Quality assessment of publicly-managed and PPP Portuguese hospitals

Margarida Matos Lopes

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisors: Prof. José Rui de Matos Figueira
Prof. Diogo Filipe da Cunha Ferreira

Examination Committee

Chairperson: Prof. Cláudia Alexandra Martins Lobato da Silva
Supervisor: Prof. Diogo Filipe da Cunha Ferreira
Member of the Committee: Prof. Ana Sara Silva Rodrigues da Costa

December 2020

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Preface

The work presented in this thesis was performed at the Institute for Bioengineering and Biosciences of Instituto Superior Técnico (Lisbon, Portugal), during the period February-December 2020, under the supervision of Prof. José Rui de Matos Figueira and Prof. Diogo Filipe da Cunha Ferreira and within the frame of the hSNS FCT - Research Project (PTDC/EGEOGE/30546/2017): Portuguese public hospital performance assessment using a multicriteria decision analysis framework.

Acknowledgments

Besides the (endless) months working on this dissertation, I spent the last five years in Instituto Superior Técnico which I could not, at all, have done alone.

Then, I will take this opportunity to show appreciation for some of my main support systems.

Firstly, thank you to my family for being the most loving, supportive and united family anyone could ask for. Thanks mom and dad for always being good examples and role models for me to follow and become the best person I know how to be. I am proud of all the values you have passed on to us and intend to keep applying them to every sphere of my life.

I am so lucky to have amazing friends to which I am grateful. Margarida, Sofia, Leo, Mada, Carlota, Mila, Rosa, thank you for the 5 years of memories and moments (especially the endless crying, laughter and conversations), I will fondly keep it all with me forever and hope to keep sharing it now we are Técnico-free. ”As Amigas”, grazie, for the best, crazy Erasmus experience I always wanted and the beautiful, genuine friendship I know will last a lifetime. Isabel and Sofia, not only for being my gym buddies and making me laugh but also for the huge support in everything during the last year, I'm so glad you showed up in my life and became part of my inner circle. Filipe, Diogo, Padrinho, Manel, André, for showing me how meaningful and important male friendships can also be. Júlia, Ana, Ritinha, Renata, Carolina, because ”home isn't a place” and I am thankful to have you to come home to every time. It was a wild ride and I'm glad I got to spend it with all of you!

Thank you Prof. José Rui Figueira and Prof. Diogo Ferreira for the guidance, help and patience during the whole process of writing this thesis.

Finally, thank you Dra Sónia Neves for, among many other things, helping me figure out every little step of this journey.

Abstract

The Portuguese SNS (*Serviço Nacional de Saúde*) was created to provide all citizens with universal, equitable, and tendentiously free access to healthcare services. However, these services face big challenges, such as access barriers and the increase of the public health expenses. Several reforms were implemented to counteract these issues, including the creation of Public-Private Partnerships (PPPs) in healthcare, which implied the participation of private parties in the public health sector. This led to questions about their ability to deliver health services with quality, which, in healthcare, is a complex concept, but can be measured through different variables such as access, safety, and care appropriateness. It is also important to assure disinvestment in certain areas does not jeopardize others. Hence, it became apparent the need to compare PPP and publicly-managed hospitals regarding quality. This study relies on a multiple criteria decision analysis approach, using the ELECTRE TRI-nC method. The sample contains data from 2018, covering 30 hospitals, where three are PPPs, and ten criteria under five points of view. Two different models were carried out, one including efficiency criteria and other without, since some of the PPP hospitals did not provide this information. Under the first model, two scenarios were analyzed, one with a social-oriented goal and the other with an efficiency-oriented goal, where the weights of the criteria differed. The results obtained with the method show that there is no evidence that one group of hospitals outperforms the other. In fact, there is margin for improvement for both groups since their performances are not outstanding.

Keywords: Quality, SNS, Hospitals, Public-Private Partnerships, Multicriteria Decision Aiding, ELECTRE TRI-nC

Resumo

O SNS (Serviço Nacional de Saúde) foi criado para proporcionar a todos os cidadãos acesso universal, equitativo e tendencialmente gratuito aos serviços de saúde. Esses serviços enfrentam grandes desafios e várias reformas e políticas foram implementadas para enfrentar esses problemas, incluindo a criação de Parcerias Público-Privadas (PPPs) na área da saúde. A entrada de entidades privadas no setor público de saúde suscitou questões sobre a sua capacidade de prestar serviços de saúde com qualidade. O conceito de qualidade em saúde é complexo, mas pode ser medido através de diversas variáveis como acesso, segurança e adequação dos cuidados prestados. Também é importante garantir que o desinvestimento em certas áreas não prejudique outras. Tornou-se, portanto, evidente a necessidade de comparar PPPs e hospitais públicos nesse âmbito. Este estudo baseia-se numa abordagem multicritério, usando o método ELECTRE TRI-nC. A amostra contém dados de 2018, abrangendo 30 hospitais, dos quais três são PPPs, e dez critérios sob cinco pontos de vista. Foram usados dois modelos distintos, um com critérios de eficiência e outro sem, uma vez que alguns hospitais PPP não forneciam essa informação. No primeiro modelo, foram analisados dois cenários, um com objetivo social e outro com orientação para a eficiência, onde os pesos dos critérios diferiram. Os resultados obtidos com o método ELECTRE TRI-nC mostram que não há evidências de que um grupo de hospitais supere o outro, havendo margem para melhorias para os dois tipos de hospitais, uma vez que os seus desempenhos são mediocres.

Keywords: Qualidade, SNS, Hospitais, Parcerias Público-Privadas, Apoio à Decisão, ELECTRE TRI-nC

Contents

List of Tables

List of Figures

Acronyms

ACSS *Administração Central do Sistema de Saúde*.

ADSE *Assistência na Doença aos Servidores civis do Estado*.

AHP Analytic Hierarchy Process.

AHRQ Agency for Healthcare Research and Quality.

DEA Data Envelopment Analysis.

DM Decision Maker.

EPE *Entidade Pública Empresarial*.

ERS *Entidade Reguladora da Saúde*.

FTE Full Time Equivalent.

GDP Gross Domestic Product.

HC Hospital Centre.

IOM Institute of Medicine.

JRC Joint Research Centre.

MCDA Multiple Criteria Decision Aiding.

NPM New Public Management.

OECD Organization for Economic Cooperation and Development.

PFI Private Finance Initiative.

POV Point of View.

PPP Public-Private Partnership.

RHA Regional Health Administration.

SA *Sociedade Anónima*.

SAMS *Serviços de Assistência Médico Social*.

SNS *Serviço Nacional de Saúde*.

SPA *Sector Público Administrativo*.

SRF Simos Roy Figueira.

UCFTP Unifetal, Cephalicand Full-Term Pregnancy.

UK United Kingdom.

WHO World Health Organization.

Chapter 1

Introduction

This introduction comprises five sections: the context, the motivation to develop this work, the objectives set for this study, the methodology followed to achieve these goals and finally, the outline of the dissertation.

1.1 Context

The ultimate goal of healthcare services is to improve patients' satisfaction and quality of life (Ferreira et al., 2018b).

Healthcare services in Portugal follow a Beveridge model and are provided by three coexisting systems: the National Health Service (SNS, standing for the Portuguese words *Serviço Nacional de Saúde*), the private sector and the social sector, with whom the public sector has collaboration agreements. When it comes to financing, the Portuguese health system also has a mix of public and private methods, being predominantly based on taxes, but with special social health insurance schemes for certain professions and voluntary private health insurance also taking part (Simões et al., 2017).

The Portuguese SNS was created in 1979 and is composed of all public entities delivering primary and secondary healthcare services to the population (Ferreira and Marques, 2019). It aims to provide universal, appropriate, and equitable care to all citizens, regardless of their ability or willingness to pay (Barata et al., 2012). However, these services face big challenges, such as access barriers (for instance concerning waiting lines or availability of resources) (OECD/European Observatory on Health Systems and Policies, 2017) and increase of the consumption of the public health expenses by the hospitals, as a result of different factors (like aging population or technology evolution) (Ferreira and Marques, 2015).

For instance, in the year of 2018, about 9% of Portugal's Gross Domestic Product (GDP) regarded health expenses, more than the average of the Organization for Economic Cooperation and Development (OECD), 8.8%, and approximately 18.3 billion€in total[1] [2]. The evolution of these expenditures, that will be further discussed in more detail, justifies this type of studies and the adoption of new health measures. In general, healthcare services are not meeting the needs of the population when it comes to the services provided. Hence, trying to face some of these challenges led to the implementation of health reforms in Portugal, such as the attempt of employing private management tools in the public sector, with the creation of Public-Private Partnerships (PPPs) (Nunes, 2018).

[1] PORDATA - Despesa Corrente em Cuidados de Saúde em % do PIB (https://www.pordata.pt/Portugal/Despesa+corrente+em+cuidados+de+sa%c3%bade+em+percentagem+do+PIB-610). Accessed: 03/02/2020.

[2] PORDATA - Despesa Corrente em Cuidados de Saúde (https://www.pordata.pt/Portugal/Despesa+corrente+em+cuidados+de+sa%c3%bade-3010). Accessed: 03/02/2020.

PPP hospitals have, then, emerged, and have become a well-known alternative to public service delivery (Cruz and Marques, 2013b). As well as publicly-managed hospitals, these entities belong to the SNS (although they are not financed using the same contracting terms) and, consequently, must also deliver tendentiously free and universal care to any citizen. However, PPPs present disadvantages (such as composing a big investment and being subject to demanding and uncertain forecasts) (Galea and McKee, 2014), besides the advantages of shared roles, and, in Portugal, have been linked to conflict of interests. Several questions, drawn by authors and government parties, have, then, risen. *Are these entities capable of providing the same level of quality and access as publicly-managed hospitals, considering that PPPs are managed by private partners, whose pursuit of profit can allegedly compromise social performance?*

Note that social performance has been defined as "*a business organization's configuration of principles of social responsibility, processes of social responsiveness, and policies, programs, and observable outcomes as they relate to the firm's societal relationships*" by Wood (1991).

The State is a major stakeholder of both publicly-managed and PPP hospitals, and its interest should focus on the improvement of the citizens' health status (Ferreira and Marques, 2019).

However, hand in hand with the aforementioned challenges and questions, comes the concern that, while trying to apply measures that improve efficiency and reduce the waste of resources, the quality of services provided can be compromised, which is critical for achieving said citizens' improved health status.

1.2 Motivation

Ensuring access, quality, and efficiency should be sufficient to safeguard the sustainability of the SNS (Nunes and Ferreira, 2019a). In fact, when it comes to resources allocation and payments, quality and access should be considered, besides the efficiency of providers (Ferreira et al., 2019).

Several authors have evaluated the evolution of the efficiency of Portuguese hospitals, even though it is not trivial (Ferreira and Nunes, 2019). Successive governments have been implementing measures and new management rules. For instance, during the financial crisis, measures were implemented to, among other objectives, achieve efficiency gains, which did not fully work (Nunes and Ferreira, 2019b). Nevertheless, after the crisis, another government was capable of improving efficiency levels (Nunes and Ferreira, 2019b).

However, according to Ferreira and Nunes (2019), although, in general, Portuguese hospitals are considered efficient, some of them (particularly in the countryside) show lower indicators, with deficits of hospital units and human resources, which is due to factors such as lower health literacy, lower access to information/internet, and lower average income, resulting from geographical isolation. There is a limited supply of medical and surgical services, in comparison to the Portuguese coastline, which leads to a restricted access to products/drugs and services that are not covered by the SNS, promoting inequities not only on access, but also in inefficiency levels (Nunes and Ferreira, 2019b).

In addition, the aforementioned increase of health expenditures and efficiency levels does not necessarily mean high quality. In the same way that the increase in hospital costs can result from inefficiency, a hospital can be efficient when it comes to using resources, but faulty when it comes to quality (for instance, needing improvements in infrastructure or technology) (Ferreira et al., 2020). While investing in promoting efficiency, sometimes comes a disinvestment in access, equipment and infrastructures (Nunes and Ferreira, 2019b).

Hence, it is mentioned a seemingly unavoidable "trade-off" between the efficiency and the quality of services, as noted by Ferreira and Marques (2019). To answer this problem, we consider it is of utmost importance that the quality assessment of the Portuguese hospitals is put under scrutiny to find out weaknesses and to find a way to, in the future, overcome the problems faced by the SNS.

Additionally, comparing publicly-managed hospitals with PPP hospitals in terms of quality seems crucial to identify if these entities present, in fact, advantages over each other and if it makes sense to either build more or end these partnerships. In fact, the process of the creation of PPP hospitals is linked to uncertainty and risks and involves high costs, monetary and time-related, as well as the frequent need of renegotiations due to unpredictability, which can result in additional charges for the State (Pereira et al., 2020). Hence, it is relevant to understand if the results concerning their performance make up for this investment.

Enlightening the limitations of the SNS and understanding where hospitals stand in terms of quality and access is a start to help them to find better practices.

On the one hand, although there have already been previous studies concerning the assessment of healthcare quality and performance, the majority of them have been tendentiously focused only on outcomes or a few indicators, such as the ones developed by Benbassat and Taragin (2000) and Park et al. (2016), or only on publicly-managed hospitals, such as Ferreira and Marques (2019), which turns out not to be completely illustrative. On the other hand, while, for instance, Akdag et al. (2014) applied Multiple criteria Decision Aiding (MCDA) to evaluate hospital's service quality, but only in Turkey, Ferreira and Marques (2020) performed the comparison between Portuguese hospitals, but with another methodology, Data Envelopment Analysis (DEA). Then, to the best of our knowledge, an assessment that joins MCDA and both groups of Portuguese hospitals has yet to be done, to try to solve the evidenced problems and compare them to past results.

This thesis, then, aims to contribute to the literature in the topic of PPPs and the discussion of whether they are a good alternative or not. It is inserted in the hSNS Project, whose main goal is to develop and employ coherent healthcare provision services performance assessment models. With the help of a polyvalent research team, it aims to contribute to the literature and provide robust tools to improve the quality of the delivered Portuguese healthcare services. Consequently, it intends to support management by monitoring and controlling performance indicators, to contribute to the benchmarking of healthcare providers and better financing of the healthcare providers (according to their performance) and to assign accountability, among others[3].

1.3 Objectives

It has been increasingly important, from a political perspective, to evaluate several countries' healthcare systems' performance and quality, including Portugal's.

In fact, citizens want to achieve the best health state possible, since it is expected that their quality of life improves consequently. Moreover, governments want healthier populations, since they are presumably more productive (Mosadeghrad, 2014).

"Healthcare performance" can be divided into social performance (ability to provide the best healthcare to the population, focusing on quality and access to healthcare) and economic-financial performance (reducing waste and/or increasing the volume of services provided to the population).

Particularly, the aim of this dissertation is to evaluate and compare the quality of Portuguese hospitals - publicly-managed and PPPs -, using decision support techniques, following on previous work developed by Rocha (2019), while simultaneously contributing to the discussion on the topic of PPPs.

For this comparison to be accomplished, the ELECTRE TRI-nC multi-criteria model is chosen as the supporting tool of this thesis. This method allows the incorporation of quantitative and qualitative

[3] hSNS Project - About hSNS (https://hsns.eu/about/). Accessed: 25/03/2020.

attributes, the attribution of different weights to criteria, different types of scale, several reference actions, among others, which is very useful when assessing quality in health, since this is a hard subject to evaluate that includes a lot of information (Almeida-Dias et al., 2012). Thus, various indicators regarding health can be taken into account, together with families of criteria and variables that are found relevant in this kind of evaluation. Bottom line, with the adoption of this method, flexibility and variability of options are provided, helping the DM make appropriate decisions, and making the process easier and more fluid.

Besides this, during the realization of this work and particularly regarding data acquisition, the goal is to involve different parties to help to better understand and contextualize the problem, namely actors that are a part of the ministry or of the regional health directorate of Lisbon and Vale do Tejo.

1.4 Methodology

To achieve its aim, a few steps were followed while developing this study, schematized in Figure 1.1.

Firstly, it is important to provide background regarding the Portuguese healthcare system, to better define, understand and contextualize the problem. This is done with the support of past studies focusing on this theme, including how the SNS works and the emergence of PPPs.

Afterwards, an extensive literature review, examining existing literature related to the main dimensions that characterize quality and access in hospitals, and how it can be measured, will be conducted.

To establish the case study, books and papers will be consulted, regarding the concepts and goals of MCDA, its application to healthcare, and some respective methods, including specifically the ELECTRE TRI-nC method, which is the one used in this study.

Finished the theoretical basis, the collection of data relevant to the Portuguese case will be carried out, along with the definition of other parameters, such as criteria and indicators. This is mandatory for the next step, the creation of the model, using the ELECTRE TRI-nC method. Hence, composite indicators that summarize the social and financial performance of the publicly-managed and PPP hospitals in Portugal will be built, describing its development.

Finally, the analysis and discussion of the results obtained in the Portuguese case will be done, allowing the withdrawal of conclusions and to point out final remarks, including future work.

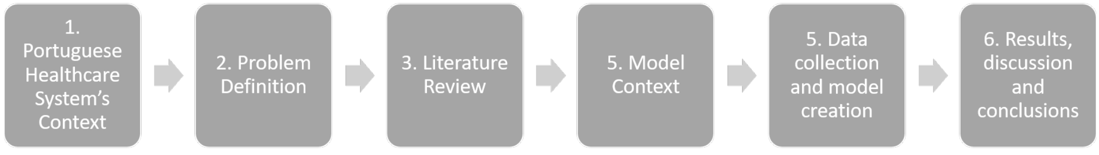

Figure 1.1: Methodology's Sequential Steps.

1.5 Structure

In this work, eight chapters are presented, aligned with the aforementioned objectives.

