs, while the elective ones decreased their proportion. These results must be complimented with the difference in attended patients between policies of 0.23%, meaning that we are attending almost the same amount of patients for both policies. As a result of this change in Case-Mix, we saved an average of 0.44 days. Combining the changes in LOS and Case Mix, we achieved a total reduction of 2.91 service days on average (an average reduction of 16.86%), per patient

6.4.3. Understanding the Q-learning policy

Comprehending the decisions behind the model is essential. This helps the decision-maker validate and understand the decisions behind the model. Furthermore, understanding the general guidelines of the Q-learning policy can assist in developing an easily interpretable and simplified policy. This simplification can increase the chances of real-world applications due to its ease of understanding and low complexity. Although there are multiple ways of approaching this policy, we will focus on analyzing the average pre-decision state and post-decision state achieved, as well as the transition among them.

6.4.3.1. Occupation Levels

According to research by González et al. (2018) and Marquinez et al. (2021), implementing proactive policies can help decongest the system and improve performance by providing available resources to handle high-demand events. In the latter case, the optimal utilization level is approximately 80%. Although not official, studies conducted by the OECD (2023), as well as the National Guideline Centre - UK (2018), support this value, being validated by renowned entities in the medical context.

For the post-decision state, we have the following rates for policies $\pi_{\mathbb{M}}$ and $\pi_{\mathbb{Q}}$, respectively.

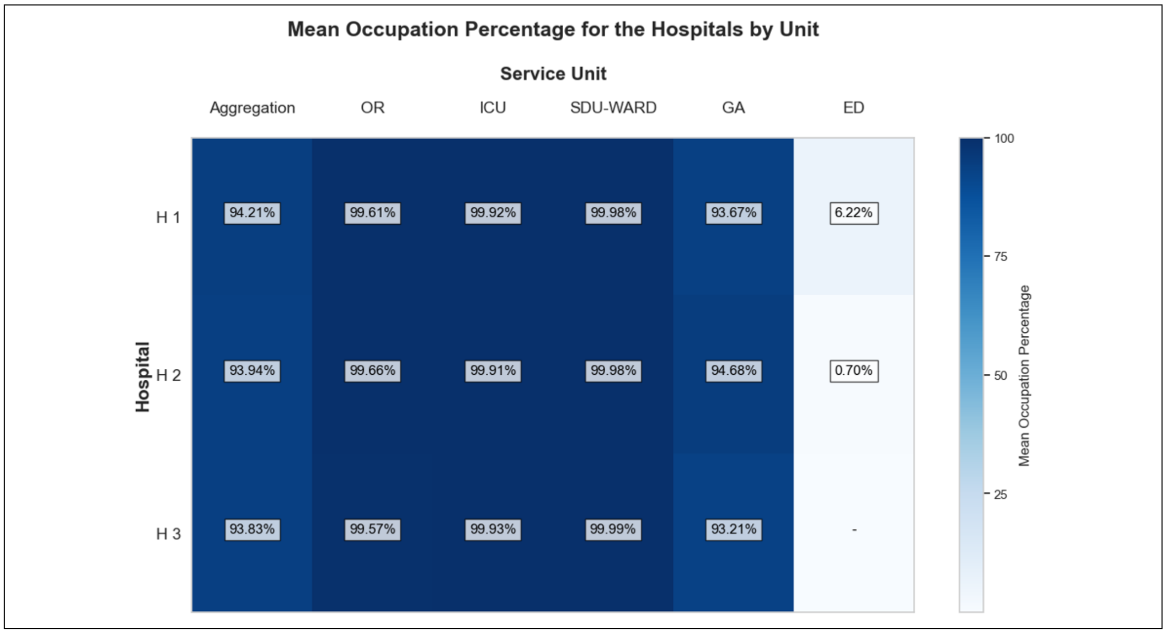

Figure 6.6. Occupancy Rates in the average post-decision state for Policy π_M

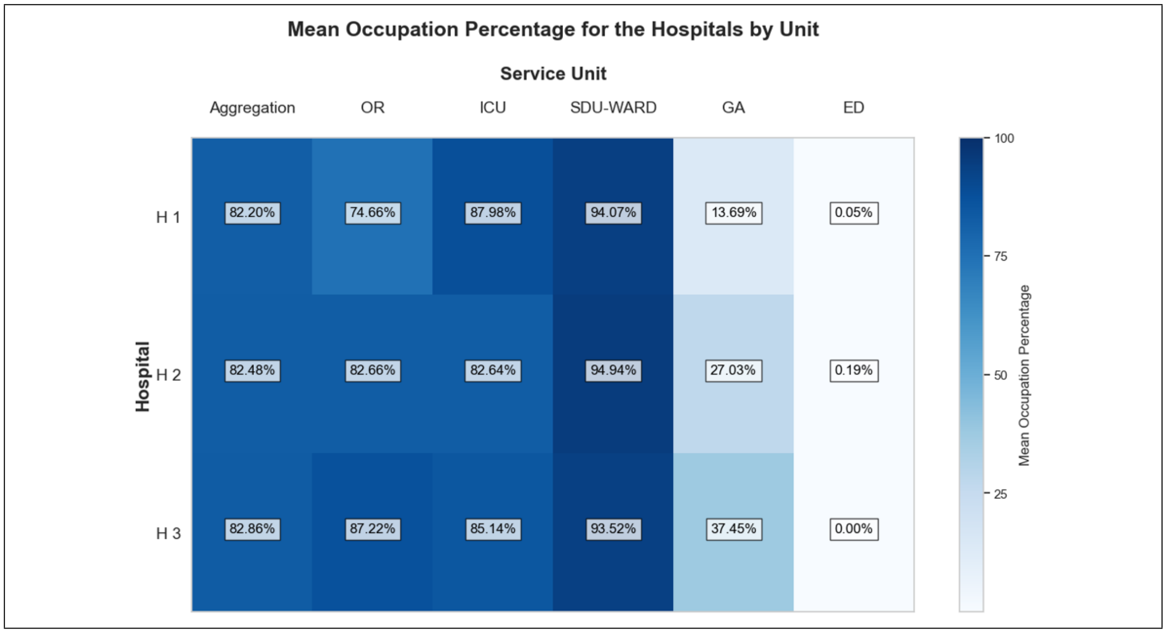

Figure 6.7. Occupancy Rates in the average post-decision state for Policy $\pi_{\mathbb{Q}}$

From Figure 6.6 it is concluded that the system is, in fact, saturated. Given each service unit occupancy rate, there is almost no space to make an efficient, impatient management, generating the high social cost presented in table 6.21. This situation is worsened by highly congested admission units, resulting in excessive waiting times.

On the other hand, from Figure 6.11, we can conclude that the system is effectively operating at the recommended occupancy rates. Not only the GA unit is decongested, but there are available resources to accommodate the incoming demand, avoiding a high social cost due to waiting times and the misuse of beds in the network.

We can conclude that Hospital 3 specializes in elective demand based on the GA occupancy rates (37.45%, compared to 13.69% and 27.03% from hospitals 1 and 2, respectively). This means that the hospital deals with less uncertain demand, which is on the waiting list, with several requests already known. Therefore, the hospital can reduce its beds for safety stock purposes (i.e., beds available to attend high-demand events).

On the other hand, Hospital 1 has a higher emergency demand composition, as indicated by the lower GA occupation levels, which are lower compared to Hospitals 2 and 3. Due to the higher expected demand that it must attend, with a higher volatility Case-Mix, it is reasonable to assume that the system operates with more beds for safety stock.

As the unit complexity increases, the safety stock of beds also increases. This is because a stockout of beds for complex medical requirements can result in diversions with high costs. Also, this avoids costly waiting times for inpatients at other units that need a higher complexity bed.

It is important to note that the policy $\pi_{\mathbb{Q}}$ maintains the same occupancy levels after admissions, diversions, and transfers. This highlights the significance of the safety stock policy in terms of beds, as well as the fact that actions are being taken to maintain this state. This contrasts the policy π_M, where the system concludes at high occupancy levels, indicating that diversions and transfers are made to satisfy the demand that cannot be reactively accommodated within the system.

6.4.3.2. Case-Mix

Given the insights, where there is evidence of hospital specialization, it is important to study the Case-Mix at the network. Similarly to the occupation study, we will focus on the average post-decision state.

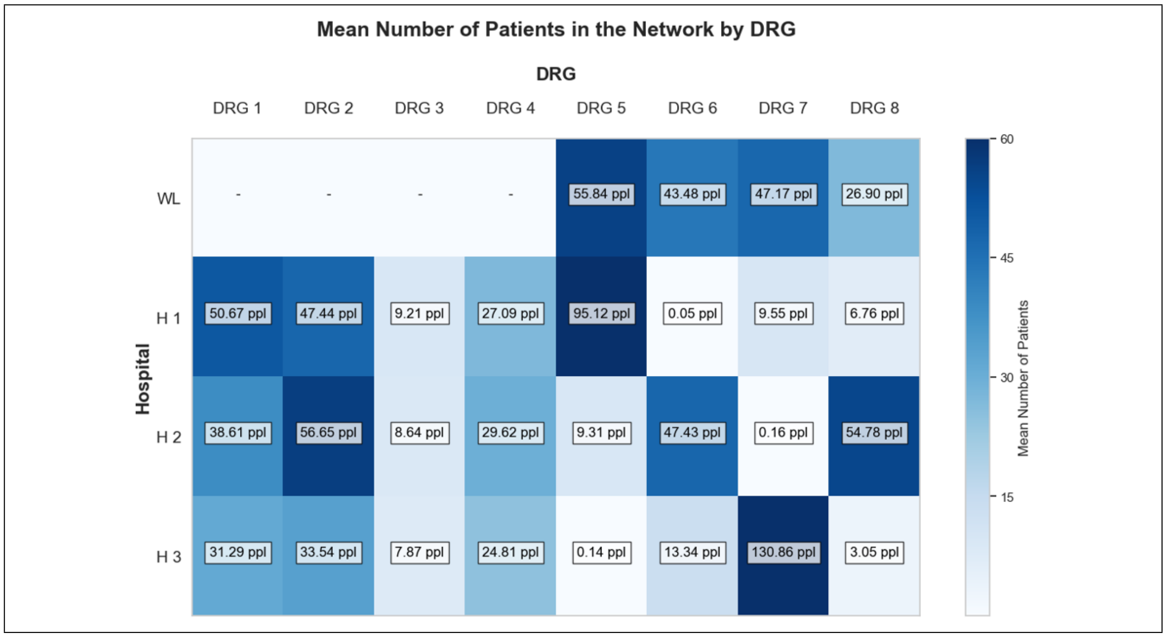

Figure 6.8. Case-Mix in the average post-decision state for Policy π_M

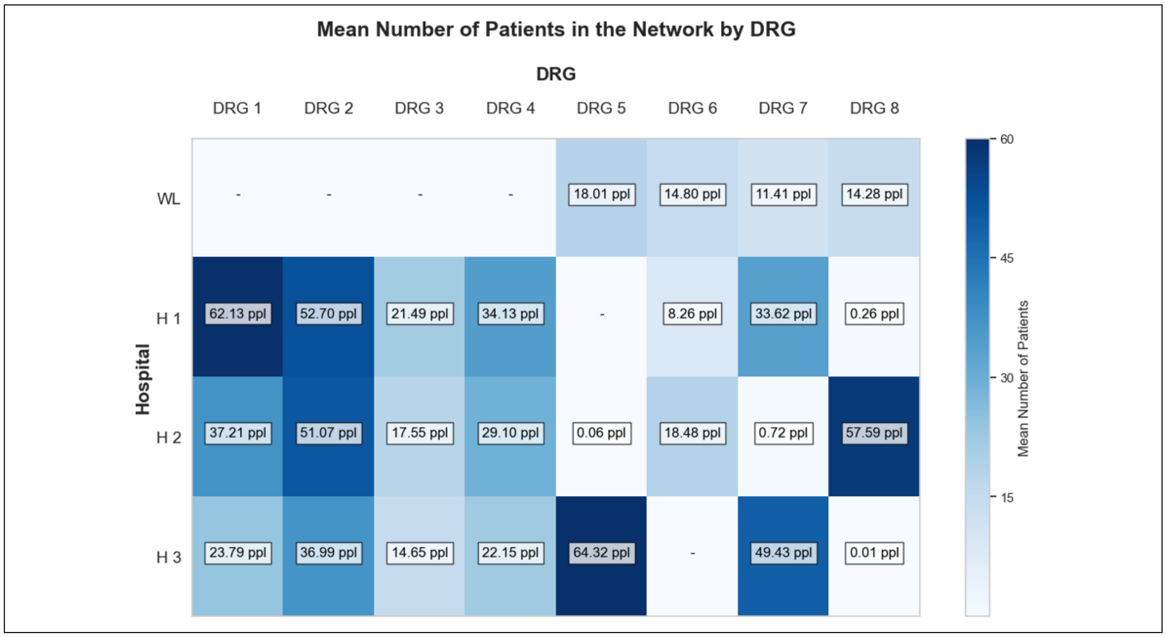

Figure 6.9. Case-Mix in the average post-decision state for Policy $\pi_{\mathbb{Q}}$

After analyzing the results, it is determined that there is a significant change in the Case-Mix. The policy $\pi_{\mathbb{Q}}$ has effectively specialized the hospitals, as suggested in the previous analysis. Based on the number of patients, Hospital 1 focuses on handling more emergency cases, while Hospital 3 specializes in elective cases. Hospital 2 falls somewhere between the other two hospitals. The specialization on DRGs is so important that there are situations where hospitals do not service certain DRGs. For instance, Hospital 1 does not attend to DRG 5, and even Hospital 2 rarely attends to them.

It is noteworthy that emergency DRGs, particularly DRG 3, are becoming more common in the network. Meanwhile, elective DRGs are experiencing a significant reduction in the average number of active cases in the network, with DRG 8 being the least affected. In terms of medical complexity, it is interesting to observe a shift towards more complex cases in the network due to the increase in the average number of active emergency cases. This specialization strategy is validated by Tiwari and Heese (2009), which determined that this phenomenon produces better management in terms of costs, as well as a higher quality service, just as our results suggest.

In this scenario, there is not enough capacity to serve all the incoming demand, so it is necessary to divert some patients. As the number of beds is a fixed resource that does not cost us to maintain, it is essential to use them to attend the cases that are expensive to divert and promote the diversion of cheap ones. The study by Marquinez et al. (2021) concluded that this strategy is extremely effective. According to the data presented in Table 6.14, it is advisable to divert elective DRGs. This strategy is being implemented by $\pi_{\mathbb{Q}}$, leading to a significant reduction in active cases for DRG 5 by 40.70%, DRG 6 by 56.03%, and DRG 7 by 40.41%. Furthermore, this applies to the WL on its own, resulting in a reduction of its size by 66.26%.

6.5. Learning a New Policy

Research has shown that even though algorithms can improve decision-making outcomes, decision-makers tend to undervalue them. This phenomenon is known as 'algorithm aversion' (Dietvorst et al., 2015) and has been observed across various domains, including economics, finance, job hiring, medicine, and even moral decisions, for both subjective and objective tasks, and assessments that are both familiar and unfamiliar. Human advice is often preferred over algorithmic advice.

Given that the $\pi_{\mathbb{Q}}$ policy uses a Machine Learning model, it can be challenging to interpret, especially outside the engineering field. It becomes more worrying if we consider the 'algorithm aversion' phenomenon, where the actions recommended by the model must influence a medical team. The situation prompts the idea of utilizing the knowledge provided by the model, along with the medical guidelines discussed earlier, to create a practical, resilient, and effective policy in real-world situations, thereby overcoming the phenomenon of 'algorithm aversion'.

Although the model could be explored further, we will focus on the findings of occupation and specialization in the network. We will call this policy 'Knowledge Assembly' (π_K), since it unites the learning of the Q-learning proposal with expert knowledge of literature in the health field. The guidelines consider the following set of rules:

(i) **Guideline 1 - Occupation Levels:** Each unit $n \in N$ at every hospital h in the network will have defined an ideal occupation level. If a unit has an occupation level below this threshold, there will be no changes regarding the myopic policy. However, if these levels are exceeded, it will be necessary to redirect the demand through transfers and diversions to decongest the saturated unit. For the test scenario, based on the $\pi_{\mathbb{Q}}$ policy results, these values are:

Table 6.19. Ideal Occupation Level by Hospital and Unit

	OR	**ICU**	**SDU & WARD**	**GA**	**ED**
H1	74.66%	87.98%	94.07%	13.69%	0.05%
H2	82.66%	82.64%	94.94%	27.03%	0.19%
H3	87.22%	85.14%	93.52%	37.45%	0%

To emulate this rule in the system, let us denote these values as θ_h^n. We can mathematically model if any unit exists over these levels using a continuous variable, u_h^n, to represent the amount of occupation above the desired level. Using the notation from Section 4, we propose the following set of constraints:

$$\frac{\alpha_2^{*hn}}{\Phi_h^n} \leq \theta_h^n + u_h^n \quad \forall (h,n) \in \{1,\ldots,H\} \times N$$

$$u_h^n \geq 0 \quad \forall (h,n) \in \{1,\ldots,H\} \times N$$

Additionally, to incentivize the decongestion, we must add the term $O_h^n u_h^n$ for each unit in the system, where O_h^n is a calibrated parameter.

(ii) **Guideline 2 - Waiting List Size:** We have determined that each elective DRG should have a maximum number of requests at the WL. The maximum requests allowed for each DRG are 18 for DRG 5, 15 for DRG 6, 12 for DRG 7, and 15 for DRG 8. These values are obtained from the average amount of patients by DRG in the WL. For completitude, the emergency DRGs coefficient is 0. If we have more patients than the maximum number of requests allowed, admissions should be made only if the first guideline allows it. If that is not the case, we need to take diversion actions. To emulate this rule in the system, let us denote these values as $\bar{\omega^g}$ for $g \in \{1, \ldots, G\}$. To emulate this rule in the system, using the notation from Section 4, we propose the following set of constraints:

$$\sum_{u \in \mathbb{U}(g)} \omega_2^u \leq \bar{\omega^g} \quad \forall g \in \{1, \ldots, G\}$$

Where:

$$\mathbb{U}(g) = \{u \in P : g(u) = g\} \quad \forall g \in \{1, \ldots, G\}$$

(iii) **Guideline 3 - Case-Mix Specialization:** Lastly, it is important to encourage hospitals to specialize in DRGs (Fetter, 1991), which provide essential guidelines for admissions and transfers within the network. To achieve this, a parameter $\xi_h^g > 0$ can be calibrated for each hospital h in the network and DRG $g \in \{1, \ldots, G\}$, based on the average number of active cases. The purpose of ξ_h^g is to penalize admission actions, thus promoting DRG specialization. A higher value of ξ_h^g means that hospital h is less likely to admit a patient of DRG $g \in \{1, \ldots, G\}$. Using the notation from Section 4, we must add these terms to the objective function:

$$\xi_h^g \sum_{u \in \mathbb{U}(g)} \left(\kappa_u^h + \sum_{n \in N_2} x_u^{hEDn} \right) \quad \forall (h, g) \in \{1, \ldots, H\} \times \{1, \ldots, G\}$$

Where:

$$\mathbb{U}(g) = \{u \in P : g(u) = g\} \quad \forall g \in \{1, \ldots, G\}$$

The proposed term considers the sum of admissions from the WL, as well as those from the ED.