The first one is composed by the introduction, concerning the motivation and objectives of the development of this thesis.

The second chapter contextualizes the problem by understanding and giving needed background of the Portuguese Healthcare System. It focuses on how it is structured, the SNS and PPPs.

A literature review is performed on the third chapter, providing theoretical basis on the definitions of quality in healthcare, which is a non-consensual concept, as well as access to this type of services. It then focuses on the importance of quality measurement and how it can be accomplished. The literature review is expanded to the fourth and fifth chapters to provide context for the model used for this dissertation. They focus on composite indicators, the concept of MCDA, its contribution to healthcare and the details of the ELECTRE TRI-nC method, which is the chosen one.

The sixth chapter comprises the case study in regards of the DM, the database and sample chosen to carry out this analysis and the definition of variables needed for the method to be applied.

The seventh chapter is dedicated to the model implementation, providing examples of its application, and the consequent results.

Finally, the eighth chapter is reserved to the conclusions drawn from the implementation of this method to the case of Portuguese hospitals' performance assessment and the limitations of this study.

Chapter 2

The Portuguese Healthcare System

This chapter introduces the context where the problem is inserted, from general to specific. Firstly, there is a description of the Portuguese healthcare system, how it is organized, financed, and the evolution of expenditures regarding this sector through the years. Secondly, the characteristics of the Portuguese SNS are described, including its major reforms. Finally, this chapter focuses on one of these reforms: the creation of PPPs.

2.1 Overview

Health systems in developed countries are defined by three main frameworks: the national health model (Beveridge model), the social insurance model (Bismarck model), and the private insurance model. In the European Union, 77% of health spending is financed by public resources and compulsory insurance (Meleddu et al., 2019).

In 2015, the United Nations declared the goal to achieve universal health coverage - the central theme of global health policy nowadays - by 2030, "*so that all people and communities receive the quality services they need, and are protected from health threats, without suffering financial hardship*" (National Academies of Sciences, Engineering, and Medicine, 2018). The ultimate goal of healthcare services is to improve patients' satisfaction and quality of life (Ferreira et al., 2018b).

In Portugal, there is a mix of public and private health service providers, divided in three coexisting systems: the National Health Service (SNS), the private sector (for instance, clinics), and the social sector, with whom the public sector has collaboration agreements (Oliveira and Pinto, 2005; OECD/European Observatory on Health Systems and Policies, 2017).

According to PORDATA[1], in 2018, about 9% of the Portuguese GDP regarded health expenses, more than the average of the OECD, 8.8%. In fact, 17,839€per capita were spent in this sector in 2018, making a total of approximately 18.3 billion €[2].

Figures 2.1, 2.2 and 2.3 exhibit the evolution of health expenditures from 2000 to 2017 (in the case of hospitals) and from 2000 to 2018 (total health expenditure and percentage of GDP), using PORDATA as the data source.

As shown in Figure 2.1, there was a fluctuating variation in healthcare spending as a percentage of GDP since 2000, but it is mostly a sustained increase. After 2010, due to economic recession and fiscal consolidation measures, it was reduced by nearly one percentage point, from 9.8% to a steady 9%, that has been constant since 2014 (OECD/European Observatory on Health Systems and Policies, 2017).

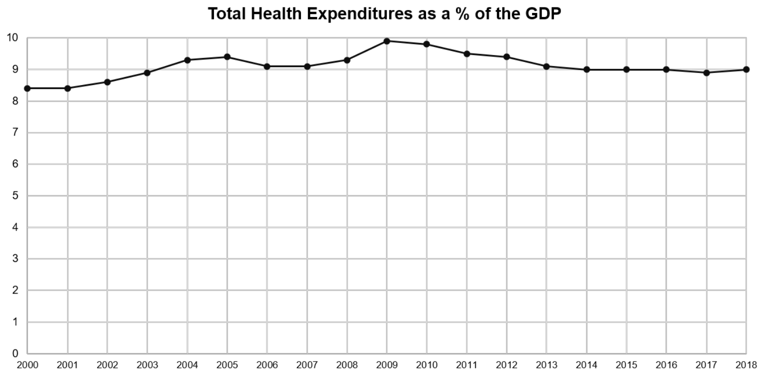

Figure 2.1: Total expenditures spent in healthcare as a percentage of GDP, from 2000 to 2018.

Figure 2.2 shows the total expenditures in the health sector, private and public. From 2000 to 2010 there was a significant increase in these expenditures. This period corresponds to the adoption of the New Public Management (NPM) paradigm - which will be discussed again subsequently -, that aimed to promote innovative private management models, but failed to achieve the reducing of costs within health services (Ferreira et al., 2018a).

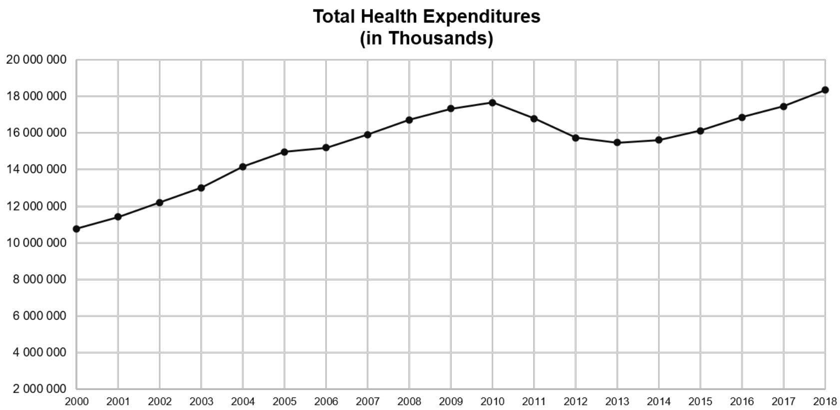

Figure 2.2: Total expenditures spent in healthcare, in thousand €, from 2000 to 2018.

Between 2011 and 2015, there was a reduction of those expenditures, linked to the signature of the Memorandum of Understanding and implementation of the austerity measures, that brought several cost containment measures, including in the health sector, aiming to control the expenditures of public hospitals (which is also noticeable in Figure 2.3), to enhance efficiency (Nunes and Ferreira, 2019b).

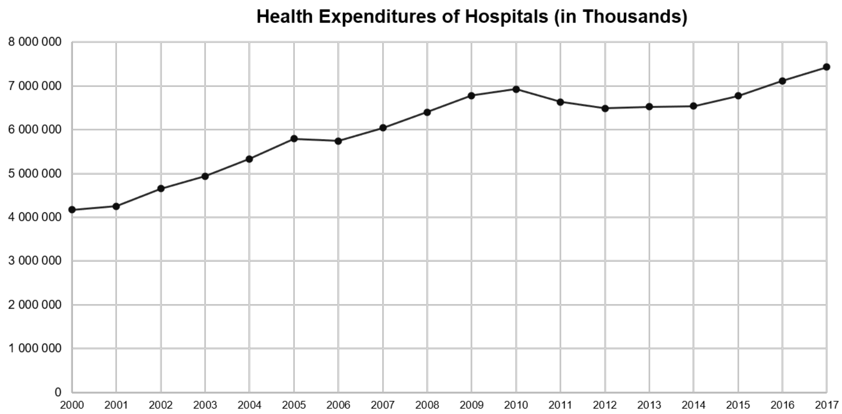

Figure 2.3: Total expenditures spent in hospitals, in thousand €, from 2000 to 2017.

The troika agreement ended in December of 2015, leaving a wider budget available for the healthcare sector. In 2016, the Portuguese government introduced some health policies that resulted in the increase of expenses with healthcare, holding the same share of GDP (Nunes et al., 2019).

Since then, these expenditures have kept growing, as a result of different factors such as demographic changes, technology evolution, healthcare advances, and aging population (Ferreira and Marques, 2019). Hospitals consume more than half of the public health expenses (Ferreira and Marques, 2015). For instance, in 2017, the value spent in healthcare was approximately 17.5 billion €, being that around 11.6 billion €- a little over 66% - are public expenses. Hospitals accounted, in that year, around 7.4 billion €of the latter expenditures, almost 64%[1].

When it comes to financing the healthcare system, there are four main sources: taxes, social contributions, copayments, and private insurance. Copayments and private insurance are voluntary payments, as opposed to taxes and social contributions (Ferreira et al., 2018b; Simões et al., 2017).

There are special social health insurance schemes for certain professions (health subsystems, like *Serviços de Assistência Médico Social*, SAMS, or *Instituto de Proteção e Assistência na Doença*, ADSE) and voluntary private health insurance (like *Médis*) (Simões et al., 2017). Around 25% of the Portuguese population is covered by health subsystems, 10% by private insurance and 7% by mutual funds (Barata et al., 2012).

The public share of health expenditure accounts for 66% of total health financing, leaving approximately 34% to be privately financed. Hence, it can be said that the Portuguese health system also presents a mix of public and private financing, and that, in the European Union and the OECD, Portugal constitutes one of the highest spenders in this sector (Oliveira and Pinto, 2005).

2.2 The Portuguese SNS

The Portuguese SNS was created in 1979, after a period of dictatorship that ended in 1974. SNS is composed of all public entities delivering primary and secondary healthcare services to the population (Ferreira and Marques, 2019).

The secondary healthcare services are constituted by singular hospitals, hospital centers (resulting from horizontal merging), local health units (resulting from vertical merging), PPPs, oncology centers, maternities, and psychiatric hospitals (Ferreira and Marques, 2020).

[1] PORDATA - Despesa Corrente em Cuidados de Saúde: total e por tipo de prestador (https://www.pordata.pt/Portugal/Despesa+corrente+em+cuidados+de+sa%c3%bade+total+e+por+tipo+de+prestador-2958-248067). Accessed: 03/02/2020.

In line with what was previously stated, the goal of the Portuguese SNS is to provide universal, appropriate, and equitable care to all citizens, regardless of their gender, religion, ethnic origin, social status, ability or willingness to pay. The Portuguese public healthcare services are also considered tendentiously free since, although some co-payments have been introduced to reduce some unnecessary demands for care services, the poorest population is free of charges (Ferreira and Marques, 2015; Crisp et al., 2014). SNS follows a Beveridge model and is financed by general taxation (Ferreira et al., 2018b).

There were positive changes in the evolution of some health indicators in the following years after the creation of the SNS, which, together with the overall improvements in the Portuguese population's general living conditions and the increased availability of healthcare brought impressive progress to the country (Simões et al., 2017; Oliveira and Pinto, 2005). In the last 40 years, Portugal not only reduced infant mortality rates remarkably but also increased life expectancy. As examples, in 1974 (before SNS), the infant mortality rate was 37.9% and, in 2019, it was 2.8%[2]. Life expectancy went from 68.2 average years of life in 1974 to 80.9 years in 2018[3]. Finally, the percentage of GDP spent in the health sector in 1974 was 4.1%, and, as previously seen, nowadays it rounds 9% (OECD, 2001; OECD/European Observatory on Health Systems and Policies, 2017).

According to Ferreira and Marques (2015), the Central Government collects funds from citizens via taxes, and distributes those funds by the different ministries, including the health one, which allocates them to SNS institutions.

The Ministry of Health is a governmental department whose mission includes assuring the application and sustainable utilization of the resources available and evaluation of the results. It is in charge of the planning, organization and regulation of the health sector in Portugal, which includes the development of health policies and running the SNS[4]. The activity of establishments providing healthcare is regulated by *Entidade Reguladora da Saúde* (ERS) [5].

Consequently, it would be expected for the Portuguese health system to be financially sustainable, especially when it comes to hospitals, which are included in the SNS institutions being financed and consume more than half of the public health expenses, as seen above. Nonetheless, considerable levels of waste, poor allocation of resources and inefficiency have already been reported among public hospitals, exposing a theoretical financial unsustainability of the system (Ferreira and Marques, 2015; Ferreira et al., 2018a; Ferreira and Marques, 2018; Ferreira et al., 2020).

Hospitals, in fact, are based on program-contracts (*contractos-programa*, in Portuguese), that are negotiated between the ERS and the Ministry of Health, the Regional Health Administrations (RHA) and the hospital in question (Ramos, 2006). These contracts follow a prospective payment system where hospitals are financed according to their dimension, past expenditures and production (volume of delivered services), and the model has shown structural weaknesses and constraints (Comissão para a Sustentabilidade do Financiamento do Serviço Nacionalde Saúde, 2007). The service prices are defined by clustering hospitals according to a set of size and complexity related variables. First, the most efficient hospital among each cluster is found, and then, their associated unitary cost is assessed, which is used as a reference value for all the others within the cluster (Nunes et al., 2019). However, Ferreira et al. (2020) consider that the concept of efficiency used is not clear and that the criteria used for clustering does not correctly reflect neither quality or environmental information of the services provided by the hospitals, likely leading to inefficient payments.

The system is managed by the Central Administration of the Health System (*Administração Central*

[2] PORDATA - Crude death rate and infant mortality rate (https://www.pordata.pt/en/Portugal/Crude+death+rate+and+infant+mortality+rate-528-2950). Accessed: 28/05/2020.

[3] PORDATA - Life expectancy at birth: total and by sex (https://www.pordata.pt/en/Portugal/Life+expectancy+at+birth+total+and+by+sex+(base+thre 418). Accessed: 28/05/2020.

[4] SNS - Ministério da Saúde (https://www.sns.gov.pt/institucional/ministerio-da-saude/). Accessed: 24/05/2020.

[5] SNS - Entidade Reguladora da Saúde (ERS) (https://www.sns.gov.pt/entidades-de-saude/entidade-reguladora-da-saude/). Accessed: 24/05/2020.

do Sistema de Saúde, ACSS) and by the five RHA (Norte, Centro, Lisboa e Vale do Tejo, Alentejo and Algarve). Each of the RHA have their own health administration council, whose responsibilities include strategic management of the services, control and supervision of hospitals, and they answer to the Ministry of Health. Hence, all Portuguese public hospitals are under the authority of this Ministry, which means that the hospitals are not autonomous in several issues, such as the purchase of innovative new technologies, or the hiring of personnel. They are subject to the commercial/private law and financed via annual prospective budgets [6] (Nunes, 2018).

As seen by Ferreira and Marques (2019), healthcare services in Portugal are also considered rivalrous due to the limited availability of staff, beds and other hospital resources. Hence, the access of one citizen may jeopardize the access of another one, increasing waiting lists and waiting times and contributing to the existence of some access barriers. According to OECD/European Observatory on Health Systems and Policies (2017), geographical disparities in the availability of services - since these resources are unequally distributed across the country -, were the main barrier to access in Portugal in 2017, joined by the aforementioned waiting times. They also refer to factors such as the increase of average life expectancy, the aging population, and the higher incidence of chronic diseases as impellers of big challenges to the healthcare services provided.

It is relevant to mention that several health reforms were introduced in the Portuguese SNS to reduce the high expenditures and waste of public funds, and simultaneously improve the value for money, the efficiency, and the effectiveness of healthcare providers. Some of these reforms include the corporatization, the vertical and the horizontal merging of public healthcare providers, and the creation of PPPs (Ferreira and Marques, 2015).

The first two, corporatization and hospital mergers, demonstrated the introduction of the NPM measure previously mentioned (Nunes and Ferreira, 2019b). The PPP reform will be discussed in the following section.

Particularly, the corporatization was implemented in response to the increase of public health expenditures, which challenged the traditional hospital model's sustainability. As reported by Rego et al. (2010), in 2002, 36 public hospitals (SPA) were converted into 31 State-owned hospital enterprises (SA), which had more autonomy than SPA - especially in the contracting and acquisition of health equipment, drugs, and human resources - and could use some private management related instruments, while still being subjected to regulatory intervention by the Ministries of Health and Finance. Later, in 2005, all SA companies were converted into corporate public entities (EPE) and, until 2009, the remaining SPA hospitals gradually followed their example. This enabled the guarantee of the public nature of the SNS hospitals, while also ensuring a greater level of strategic and supervisory intervention by the Ministries of Health and Finance, reducing, consequently, the autonomy of these hospitals.

When it comes to the merging measure, it was instigated by the lack of communication between the different levels of healthcare (primary and secondary) suffered by the Portuguese SNS (Nunes and Ferreira, 2019a). To articulate this and make this link more efficient and effective, there was a restructuration in the organizational model (Ferreira et al., 2018a). Thus, some public hospitals were integrated with primary or secondary healthcare centers, creating hospital centers (horizontal merging) and local health units (vertical merging) (Entidade Reguladora da Saúde, 2015). In this period, the post-hospital care (continued care) and end-of-life care (palliative care) were also created (Nunes and Ferreira, 2019b).

However, as presented in Figure 2.2, these reforms did not achieve the desired results in the healthcare sector, having even generated increased costs and high debt. This situation has also contributed to the financial crisis that came in the following years (Nunes and Ferreira, 2019b).

According to Nunes (2018) and Perelman et al. (2015), the financial crisis (2009-2015) was characterized by a poor definition of public health policies and a lack of strategic vision, reforms and investment

[6] Portuguese Government - Decree Law number 22/2012.

in the healthcare sector. The Portuguese population suffered inequalities regarding accessibility, worsening of their health status and loss of quality of services. The government had to implement and redefine measures promoting access, quality, and efficiency in the use of resources, to recover the country's SNS. Austerity-based policies included in the troika agreement also introduced a set of measures to reduce costs and waste in the health system, as it was already seen in the previous section, to increase the regulation of the pharmaceutical market, and to enhance the hospitals' management, by reducing contracted budgets with these entities (Sakellarides et al., 2014).