Given the set of guidelines described previously, the policy π_K can be formulated by modifying the MDP proposition as follows:

$$C(\vec{s_1}) = \min_{\vec{a}\in\mathbb{X}(\vec{s_1})} \left\{ c(\vec{s_1},\vec{a}) + O^*_* u^*_* + \sum_{h=1}^{H}\sum_{g=1}^{G} \left[\xi^g_h \sum_{u\in\mathbb{U}(g)} \left(\kappa^h_u + \sum_{n\in N_2} x^{hEDn}_u \right) \right] \right\} \quad \forall \vec{s_1} \in \mathbb{S}_1$$

$$\frac{\alpha_2^{*hn}}{\Phi^n_h} \leq \theta^n_h + u^n_h \quad \forall (h,n) \in \{1,\dots,H\} \times N$$

$$u^n_h \geq 0 \quad \forall (h,n) \in \{1,\dots,H\} \times N$$

$$\sum_{u\in\mathbb{U}(g)} \omega^u_2 \leq \bar{\omega}^g \quad \forall g \in \{1,\dots,G\}$$

$$\mathbb{U}(g) = \{u \in P : g(u) = g\} \quad \forall g \in \{1,\dots,G\}$$

The first new term of the Objective Function is the penalization of congestion over the ideal levels. In contrast, the second term is the penalization of the admissions to promote DRG specialization. It is important to remark that, in this case, there is no $\mathbb{Q}(\vec{s_2})$ term.

6.5.1. New Policy Costs

The simulation results for 50 independent runs throughout 1,300 periods are presented in Figure 6.10. The sampling is the same as that of the previous two policies studied. In terms of model complexity, the changes are negligible compared to the myopic formulation. They have the same characteristics in execution times.

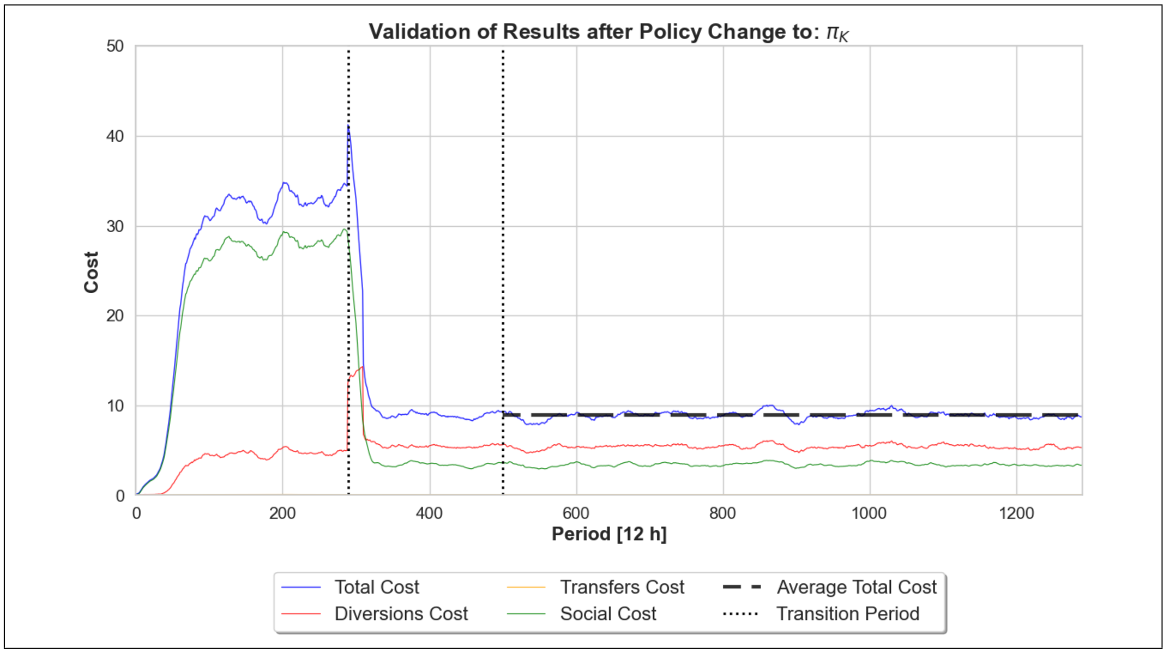

Figure 6.10. Results for Policy π_K

From the figure, it can be concluded that transfer costs are negligible compared to diversions and social costs. Policy π_K experiences a similar rough start during the initial phase of the transition, compared to policy $\pi_{\mathbb{Q}}$. Although there is an increase in the diversions cost, the social cost is reduced at the point where the total cost is significantly lower, compared to the policy π_M. However, results are not as good, when compared to policy $\pi_{\mathbb{Q}}$.

The following table summarizes the results, as well as a statistical comparison between the different cost items between $\pi_{\mathbb{Q}}$ and π_K. Each policy results are presented as means, with coefficients of variation in parentheses.

Table 6.20. Summary of Policies $\pi_{\mathbb{Q}}$ and π_K Results

	Total Cost	**Diversions Cost**	**Transfers Cost**	**Social Cost**
$\pi_{\mathbb{Q}}$	7.9148 (0.0606)	5.1354 (0.0334)	0.0571 (0.0455)	2.7225 (0.1314))
π_K	8.8825 (0.045)	5.4064 (0.0462)	0.0767 (0.0222)	3.3994 (0.0584)
Difference	[-1.0439, -0.9098]	[-0.3225, -0.2198]	[-0.0201, -0.0191]	[-0.7256, -0.6464]

From the table, we can conclude that there is a statistical increase (95% Confidence) in total costs, with an average variation of 12.22%. This is due to the increase in the diversions cost of 20.73%, coupled with an increase in the social cost of 24.86%. In general terms, the results are promising, considering that the new policy is extremely simple, compared to the one with the Q-learning methodology.

The following table summarizes the results, as well as a statistical comparison between the different cost items between π_M and π_K. The results of each policy are presented as means, along with their corresponding coefficients of variation shown in parentheses.

Table 6.21. Summary of Policies π_M and π_K Results

	Total Cost	**Diversions Cost**	**Transfers Cost**	**Social Cost**
π_M	31.1043 (0.0504)	4.2535 (0.1012)	0.0806 (0.0422)	26.7708 (0.0438)
π_K	8.8825 (0.045)	5.4064 (0.0462)	0.0767 (0.0222)	3.3994 (0.0584)
Difference	[22.096, 22.4436]]	[-1.2267, -1.088]	[0.0032, 0.0044]	[23.2914, 23.5562]

From the table, we can conclude that, effectively, there is a significant statistical reduction (95% Confidence) in total costs, with an average reduction of 71.44%. This is achieved by increasing the diversions cost by 27.10%, which allows a reduction of the social cost of 87.30%. It is worth noting that the benefits of policy π_k are almost the same as the benefits of policy $\pi_{\mathbb{Q}}$ when compared to the base scenario. This means simplifying the policy does not significantly impact system benefits.

6.5.2. New Policy Case-Mix Report

To better understand the results, a Case-Mix Report is proposed. We will consider the reference system the base scenario with policy π_M, while the case study will consist of the system with policy π_K.

Table 6.22. Case-Mix Report

	A	P	a	p	a-A	p-P	Δ LOS	Δ CM
DRG 1	16.54	0.1497	16.31	0.1587	-0.23	0.009	-0.04	0.15
DRG 2	19.92	0.1434	19.69	0.1508	-0.23	0.0074	-0.03	0.15
DRG 3	7.98	0.0722	7.63	0.1409	-0.35	0.0687	-0.04	0.54
DRG 4	14.04	0.1244	13.93	0.1353	-0.11	0.0109	-0.01	0.15
DRG 5	17.11	0.1641	13.54	0.1149	-3.57	-0.0492	-0.50	-0.75
DRG 6	14.28	0.1305	12.1	0.0747	-2.18	-0.0558	-0.22	-0.74
DRG 7	23.94	0.1394	17.74	0.1451	-6.2	0.0057	-0.88	0.12
DRG 8	20.96	0.0763	17.59	0.0796	-3.37	0.0033	-0.26	0.06
Total							**-1.99**	**-0.32**

Based on the table provided, it can be concluded that we achieved an estimated average of 1.99 days saved per patient in terms of operational efficiency, which is 0.48 days lower when compared to policy $\pi_{\mathbb{Q}}$. In column $a - A$, it can be observed that all DRGs reduced their ALOS, which contributed to the Δ LOS sum. Regarding Case Mix, column $p - P$ shows a tendency for the system to increase the proportion of emergency DRGs. Compared to policies π_M and $\pi_{\mathbb{Q}}$, the new policy diverts, on average, 1 extra patient per period, thus explaining the increase in operational expenses. As a result of this change in Case Mix, we saved an average of 0.32 days, pretty similar to policy $\pi_{\mathbb{Q}}$. Combining the changes in LOS and Case Mix, we achieved a total reduction of 2.31 service days with the new policy on average (an average reduction of 13.38%) per patient, which is 20.61% lower, compared to the 2.91 saved by policy $\pi_{\mathbb{Q}}$. This could be explained by the fact that policy π_K guidelines do not explicitly consider patient waiting time and in-hospital management improvements.

In terms of the occupation levels, as well as the active cases, the following figures present that information:

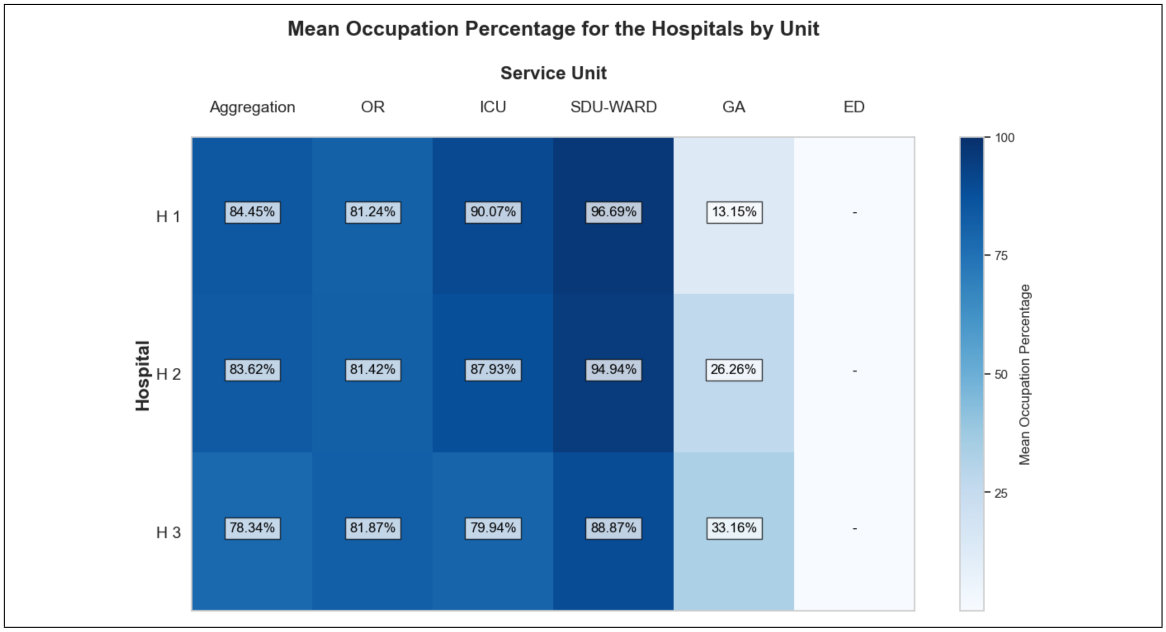

Figure 6.11. Occupancy Rates in the average post-decision state for Policy π_K

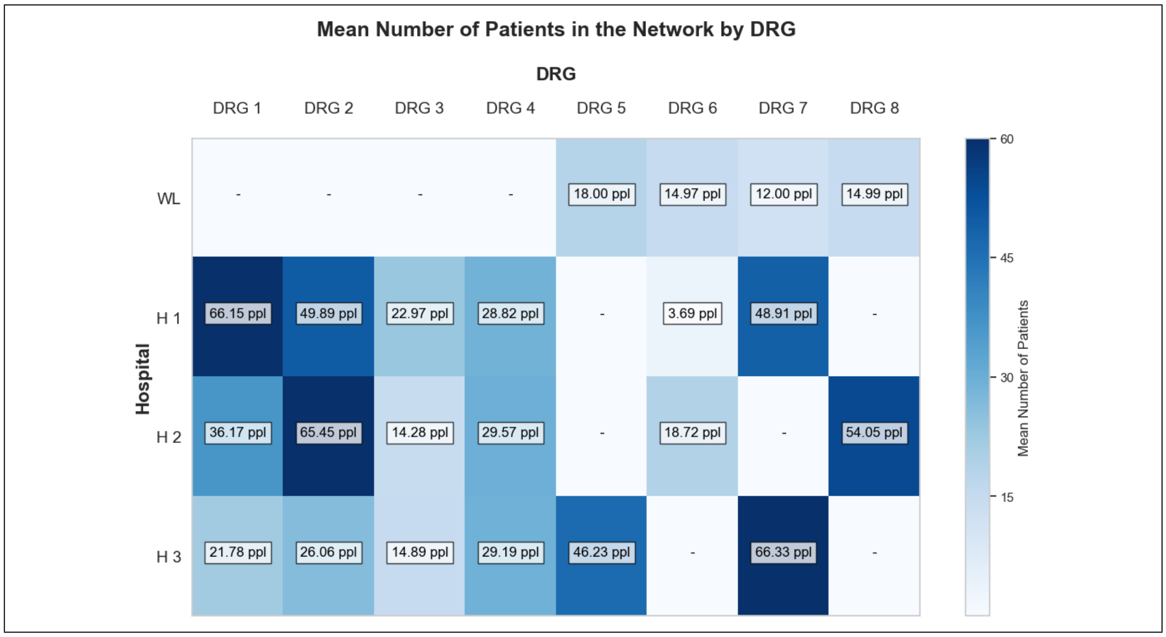

Figure 6.12. Case-Mix in the average post-decision state for Policy π_K

In summary, it can be concluded that the new policy is effectively achieving the desired bed safety stock as recommended by the Q-learning policy. Although there are some differences in occupancy rates, it is still observed that higher-complexity units have lower occupancy rates. On the other hand, regarding the Case-Mix, the specialization of hospitals becomes more evident, and there are some changes in the average number of cases. However, the phenomenon where Hospital 3 specializes in elective DRGs and Hospital 1 specializes in emergency DRGs remains unchanged.

6.5.3. Validating the New Policy

One question regarding the new policy is whether it inherits proactiveness from the Q-learning policy despite being simplified. Although the cost results, occupation levels, and DRG specialization suggest that we have a proactive policy π_K, it is necessary to take a closer look at this issue.

Given a sample of pre-decision states $\vec{s_1}$, retrieved from the simulation process, we will analyze how certain KPIs behave. The first one is the $\mathbb{Q}$ of the post-decision state $\vec{s_2}$ that is achieved, given $\vec{s_1}$. Figure 6.13 presents the results.

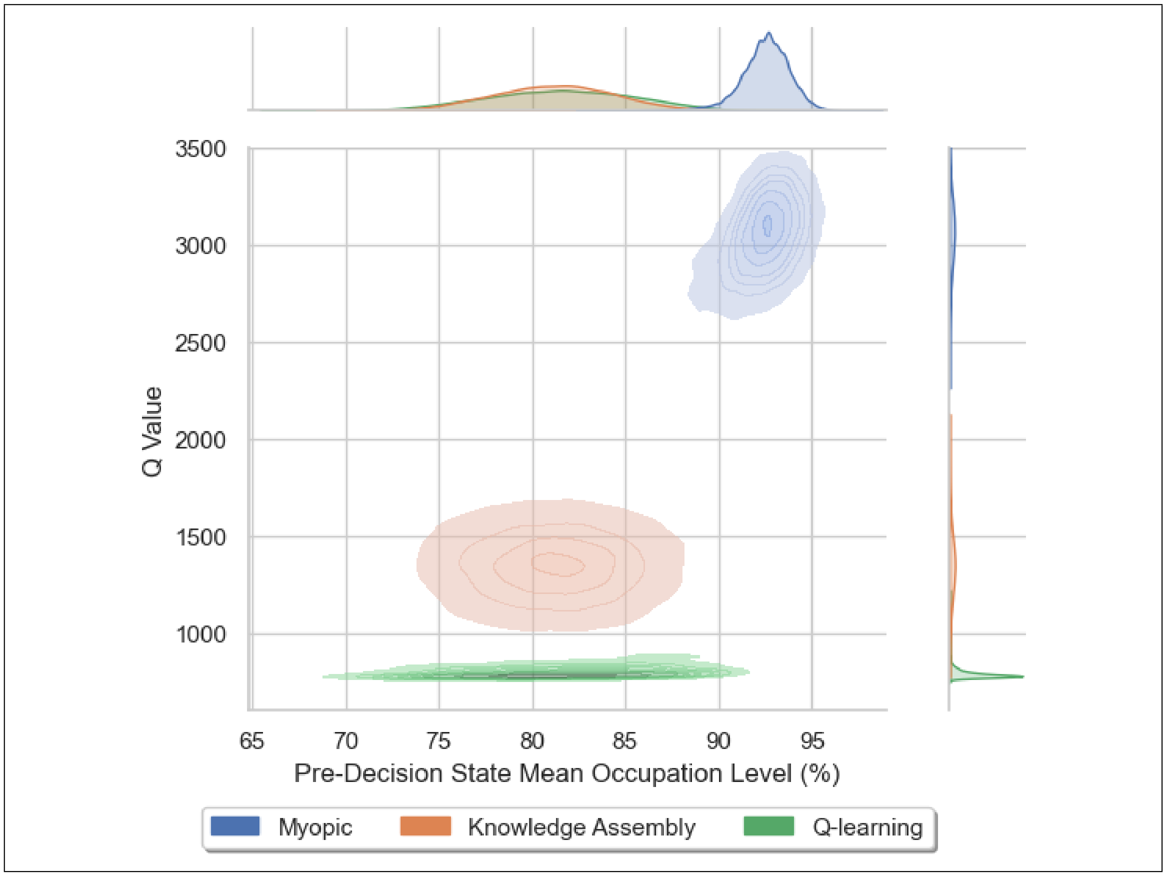

Figure 6.13. $\mathbb{Q}(\vec{s_2})$ Values Achieved from a Sample of Pre-Decision States $\vec{s_1}$

Based on the figure, it is clear that the policy with the highest congestion is π_M. However, policies $\pi_{\mathbb{Q}}$ and π_K have a similar distribution of occupation, with π_K having a lower dispersion. This can be explained by the flexibility that policy $\pi_{\mathbb{Q}}$ provides due to the sophistication of the model. Upon analyzing the $\mathbb{Q}$ predicted by the trained XGBoost model, we can observe that the states from the myopic policy have the highest values. Although the situation for policy π_K suggests worse values as compared to policy $\pi_{\mathbb{Q}}$, we can still see that both policies are closer in terms of values when we compare them with π_M, even for pre-decision states with the same occupation level in the system.