On the one hand, this resulted on a rationalization of resources, with efficiency gains namely in the drug market and in the reduction of the debt of the SNS. On the other hand, these cuts in spending had negative impacts, not solving the problems pre-existing in the SNS. Some were even aggravated, such as loss of responsiveness, increment of barriers to access to healthcare, lack of investment in equipment and infrastructures, and worsening of population health status (Nunes et al., 2019; Doetsch et al., 2017). This led to the dissatisfaction of the patients and demotivating of health professionals, along with the loss of quality of services.

Given the urgency of the situation, a new four-year strategy (2016–2019) was implemented, to reform the SNS and ensure its long-term sustainability. This strategy had measures regarding the improvement of the capacity to respond to citizens, the enhancement of the access of citizens to the appropriate and timely care and the extending of the supply of continuous and palliative care (Nunes, 2018; Nunes and Ferreira, 2019b). Since this period is the most recent and the measures are, at the time of writing, still being applied, there is not enough information to conclude on their effectiveness, even though the results are beginning to show. However, the SNS still faces many challenges, for instance being underfinanced since the time of the crisis, which might compromise its efficiency, equity, access and quality (Nunes et al., 2019).

2.3 Public-Private Partnerships

The concept of a Public-Private Partnership has emerged and become a well-known alternative to traditional ways of public service delivery in the past three decades. In recent years, the use of PPPs has become increasingly popular, including in the healthcare sector (Torchia et al., 2015). However, because of the little available activity history and of its long-term contracts, not many definitive evidence-based conclusions can be drawn about their effectiveness (Parker et al., 2019; McKee et al., 2006).

According to the Portuguese decree law number 111/2012, a PPP is defined as follows: "*Public-private partnership means the contract or the merger of contracts whereby private entities, designated by private partners, are obliged, in a lasting way, to a public partner to ensure, given previously negotiated conditions, the development of an activity in order to satisfy a collective need, in which responsibility for the investment, financing, exploitation, and associated risks is incumbent, in whole or in part, on the private partner.*"

This broad definition is applied not only for the health sector, but for a set of different ones such as the security, transport (railways, roads, ports or airports), power (generation and distribution), social, and governmental infrastructures (health, education, housing) sectors [7].

Firstly, a PPP establishes a long-term contract between the two parties. The private party is responsible for maintaining the infrastructure, and for the management and control of the project, which adds pressure not only to ensure a careful design and construction but also in terms of performance regarding the administration of hospitals (Pereira et al., 2020). These are preliminary indicators that PPPs are

[7] PPP Knowledge Lab - PPP Reference Guide (https://pppknowledgelab.org/guide/sections/8-how-ppps-are-used-sectors-andservices/). Accessed: 02/03/2020

expected to do better than regular public management in this sphere.

When it comes to Portugal, the Portuguese Government adopted PPP models for the same reasons as other countries, such as dealing with financial constraints in the public sector while having to increase health service spending, cost overruns in infrastructure investments and a political directive to increase private sector involvement in the delivery of public services (Cruz and Marques, 2013b; Torchia et al., 2015). In this way, it was believed that efficiency and effectiveness could be improved, and overall costs in construction and operation of hospitals could decrease (Cruz and Marques, 2013a).

The Portuguese model of PPPs on the health sector is based on the United Kingdom version of these partnerships, Private Finance Initiative (PFI). The main differences between the two are that PPPs express a joint ownership, while PFI denotes a contractual arrangement, and that the Portuguese version includes clinical services in the contract (healthcare provision and the equipment renovation), contrary to the UK model, that only covers the infrastructural and logistic services (Barros and Martinez-Giralt, 2009; Pereira et al., 2020).

These contracts comprise two managerial components: the buildings' development, maintenance, and management are provided and operated by one company – in Portuguese, *Entidade Gestora do Edifício* -, over 30 years, while another company – in Portuguese, *Entidade Gestora do Estabelecimento* - is responsible, over 10 years, for the provision of clinical services. These are, respectively, the entity in charge of the hospital building that administers the infrastructures, and the entity responsible for the hospital establishment that manages the clinical care provision [8]. Payments are made by the public sector to its private partners during the contract time. Since these two different entities have different objectives, they will have distinct payment methods which will lead to two independent payment streams (Simões et al., 2017).

The Portuguese PPP model in healthcare was applied to all of the four PPP hospitals from the first wave. There are, currently, only three hospitals under a PPP format in Portugal, although the construction of additional ones (under a different framework) has already been planned. These are Hospital de Cascais, Hospital de Vila Franca de Xira, and Hospital Beatriz Ângelo (Loures). It is planned that Vila Franca de Xira Hospital will no longer be under the responsibility of the private entity for clinical exploration in 2021, when their contract ends.

As mentioned above, PPPs can simply privately manage an existing hospital or be a full-service provision at all levels of care (European Commission, 2014). In the Portuguese case, the clinical management (e.g., staff, clinical devices, drugs) and soft facilities (e.g., laundry) are included in the project, besides infrastructure and hard facilities management (Barros and Martinez-Giralt, 2009). This triggers a few questions regarding its effectiveness. As publicly-managed hospitals, PPP hospitals also belong to the Portuguese SNS, but they are not financed in the same contracting terms (Ferreira and Marques, 2020).

2.3.1 Advantages and Disadvantages

It is possible to outline advantages and disadvantages of the PPP measure.

In the managing context, the private sector is usually known for its competence (allegedly increasing the quality of services provided) and efficiency in delivering projects (there is an earlier delivery of the investment program, since PPPs can provide further funding) (Galea and McKee, 2014).

Hence, as stated by Ferreira and Marques (2020), Buse and Harmer (2004) and Torchia et al. (2015), in this shared risk, rewards and roles between both public and private partners, the former (State) should take advantage of the competitive potential, strategic skills and the higher efficiency in risk management

[8] ACSS - Parcerias Público-Privadas na Saúde (http://www.acss.min-saude.pt/2016/10/12/parcerias-publico-privadas).Accessed: 20/03/2020.

of the latter; yet, it has shown a weak ability to do this, reflecting poor outcomes. As a consequence of the innovation, know-how, and managerial/ technical efficiency provided by the private partner, these partnerships are also expected to make a more rigorous and efficient selection of projects possible, create jobs and incentives, deliver high quality secondary healthcare services to the citizens, among other benefits.

Marques and Silva (2008) expose the advantages in the financial and economical context, including the reduction and optimization of costs and time for project delivery. While during the planning of a public hospital construction the price of the project is, in average, underestimated by around one fifth, and the construction time is usually extended, for the case of PPPs these estimations are done much more carefully since most of the risk falls on the private company, leading to significant loss in revenue. Also, because there is a team dedicated to monitoring the fulfilment of the contracted services, the private company must ensure the quality of such services, otherwise they incur in a financial penalty stipulated by the contract.

Concerning disadvantages, they comprise a big investment (sometimes out of hands for the governments), frequently based on economic interests rather than population needs and public interest, being subject to opportunistic, unreliable behaviours (Johnston and Gudergan, 2007; Galea and McKee, 2014).

The fact that their forecasts, besides being demanding, are subject to great uncertainty, being very hard to predict, is also an issue (Ferreira and Marques, 2020). There is, usually, an overestimation of demand for healthcare, which sometimes is not met, imposing compensations to the private partner (Cruz and Marques, 2013a).

These projects are also likely to face skimming effects, which happens when providers maximize financial gains by selecting or transferring patients according to the risks involved (Yang et al., 2020). Yang et al. (2020) hint that publicly-managed hospitals are less likely to fall under these circumstances than profit-motivated hospitals. In addition, they are under rigid and incomplete contracts that are difficult to elaborate and redact, since they need to be extremely clear and well defined for both parties, especially since contracts are created from scratch without many previous references (Marques and Silva, 2008).

2.3.2 Assessment

In some Portuguese political sectors, there have been some objections and maneuvering of public opinion about the long-term consequences of PPPs. When it comes to the healthcare sector, specifically, some parties believe that PPPs are not capable of providing equitable care as publicly-managed hospitals, and these entities are often seen as a threat to accessibility.

Besides this, with the recent development of new technologies and medicines, renegotiations between the public sector and the private partner of the PPP are constantly happening. There is a regular need to update tenders and policies, which turns out very costly (Ferreira and Marques, 2020).

A timely evaluation of management tools such as PPPs might bring benefits, since it allows the comparison between public-managed hospitals and PPPs, which might later allow for corrections of protocols and adjustment of measures, among others.

Since 2013, hospitals' benchmarking dashboards have been used, to enhance the transparency and sharing of information with the community to build a public market of healthcare providers, which aims to improve the access and quality of services, and better detect aspects needing improvement. This is based on the model developed by ACSS, and assesses the quality, productivity, and economic-financial performance. Data are provided every month by each institution (Simões et al., 2017).

According to a study done by ERS in May of 2016, where the institutions analyzed were both publicly managed hospital entities (with the exception of psychiatric hospitals) and PPPs, the parameters most complained about were the same for all types of hospitals. All the complaints are registered by the SNS and, at the top of both lists, were waiting times followed by the focusing, or lack thereof, on patients as well as their safety. However, it is important to notice that PPP hospitals are not obliged to reveal data regarding expenditures, something EPEs are, which makes the comparison less accurate than it should be[9].

Therefore, if quality of the delivered care cannot be periodically and reliably monitored, a lot of negative outcomes can happen, which can even harm patients' health status. Ferreira and Marques (2020) point this as a reason for questions to arise as far as the providing of clinical services by the private partner is concerned. In fact, since private partners are known for wanting to maximize their profit, there is a great concern that it can compromise the good conduct and quality of the services provided, especially when considering such an important and demanding sector as healthcare.

In line with Cruz and Marques (2013b), to ensure the success of these reforms, it is possible to define quality indicators and monitoring plans. However, it is complex to measure health services, especially regarding quality standards, as will be discussed further, making it hard to keep these partnerships under control. In general, PPPs have been under some criticism as evidence seems to suggest that they did not achieve the desired goals, namely the public expenses reduction (Ferreira and Marques, 2019).

In general, what customers (patients) demand from the healthcare system, and should be the main goal, includes: the delivery of effective and efficient services, equity on access, low waiting times, and the exceeding of their expectations. Unfortunately, while presenting very good characteristics, the Portuguese SNS has also been commonly associated with lack of quality and patient's dissatisfaction, which needs to be addressed (Oliveira and Pinto, 2005; Ferreira et al., 2018b).

However, an opportune question arises: what does a high-quality healthcare service constitute?

2.4 Summary

This chapter presented the background context where the problem under scrutiny is defined. First by characterizing the overall Portuguese health sector, then, entering the domain of the SNS, and, finally, focusing on PPPs.

The Ministry of Health is responsible for the planning, organization, and regulation of the health sector in Portugal. The provision of healthcare is made by public entities (primary care centers, hospitals, continued, and palliative care) and by some private partners (consultations, diagnostic and therapeutic examinations, hospitals, and other private clinics). The financing of healthcare in Portugal is a mix of public and private funds, being predominantly based on taxes – as a typical Beveridge model.

The SNS itself is characterized as a complex organizational structure, which provides universal, appropriate, and equitable care to all citizens and has faced up and downs since its creation, in 1979.

The system is managed by the ACSS and by the five RHAs. Hospitals are financed under program contracts and are the main healthcare providers, consuming more than half of the total of public expenditures.

PPPs are long-term contracts between two parties: the public one, governments, and the private partners. They emerged with the objective, among others, of reducing the financial pressure and constraints in the public treasury by creating shared roles, risks and rewards between parties.

In Portugal, PPP hospitals include, besides infrastructure and hard facilities management, clinical services. They also belong to the SNS but are not financed in the same way as publicly-managed hos-

[9] Entidade Reguladora da Saúde - Estudo de avaliação das Parcerias Público-Privadas na saúde (https://www.ers.pt/pt/regula%C3%A7%C3%A3o/estudos/estudos/avaliacao-das-parcerias-publico-privadas-na-saude). Accessed: 30/03/2020.

pitals. Besides the potential benefits to the health sector that these entities offer, many concerns have been raised. It has been discussed whether PPP hospitals can deliver healthcare services with, at least, the same level of social performance as publicly-managed hospitals, considering their management's private nature.

Obviously, it is of high importance to ensure the sustainability and efficient use of hospital resources, while guaranteeing that patients can access and receive appropriate and timely care, with maximum security, and equitable manner. Thus, it is extremely important to contribute to closing the gap in literature related to the capacity of PPP hospitals.

This makes understanding and assessing the quality of hospitals, publicly-managed and PPPs, considering their performance, a priority.

Chapter 3

Quality in Healthcare

The concept of quality has been used in many different contexts, with different meanings, by different authors. It is a complex concept in literature, especially when it is linked to healthcare, not existing an official single definition to it. In this chapter, a literature review is carried out, exposing the multiple definitions one comes across when trying to find out how to characterize quality in the area of healthcare. The final goal is to settle on a definition of quality and its many dimensions in this domain, so that later a comparison can be done between the Portuguese hospitals regarding this concept.

3.1 Definition of Quality in the Healthcare Sector

Quality, access, and efficiency should always be ensured for a health system to be sustainable. It is hard for healthcare providers to maximize the quality and increase the quantity of their services and at the same time contain their costs. Therefore, health systems capable of delivering services equitably and efficiently are critical for achieving improved health status (World Health Organization, 2010; Ferreira and Marques, 2019).

According to World Health Organization (2010), a health system consists of all the organizations, institutions, resources and people whose primary purpose is to improve health and delivers preventive, promotive, curative and rehabilitative interventions. The actions of the health system should be responsive and financially fair, while treating people respectfully. It needs staff, funds, information, supplies, transport, communications, and overall guidance and direction to function.

Interest in measuring and improving the quality of healthcare has been increasing as a consequence of factors such as growing demand for healthcare, rising costs, constrained resources, an increasing number of medical specializations, complex therapies and equipment, and disease burden, to name a few (Campbell et al., 2000; Talib et al., 2015). As previously mentioned, a considered "efficient" health provider does not necessarily imply a high quality delivered service, so it is important to understand what "quality" means, in this context.

As found in literature, "quality", "care", and "quality of care" are non-consensual, multi-faceted and universally-hard-to-define concepts.

One of the factors that make the definition of quality so difficult is the "distinct" feature, when it comes to the healthcare industry characteristics - such as intangibility, heterogeneity and simultaneity. "Heterogeneity", for instance, can occur because different healthcare professionals provide services differently (according to their individual experience, abilities and personalities) to patients with varying needs (Mosadeghrad, 2014).

Besides, understanding these concepts can lead to a reduction of waste and delays, lower costs, higher market share, and a positive company image, which, by its turn, results in the increase of productivity and profitability. In addition, satisfied patients are more likely to follow medical advice, which is expected to improve their own quality of life (Mosadeghrad, 2014; Ferreira et al., 2018b). In general, defects in the quality of healthcare deny populations the potential benefits of effective care (National Academies of Sciences, Engineering, and Medicine, 2018).

For instance, National Academies of Sciences, Engineering, and Medicine (2018) defined quality as "*the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge*".

Other authors, instead of giving straightforward generic definitions, that are not easily operationalizable since they lack sensitivity and specificity, follow disaggregated approaches which take into account the complexity and multidimensionality facets of quality. In these approaches, each individual component (dimension) provides, per se, an evaluation of quality. When combining all these dimensions, results a more specific evaluation (Campbell et al., 2000).

For instance, Donabedian (2005) uses a Structure-Process-Outcome model to define quality, in which structure denotes the attributes in which care occurs, like infrastructures and equipment, process stands for the actions performed by the staff and their interactions with the patients, to deliver care services, and outcomes express the effects of the care services delivered on patients' quality of life. There are subcategories, or dimensions, that can be introduced within outcomes and process, which are **care appropriateness**, and **clinical safety**.

According to Ferreira and Marques (2019), on the one hand, care appropriateness regards the ability of delivering patient-centered care services supported by evidence-based guidelines (the services provided are not considered appropriate unless they follow both of these mandatory characteristics). The disrespect of this dimension can result in avoidable re-admissions after inpatient discharge and excessive staying, i.e., long periods of permanence in hospital wards, which, in its turn, can increase the probability of other diseases developing, such as septicemia or nosocomial infections. On the other hand, the patients' clinical safety is the capacity of preventing and reducing the risk of unnecessary complications, harm or even deaths, during the process of care. There are medical complications that can be considered preventable (e.g. bloodstream infections and post-op pulmonary embolisms), so its occurrence and consequent effects on a patient are often linked to staff errors, which presupposes lack of clinical safety during care. This leads to multiple problems both for the patients and for the provider entity: patients are undeniably left uncomfortable and dissatisfied, there is a consequent loss of trust in the healthcare system and, apart from all of that, additional care is needed to fix these problems, which also can lead to more costs and, consequently, waste.

Care appropriateness and clinical safety are likely linked, since inappropriate provided care can result in negative outcomes, which regards the safety of the patient.

Zineldin (2006), in its turn, suggests that "service quality" can be broken down into two distinct dimensions, which are the technical dimension and the functional dimension. When related to the healthcare sector, the technical dimension is defined primarily on the basis of the technical accuracy of the medical diagnoses and procedures, while the functional dimension refers to the way the healthcare service is delivered to the patients, concerning the quality of the relationship between patients and the organization (Talib et al., 2015).

Furthermore, Campbell et al. (2000) starts by defining healthcare, which consists on "*a composition of healthcare systems and actions taken within them, designed to improve health or well-being*". In line with Donabedian (2005), those authors also distinguish between the structure of healthcare, the actual care given (process), and the consequences of the interaction between individuals and a healthcare system (outcome). In this case, however, it is suggested that outcome is not a component of care, but a consequence of care. Similarly, structure is not considered a component of care but the conduit through which care is delivered and received.

When it comes to the definition of quality of care for individuals, however, Campbell et al. (2000) refers to it as "*whether individuals can access the health structures and processes of care which they need, and whether the care received is effective*", being suggested two dimensions of quality: **effectiveness** and **access**.