The following figures aim to analyze how the transfers and diversions are being made, considering the expenditure in these actions.

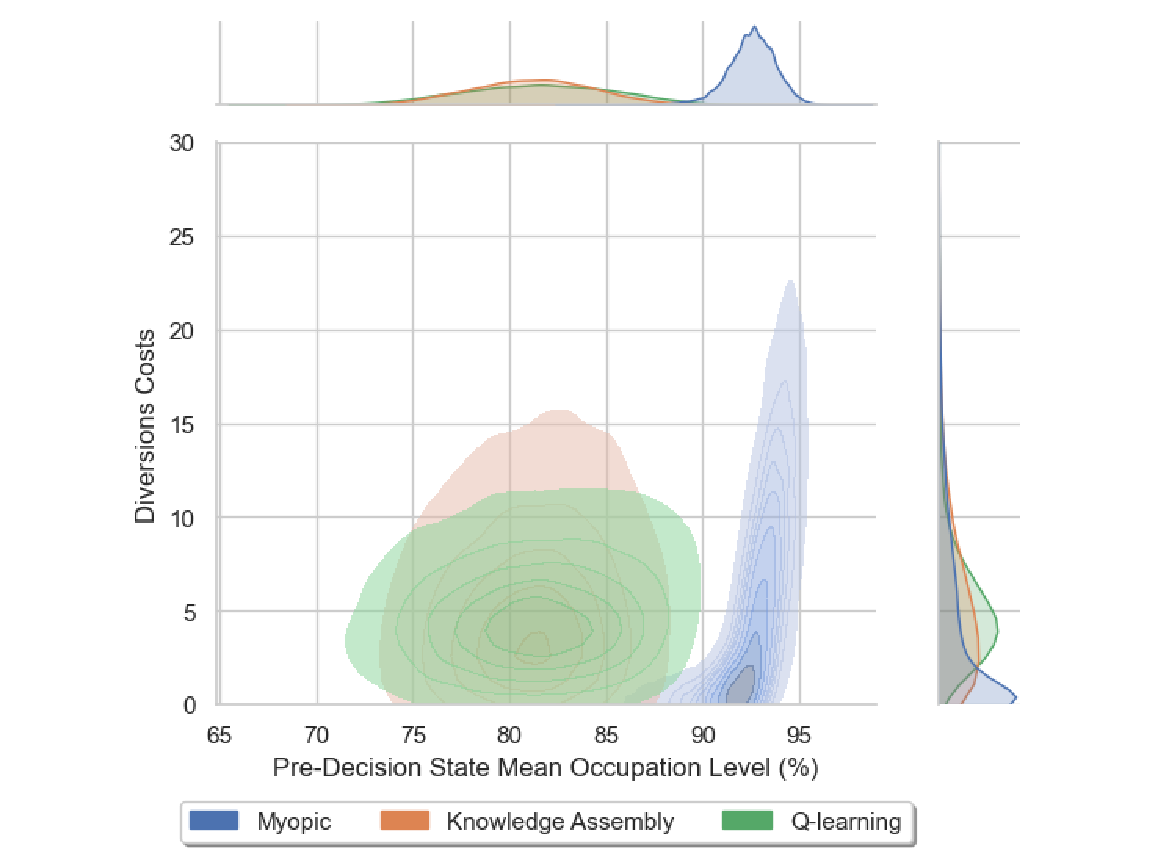

Figure 6.14. Diversions Cost Incurred from a Sample of Pre-Decision States $\vec{s_1}$

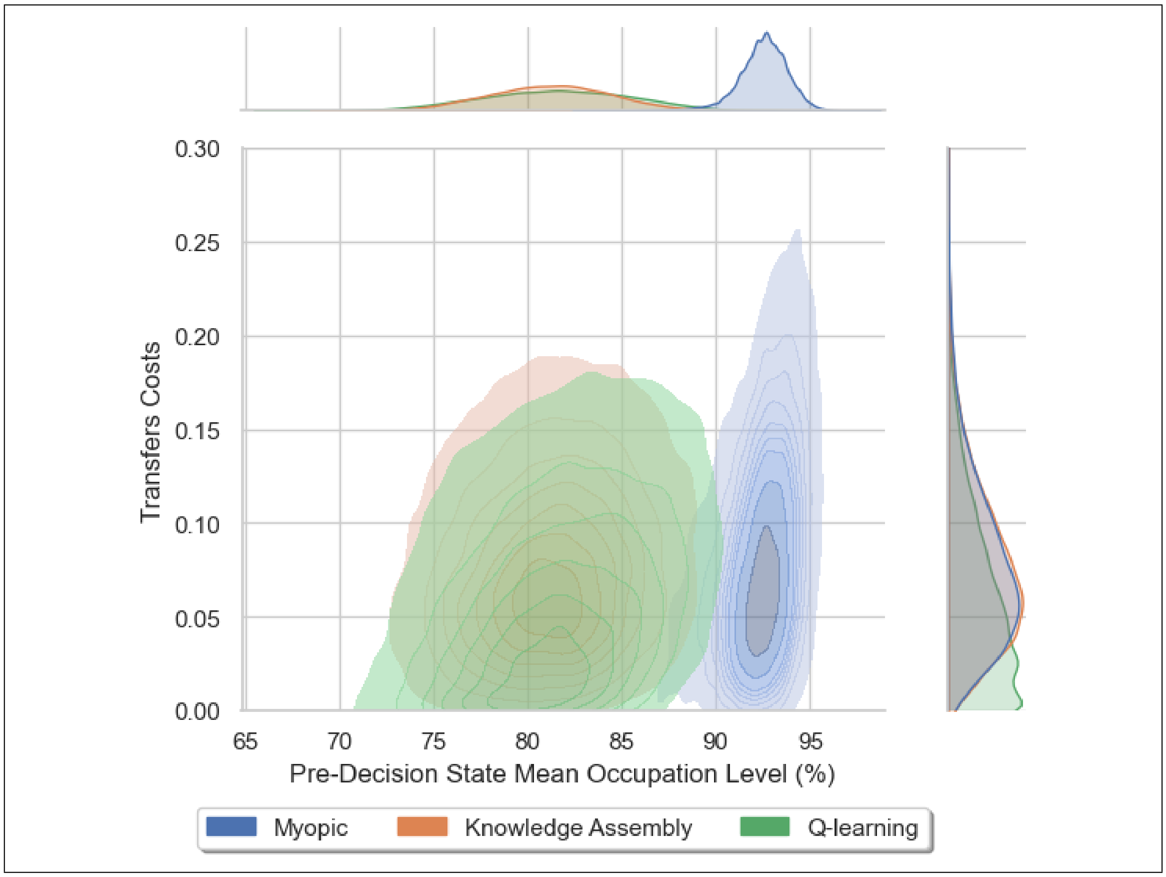

Figure 6.15. Transfer Cost Incurred from a Sample of Pre-Decision States $\vec{s_1}$

From both figures, we can determine that the expenditure for both policies $\pi_{\mathbb{Q}}$ and π_K follow a similar behavior. Obviously, given the simplicity of policy π_K, it tends to expend more to manage the system. On the other hand, the myopic policy avoids spending on these actions until there is now space left to accommodate the demand, evidencing an abrupt curve at higher occupation levels.

It is correct to conclude that policy π_K has inherited a great amount from policy $\pi_{\mathbb{Q}}$ proactiveness, as evidenced by their similarities presented in the results section, as well as the analysis in this one.

6.6. Some Extensions to the Obtained Policy

After testing and concluding that π_K is a great policy, considering the proactiveness behind it, as well as the results we obtain with it, two major questions arise in terms of resilience, which are going to be briefly analyzed and discussed in this section.

6.6.1. Policy Resilience Over Time

Our first question is: ¿How does the policy π_K behave over time? To test it, we considered a scenario with increasing demand, where the expected arrivals increase 5% annually. At the transition period, $t = 0$, the initial values are taken from the stationary execution of the policy. The results are as follows:

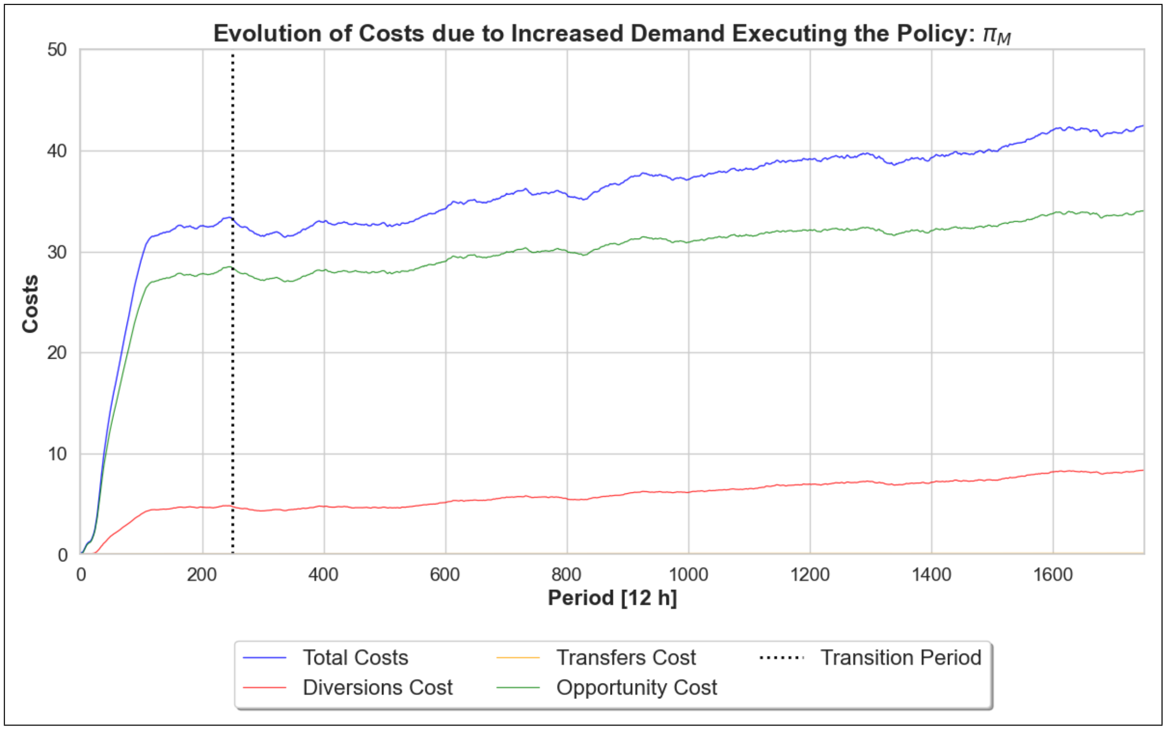

Figure 6.16. Evolution of Cost Over Time for Policy π_M

Table 6.23. Evolution of Cost Over Time for Policy π_M

	$t = 0$	$t = 200$	$t = 400$	$t = 600$	$t = 800$	$t = 1000$	$t = 1200$	$t = 1400$	$t = 1500$
Diversions	4.25	4.71	5.12	5.51	6.13	6.95	7.14	8.14	8.33
Transfers	0.08	0.08	0.08	0.09	0.09	0.10	0.10	0.11	0.11
Social	26.77	28.11	29	30.05	30.91	32.13	32.07	33.75	34.06
Total	31.10	32.90	34.20	35.65	37.13	39.18	39.31	42	42.5

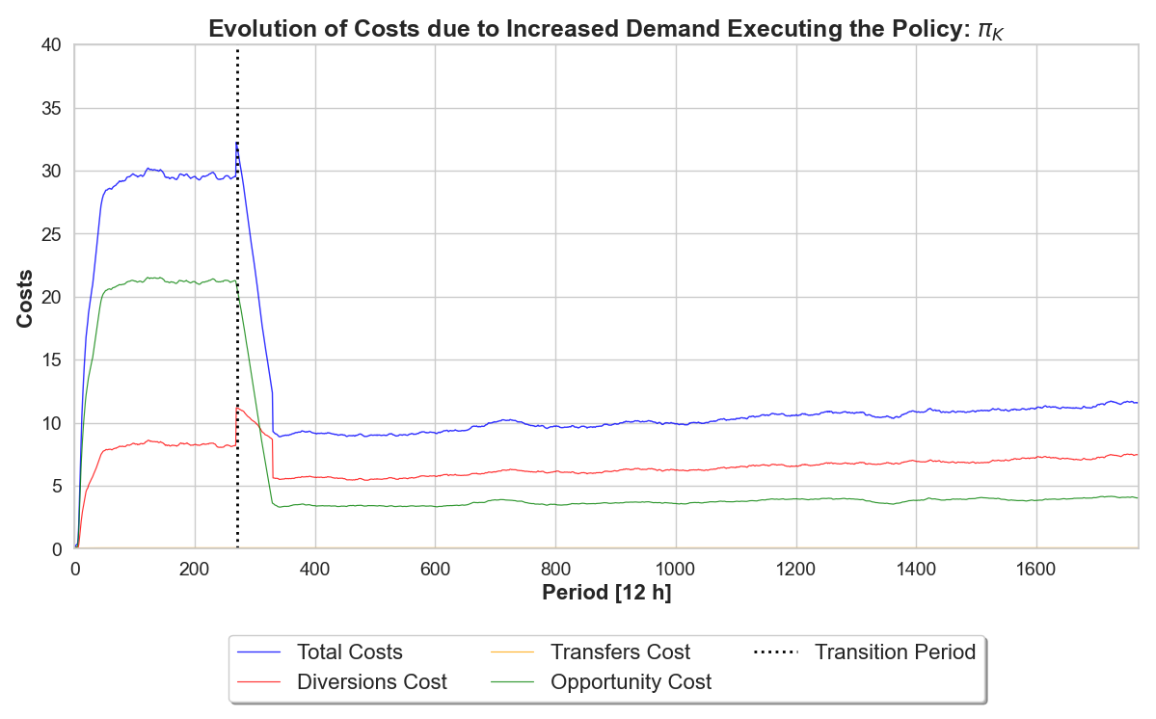

Figure 6.17. Evolution of Cost Over Time for Policy π_K

Table 6.24. Evolution of Cost Over Time for Policy π_K

	$t = 0$	$t = 200$	$t = 400$	$t = 600$	$t = 800$	$t = 1000$	$t = 1200$	$t = 1400$	$t = 1500$
Diversions	5.41	5.67	5.75	6.11	6.23	6.56	6.93	7.25	7.46
Transfers	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Social	3.39	3.41	3.34	3.52	3.59	3.97	3.84	3.84	4.05
Total	8.88	9.16	9.17	9.71	9.90	10.61	10.85	11.17	11.59

It is determined that the increase in demand affects both policies. Interestingly, although π_K starts with a higher diversion cost, after 1,000 periods (500 days), policy π_M starts spending more on these actions.

If we study the variation (with respect to $t = 0$), we obtain the following tables:

Table 6.25. Evolution of Cost Variation Over Time for Policy π_M

	$t = 0$	$t = 200$	$t = 400$	$t = 600$	$t = 800$	$t = 1000$	$t = 1200$	$t = 1400$	$t = 1500$
Diversions	100%	111%	120%	130%	144%	164%	168%	192%	196%
Transfers	100%	100%	100%	113%	113%	125%	125%	138%	138%
Social	100%	105%	108%	112%	115%	120%	120%	126%	127%
Total	100%	106%	110%	115%	119%	126%	126%	135%	137%

Table 6.26. Evolution of Cost Variation Over Time for Policy π_K

	$t = 0$	$t = 200$	$t = 400$	$t = 600$	$t = 800$	$t = 1000$	$t = 1200$	$t = 1400$	$t = 1500$
Diversions	100%	105%	106%	113%	115%	121%	128%	134%	138%
Transfers	100%	100%	100%	100%	100%	100%	100%	100%	100%
Social	100%	101%	99%	104%	106%	117%	113%	113%	119%
Total	100%	103%	103%	109%	111%	119%	122%	126%	131%

From these tables, we determine that the policy π_K can persist over time in an extremely saturated environment. This gives us the intuition that, for this problem, it is not crucial to be constantly updating the policy if we have a steadily increasing demand. This is important, considering that updating the new policy could take more than a couple of days.

6.6.2. Policy Resilience Over Sub-Optimal Actions

Another question is: ¿How does the policy π_K behave over time? To test it, we considered different simulation scenarios, where each one had a probability of failing to implement the π_K policy and ended up implementing π_M at each period. The following figure presents the results obtained:

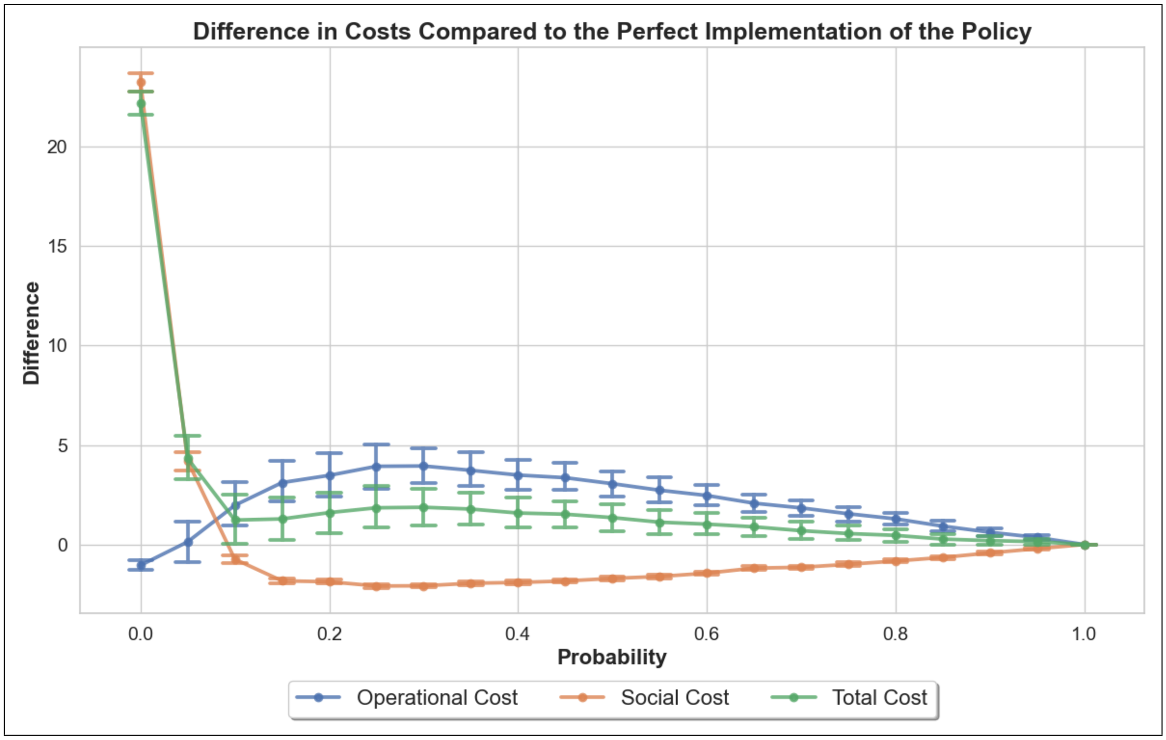

Figure 6.18. Difference in Costs Compared to the Perfect Implementation of the Policy π_K

Firstly, the volatility of the results is dependent on the proportion of times we implement actions from policy π_K. A higher proportion of correct implementation leads to lower volatility.

In terms of social cost, is interesting the effect that some incipient implementation of the guidelines from π_K could have in the network. There is an opportunity to understand, in further research, how we could balance these policies in order to minimize the opportunity cost, which is better than the strictly π_K policy in the range from $[0.1,\ 1)$, presented in the negative values.

On the other hand, the operational cost (i.e., transfers plus diversions) shows us an undesirable situation where the π_K is trying to reconfigure the system, but the policy π_M does not allow it. A higher proportion of correct implementation of policy π_K leads to lower differences with respect to the perfect implementation.

Finally, based on the total cost π_K, we can determine that the policy is robust against undesirable and incorrect actions, where errors up to 10% (i.e., an implementation of the actions from π_K 90% of the time) do not have significant impacts in the objective function.

7. CONCLUSIONS

7.1. Contributions

This thesis proposes a new approach to address the Approximate Dynamic Problem that arises in the context of a complex hospital network, using reinforcement learning. The problem is formulated mathematically as a Markov Decision Process (MDP), which enables handling transfers, diversions, and admissions across the network as well as in-patient management at each hospital. By implementing this methodology, we were able to move from a reactive policy, where actions are taken only when there is no room left, to a proactive policy. This transition helped to reduce both operational and social costs, improving patient experience.

Furthermore, based on expert knowledge and the results from the proactive policy learned, a simplified policy proposal was made based on 3 principal guidelines: (i) Recommended Occupation Levels, (ii) Waiting List Size Management, and (iii) Case-Mix Specialization by Hospital. The results from this new policy are pretty similar, compared to the reinforcement learning one. Also, the simplified policy was tested and evaluated in terms of resilience, concluding it has satisfactory characteristics. Although designed for the Chilean reality, this model and related methodology can be applied to other situations where the healthcare system aims to operate similarly.

In conclusion, this thesis contributes in the following aspects:

(i) A model of a hospital network is presented, taking into account interactions between a Waiting List, a public hospital network with different types of units, and private clinics.

(ii) A novel solution approach for Approximate Dynamic Programming, which can handle the non-linear relations among variables and the future expected costs.

(iii) Validation of hospital management guidelines, as well as an opportunity to couple algorithms techniques with expert knowledge to achieve simple but effective policies to manage the healthcare system.

7.2. Limitations

The most significant limitation regarding this work is the assumption that all the system information is available. This situation challenges the CCMU labor, requiring a centralized effort among the hospital network to share and maintain updated information and the disposition to collaborate as a network. In the Chilean context, not only is this required, but a major overhaul of technology is needed to quickly and reliably determine crucial patient information, such as the patient DRG.

Another limitation is the availability and correctness of the information. As presented in this research, specific parameters require expert knowledge. This situation raises the question of how particular parameters, if incorrectly estimated, could alter the results. For instance, regarding the danger curve due to waiting times, we could promote hazardous situations for certain types of patients if poorly estimated.

Based on the previous point, it is essential to collaborate with medical staff. Some of the guidelines that can be obtained could have moral implications because the decisions being made impact people's integrity. Just consider the case where the policy determines that a person in a risky situation must wait for the benefit of the whole system. In this situation, interdisciplinary collaboration with healthcare professionals is essential from the beginning of the model conceptualization and data gathering to the evaluation and judicious examination of the policy obtained.

Finally, the proposed workflow requires vast computational resources to train and test the results effectively. The computational implementation of this research did not consider parallelization in the simulation and optimization runs, giving up space to important reductions in time, if implemented. Additionally, we tested the model considering only 8

DRG groups, whereas there are more than 300 different groups. Although execution times are promising, working with aggregation levels, such as the Major Diagnostic Categories (Fetter, 1991), could be an option to address this issue.

7.3. Further Research

Based on the results and challenges, new development opportunities are:

- In this model, we are only considering transfers between EDs. It is possible to consider transfer between other units, even considering that certain types of patients must return to their original hospital.
- In terms of transfers, we have considered an infinite number of resources (i.e., ambulances) to make transfers and diversions in the network. We may wonder how the policy changes if we add this type of constraint to the model.
- In terms of patient welfare, there is the possibility to consider patient preferences or complications that affect the decision criteria for diversions and transfers.
- In this problem, we are not considering demand seasonality. An extension to the model could be done in terms of states, or one policy could be developed for each season.
- The model does not consider changes in the demand preferences. As a hospital becomes efficient in certain DRGs, it is reasonable to consider that the demand from the network would try to be treated at that hospital as a result of the best service it provides, generating changes in the probability distribution of demand generation.
- The model considers that the beds from the system are always available. It raises the challenge of considering the variation in bed availability over time for various reasons (staffing, resource increase, resource failures, among others).
- The proposed model could be extended to evaluate capacity increases, either in the number of beds for a hospital or the incorporation/construction of a new facility in the network.

- In this model, we are optimizing the considering the expected discounted value. New developments have explored the possibility of achieving a distributional approach, in which the distribution over returns is modeled explicitly (Dabney et al., 2018). This could enable robust optimization, as well as optimization with risk aversion.

7.4. Final Remarks

Universal health coverage is crucial for the world, and we must optimize healthcare resources to achieve this by 2030. Chile's healthcare system faces significant systemic issues, as reflected in appalling statistics on bed availability and patient waiting times. Although the CCMU is a step in the right direction, much work remains to be done. We must act innovatively and prioritize people's well-being by forging partnerships and embracing cutting-edge research. Success is not just about efficient operations but also about tangible impacts on people's lives. We cannot leave anyone behind in their hour of need.

REFERENCES

Ahalt, V., Argon, N. T., Ziya, S., Strickler, J., & Mehrotra, A. (2016). Comparison of emergency department crowding scores: a discrete-event simulation approach. *Health Care Management Science*, *21*(1), 144–155. doi: 10.1007/s10729-016-9385-z

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining.* doi: 10.1145/3292500.3330701

Astaraky, D., & Patrick, J. (2015). A simulation based approximate dynamic programming approach to multi-class, multi-resource surgical scheduling. *European Journal of Operational Research*, *245*(1), 309–319. doi: 10.1016/j.ejor.2015.02.032

Banditori, C., Cappanera, P., & Visintin, F. (2013). A combined optimization-simulation approach to the master surgical scheduling problem. *IMA Journal of Management Mathematics*, *24*(2), 155–187. doi: 10.1093/imaman/dps033

Batista, A., Pozo, D., & Vera, J. (2021). Managing the unknown: A distributionally robust model for the admission planning problem under uncertain length of stay. *Computers amp; Industrial Engineering*, *154*, 107041. doi: 10.1016/j.cie.2020.107041

Bertsimas, D., & Dunn, J. (2017). Optimal classification trees. *Mach. Learn.*, *106*(7), 1039–1082. doi: https://doi.org/10.1007/s10994-017-5633-9

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi: 10.1023/a:1010933404324

Burdett, R. L., Kozan, E., Sinnott, M., Cook, D., & Tian, Y.-C. (2017). A mixed integer linear programing approach to perform hospital capacity assessments. *Expert Systems* *with Applications*, *77*, 170–188. doi: 10.1016/j.eswa.2017.01.050

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.* ACM. doi: 10.1145/2939672.2939785

Colegio Médico de Chile. (2019). *Falta de camas hospitalarias: Un mal endémico.* Retrieved from `https://www.regionalsantiago.cl/falta-de-camas-hospitalarias-un-mal-endemico/` ([Accessed 12-12-2023])

Dabney, W., Rowland, M., Bellemare, M., & Munos, R. (2018). Distributional reinforcement learning with quantile regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). doi: 10.1609/aaai.v32i1.11791

de Bruin, A. M., van Rossum, A. C., Visser, M. C., & Koole, G. M. (2007). Modeling the emergency cardiac in-patient flow: an application of queuing theory. *Health Care Management Science*, *10*(2), 125–137. doi: 10.1007/s10729-007-9009-8

Devapriya, P., Strömblad, C. T. B., Bailey, M. D., Frazier, S., Bulger, J., Kemberling, S. T., & Wood, K. E. (2015). Stratbam: A discrete-event simulation model to support strategic hospital bed capacity decisions. *Journal of Medical Systems*, *39*(10). doi: 10.1007/s10916-015-0325-0

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. doi: 10.1037/xge0000033

Feng, Y.-Y., Wu, I.-C., & Chen, T.-L. (2015). Stochastic resource allocation in emergency departments with a multi-objective simulation optimization algorithm. *Health Care Management Science*, *20*(1), 55–75. doi: 10.1007/s10729-015-9335-1

Fetter, R. B. (1991). Diagnosis related groups: Understanding hospital performance. *Interfaces*, *21*(1), 6–26. doi: 10.1287/inte.21.1.6

Fetter, R. B., Shin, Y., Freeman, J. L., Averill, R. F., & Thompson, J. D. (1980). Case mix definition by diagnosis-related groups. *Medical Care*, *18*(2), i–53. Retrieved 2023-10-02, from `http://www.jstor.org/stable/3764138`

Gartner, D., Kolisch, R., Neill, D. B., & Padman, R. (2015). Machine learning approaches for early drg classification and resource allocation. *INFORMS Journal on Computing*, *27*(4), 718–734. doi: 10.1287/ijoc.2015.0655

González, J., Ferrer, J.-C., Cataldo, A., & Rojas, L. (2018). A proactive transfer policy for critical patient flow management. *Health Care Management Science*, *22*(2), 287–303. doi: 10.1007/s10729-018-9437-7