Firstly, according to Campbell et al. (2000), it is suggested that effectiveness, when related to individual patients, should refer to maximizing care and desired processes and outcomes based upon need.

Secondly, it is considered that a citizen has access to a service if he/she can use it whenever necessary and at his/her will. The most basic dimension of access to a health structure considered is geographic/physical access, to which rurality or difficulty of access by disabled or elderly people are examples of barriers that can arise. This is the issue of some regions in Portugal, as mentioned in the previous chapters.

Hospitals, as complex structures composed of multiple services whose aim is to deliver timely, equitable, patient-centered, safe, efficient, and effective secondary healthcare services, have the duty to reduce this type of barriers. In fact, the lack of access is likely to deteriorate the quality of those services. Thus, it is equally important to guarantee access to healthcare and overcome these barriers as it is to provide the best quality healthcare services (Ferreira and Marques, 2019).

Nonetheless, Campbell et al. (2000) adds that, even if patients show no problems physically accessing the facility, issues regarding availability can still be faced, i.e., the healthcare system can be incapable of providing facilities and services that meet the needs of individuals.

Then, access can, per se, present multiple dimensions. Gulliford et al. (2002), for instance, takes into account four different dimensions of access: **service availability** (available hospital resources); **personal barriers** (capability of recognizing the need to seek specialized care); **financial barriers** (e.g. cost and the distance between home and the healthcare service) and **organizational barriers** (e.g. long waiting lists and waiting times). Meanwhile, Ferreira and Marques (2019), in their turn, divide it into three categories: **the timeliness of services** (capacity of delivering healthcare services at the right time whenever required); **the services availability** (existence of disposable resources to be used when necessary) and the **characteristics of the population at risk** (propensity of individuals to use services, own income/insurance coverage, community attributes, and need for healthcare).

As mentioned by Donabedian (2005), aspects such as prevention, rehabilitation, coordination, continuity of care, and patient-physician relationships are often forgotten when defining "quality". It can be concluded that healthcare quality is, indeed, a subjective, complex, and multi-dimensional concept.

3.2 Frameworks for Quality Assessment

Hibbard and Pawlson (2004) bring attention to the fact that, even though patients worry and care about quality of medical care, they struggle to understand what it should entail when searching for it. The authors believe that providing frameworks for understanding, measuring, and evaluating the quality of medical care makes it easier for patients to understand the meaning and relevance of high-quality care. Besides this, frameworks help the communication between healthcare providers and consumers, bringing them together in the way that they understand quality.

Some of the existing studies rely on Parasuraman et al.'s (1985) model (SERVQUAL) to study healthcare quality in healthcare facilities around the world. According to Badri et al. (2009), Akdag et al. (2014) and Talib et al. (2015), this model is a five-dimensional concept of service quality, the five concepts being:

Reliability: Ability to provide services accurately and dependably.

Responsiveness: Willingness or promptness in responding to customers' needs.

Assurance: Courtesy and knowledge of the employees and their ability to convey trust and confidence (trustworthiness).

Empathy: Caring and individualised attention provided to customers.

Tangibles: Physical evidence in a service facility (e.g., personnel, equipment).

Zineldin (2006) expanded the aforementioned technical-functional and SERVQUAL quality models into a five-quality dimension framework, the 5Qs model, which includes:

Quality of object - the technical quality;

Quality of processes - the functional quality;

Quality of infrastructure - the basic resources;

Quality of interaction - the measurement of the quality of information exchange;

Quality of atmosphere - the relationship and interaction process between the parties are influenced by the quality of the atmosphere in a specific environment where they cooperate and operate.

This method is, according to that author, more comprehensive than SERVQUAL, incorporating essential and multidimensional attributes that the latter does not (Badri et al., 2009; Talib et al., 2015).

However, it is believed that more appropriate optimization models shall be used instead.

In the 2018 reviewed version of Crossing the Global Quality Chasm, the Institute of Medicine (IOM) put forth a framework where the quality dimensions identified are:

Safety: Avoiding harm to patients from the care that is intended to help them.

Effectiveness: Providing services based on scientific knowledge to all who could benefit, and refraining from providing services to those not likely to benefit (that is, avoiding both overuse of inappropriate care and underuse of effective care).

Person-centeredness: Providing care that is respectful of and responsive to individual preferences, needs, and values and ensuring that people's values guide all clinical decisions. Care transitions and coordination should not be centered on healthcare providers, but on recipients.

Accessibility, Timeliness, and Affordability: Reducing unwanted waits and harmful delays for both those who receive and those who give care; reducing access barriers and financial risk for patients, families, and communities; and promoting care that is affordable for the system.

Efficiency: Avoiding waste, including waste of equipment, supplies, ideas, and energy, as well as the waste resulting from poor management, fraud, corruption, and abusive practices. Existing resources should be leveraged to the greatest degree possible to finance services.

Equity: Providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, race, geographic location, and socioeconomic status (National Academies of Sciences, Engineering, and Medicine, 2018).

Quality attributes such as timeliness, consistency, and accuracy are hard to measure beyond a subjective assessment by the patient (Mosadeghrad, 2014).

This committee concludes that the global quality chasm at the time of the study was even larger than the quality chasm identified by the Institute of Medicine in the United States in 2001, and much more disturbing in terms of human costs.

It is possible to see that all the definitions and points of view presented in this section are correlated, some of the dimensions even coinciding, suggesting their importance. Overall, high quality of care needs to be safe, effective, centered on patient's needs, accessible and given in a timely, equitable and affordable manner.

Relevance is added to the existing need of effectiveness, access and quality to go "hand in hand". However, to achieve "universal quality care", investment, responsibility, and accountability on the part of health system leaders are required.

As stated previously, earlier studies showed that quality improvements may lead to efficiency deterioration and hospital costs increasing. Trying to counteract this can jeopardize the patients' clinical safety.

In fact, hospitals consume financial resources and physical resources, which are considered inputs. The first comprises, among others, expenses with staff, drugs, pharmaceutical products, clinical consumables and outsourcing, and the second accounts for the workload of clinical staff, especially doctors and nurses (Ferreira and Marques, 2019). For them to be considered "technically efficient", they must use the minimum number of inputs for a given number of outputs. This means that these entities can be technically efficient simply by achieving a higher number of treated patients, no matter the damages, whether it is disinvestment on the safety of the patients or in infrastructures and equipment. This results in poor treatment, access barriers like long waiting lists and even more costly treatments (Nunes et al., 2019).

This makes apparent, once again, the need for efforts to be made to improve efficiency without sacrificing quality and the other way around (Ferreira and Marques, 2019).

3.3 Measurement of Quality in the Healthcare Sector

Measuring quality is essential and is a type of evaluation used in many industries. Currently, it has also been increasingly common in the healthcare sector. It is considered crucial since, with the information provided, it can help to evaluate the appropriateness of the health policies currently followed by health systems (i.e, if the best results for the population are being achieved).

Although there is an extensive body of literature regarding healthcare quality dimensions, few tools exist for assessing and managing healthcare quality (Badri et al., 2009). When it comes to this subject, there is also no consensus. This kind of assessment requires the ability to measure each of the components of the system and their relationships with each other (Handler et al., 2001).

According to the Agency for Healthcare Research and Quality (AHRQ), a quality measurement tool must be objective, be based on scientific evidence and not affect or distort results. Besides that, AHRQ reinforces that, to use quality measurements, there needs to be a responsible entity that can be identified, held accountable for its observed behaviour, and has a reasonable degree of control over the aspect of care being evaluated. They also refer that quality measurements imply that the approach followed is rigorous, systematic and quantifiable, and that these measurements must be tested to ensure its reliability, validation, and standardization. Ultimately, expert consensus is often used to achieve agreement on definitions and measurement specifications when considering quality measurement in healthcare[1].

According to Talib et al. (2015), measuring the quality of these services is important to understand the clients' (patients') preferences, since they select the hospital to attend based on their previous experiences. In fact, it is not only the latest medical equipment that matters to draw patients to the facility, but indicators like general health services and the quality satisfaction are what matters most to build their trust. Factors and events that affect the patients' perception include technical, functional, infrastructural, interaction, political environment, healthcare quality, social perceptions, and information technology. Typical patient complaints include long waiting times, high costs, and unfriendly, apathetic and uncaring staff. Consequently, most of healthcare institutions are going for a patient-centered attitude.

Meleddu et al. (2019) mentions Qin and Prybutok (2013), Anbori et al. (2014) and Kitapci et al. (2014) as authors who found that five indicators (installations, product/service quality, staff professionalism, and economic and non-economic costs) had a significant influence on consumer trust and satisfaction.

[1] Agency for Healthcare Research and Quality - Understanding Quality Measurement (https://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/chtoolbx/understand/index.html). Accessed: 09/05/2020.

The control and revision on the domain of quality should be conjugated with management and strategy. Nations, regions, and healthcare organizations should measure and report it routinely and transparently to support learning, and promote accountability and trust in the healthcare system (Akdag et al., 2014; National Academies of Sciences, Engineering, and Medicine, 2018).

Standardized indicators allow comparisons between countries and can help mutual learning (World Health Organization, 2010).

3.3.1 The use of outcomes

Outcomes can be used to provide information about the system's overall performance, including its efficiency, effectiveness, and ability to achieve equity between populations, and have been frequently used as an indicator of the quality of medical care (Handler et al., 2001; Donabedian, 2005).

For instance, some studies have found an association between hospital readmission (i.e., rehospitalization within one, two, four, or 12 months of discharge) rates and inappropriate care during the hospitalization, since better hospital and post-discharge care are associated with fewer readmissions. However, data on global readmissions have a limited value in the assessment of quality of care, since the length of stay, death, and unplanned readmissions were proved to be predicted mainly by age, severity, and comorbidity. Nevertheless, it raises concern, and efforts should be done to reduce those rates (Benbassat and Taragin, 2000).

More recently, many studies have used patient satisfaction as an outcome to measure the performance of healthcare institutions, since the main goal of service providers is to deliver high quality service and stay competitive (Talib et al., 2015). The aim of measuring patients' satisfaction is that providers can then strategically reallocate resources with the goal of improving the patients' experience. It constitutes, nonetheless, a very subjective and emotional measurement, since their satisfaction can be influenced by past bad experiences, their own personality, severity/complexity of their illness, too high expectations and a number of non-health related factors that might not be connected to the quality of services. It is, then, apparent that satisfaction itself is not a good proxy for quality, since one is unable to know whether rating differences come from individual expectations or, indeed, the hospital performance (Ferreira et al., 2018a).

Moreover, most of the information is obtained through surveys and questionnaires which, while representing feedback and allowing an idea of overall healthcare statement at the hospital level, can also show irregular behaviour of the patient while ranking the service. Hence, the scores obtained are usually not illustrative. In addition, there is the risk of loss of compliance, which can only be avoided if the perception of quality of care for the hospital staff and patients is similar (Akdag et al., 2014; Kisliakovskii et al., 2017).

The nature of a typical healthcare service also contributes to these statements. The customer experience consists of a series of service encounters or touchpoints, which involves setting up appointments, reception, waiting, physician interaction, interactions with healthcare personnel, pharmacy, billing, and other administrative personnel (Park et al., 2016). This can result in failures of communication and coordination between the different providers and facilities, causing important health information to get lost along the way and impairing patient experiences (National Academies of Sciences, Engineering, and Medicine, 2018). Therefore, a negative experience at one of these multiple touchpoints can result in an overall negative evaluation, even if all the other touchpoints resulted in a positive experience (Park et al., 2016). This makes the healthcare system seem more of a collection of separate activities than a coherent, self-conscious, purposefully designed system, in addition to giving more confirmation of the possible unreliability of patients' evaluation (National Academies of Sciences, Engineering, and Medicine, 2018).

One must also notice that outcomes (and other related factors) are affected by the context in which the system operates, i.e., its social, economic, and political environment, which are unrelated to medical care (Handler et al., 2001). Besides this, while some of these metrics are easy to measure (death, for instance), most of them are harder and depend on a wide set of external and nondiscretionary events, including the health status prior hospitalization and the follow-up of medical indications and prescriptions after discharge (Ferreira and Marques, 2020).

In summary, although outcomes can work as validators of the quality of medical care and have been regularly applied, they must be used with precautions and discrimination if considered quality indicators (Donabedian, 2005). Other, probably more relevant, indicators must be considered.

3.3.2 Other relevant variables

Considering the information present in the reviewed literature, it is evident that the mere supply of staff, space, and other resources is essential, but it does not guarantee good quality. As seen above, factors like the adequacy and availability of facilities, technology and equipment, the qualifications, actions and organization of clinical and non-clinical staff, the administrative structure and operations of programs/institutions providing care, might also influence the quality of medical care (Donabedian, 2005; Ferreira and Marques, 2020).

Each country and health system needs to accommodate its approaches according to the needs of its populations and its health strategy objectives (National Academies of Sciences, Engineering, and Medicine, 2018).

The evaluation can be done using a set of indicators, where each of them is related to one of the quality dimensions seen above. However, existing measures address some domains more extensively than others. For instance, effectiveness and safety are commonly addressed, but only a smaller number of measures assess timeliness and patient-centeredness, and even less consider the efficiency or equity of care.

Different entities, such as AHRQ, IOM, OECD and WHO, have developed efforts towards validating indicators to measure the quality of medical care provided by healthcare entities. In fact, more than 650 hospitals in OECD countries have adopted standard indicators capable of measuring quality, that must be adapted to consider the appropriate context and provide helpful insights for patients, providers, and policy makers. This is important since, even though having many quality indicators is good, the conclusions might be wrong/biased if they are not contextually relevant (Ferreira and Marques, 2020).

Most current indicators are, then, mainly focused on structural and process inputs instead of outcomes, since it was already seen that it is more difficult to have access to this information.

Some of the common quality related variables that are used are: the failure to rescue and risk-adjusted in-hospital mortality, (acquired) infections due to medical care, readmissions, caesarean rate, patients' satisfaction, postoperative respiratory failure and sepsis, and inappropriate discharges (Ferreira and Marques, 2020). For instance, in this area of quality, the indicators assessed by ACSS are: percentage of readmissions within 30 days, percentage of patients discharged with a length of stay above the maximum threshold, percentage of outpatient ambulatory surgical procedures, percentage of hip surgeries performed within the first 48 hours, percentage of deliveries by caesarean section, the mean adjusted delay rate, and the adjusted mortality rate (Simões et al., 2017).

Ferreira et al. (2018b) consider hospital image (measured by its trustworthiness, know-how, concern about patients and technological progress), facilities quality (in terms of cleanliness, comfort, conservation, and privacy protection) and waiting time for (and before) a medical appointment as the most valued criteria for patients to be satisfied.

3.4 Summary

In this chapter it was possible to see how many definitions and different dimensions the concept of quality has been associated with, in literature. Many authors have paired quality with effectiveness and access which also presents different dimensions within itself. It was also found in literature many models regarding the study of quality, being the SERVQUAL model very common since 1985. However, it has been expanded since then and, in agreement with the National Academies of Sciences, Engineering, and Medicine (2018), quality can now be composed by even more dimensions. These are: safety, effectiveness, person-centeredness, accessibility, timeliness and affordability, efficiency and equity. These dimensions can be directly linked to the assessment of hospital's quality, being very important.

When considering how such a complex concept can be measured, multiple variables have been considered in literature. First, outcomes, which, although useful, can also be very hard to measure, subjective and affected by external aspects, and then, other variables, standard adaptable indicators shared by OECD countries.

Chapter 4

Composite Indicators

In this chapter, the concept of composite indicators is presented, along with its definition and the main methods used to create them. This is done with evidence found in literature of how they can be considered as a tool to carry a proper evaluation of quality in healthcare organizations.

4.1 Overview

Composite indicators are becoming increasingly important and accepted as a useful tool in a wide spectrum of fields such as the environment, economy, technology, and education, including monitoring and comparing performances (Veillard et al., 2009). However, statisticians, economists, and other groups of users remain skeptical and critical when it comes to its use, due to the lack of transparency of some existing indicators, especially as far as methodologies and basic data are concerned, which affects its relevance, credibility, and interpretability (OECD and JRC, 2008).

Overall, an indicator is a quantitative or a qualitative measure derived from a series of observed facts that can reveal relative positions (e.g. of a country) in each area. When evaluated at regular intervals, an indicator can point out the direction of change across different units and through time (El Gibari et al., 2019).

A composite indicator is defined by OECD (2004) as what is formed "*when individual indicators are compiled into a single index, on the basis of an underlying model of the multi-dimensional concept that is being measured*". Hence, since they are the mathematical combination of single simple indicators that represent different dimensions of a concept (whose description is the objective of the analysis), they are considered much easier to interpret than identifying a common trend across many separate indicators. This single resulting index captures policy attention more easily, facilitates communication – potentially enhancing accountability -, and ensures that, instead of focusing on a single aspect, a rounded assessment of performance is presented (Veillard et al., 2009). In summary, they measure multi-dimensional concepts that single indicators are not capable of capturing, providing a useful overall summary of performance that can as well facilitate the comparison between similar organizations and track particular services and units (Shwartz et al., 2015). Nevertheless, if poorly constructed and not carefully and transparently treated, composite indicators can lead to misinterpretation, misleading policy messages and potential manipulation (OECD and JRC, 2008; El Gibari et al., 2019).

The construction of a composite indicator involves making choices when combining criteria of different natures, and it requires several steps in which the DM must make decisions, namely processes of normalization, weighting, and aggregation. All the methodological choices made at each stage of construction will influence their accuracy, reliability, and appropriateness, possibly even having a dramatic impact on the final results (Jacobs et al., 2005).