Green, L. (2005). Capacity planning and management in hospitals. In *Operations research and health care* (p. 15–41). Springer US. doi: 10.1007/1-4020-8066-2_2

Green, L. (2006). Queueing analysis in healthcare. In *Patient flow: Reducing delay in healthcare delivery* (p. 281–307). Springer US. doi: 10.1007/978-0-387-33636-7_10

Gupta, P. D. (1978). A general method of decomposing a difference between two rates into several components. *Demography*, *15*(1), 99–112. doi: 10.2307/2060493

Gurobi Optimization. (2023). *Gurobi machine learning documentation.* Retrieved from `https://gurobi-machinelearning.readthedocs.io/en/stable/` ([Accessed 18-10-2023])

Humphreys, P., Spratt, B., Tariverdi, M., Burdett, R. L., Cook, D., Yarlagadda, P. K. D. V., & Corry, P. (2022). An overview of hospital capacity planning and optimisation. *Healthcare*, *10*(5), 826. doi: 10.3390/healthcare10050826

Instituto de Sistemas Complejos de Ingeniería. (2020). *Regulación de flujo de pacientes.* Retrieved from `https://isci.cl/wp-content/uploads/2020/04/Reporte_Regulaci%C3%B3n-de-flujo-de-pacientes-ISCI_VMarianov-2.pdf` ([Accessed 12-12-2023])

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association*, *50*(272), 1168. doi: 10.2307/2281213

Lee, E. K., Atallah, H. Y., Wright, M. D., Post, E. T., Thomas, C., Wu, D. T., & Haley, L. L. (2015). Transforming hospital emergency department workflow and patient care. *Interfaces*, *45*(1), 58–82. doi: 10.1287/inte.2014.0788

Lin, L.-J. (1992). Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, *8*(3–4), 293–321. doi: 10.1007/bf00992699

Liu, J., Capurro, D., Nguyen, A., & Verspoor, K. (2021). Early prediction of diagnostic-related groups and estimation of hospital cost by processing clinical notes. *npj Digital Medicine*, *4*(1). doi: 10.1038/s41746-021-00474-9

Madeira, A., Moutinho, V., & Fuinhas, J. A. (2021). Does waiting times decrease or increase operational costs in short and long-term? evidence from portuguese public hospitals. *The European Journal of Health Economics*, *22*(8), 1195–1216. doi: 10.1007/s10198-021-01331-y

Maragno, D., Wiberg, H., Bertsimas, D., Birbil, , den Hertog, D., & Fajemisin, A. O. (2023). Mixed-integer optimization with constraint learning. *Operations Research*. doi: 10.1287/opre.2021.0707

Marquinez, J. T., Sauré, A., Cataldo, A., & Ferrer, J.-C. (2021). Identifying proactive ICU patient admission, transfer and diversion policies in a public-private hospital network. *European Journal of Operational Research*, *295*(1), 306–320. doi: 10.1016/j.ejor.2021.02.045

Martinez, D. A., Zhang, H., Bastias, M., Feijoo, F., Hinson, J., Martinez, R., . . . Prieto, D. (2019). Prolonged wait time is associated with increased mortality for chilean waiting list patients with non-prioritized conditions. *BMC Public Health*, *19*(1). doi: 10.1186/s12889-019-6526-6

McManus, M. L., Long, M. C., Cooper, A., & Litvak, E. (2004). Queuing theory accurately models the need for critical care resources. *Anesthesiology*, *100*(5), 1271–1276. doi: 10.1097/00000542-200405000-00032

Min, D., & Yih, Y. (2010). Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research*, *206*(3), 642–652. doi: 10.1016/j.ejor.2010.03.014

Ministerio de Salud de Chile. (2023a). *Glosa 06 - lista de espera no ges y garantías de oportunidad ges retrasadas.* Retrieved from `https://www.minsal.cl/wp-content/uploads/2021/05/Ord.-331-Glosa-06-IV-Trim-2022.pdf` ([Accessed 12-12-2023])

Ministerio de Salud de Chile. (2023b). *Unidad de gestión centralizada de casos (ugcc), división de gestión de la red asistencial, subsecretaría de redes asistenciales.* Retrieved from `https://www.minsal.cl/unidad-de-gestion-centralizada-de-casos-ugcc-division-de-gestion-de-la-red-asistencial-subsecretaria-de-redes-asistenciales/` ([Accessed 12-12-2023])

National Guideline Centre - UK. (2018). *Emergency and acute medical care in over 16s: Service delivery and organisation* (No. 94). London: National Institute for Health and Care Excellence (NICE). Retrieved from `https://www.ncbi.nlm.nih.gov/books/NBK564911/` ([Accessed 10-03-2024])

OECD. (2023). *Health at a glance 2023: Oecd indicators.* Author. doi: 10.1787/7a7afb35-en

Plunkett, P. K., Byrne, D. G., Breslin, T., Bennett, K., & Silke, B. (2011). Increasing wait times predict increasing mortality for emergency medical admissions. *European Journal of Emergency Medicine*, *18*(4), 192–196. doi: 10.1097/mej.0b013e328344917e

Powell, W. B. (2011). *Approximate dynamic programming: Solving the curses of dimensionality* (2nd ed.). Wiley-Blackwell. doi: 10.1002/9781118029176

Powell, W. B. (2012). Perspectives of approximate dynamic programming. *Annals of Operations Research*, *241*(1–2), 319–356. doi: 10.1007/s10479-012-1077-6

Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., & Rademakers, F. E. (2016). Scheduling operating rooms: achievements, challenges and pitfalls. *Journal of Scheduling*, *19*(5), 493–525. doi: 10.1007/s10951-016-0489-6

Seidel, J., Whiting, P., & Edbrooke, D. (2006). The costs of intensive care. *Continuing Education in Anaesthesia Critical Care amp; Pain*, *6*(4), 160–163. doi: 10.1093/bjaceaccp/mkl030

Sun, B. C., Hsia, R. Y., Weiss, R. E., Zingmond, D., Liang, L.-J., Han, W., … Asch, S. M. (2013). Effect of emergency department crowding on outcomes of admitted patients. *Annals of Emergency Medicine*, *61*(6), 605–611.e6. doi: 10.1016/j.annemergmed.2012.10.026

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*(1), 9–44. doi: 10.1007/bf00115009

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning* (2nd ed.). Cambridge, MA: Bradford Books.

Tan, S. S., Bakker, J., Hoogendoorn, M. E., Kapila, A., Martin, J., Pezzi, A., … Hakkaart-van Roijen, L. (2012). Direct cost analysis of intensive care unit stay in four european countries: Applying a standardized costing methodology. *Value in Health*, *15*(1), 81–86. doi: 10.1016/j.jval.2011.09.007

Tiwari, V., & Heese, H. S. (2009). Specialization and competition in healthcare delivery networks. *Health Care Management Science*, *12*(3), 306–324. doi: 10.1007/s10729-008-9096-1

Truong, V.-A. (2015). Optimal advance scheduling. *Management Science*, *61*(7), 1584–1597. doi: 10.1287/mnsc.2014.2067

Ulmer, M. W. (2017). *Approximate dynamic programming for dynamic vehicle routing* (1st ed.). Cham, Switzerland: Springer International Publishing.

United Nations. (2020). *The 17 goals — sustainable development.* Retrieved from `https://sdgs.un.org/goals` ([Accessed 12-12-2023])

Van Hasselt, H., Guez, A., & Silver, D. (2016). Deep reinforcement learning with double q-learning. *Proc. Conf. AAAI Artif. Intell.*, *30*(1).

Van Riet, C., & Demeulemeester, E. (2015). Trade-offs in operating room planning for electives and emergencies: A review. *Operations Research for Health Care*, *7*, 52–69. doi: 10.1016/j.orhc.2015.05.005

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3–4), 279–292. doi: 10.1007/bf00992698

Wendt, C. (2009). Mapping european healthcare systems: a comparative analysis of financing, service provision and access to healthcare. *J. Eur. Soc. Policy*, *19*(5), 432–445. doi: https://doi.org/10.1177/09589287093442

Zapata, M. (2018). Importancia del sistema grd para alcanzar la eficiencia hospitalaria. *Revista Médica Clínica Las Condes*, *29*(3), 347-352. (Tema central: Enfermería) doi: https://doi.org/10.1016/j.rmclc.2018.04.010

Zhang, J., Dridi, M., & Moudni, A. E. (2021). An approximate dynamic programming approach to the admission control of elective patients. *Computers &amp Operations Research*, *132*, 105259. doi: 10.1016/j.cor.2021.105259

Zhang, X., Barnes, S., Golden, B., Myers, M., & Smith, P. (2019). Lognormal-based mixture models for robust fitting of hospital length of stay distributions. *Operations Research* *for Health Care*, *22*, 100184. doi: 10.1016/j.orhc.2019.04.002

Zhang, X., Dupre, M. E., Qiu, L., Zhou, W., Zhao, Y., & Gu, D. (2017). Urban-rural differences in the association between access to healthcare and health outcomes among older adults in china. *BMC Geriatr.*, *17*(1), 151. doi: 10.1371/journal.pone.0240194

Zhou, L., Geng, N., Jiang, Z., & Wang, X. (2018). Multi-objective capacity allocation of hospital wards combining revenue and equity. *Omega*, *81*, 220–233. doi: 10.1016/j.omega.2017.11.005

Zhu, S., Fan, W., Yang, S., Pei, J., & Pardalos, P. M. (2018). Operating room planning and surgical case scheduling: a review of literature. *Journal of Combinatorial Optimization*, *37*(3), 757–805. doi: 10.1007/s10878-018-0322-6