OECD and JRC have developed, together, a set of recommendations on how to design, develop and disseminate a composite indicator. It is important to note that the quality of a composite indicator depends not only on the methodology used in its construction, but primarily on the quality of the framework and the data used (OECD and JRC, 2008). They should be based on a well-grounded theoretical definition/framework which allows the selection, combination, and weighting of individual indicators to be carried out in the appropriate manner for each measurement (OECD, 2004).

The steps to be followed, according to OECD and JRC (2008), involve, besides the theoretical framework, the data selection, the imputation of missing data - to provide a complete dataset -, a multivariate analysis - to study the overall structure of the dataset, assess its suitability, and guide the next step -, normalization, weighting and aggregation, a robustness and sensitivity analysis, and finally the identification of correlations to other variables, the visualization of results and the assurance of transparency, by going back to "real data".

It makes sense to extend this approach to the healthcare sector since healthcare performance quality is, as it was already seen, multi-dimensional and thus not easily captured by a single measure, benefiting from a composite/aggregate score that is concise and easy to report.

If quality scores are difficult to interpret, health professionals and managers might not be interested in engaging with this approach. The truth is, unlike individual indicators, there is no single, simple way of developing composite indicators for a group of quality indicators, existing several different techniques to achieve it (Reeves et al., 2007). It is very important to study these different methodologies, since each of them can yield different scores or produce different rank orders for provider organizations, making the choice of method a critical step.

4.2 Methods to construct composite indicators

There are multiple methods to combine individual measures or domains into composite ones. When narrowing the research to the healthcare sector, specifically evaluation of healthcare organizations, the methods one comes across more often are linear combinations, latent trait modeling, opportunity scoring, and the creation of patient-level composite end points such as the all-or-none composite. All methods have advantages and disadvantages.

Just as an example, The US News & World Report Index of Hospital Quality - the basis for evaluating hospitals in the magazine's annual "Best Hospitals" report – evaluated 4,656 facilities, in which hospitals receive a composite score based on data from multiple sources in 12 of the 16 adult specialty rankings (exceptions are ophthalmology, psychiatry, rehabilitation, and rheumatology). They use indicators for different components like volume for structure, reputation for process, 30-day mortality for outcomes, a patient safety score, among others. Weights are determined using different methods for each type of measure (for instance equal weighting or relying on an expert panel) and each component has its own weight assigned as well, being obtained a raw score as a simple weighted sum of the four ranking components (Olmsted et al., 2018). Other ways, found in literature, to compound a composite indicator, are tabled below, from Table 4.1 to Table 4.7

Table 4.1: Literature review of Composite Indicators in Healthcare facilities

Author, Year	Sample	Variables	Methods	Conclusions	Strengths	Weaknesses
Reeves et al. (2007)	Data set A: 16 UK family practices Data set B: 60 UK family practices	Data set A: 129 indicators, covering 23 different conditions included in acute, chronic, and preventative care. Data set B: 25 indicators, covering 3 conditions (coronary heart disease, asthma, and type 2 diabetes)	"All-or-None," "70% Standard," "Overall Percentage," "Indicator Average," "Patient Average."	• Different methods of computing composite quality scores can lead to different conclusions being drawn about both relative and absolute quality among health care providers. • Different methods are suited to different types of applications. • There were large differences between the all-or-none ranks and the ones using the other methods. • The resulting scores differed between methods, but also depended on the nature of the indicator set to which they were applied. • Dataset B: all methods produced reliable scores Dataset A: most scores had low reliability, rank correlations were low, and the rank ordering of practices was unstable.	**All-or-none**: - It assesses quality with respect to the care given to each patient; - It is appropriate when the desired outcome depends on completing a full set of tasks; - It is suitable when individual patient scores are required; - It allows standard errors to be estimated using standard distributional methods. **Opportunity-based**: - It assesses quality with respect to each opportunity to provide care; - The composite scores are not affected by the number of measures; - It can be used with a smaller sample of patients (ie, smaller standard errors). **Indicator Average**: More suitable indicator for comparing health care organizations with heterogeneous populations. **Patient Average**: The best choice for evaluating the average quality of care provided to a specific set of patients. **Overall Percentage method**: Suitable for comparing scores from a range of health care organizations serving relatively homogeneous populations.	Dataset A: Samples incomplete. **All-or-none**: scores very much lower than the others, not recommended except for situations where success depends upon meeting all indicators. **70% standard**: arbitrary % so changing the threshold could make a considerable difference in the results. **Indicator Average**: - standard errors might be higher, when used to compare health care organizations, unless large numbers of records have been sampled; - the indicators triggered may vary from one provider to another.

Table 4.2: Literature review of Composite Indicators in Healthcare facilities (continued)

Author, Year	Sample	Variables	Methods	Conclusions	Strengths	Weaknesses
Eapen et al. (2011)	194,000 acute myocardial infarction (AMI) patients from 334 334 Hospitals	Aspirin use within 24 hours of arrival; Aspirin use at discharge; blocker use at discharge; Angiotensin-converting enzyme inhibitors or Angiotensin receptor blocker use at discharge for left ventricular systolic dysfunction; Smoking cessation counseling; Lipid-lowering medication use at discharge for patients with a low-density lipoprotein level of greater than 100 mg/dL.	"Opportunity-based score", "All-or-none approach"	• Both methods were significantly correlated with one another. • Over time, the number of hospitals in the top and bottom quintiles of hospital rankings diminished similarly for both composite measures. • When including additional performance measures into the composite score, both methods produced similar changes in rankings. • Both methods provide similarly correlations with 30-day mortality, but not readmission.	The opportunity-based and all-or-none composite measures may be similarly capable of stratifying hospitals according to process of care. An advantage of the all-or-none scoring method is that it is structured at the patient-level.	data were restricted to the available in the GWTG–CAD registry (a well-known quality improvement program). The degree of selection bias of submitted patient records is unknown (some hospitals are more likely to submit patient records that reflect better acute treatment). These data are dependent on the quality of medical record documentation and chart abstraction.
Schang et al. (2016)	14 Health Boards in Scotland	I: Number of patient journeys from referral to treatment over 18 wks per 100,000 RTT patient journeys from referral to treatment; Number of recorded A&E waits lasting ¿4h per 100,000 A&E attendances; Number of emergency admissions among +75 yrs per 100,000 population; Number of MRSA/MSSA infections per 100,000 population; Number of Clostridium difficile infections per 100,000 population; Number of bed days lost due to delayed discharges per 100,000 occupied bed days discharges. II: Number of Clostridium difficile infections; Number of MRSA/MSSA infections; Resident population; Number of occupied bed days.	"Ratio-based efficiency analysis technique" (REA)	• When one assumes complete ignorance about the relative weights assigned to different indicators, it is impossible to differentiate the performance of the Health Boards. • When some reasonable ordinal and proportional weight restrictions are applied, organizational performance appears more clarified. • Ranking intervals and dominance relations appear to offer relatively intuitive ways to synthesise key messages contained in disparate indicators.	**REA**: - Calculates pairwise comparisons between DMUs rather than comparing each DMU to an efficiency frontier. - Is based on pairwise comparisons, requiring a minimum of only two DMUs and providing confiable comparative performance assessments. - Wide and overlapping intervals visualize in a transparent way the existing uncertainty.	- The use of a ratio function does not account for structural differences (such as a higher share offixed costs) between organisations. - Ratio measures may be preferred when there is primarily a concern with evaluation rather than explanation.

Table 4.3: Literature review of Composite Indicators in Healthcare facilities (continued)

Author, Year	Sample	Variables	Methods	Conclusions	Strengths	Weaknesses
Shwartz et al. (2008)	3275 Hospitals	AMI: Aspirin at arrival; Aspirin prescribed at discharge; Beta blocker prescribed at discharge; Beta blocker at arrival; CHF: Left ventricular function assessment; Discharge instructions; Pneumonia: Oxygenation assessment; Pneumococcal vaccination status assessment; Initial antibiotic received within 4h of hospital arrival; Blood culture performed in emergency department before first antibiotic received in hospital; ACE inhibitor or ARB for left ventricular systolic dysfunction (LVSD) (AMI, CHF); Adult smoking cessation advice/counseling (AMI, CHF, PN)	"Denominator-based weights" (DBWs), "Bayesian hierarchical latent variable model" (BLVM)	• Ranks were highly correlated, but different approaches result in differences in the top/bottom deciles. • In terms of assessing organizational performance and monitoring changes in performance over time, the choice of approach is unlikely to make much difference. • Both indicate that the larger hospitals have higher quality and the smaller hospitals lower quality.	**DBW**: Allows the comparison of weights within the sets of hospitals/to weights from other approaches. **Bayesian model**: - The shrinkage estimates appropriately take into account differences in the reliability of estimates from hospitals of different sizes. - Allows estimates of policy-relevant performance in ways that other approaches do not.	- Assumption that all of the Quality Indicators (QIs) are of equal clinical importance. - The particular set of QIs considered is limited to 3 conditions, 2 of which are heart related. **DBW**: - Does not take into account differences in the reliability of estimates from hospitals of different sizes.
Moore et al. (2013)	19,853 patients from 59 trauma centers	Transfer of traumatic brain injury patients with GCS $\leq$12 from Level III/IV center to Level I/II center; Transfer of patients with open/depressed skull fracture from Level III/IV center to Level I/II center; Transfer of spinal injury patients to acute spinal center; Reduce dislocation of major joint in $<$1h; Airway secured in emergency department (ED) for GCS score ¡9; Stabilize/embolize unstable pelvic fracture; Open long-bone fracture surgery $<$6h; Epidural hematoma surgery $<$1h (Level I/II); Fractured femur surgery $<$24h (femoral shaft fractures); ED stay $<$1h for patients with GCS score $<$9 or intubated (Level I/II); Deaths $>$1h after arrival occur on ward; ED stay $<$4h for patients with ISS $\geq$15; Therapeutic laparotomy; Delay for abdominal, thoracic, brain surgery $<$24h; No reintubation within 48h after extubation.	"The Indicator Average", "Opportunity Model", "latent variable model"	• Results based on the Indicator Average did not agree with those based on the others. • Only the Indicator Average showed evidence of content validity and predictive criterion validity. • The results suggest that the Indicator Average is the most appropriate for calculating a composite score to evaluate trauma care performance. • The best method depends on the context of use.	**Indicator average method**: Gives equal weight to each indicator, ensuring that indicators related to highly specialized processes of care are accounted for. **Latent variable model**: Provides greater opportunity to discriminate among hospitals. Weights based on the opportunity and latent trait models will vary across data sets, precluding comparisons between trauma systems and over time.	**Indicator average method**: Equal weights on different sample sizes lead to an inefficient estimate of global performance. **Opportunity model**: Is based on weighting by the number of eligible cases, which is problematic in the context of trauma. **Latent variable model**: - High variation may reflect weaker consensus on the clinical importance of each indicator; - Complex, difficult to interpret and involve model assumptions that are not easy to respect. The authors had difficulty obtaining consensus on weighting indicators, and they may not all be applicable in all trauma systems.

Table 4.4: Literature review of Composite Indicators in Healthcare facilities (continued)

Author, Year	Sample	Variables	Methods	Conclusions	Strengths	Weaknesses
Jacobs et al. (2005)	117 English acute Hospitals	Deaths within 30d of surgery; Emergency readmissions to hospital after treatment; Percentage of patients returning home after hospital treatment; Percentage of patients waiting ¡6mo for an inpatient admission; Percentage of patients seen within 13 wks of general practitioner written referral for 1st outpatient appointment; Satisfaction surveys; Summary measure of Hospital Episode Statistics data quality for NHS trusts with inpatient activity; Percentage of junior doctors working hrs complying in full with the new deal on junior doctors' working hrs; Amount of time lost through absences as a percentage of staff time available for directly employed NHS staff.	"Simple linear fashion" using equal weights	• The analysis presented in this article suggests that a degree of caution is required when interpreting the results of composite performance measures. • The conclusions drawn about the comparative performance of healthcare organizations are subject to a substantial degree of uncertainty.	**Linear aggregation and equal weighting**: Useful when all sub-indicators have the same measurement unit and when further ambiguities, due to the scale effects, have been neutralized.	- Difficult to draw definitive conclusions about the precise level of attainment of an organization when measures are vulnerable to uncontrollable random factors. - Existence of overlap along much of the distribution of composite scores when uncertainty intervals around them are estimated by repeated simulations. - Varying the weights on individual indicators has a substantial impact on the rankings of many hospitals.
Shwartz et al. (2011)	1,006 Hospitals	Adherence to processes of care, 30-day readmission rates, In-hospital mortality, Efficiency, Patient satisfaction (For acute myocardial infarction, congestive heart failure, and pneumonia). For a subset of hospitals, two other measures are examined: survey-based assessments of patient care quality by hospitalchief quality officers and by frontline clinicians.	Summing the quintiles of the ranks on the individual measures	• There was little correlation among the publicly available measures. • There was notable correlation between objective measures and survey-based measures. • Hospitals that performed well on a composite measure were often not in the top quintile on most individual measures. • For some purposes, a composite measure is necessary.	- This type of construct minimizes the impact of outliers; - It allows comparisons between performance on the composite measure and performance on the individual indicators more easily. - The results of this study lend some external validity to patient satisfaction measures, showing it as a more valid measure of performance than any of the objective measures.	- Questionable quality of the input data and the adequacy of (or lack of) risk adjustment. - For the mortality and readmissions measures, the authors did not use the risk-adjustment models most widely used to generate the data. - 30-day mortality data were not available but it would be better than using in-hospital mortality measures in the mortality models.

Table 4.5: Literature review of Composite Indicators in Healthcare facilities (continued)

Author, Year	Sample	Variables	Methods	Conclusions	Strengths	Weaknesses
Rosenthal et al. (2007)	141 Hospitals (From Boston, Miami, Phoenix, Seattle, and Syracuse)	Quality Measures: Percent of patients within each admission type that died in the hospital; Percent of patients within each admission type with any type of major complication; Percent of patients with major complications that died in the hospital; Number of patients discharged for specific procedure or diagnosis; Intensive care unit intensivist staffing and computerized physician order entry; Computed cost measures.	Sum of standardized quality and cost scores to create a single index; Application of a separate threshold for cost and/or quality	• Both approaches led to different results. • The combined index of cost/quality demonstrated moderate agreement with rankings based only on quality, with overlap in preferred hospitals. • The threshold method produced a set of preferred hospitals with little overlap with hospitals that were preferred in terms of overall medical/surgical quality. • There is a need for clear prioritization and the application of more rigorous methods to identify high-value hospitals.		- Not appropriate weighing of the magnitudes of incremental cost and quality differences. - There is not a guarantee that the highest value hospitals are identified. - The quality and cost measures are based on administrative data. - Data on actual payments to hospitals were unavailable, so it was used an adjusted cost measure. - Risk-adjustment approach used not ideal for examining cost–quality tradeoffs; better risk adjustment could lead to different results.
Grigoroudis et al. (2012)	GHD (Governmental Hospital of Didimoticho), a public secondary-level hospital	Net profit margin; Operating revenues to assets ratio; Current ratio; Debt ratio; Inventory turnover; Operating expenses to operating revenues ratio; Patient satisfaction index; Number of patient complaints; Avg waiting time; Hospital beds per 1000 people; % of cases transferred to other hospitals; % of readmissions; Avg duration of hospitalization; Employee satisfaction index; Employee retention index; Employee absenteeism index; Surplus inventory; Bed occupancy ratio; Number of projects with other organizations; % of budget used for purchase of new technology; Resource allocation to information technology/capital; % of employees trained; % of medical staff participating to conferences.	BSC Methodology: “UTASTAR method”	• There are significant improvement margins regarding the implementation of the hospital’s strategy. • The indicators are an important part of the developed BSC system. • The developed BSC system may be used as an effective measurement tool of corporate performance, as well as an evaluation approach of the successful implementation of corporate strategy.	- Necessary information easily collected. - The preferences of the DM are expressed through different scenarios presenting different outcomes and can be modeled by a set of completed results provided by the method. - It is flexible and has the ability to handle ordinal data, without arbitrarily quantifying the information. - The BSC methodology focuses on a set of performance measures and is able to provide overall and marginal evaluation scores for it.	- Difficult to model performance indicators as a consistent family of criteria, particularly testifying the property of exhaustiveness. - There are compensation assumptions regarding the performance of the assessed criteria.

Table 4.6: Literature review of Composite Indicators in Healthcare facilities (continued)

Author, Year	Sample	Variables	Methods	Conclusions	Strengths	Weaknesses
Canadian Institute for Health Information (2001) (Jacobs et al., 2004)	Healthcare organizations in 54 regions in Canada	Outcomes: Life expectancy at birth Heart attack survival; Prenatal care: Proportion low birthweight babies under 2500g, Percentage of babies born by Caesarean section, Percentage of vaginal births after Caesarean section; Community health: Hip fractures, Pneumonia and flu hospitalisation of persons over 64; Elderly services: Hip replacements, Knee replacements; Efficiency: Possible outpatients – hospitalisations for conditions not requiring admission, Early discharge – variation from expected length of stay, Preventable admissions - hospitalisations for conditions considered preventable by appropriate ambulatory care; Resources: Physicians per 100,000; Specialists per 100,000; Local services – percentage of hospitalisations generated by local residents.	Combinations of the scores on the performance indicators using weights 'based on expert opinion'; Combination of the six categories using determined weights	There is very little variation in some of the indicators, however the authors discuss these variations as if clear differences exist which reflect real differences in health services.	The sequential approach to assigning weights allows a more careful treatment of priorities.	- Weighting scheme very rudimentary - 'Expert' references may not necessarily reflect those of the public. - No psychiatric performance indicators included, which is a major aspect of health care services. - The inclusion of efficiency indices leads efficiency to be treated simply as another objective that contributes to the concept of performance
Appleby, J. & Mulligan, J. (2000) (Jacobs et al., 2004)	English and Welsh health authorities and Scottish health boards	Number of deaths from cancer (per 100,000); Number of deaths from heart disease (per 100,000); Number of people on hospital waiting lists (per 1,000); Percentage of people on waiting lists for ¿ 12mo; Number of hip operations (per 100,000); Number of deaths from 'avoidable' diseases (per 100,000)	Normal ranking, "Budget allocation", "Conjoint analysis"	• All three methods produced very similar overall rankings of health authorities and health boards based on the final composite indicator. • The rankings generated from transformed and standardised indicators differed from the original ranks, but not significantly.	Effort to assign meaningful weights to increase legitimacy of the composite indicator.	Limited choice of indicators and their coverage.

Table 4.7: Literature review of Composite Indicators in Healthcare facilities (continued)

Author, Year	Sample	Variables	Methods	Conclusions	Strengths	Weaknesses
Shwartz et al. (2009)	32 Department of Veterans Affairs (VA) nursing homes	Pressure Ulcer Development (if a patient who was ulcer free at one assessment had a stage 2 or deeper pressure ulcer at the subsequent assessment); Functional Decline (change between assessments in a score measuring limitations in eating, toileting, and transferring); Behavioural Decline (change in a score measuring extent of verbal disruption, physical aggression, and socially inappropriate behaviour); Mortality (if there was a death within 6 months of an assessment regardless of location); Preventive Hospitalizations (if the patient was admitted to an acute medical unit within 6 months of an assessment for one of 13 conditions identified as a potentially preventable hospitalization).	”Benefit-of-the-doubt” (BOD) approach using Simple ”Linear Programming” (LP) models and ”Data Envelopment Analysis” (DEA)	• No evidence that a composite measure calculated from one set of weights is better than another. • DEA has a larger impact on ranks than LP. • In this kind of environment, BOD approaches seem a considerable option. • Under BOD approaches, providers are assured baseline weights will be adjusted.	**DEA**: - No need for normalization - Identifies for each facility performing below its potential benchmark facilities - BOD weights obtained through it have nice theoretical properties. **LP**:Transparency (it is easy to justify the resulting weights to policy makers and facility managers)	For many indicators, the link between processes and outcomes is weak. **DEA**: - Number of facilities to small to reliably identify appropriate benchmark facilities - Difficult to determine exactly what it is about a facility's performance that results in the specific weights identified and the resulting composite score. **LP**: Can not handle indicators measured on different scales without normalizing the scores.
Shwartz et al. (2016)	112 VA nursing homes	Pressure Ulcer Development; Functional Decline; Behavioural Decline; Mortality; Preventive Hospitalizations.	”Opportunity-Based Weights” (OBWs, the baseline), DEA, BLVM with shrunken estimates	• Based on the composite measures, the same facilities appear in the top decile. • Shrunken estimators are an improvement over using raw rates to calculate performance measures, but there is disagreement over the ways to calculate them.	- Shrinkage estimators using Bayesian thinking can be of value when dealing with sample sizes. - Providers are given the BOD and allowed flexibility in selecting weights.	**DEA**: - Every facility is given the flexibility to set its own weights, so it is much harder to distinguish the high performers. - The low performers were identified from a model designed to identify high performers and provide the BOD in that direction. **BLVM**: Slow to converge. **OBWs** and **DEA**: The sample size differences underlying particular point estimates do not reflect differences in reliability across providers.

It is possible to point out the common use of opportunity-based models, the indicator average model, and latent variable models. However, since several variations of each method exist, it is difficult to evaluate them flawlessly. Besides this, along this research, it was made obvious that the application of each method depends strongly on the characteristics of the dataset used, whether it is homogeneous or heterogeneous, if it is made of small or big samples, and even the kind of care provided.

When it comes to weighting, whatever method is used, there is no consensus, since all of them present advantages and disadvantages. According to Nardo et al. (2005), it is very important that assumptions used for each weighting system are made clear and transparent, posteriorly undergoing a robustness analysis. Usually, when there is a lack of consensus, lack of statistical or empirical grounds, or ignorance about the correct model to apply, simpler methods are chosen, such as equal weighting, that can yield unreliable results (Greco et al., 2019). Other widely used methods, pointed out both by Nardo et al. (2005) and Greco et al. (2019), were the expert judgement and the Analytic Hierarchy Process (AHP), which can lead to cognitive stress if there are many indicators to consider or do not reflect the "general" opinion. Other weighting methods found were, for instance, the Benefit of the Doubt, Budget allocation and Conjoint analysis, from which conclusions were hard to draw (Saisana and Tarantola, 2002; Wind and Green, 2013).

Then, aggregation techniques have to be considered, next to the weighting methods. The most used, additive aggregation, can range from simply summing up indicators (linear aggregation) to aggregating weighted transformations of them. However, the most suggested aggregation methods fall into geometric aggregations or non linear aggregations, like MCDA. In fact, in additive aggregations, weights do not necessarily demonstrate the importance of the correlated indicator, which implies a compensatory logic, not being highly recommended (Podinovskii, 1994). In both linear and geometric aggregation, weights express trade-offs between indicators, with the former presenting constant compensability, while with the latter, compensability is partial. When one is faced with a situation where different goals are evenly legitimate and important, these methods may not be suitable (Nardo et al., 2005).

Some authors claim that, since MCDA - which will be discussed in further detail in the next chapter - is a well-known branch of decision making that takes into account different criteria simultaneously, while analyzing different alternatives, these techniques are highly suitable in multidimensional frameworks to construct a composite indicator by aggregating single ones (OECD and JRC, 2008).

To do this, there are multiple methods available, and the final value provided by them is taken as the composite indicator. The literature research conducted showed that most of the MCDA methods used to construct composite indicators have been applied in the fields of sustainability and environment, with TOPSIS being the most widely used method (El Gibari et al., 2019). For the healthcare field, Tables 4.5 and 4.7 show the application of DEA based methods and a value and utility based method (UTASTAR). However, UTASTAR and DEA allow for full compensation among the criteria, i.e., there are trade-offs among the different criteria, and adopt additive aggregation for constructing composite indicators. For the weights to be interpreted as "importance coefficients", one must rely on non-compensatory aggregation procedures to construct composite indicators, which can be done using a non-compensatory multi-criteria approach, based on the DM's preferences (El Gibari et al., 2019). The examples that have already been studied with this objective were the methods of the ELECTRE family and PROMETHEE, which limit or completely prevent compensation since they allow veto thresholds and only ordinal comparisons among alternatives. However, there was not evidence found in literature of its application to create composite indicators for the assessment of healthcare organizations' quality.

ELECTRE was introduced by Roy (1991), and is applied to three main problems: choosing, ranking and sorting. The concordance and discordance indices and threshold values are used in this method, which will be further explained.

To construct a composite indicator using this method, the decision maker must associate the corresponding parameters (thresholds or pseudo-criteria) to each indicator, in addition to the weights. They do not require a normalization before aggregating variables, since they use the original data for the comparisons. Afterwards, each indicator is converted to a 0–1 scale (El Gibari et al., 2019).

4.3 Summary

A composite indicator is a mathematical combination of single simple indicators that can represent and measure multi-dimensional concepts in a way single indicators can not, being much easier to interpret than identifying a common trend across many separate indicators. However, they are also victims of criticism, their relevance and credibility questioned, due to the lack of transparency of some poorly constructed composite indicators, likely leading to misinterpretation, misleading policy messages and potential manipulation.

OECD and JRC (2008) have developed a set of steps to be followed to enhance the building of this type of measures, and they have already been used in multiple fields, namely the environment, economy, technology and education.

There are several methods to combine individual measures into composite ones and for the weighting and aggregating steps, each with advantages and disadvantages and rarely existing consensus regarding the ”best” one. Examples of techniques found in literature, applied to healthcare, consist on linear combinations, latent variable models, opportunity based models, the all-or-none approach and the indicator average model. Multi-criteria approaches such as DEA and UTASTAR were also found in literature, but they allow for full compensation, which may not be always suitable. Hence, in situations where there are equally important different goals, one should rely on non-compensatory aggregation procedures to construct composite indicators, such as methods of the ELECTRE family and PROMETHEE.

Composite Indicators also give a needed background for the next chapter, that introduces multi-criteria decision aiding and the method used throughout this dissertation, the ELECTRE TRI-nC.

Chapter 5

MCDA and ELECTRE methods

In this chapter, the definition and some of the most important concepts to characterize and understand Multiple Criteria Decision Aiding are introduced, as well as the methodology and its application in the healthcare sector. Moreover, it comprises a presentation of the method chosen for this dissertation, ELECTRE TRI-nC, and its roots, followed by two sections focusing on information necessary for its application, namely its concepts, definitions and notation and the assignment procedure. It is also mentioned which software platform was used to help the application of the ELECTRE TRI-nC method.

5.1 Multiple Criteria Decision Aiding

For many years, problems regarding decision making were solved using a single objective function, meaning only one point of view was considered, not taking into account that a situation involving a decision typically presents a multidimensional character. Recently, interest in multiple criteria decision aiding/making methods (MCDA/MCDM) has been growing rapidly, being applied to solving real world problems in different areas, such as the financial area, energy planning and the creation of a telecommunications network (Greco et al., 2005). They are, nowadays, slowly starting to be more applied to the healthcare sector. In this area, it is considered a natural extension of Evidence Based Medicine (EBM), making use of different points of view and taking into account, in addition, the stakeholders' preferences (Marsh et al., 2017).

In this dissertation the term being considered is 'decision aiding', since, instead of the normative and prescriptive perspectives present in 'decision making', it reflects the constructive perspective of the decision process (Figueira et al., 2012).

It is important to understand what it means to be in a "multiple criteria" domain, i.e, an approach where more than one criterion is used. Firstly, a criterion can be defined as "*a real-valued function on the set A of alternatives, such that it appears meaningful to compare two alternatives a and b according to a particular point of view on the sole basis of the two numbers $g(a)$ and $g(b)$*" (Bouyssou, 1990). Using multicriteria methods, several criteria are built using many points of view, and each criterion is considered independently from the others. Criteria are used for evaluating and comparing one or more potential actions, according to a well-defined point of view and their performance according to each criterion, which corresponds to a score and can be represented by a number, a verbal statement or a pictogram. In addition, the quality of the MCDA method depends on the quality of the construction of the criteria (Roy, 2016).

The main goal of these methods is to help Decision Makers, whose name or for whom this decision aiding is to be given, make more consistent, transparent and robust choices by giving them more information, understanding and insight into the decision they face. This is done with the assistance of an analyst, who is responsible for giving the decision aiding through developing the models and all the computational skills of the process. These methods also take into account the preferences and values of the DM, while incorporating multiple considerations, alternative courses of action and data on performance. The analyst should also aid the DM in this step, and then help them understand and interpret the results (Dolan, 2010; Tànfani and Testi, 2012; Figueira et al., 2012).

As reported by Marsh et al. (2017), this approach provides support and structure to the decision making process, making explicit the criteria applied and the relative importance of it, consequently providing clarity for stakeholders, enhancing legitimacy, transparency, and accountability. Overall, MCDA joins objective measurement and value judgement and attempts to manage subjectivity, helping the DM deal with the organization and synthesis of all the complex information multiple criteria problems typically face.

This author also mentions how it is important that these methods acknowledge frames (cognitive bias that exist when individuals react differently to a criterion depending on how the information is represented), such that stakeholders can similarly understand the criterion.

Any MCDA method involves at least three steps: defining the decision problem, selecting the criteria that reflect relevant values, and constructing the performance matrix, being the latter a central element. The way the performance matrix is used defines the method as "qualitative MCDA", "quantitative MCDA" or "MCDA with decision rules" (Baltussen et al., 2019).

However, in general, these steps can always be included in two distinct phases defined as **Problem Structuring**, which consists of identifying the main points of view, the objectives, stakeholders and potential actions, developing a common understanding of the problem by all parties involved, and **Model Building**, which involves constructing consistent representations of the DM's values and value trade-offs, and developing a framework for the evaluation of alternatives (Marsh et al., 2017). A redefinition of the criteria and other aspects can also be done in this step. Afterwards, there is a translation of the results of the model's implementation into specific action plans (Belton and Stewart, 2002).

According to Figueira et al. (2012), the three major problematiques in MCDA are:

Choosing: Selecting a restricted number of the most interesting potential actions, as small as possible, which will justify to eliminate all others.

Sorting: Assigning each potential action to one of the categories among a family previously defined; the categories are ordered, in general, from the worst to the best one. e.g:

C_1: actions whose implementation is not advised;

C_2: actions whose implementation could only be advised after significant modifications;

C_3: actions whose implementation could only be advised after slight modifications;

C_4: actions whose implementation is always advised without any reservation.

Ranking: Ranking of actions from the best to the worst, with the possibility of ties and incompatibilities.

Previous studies have already shown that the implementation of MCDA is very useful in the healthcare sector, since it is a complex field that constantly faces hard decisions that involve different perspectives. Examples of this application usually fall into the categories of finding the best treatment alternative for a given patient, such as the ones carried by Chen et al. (2013) and Figueira et al. (2011), or application in different departments of health management, such as resource allocation support. However, there is still a lot of work to be done on the development of comprehensive MCDA approaches, since, besides the fact that most of the models do not reflect all the important concerns and values, they are still not that frequently used (Tànfani and Testi, 2012).

It is also worth noting that the literature found only includes studies that are self-described and labeled as MCDA, which can explain the small number of evidence found of these methods being applied in healthcare. This shows that the existing literature should be redefined for it to be a more coherent and complete set of studies.

According to Thokala and Duenas (2012), three broad methodologies are often distinguished: Value Measurement, Goal Programming and Outranking Methods. These authors consider Value Measurement as the most widely used and prevalent method, which develops quantitative measures and aggregates preferences across criteria, to allow the DM to characterize the degree to which one alternative program is preferred to another, to achieve the goal of the decision. Marsh et al. (2014) mention two important steps that are part of each of these methodologies, which are the scoring and weighting steps. Again, there is a panoply of methods to draw scores and weights, such as direct rating, swing weighting, AHP, and Discrete Choice Experiments (DCEs). This is another gap found in the healthcare literature, since the selection of techniques for scoring and weighting is not frequently justified, nor does guidance for this selection exist.

MCDA methods can be either compensatory or non-compensatory. The former incorporate information from all the decision criteria whereas the latter do not (Dolan, 2010).

Some of the methods mentioned in previous studies in the healthcare sector are, for instance, MACBETH, PROBE and TOPSIS.

For instance, regarding MACBETH, its distinctive characteristic is that it requires only qualitative judgements from the DM about differences in attractiveness (or value) between performance levels to score programs on each criterion and to weight criteria. PROBE, in its turn, allows the DM to take into consideration, in the selection of a portfolio, different forms of interdependence between programs and also to analyze the robustness of a selection in face of uncertainty phenomena (such as imprecise performance estimations or doubts about values and weights) (Tànfani and Testi, 2012). According to Akdag et al. (2014), Hwang and Yoon have developed the TOPSIS method based on the assumption that the best alternative should have the shortest distance from the positive ideal solution and the farthest distance from the negative ideal solution.

Even though there have been successful applications of MCDA methods to healthcare problems and medicine, they have not yet been widely applied to measure quality.

5.2 The ELECTRE TRI-nC method

5.2.1 Overview

There was a family of methods - ELimination and Choice Expressing the REality (ELECTRE) - developed to answer the aforementioned main problematiques of MCDA. This family of methods was developed by Roy and associates of the University of Paris and belongs to the set of methods based on outranking relations. These models tend to focus in a small number of key criteria. However, each of these criteria can be derived from a complex building of several subcriteria. These methods, although more fragile than value functions, are considered more easily built (Belton and Stewart, 2002).

As reported by Figueira et al. (2012), the ELECTRE methods are based, in a certain sense, on the reasons for (concordance) and the reasons against (discordance) of an outranking between two actions and enable a veto threshold to be used. Some characteristics of this family of methods, that are also strong features in comparison to others, are, for instance, the fact that they are able to handle qualitative performance scales of criteria, they can deal with heterogeneous scales, preserving the original performances of the actions on the criteria, without the need of normalizing them, and they do not allow for compensation of performances among criteria, that is, the degradation of performances on certain criteria cannot be compensated by improvements of performances on other criteria (non-compensatory methods). They also allow taking into account the imperfect knowledge of the data and some arbitrariness when building the criteria, through the introduction of the indifference and preference thresholds.

ELECTRE TRI-C was a method designed for sorting problems and dealing with decision aiding situations where each category from a completely ordered set is defined by a single characteristic reference action. The characteristic reference actions are co-constructed through an interactive process involving the analyst and the DM. ELECTRE TRI-C has also been conceived to verify a set of natural structural requirements (conformity, homogeneity, monotonicity, and stability). The method makes use of two joint assignment rules (ascending and descending rules), where the result is a range of categories for each action to be assigned (Figueira et al., 2011).

ELECTRE TRI-C was then generalized to the ELECTRE TRI-nC method, a multiple criteria method of ordinal classification where each category is defined by a set of several reference characteristic actions, rather than one. When characterizing the categories, this provides a "particular freedom" to the DM by allowing more narrow ranges of categories to which an action can be assigned to. Thus, the objective of this new method is not to discover a pre-existing set of categories where the actions would naturally be assigned to, but rather help DMs characterize an appropriate set of categories to receive these actions (Figueira et al., 2012; Almeida-Dias et al., 2012). This is the one that will be applied to the resolution of the problem presented in this dissertation.

A few advantages are outlined by Almeida-Dias et al. (2012), when comparing ELECTRE TRI-nC to ELECTRE TRI-C. For instance:

- The DM can introduce several representative reference actions that she/he considers as appropriate to be assigned to each category.
- It is offered the possibility of "approaching" a frontier between two consecutive categories.
- To do a merging operation, the DM can simply keep the union of the characteristic reference actions of the two merged categories to define the new category.

In general, the ELECTRE TRI-nC method is ideal for situations characterized by three main aspects (Costa and Figueira, 2016):

1. The set of categories, to which the assessed actions will be assigned to, are ordered;
2. For this assessment to be done, the actions' performances are evaluated according to multiple criteria;
3. When assessing actions, they are compared with characteristic reference actions that characterize each category, providing the advantage of an absolute comparison rather than a relative comparison.

Nonetheless, this family of methods also presents limitations, not being necessarily the most adequate method to every existing situation.

For instance, according to Costa and Figueira (2016), sometimes the DM feels that it is fundamental to attribute a performance to an action, and all the scales of the criteria are quantitative. If so, it is appropriate to choose other methods instead. However, even if all criteria are quantitative, ELECTRE methods are still recommended if the user wants to use a non-compensatory method as well as the concepts of concordance and discordance. If transitivity is defined, *a priori*, as a required property for the preferences, this family of methods presents another weakness, even though, typically, methods based on outranking relations do not have to satisfy this feature.

5.2.2 Concepts, definitions and notation

In this subsection, the fundamental concepts of the ELECTRE TRI-nC method, definitions and notations are presented, according to Almeida-Dias et al. (2012) and Costa and Figueira (2016).

Let $A = \{a_1, a_2, ..., a_i, ...\}$ be the set of potential actions, which can be known *a priori*, or can be built progressively during the decision aiding process. Consider as well a certain criterion, g, which is constructed to characterize the potential actions, according to a certain point of view. The characterization of an action a, $g(a)$, represents the performance of that same action according to the considered criterion.

Thus, for this, a family of criteria is necessary, which is designated $F = \{g_1, g_2, ..., g_j, ...g_n\}$ (with $n \geq 3$, otherwhise the concept of concordance is not really pertinent) and will characterize the potential actions to assign them to a existing category, from a set of ordered categories, defined by $C = \{C_1, C_2, ..., C_h, ..., C_q\}$, where $q \geq 2$. C_1 corresponds to the worst category and C_q to the best one.

Notice that each criterion g_j can also have a set of sub-criteria, $G_j = \{g_{j,i}, ..., g_{i,j}, ..., g_{j,n_j}\}$ where $j = 1, ..., n$.

$B = \{B_1, B_2, ..., B_h, ..., B_q\}$ is the set of characteristic reference actions that allow the definition of the categories, and $B_h = \{b_h^r, r = 1, ..., m_h\}$ is a subset of characteristic actions that characterize the category C_h, such that $m_h > 1$ and $h = 1, ..., q$.

Each criterion, g_j, is considered a pseudo-criterion or criterion with thresholds, since it is associated with a threshold of preference (p_j) and a threshold of indifference (q_j), such that $p_j \geq q_j \geq 0$. Note that these thresholds are constructed to model the imperfect character of the data, as well as the arbitrariness underlying the definition of the criteria. If $p_j = 0$, any difference of performances in favour of one action over another can be considered as significant for a strict preference on criterion g_j.

- Preference threshold (p_j): between the performances of two actions, corresponds to the smallest performance difference that, when exceeded, the action with the best performance is considered strictly preferable.

- Indifference threshold (q_j): between the performances of two actions, corresponds to the greatest difference in performance considered compatible, with a situation of indifference between two actions, with different performances.

When two different actions a and a' are considered, where, for a given criterion g_j that is to be maximized, $g_j(a) \geq g_j(a')$, and taking into account the definitions of the mentioned thresholds, it is possible to establish the following binary relations for each criterion:

- $|g_j(a) - g_j(a')| \leq q_j$, where a is *indifferent* to a' according to criterion g_j, represented by aI_ja';

- $g_j(a) - g_j(a') > p_j$, where a is *strictly preferable* to a' according to criterion g_j, represented by aP_ja';

- $q_j < q_j(a) - q_j(a') \leq p_j$, where the judgment is ambiguous, and there are no sufficient reasons to conclude an indifference situation, nor a strict preference between the two actions. There is a hesitation between indifference and strict preference, meaning that a is *weakly preferable* to a', represented by aQ_ja'.

5.2.3 Definition of an outranking relation

The outranking relation, represented by aS_ja' means that "action a is, at least, as good as action a'", according to criterion g_j.

According to Costa and Figueira (2016), to create an outranking relation, there are three basic concepts that need to be addressed: concordance, non-discordance and degree of credibility:

– Concordance: For aS_ja' to be accepted, a sufficient majority of criteria must be in favor of this relation. The concordance of the aS_ja' affirmation is estimated by the global concordance index, $c(a,a')$, that associates each criterion to a weight w_j, such that $w_j > 0$, for $j = 1, ..., n$ and $\sum_{n}^{j=1} w_j = 1$ (assuming the sum of all weights is 1).
  The global concordance index is defined by the following expression:

$$c(a,a') = \sum_{j \in C(aPa')} w_j + \sum_{j \in C(aQa')} w_j + \sum_{j \in C(aIa')} w_j + \sum_{j \in C(aQa')} w_j \varphi_j$$

  where

$$\varphi_j = \frac{p_j - (g_j(a') - g_j(a))}{p_j - q_j} \in [0, 1[$$

  The fact that this index takes the weights of each criterion into account contributes to validate the statement "a outranks a'".

– Non-discordance: happens when none of the minority criteria that opposes aS_ja' exercises its power to veto this statement. The discordance is estimated by the discordance index, that associates each criterion g_j to a veto power, v_j, such that $v_j > p_j$. For each criterion, the veto effect is estimated through the partial discordance index, $d_j(a,a')$, with $j = 1, ..., n$, which is defined as:

$$d_j(a,a') = \begin{cases} 0 & if \quad g_j(a) - g_j(a') \geq -pj, \\ \frac{g_j(a) - g_j(a') + p_j}{p_j - v_j} & if \quad -v_j \leq g_j(a) - g_j(a') < -pj, \\ 1 & if \quad g_j(a) - g_j(a') < -vj. \end{cases}$$

– Credibility index: The credibility index, $\sigma(a,a')$, reflects how the statement "a outranks a'" is justified when the whole family of criteria, F, is considered. To estimate this index, the global concordance index and the partial discordance index are considered, resulting in the following expression:

$$\sigma(a,a') = c(a,a') \prod_{j=1}^{n} T_j(a,a')$$

  where $T_j(a,a') = \begin{cases} \frac{1 - d_j(a,a')}{1 - c(a,a')} & if \quad d_j(a,a') > c(a,a'), \\ 1 & otherwhise. \end{cases}$

To convert fuzzy outranking relations into crisp ones, the level of credibility, λ, is used. λ is considered to be the minimum credibility level of $\sigma(a,a')$, which is necessary for the DM to validate, or not, the statement "a outranks a'", taking into account all criteria from F. Typically, λ takes values between [0.5; 1.0[.

The level of credibility, λ, is compared to the credibility indexes of the different actions and to the set of reference actions in each category, where $\sigma(\{a\}, B_h) = max_{r=1,...,m_h} \{\sigma(a, b_h^r)\}$ and $\sigma(B_h, \{a\}) = max_{s=1,...,m_h} \{\sigma(b_h^s, a)\}$, making possible the definition of four binary relations:

- λ-outranking: $\{a\} S^\lambda B_h \Leftrightarrow \sigma(\{a\}, B_h) \geq \lambda$;
- λ-preference: $\{a\} P^\lambda B_h \Leftrightarrow \sigma(\{a\}, B_h) \geq \lambda \wedge \sigma(B_h, \{a\}) < \lambda$
- λ-indifference: $\{a\} I^\lambda B_h \Leftrightarrow \sigma(\{a\}, B_h) \geq \lambda \wedge \sigma(B_h, \{a\}) \geq \lambda$
- λ-incomparability: $\{a\} R^\lambda B_h \Leftrightarrow \sigma(\{a\}, B_h) < \lambda \wedge \sigma(B_h, \{a\}) < \lambda$

5.2.4 Assignment Procedure

The ELECTRE TRI-nC method makes use of two rules conjointly. These rules allow the assignment of a possible category or a set of possible categories to an action, a, which is compared with the subsets of the reference actions, B_h, taking into account a certain level of credibility, λ, previously chosen.

Firstly, they pre-select a category between two possible ones, and secondly they select an appropriate category by using a function, $\rho(\{a\}, B_h)$, to which an action a can be attributed (Greco et al., 2005; Costa and Figueira, 2016; Almeida-Dias et al., 2012).

The selection function is:

$$\rho(\{a\}, B_h) = min \{\sigma(\{a\}, B_h); \sigma(B_h, \{a\})\}$$

The two joint rules, ascending rule and descending rule, are presented below.

1. Ascending Rule

Choose a level of credibility, λ ($\frac{1}{2} \leq \lambda \leq 1$) and increase the value of h, from zero to the first value, k, such that $\sigma(B_k, \{a\},) \geq \lambda$. C_k is called the ascending pre-selected category.

- For $k = 1$, select C_1 as a possible category to assign action a.
- For $1 < k < (q + 1)$, if $\rho(\{a\}, B_k) > \rho(\{a\}, B_{k-1})$, then select C_t as a possible category to assign action a; otherwise, select C_{k-1}.
- For $k = (q + 1)$, select C_q as a possible category to assign action a.

2. Descending Rule

Choose a level of credibility, λ ($\frac{1}{2} \leq \lambda \leq 1$) and decrease the value of h, from $(q+1)$ to the first value, t, such that $\sigma(\{a\}, B_t) \geq \lambda$. C_t is called the descending pre-selected category.

- For $t = q$, select C_q as a possible category to assign action a.
- For $0 < t < q$, if $\rho(\{a\}, B_t) > \rho(\{a\}, B_{t+1})$, then select C_t as a possible category to assign action a; otherwise, select C_{t+1}.
- For $t = 0$, select C_1 as a possible category to assign action a.

Using these rules simultaneously, ELECTRE TRI-nC provides a possible assignment for each action. A lowest and highest possible categories to which an action can be assigned are selected.

An action a can be assigned to:

– One category, when the selected minimum and maximum categories are the same;

– Two categories, when the selected categories are consecutive;

– A range of more than two consecutive categories, delimited by the two selected categories.

5.2.5 Application

In this dissertation, the tool used to help carrying the application of the ELECTRE TRI-nC method was MCDA-ULaval, which is a software tool programmed in Java, developed at Universitè Laval, that implements multi-criteria decision analysis algorithms and supports all the ELECTRE family methods. This system is based on the concept of projects, allowing the users to create, edit and delete them, containing different types of objects, including multiple datasets, actions, and criteria, which can result in several performance tables, and decision configurations. It also has many benefits, such as allowing the association of constant, inversely or linearly variable thresholds, the use of criteria in ordinal and cardinal scales, the possibility of importing and exporting data and the possibility of automatically normalizing weights. It also enables the performance of a sensitivity analysis of the decision parameters and the presentation of the analysis of the scenario using diagrams, graphs and charts for the data, for their incorporation in the results[1].

5.3 Summary

In this chapter, the concept of MCDA methods was presented. This approach takes into account the multiple criteria dimension of a problem (more than one criterion is used), allowing different types of information to be integrated, different points of view, as well as the preferences of DMs, being very useful to solve decision making problems in a well informed way. They are a field that is increasingly growing and attracting interest. The main goal of these methods is to help DMs make more consistent, transparent and robust choices with the assistance of an analyst.

The MCDA process has two main phases: problem structuring, where the problem is identified, along with the points of view, objectives, stakeholders and potential actions, and model building, where a model that represents the problem is constructed. It also faces 3 main problematiques: choosing, sorting and ranking.

This kind of approach has been applied to multiple fields and can be very useful in the healthcare sector, since it is a complex one and constantly faces hard decisions that involve different perspectives. Lastly, there are different types of methods and methodologies, such as value measurement, goal programming and outranking methods.

This is an important introduction to provide background context for the method chosen throughout this dissertation, the ELECTRE TRI-nC, which is one of the outranking methods available.

ELECRE TRI-nC is a MCDA method of ordinal classification that belongs to the ELECTRE family and uses two rules of affectation (ascending rule and descending rule) conjointly, to select possible categories (it can be a single category or a range of them) to assign a set of actions to, according to the performance of each action in a set of criteria. This method can take into account several reference actions to characterize each category, which differentiates it from its ancestor ELECTRE TRI-C.

The application of ELECTRE TRI-nC starts with the definition of the problem and, then, outranking relations are defined by comparing the level of credibility with the credibility index, calculated between each action and the set of reference actions in each category. Afterwards, the simultaneous application of the descending and ascending rules is carried out to assign each potential action to one of the defined categories. Its application extends to several decision support contexts, in the most diverse areas, but each situation must be thoughtfully considered and obey to certain conditions to make sure it is the most adequate and recommended method.

To apply this method to a data set, a platform is required; specifically, in this dissertation, the MCDA-ULaval software tool was chosen.

[1] MCDA-ULaval (https://cersvr1.fsa.ulaval.ca/mcda-ulaval/?q=en).

Chapter 6

Case Study

In this chapter, the DM and the relationship with him throughout this dissertation is presented, alongside with details of the database and sample chosen. Moreover, the inputs to be used for the ELECTRE TRI-nC method are exposed, including the criteria, indicators and actions. It is also introduced the Simos Roy Figueira (SRF) procedure for criteria weighting and the other elements of the model, namely the reference actions and thresholds considered.

6.1 Decision Maker

Due to the COVID-19 pandemic, we were not able to resort to the DM initially planned, which was an expert working in the Ministry of Health. Thus, the DM of this dissertation was, instead, an expert in the healthcare sector, that possesses know-how in the area and has several published work concerning performance assessment of quality and efficiency.

The DM was present during different steps of this study, cooperating in tasks such as the treatment of the benchmarking data. Particularly, the DM accompanied the selection of indicators that structured the criteria, the weighting procedure and the definition of thresholds and reference actions assigned to each category.

6.2 Database

As seen in Section 2.3.2, with the objective of improving economic performance, access and quality of services, and to increase the transparency for the population by publicly providing information, there are hospitals' benchmarking dashboards available since 2013. ACSS has particularly developed a website which includes monthly reports from the hospitals of the SNS. As an official source, the data is considered reliable and substantial and can be easily assessed using the website (https://benchmarking-acss.min-saude.pt/) and exporting excel files of the data, making it suitable as the main database for this dissertation.

6.3 Sample

When it comes to the time interval, there was data available from every month of 2013 to, in some dimensions, the beginning of 2020.

The data chosen for this analysis was the year of 2018, i.e., data from every month since January of 2018 until December of 2018, since this was the most recent completed year, enabling a more exhaustive investigation and a possible comparison between each month.

The year of 2019 was excluded due to lack of data provided by some of the institutions under scrutiny and the fact that, in the efficiency dimension, data were not available for months later than September.

Besides this, it is important to note that the PPPs Loures Hospital and Cascais Hospital did not provide information for the efficiency and productivity dimensions in any year whatsoever, which includes data for indicators such as ”Inpatient per FTE (Full-Time Equivalent) doctor/nurse” and economic expenses per inpatient.

Hence, it was decided to carry through two models for the year of 2018:

- Model 1: Includes all criteria, contemplating Vila Franca de Xira Hospital as the only PPP element.
- Model 2: Excludes efficiency-related criteria, including all three PPP hospitals.

In the cases where an institution presented lack of data in one or more months in one of the other dimensions, the values were extrapolated applying a linear regression, through other high correlated variables. Each health entity in each month was considered as a separate unit from the previous month.

Concerning the healthcare providers, the benchmarking data includes 43 health institutions, so it was necessary to select the ones relevant to this study. The clustering groups in which the entities were divided and presented on the website were not considered, since one wanted hospitals to be treated as individual institutions in this analysis. We excluded:

– The three oncology centers (IPO Coimbra, IPO Porto, IPO Lisboa), since these present specific technology of production (directed to cancer);

– All eight local health units, since comparing them with PPPs, public hospitals and hospital centers in indicators of the efficiency and productivity dimensions would lead to unreliable results due to their different structure and paying system;

– Figueira da Foz District Hospital and Santa Maria Maior Hospital, since they do not perform birth deliveries which is one of the dimensions under scrutiny (see below).

Concerning the ACSS indicators, in the website there is a total of 34 indicators grouped in six benchmarking dimensions (Access, Performance Assistance, Safety, Volume and Usage, Productivity, and Economic-Financial). It was decided that the adopted method would not consider dimensions, combining indicators and handling them as criteria in the same level, in line with a similar study, Jorge (2019). To select the ones to be used in this dissertation, it was important to take into account the integrity of the data and the DM perspective. Thus, only the considered most representative indicators for this study were chosen, which means that some indicators were excluded or merged.

Nonetheless, some considerations on this point are noteworthy:

– The indicators of the dimension ”Volume and Usage” were all excluded, since these are specific to some entities, existing lack of data concerning the others;

– The *Annual occupancy rate* indicator was converted into the absolute difference in inpatient bed annual occupancy rate to a reference value of 85%, considered ideal for this rate, since the original indicator could not be maximized or minimized due to the trade-off between the Productivity and Access dimensions;

– The indicators considered the most relevant ones of the ”Efficiency” dimension, *Expenses with staff*, *Expenses with pharmaceutical products*, *Expenses with drugs*, *Expenses with clinical consumption material* and *Expenses with supplies and external services* per standard patient were merged, for practicality reasons, in an individual indicator, *Operational Expenses per standard patient*;

– Since almost all their values are null, most childbirth related indicators are not considered due to the lack of relevance of the data. The most relevant one was selected;

– The indicator *standard patient per FTE nurse* was excluded since it is highly correlated to the indicator *standard patient per FTE doctor* (with a correlation coefficient of 0.73), being enough to evaluate just one of the indicators.

6.4 Actions

In this case study, the actions are represented by the health entities (PPPs, hospitals or hospital centres) under scrutiny. The final list of the actions included in this study, with its corresponding notation, is shown on Table 6.1.

Table 6.1: Health Institutions/actions under scrutinity.

EPEs		**PPPs**
Hospitals	**Hospital Centres (HC)**	
a_{12} - Senhora da Oliveira, Guimarães Hospital	a_1 - Médio Ave	a_{28} - Vila Franca de Xira Hospital
a_{13} - Santarém District Hospital	a_2 - Oeste	a_{29} - Cascais Hospital
a_{18} - Braga Hospital	a_3 - Póvoa do Varzim/Vila do Conde	a_{30} - Loures Hospital
a_{19} - Espírito Santo de Évora Hospital	a_4 - Barreiro/Montijo	
a_{20} - Fernando Fonseca Hospital	a_5 - Leiria	
a_{21} - Garcia de Orta Hospital	a_6 - Setúbal	
	a_7 - Baixo Vouga	
	a_8 - Entre Douro e Vouga	
	a_9 - Médio Tejo	
	a_{10} - Tâmega e Sousa	
	a_{11} - Cova da Beira University	
	a_{14} - Tondela-Viseu	
	a_{15} - Trás-os-Montes e Alto Douro	
	a_{16} - Algarve University	
	a_{17} - Vila Nova de Gaia/Espinho	
	a_{22} - Lisboa Ocidental	
	a_{23} - Coimbra University	
	a_{24} - Lisboa Central	
	a_{25} - São João University	
	a_{26} - Porto University	
	a_{27} - Lisboa Norte University	

6.5 Construction of the Criteria Tree

Considering the literature review carried out in previous chapters, the benchmarking dimensions and the DM's input, it was possible to identify different points of view, under which ten criteria were later defined, described by the elected indicators.

6.5.1 Points of View (POV)

The points of view considered were:

- **Access**: Evaluates the system's ability of providing care services to any citizen whenever necessary and at his/her will. This dimension includes aspects such as availability (available hospital resources), timeliness (capacity of delivering healthcare services at the right time whenever required, presenting short waiting lists and waiting times), accessibility (the cost and distance between home and the healthcare service), among others (Institute of Medicine, 1993; Campbell et al., 2000; Gulliford et al., 2002; Doetsch et al., 2017).
- **Care Appropriateness**: Measures the ability of delivering patient-centred care services supported by evidence-based guidelines. An intervention or service is considered appropriate if the expected health benefits for the patient (e.g., increased life expectancy, pain relief, improved functional capacity) beat its expected health risks (e.g., mortality, morbidity, pain caused by the intervention, inaccurate diagnoses) by a wide enough margin to make the intervention or service beneficial (Naylor, 1998; Robertson-Preidler et al., 2017; National Academies of Sciences, Engineering, and Medicine, 2018). The disrespect of this dimension can result in avoidable re-admissions after inpatient discharge and excessive staying, which, in its turn, can increase the probability of the development of other diseases (Ferreira and Marques, 2019).
- **Safety**: Measures the capacity of preventing and reducing the risk of complications, harm, injuries or even deaths happening to patients during the process of care (Pronovost et al., 2005; Hughes, 2008; Emanuel et al., 2009; Ferreira and Marques, 2019).
- **Caesarean Appropriateness**: Just like care appropriateness, caesarean appropriateness is related to the ability of delivering patient-centered care services supported by evidence-based guidelines, in this case in caesarean sections. According to Betrán et al. (2016), Queenan (2011), Robson et al. (2013), and Robson (2018), a caesarean section, when required and medically justified, can effectively prevent and decrease maternal and perinatal mortality and morbidity. However, otherwise, there is no evidence describing the benefits of this procedure for women or infants. Caesarean sections can cause significant and sometimes permanent complications, disability or even death, especially in situations related to lack of prepared facilities and/or capacity to properly conduct safe surgery and deal with surgical complications. Nonetheless, the rates of caesarean delivery have increased over time in most OECD countries and globally. WHO reported that, annually, more than six million caesarean sections are linked to inappropriate care (Betrán et al., 2016).
- **Efficiency**: Concerns the system's capacity of treating patients with the available resources, with the minimum waste, and regardless of the occurrence of complications, deviations from evidence-based medicine guidelines, or even in-hospital mortality episodes (Jacobs et al., 2006). It usually considers the ratio between outputs (health outcomes) and inputs (expenditures, staff and other resources) (Mitropoulos et al., 2013; Bem et al., 2014; World Health Organization, 2016; Blatnik et al., 2017). The main goal of the health providers should be a financially sustainable management

of the resources, while at the same time delivering the best care possible, being cost-effective. However, a hospital being technically efficient does not necessarily mean the best practices are being followed, since it can be a reflection of lack of investment on safety, care appropriateness and access, to increase the quantity of treated patients (Guven, 2007; Ferreira and Marques, 2019).

6.5.2 Criteria and indicators

After establishing the Points of View, it was possible to outline one or more important criteria within each of them, which are described by specific indicators.

Note that several recent studies, which also intended to evaluate Portuguese healthcare services, and have already been mentioned throughout this dissertation, have relied on some of these indicators as well, adapting them according to their specific goals. For instance, take Nunes et al. (2019), Ferreira and Marques (2019), Ferreira et al. (2019), Jorge (2019), Rocha (2019), Ferreira and Marques (2020), and Ferreira et al. (2020).

In accordance with the Section 6.3, the indicators described below were selected in consonance with their relevance for this particular study, and are available to be assessed in the ACSS benchmarking website.

In the POV of Access, two main criteria were identified:

– **First medical appointments timeliness**, g_1: As previously mentioned, for a healthcare service to be considered accessible, it needs to treat its patients in a fairly timely manner, whenever required. The indicator that represents this criterion is the *Number of non-urgent first medical appointments performed in adequate time per 100 first medical appointments*, which is the amount of first medical appointments that occur in guaranteed maximum response times compared to the total number of first medical appointments, and the goal is for it to be maximized;

– **Occupancy Rate**, g_2: This criterion takes into account the occupancy rate of the healthcare services. The number of hospital beds measures the resources available to provide the necessary care to inpatients, which is associated with accessibility and equity of the services provided. 85% is referred to as the ideal value for occupancy rate. Considering this assumption, the values used to measure hospitals' performance were the absolute of the difference of the values available by the indicator *Occupancy rate* to the aforementioned ideal value. This difference is to be minimized.

In the POV of Care Appropriateness, the considered most representative criteria were:

– **Minor surgeries appropriateness**, g_3: This criterion is directly related to care appropriateness, since it concerns the possibility of some major surgeries being executed as minor procedures, without harming the patient. The corresponding indicator to this criterion is the *Number of outpatient surgeries per 100 potential outpatient procedures*, and is to be maximized;

– **Avoidable re-admission in 30 days after discharge**, g_4: This criterion is treated as a representative one under this POV since it can be a direct consequence of lack of appropriate care, provided before or after discharge. Evidently, it must be minimized, and is described by the indicator *Number of re-admissions in 30 days after discharge per 100 inpatients*.

For the Safety POV, the following three criteria were chosen, taking into account that all of them are considered preventable occurrences that harm the patients and, consequently, are seen as a sign of inadequate clinical safety. Their related indicators must, then, all be minimized.

– **Bedsores**, g_5: Can be assessed employing the indicator *Number of bedsores per 100 inpatients*, which represents the percentage of bedsore episodes related to the episodes where exclusion of bedsore has occurred;

– **Postoperative pulmonary embolisms or thrombosis**, g_6: The indicator used to measure this criterion corresponds to the *Postoperative pulmonary embolism/deep venous thrombosis cases per 100 surgical procedures*, which represents the percentage of postoperative pulmonary embolism/deep venous thrombosis episodes related to the episodes where exclusion has occurred;

– **Postoperative septicaemia**, g_7: To assess this criterion, one uses the indicator *Postoperative septicaemia cases per 100 inpatients*, which represents the percentage of postoperative septicaemia episodes related to the episodes where exclusion has occurred.

Under the POV Caesarean Appropriateness, only one criterion was selected: **Caesarean sections in UCFTPs** (Unifetal, Cephalicand Full-term Pregnancies), g_8. As previously mentioned, the performance of caesarean sections in UCFTPs can be related to inadequate care, thus, it must be minimized, through the indicator *Number of cesarean sections in UCFTPs per 100 sections in UCFTPs*, which represents in percentage the number of cesarean sections in UCFTP, considering all deliveries.

Two main criteria were identified under the POV of Efficiency, where one of them is described by the merge of various indicators.

– **Operational Expenses**, g_9: This criterion includes expenses with staff, drugs, pharmaceutical products, clinical consumables, supplies and external services, which were all grouped together. All of these expenses, both individually and joined, are related to efficiency, since they all constitute resources that must be carefully managed. Thus, the value of this criterion must be minimized, without compromising the patients' health and care. It can be measured through the indicator *Expenses with staff, drugs, pharmaceutical products, clinical consumables, supplies and external services per standard patient*, which is represented in €;

– **Inpatient per FTE doctor**, g_{10}: ACSS provides this information in the FTE unit, that measures the workload of employed people in a way that makes them comparable in different contexts[1]. The presence of doctors in healthcare services' providers is obviously critical and, for the best possible quality of care to be achieved, these doctors must exist in an adequate number and properly distributed in a hospital. This criterion could be evaluated from both the Access or Efficiency points of view. In the Access POV, it is recommended a low number of patients per doctor. However, under the Efficiency perspective (output/input), to which it is assigned in this study, doctors are seen as resources (inputs) needed to achieve the patients' health status (outputs). Thus, a higher number of patients treated per doctor is wanted, maximizing the *standard patient per full time equivalent doctor* indicator.

A Criteria Tree could, then, be created, taking all of the previous information into account and compiling it on Table 6.2.

[1] Eurostat - Glossary: Full-time equivalent (FTE) (https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Full-time_equivalent_(FTE)). Accessed: 21/09/2020.

Table 6.2: Points of view, criteria, corresponding indicators and direction preferences.

Points of View	Criteria	Indicators	Direction
Access	g_1: First medical appointments timeliness	Number of first medical appointments performed in adequate time per 100 first medical appointments	Maximize
	g_2: Occupancy Rate	Absolute difference in inpatient bed annual occupancy rate to a reference value of 85%	Minimize
Care Appropriateness	g_3: Minor surgeries appropriateness	Number of outpatient surgeries in the Total of Scheduled Surgeries (GDH) for ambulatory procedures per 100 potential outpatient procedures	Maximize
	g_4: Avoidable re-admission in 30 days after discharge	Number of re-admissions in 30 days after discharge per 100 inpatients	Minimize
Safety	g_5: Bedsores	Number of bedsores per 100 inpatients	Minimize
	g_6: Postoperative pulmonary embolisms or thrombosis	Postoperative pulmonary embolism/ deep venous thrombosis cases per 100 surgical procedures	Minimize
	g_7: Postoperative septicaemia	Postoperative septicemia cases per 100 inpatients	Minimize
Caesarean Appropriateness	g_8: Caesarean sections in UCFTPs	Number of cesarean sections in UCFTPs per 100 sections in UCFTPs	Minimize
Efficiency	g_9: Operational Expenses	Expenses with staff, drugs, pharmaceutical products, clinical consumables, supplies and external services per standard patient	Minimize
	g_{10}: Inpatient per FTE doctor	Standard patient per full time equivalent doctor	Maximize

6.6 Elements of the model

Different preference parameters need to be established before applying the method. This step is done with the cooperation of the DM, and includes defining categories and their reference actions, criteria weights and thresholds. Finally, performance tables are constructed to be part of the inputs.

6.6.1 Categories and Reference actions

For the criteria, the DM was able to establish different reference actions per category, to later apply the ELECTRE TRI-nC method.

To achieve this, the DM defined five categories *a priori*:

– C_1: Very Weak performance;

– C_2: Weak performance;

– C_3: Neutral performance;

– C_4: Good performance;

– C_5: Very Good performance.

Then, for each of them, one or more characteristic reference actions were defined, as well as their performance in every criterion (see Table 6.3). For instance, for an action to be considered ”Very Good”, at least 95 out of the 100 first medical appointments would be performed in adequate time; the absolute difference between the inpatient bed occupancy rate to the reference value would be, at most, 0.1; there would be performed at least 90 outpatient surgeries out of 100 potential outpatient procedures for ambulatory procedures; there would be no more than three re-admissions in 30 days after discharge, per 100 patients; there would be no cases of bedsores, postoperative pulmonary embolisms or thrombosis, and postoperative septicaemia per 100 inpatients, and a maximum of eight caesarean sections performed, per 100 deliveries. Finally, the operational expenses would not surpass 2700 €, and a minimum of 7.5 inpatients per FTE doctor would be expected.

Table 6.3: Categories, reference actions and their performance for each criteria.

Category		**Performance**	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}
C_1	b_1^1	Very Weak	55	15	70	10	0.8	0.85	3	45	3800	4
C_2	$b_{2,1}^1$	Weak	65	12	75	8	0.65	0.7	2	35	3500	4.5
	$b_{2,2}^1$		70	9	75	8	0.5	0.7	1.5	30	3500	4.5
C_3	b_3^1	Neutral	80	5	80	6	0.3	0.5	0.7	25	3250	5
C_4	$b_{4,1}^1$	Good	85	3	85	4	0.15	0.3	0.3	15	3000	6
	$b_{4,2}^1$		90	2	85	4	0.1	0.2	0.15	10	3000	6.5
C_5	b_5^1	Very Good	95	0.1	90	3	0	0	0	8	2700	7.5

6.6.2 Criteria Weighting: the SRF procedure

Several methods can be used to give an appropriate value to the weights of criteria.

In 1994, Simos proposed a very simple procedure that allowed the indirect determination of numerical values for weights, using a set of cards. This procedure, the Simos' deck of cards method, has been applied to different real-life contexts over the years.

In 2002, the modified SRF procedure was proposed, consisting on a revision of the aforementioned method, allowing additional information to be included, concerning the ratio between the weights of the most important criterion and the least important one in the ranking, and interval scales.

This was the method chosen in this dissertation to achieve this step of the analysis. As an implementation of the revised Simos' procedure, the SRF software can also be applied to determine the weights of criteria in the ELECTRE type methods and must allow any DM to easily think about and express a ranking of criteria and to introduce some complementary information in the software to obtain the weights of the criteria (Figueira and Roy, 2002).

The SRF procedure considers two phases. The first one consists on a meeting with the DM to collect all the information needed for the application of the method, after the definition of the criteria, and the second regards the calculation of the weights of each criterion, which will be performed in the DecSpace platform [2].

In the first phase, four steps are followed, being that the fourth one is introduced by the revised SRF procedure.

1. The user (DM) is given a set of cards, where, in each card, the name of each criterion is written. Therefore, we have n cards, n being the number of criteria of a family.

[2] DecSpace (Pre-Alpha) - A multi-criteria decision analysis framework (http://app.decspacedev.sysresearch.org/?fbclid=IwAR2hJjgp2z49JEZH78nOxM9VdMgt6jixu4INIvtQO6bLz5xGoxDslsbEV_w#/). Accessed: 09/07/2020

2. The user is asked to rank these cards (or criteria) from the least important to the most important. So, the user will rank in ascending order, the first criterion in the ranking is the least important and the last criterion in the ranking is the most important one. In the case of criteria having the same importance (same weight), they are grouped together.

3. The user is asked to introduce white cards between two successive cards. The greater the difference between the mentioned weights of the criteria, the greater the number of white cards. No blank card added means that the difference of two consecutive levels is one unit; one blank card means the difference of importance is two units, and so on.

4. The user is asked to state how many times the last criterion is more important than the first one in the ranking. The value of this ratio is designated z. This software allows the user to introduce different values concerning the ratio z (between the weight of the most important criterion and the weight of the least important one in the ranking) since it is very difficult to express this ratio using a single constant value (Figueira and Roy, 2002).

Moving on to the second phase, the DecSpace website is used as support to execute the SRF procedure. Firstly, a DCM-SRF project is created, enabling the implementation of the revised Simos' procedure. The information previously gathered with the DM, including the criteria, ranking of the criteria, the blank cards and the value of the ratio-z, is then inserted. The software allows the user to choose the number of decimal places (one or two) and the weight type (normalized, non-normalized or both displayed), in our case we chose one decimal and the normalized weight type.

One has to address, once again, the two models already mentioned: the one that includes all criteria, Model 1, and the other which excludes efficiency-related criteria ("Operational Expenses", g_9, and "Inpatient per FTE doctor", g_{10}), Model 2.

Besides this, within Model 1, two scenarios were examined, one where the objective of the DM was social, i.e, the goal was to minimize adverse effects and improve access, thus the criterion "Operational expenses" was considered one of the least important ones; and another where this criterion was, instead, considered the most important one, since the goal was efficiency-oriented. The orders chosen by the DM in both scenarios of Model 1 are represented on Tables 6.4 and 6.5, while the order chosen considering Model 2 is represented on Table 6.6. The value picked for the ratio z was 2.5. Note that the ranks were ordered numerically, so the higher number (6/7) represents the most important ranked criteria and the lower number (1) represents the least important ranked criteria.

Table 6.4: Ranking of criteria and white cards considering Model 1 and the social-oriented goal scenario.

Rank	Cards and White Cards
7	g_6, g_7
	0
6	g_4
	0
5	g_5
	1
4	g_2, g_3, g_{10}
	1
3	g_1
	1
2	g_9
	2
1	g_8

Table 6.5: Ranking of criteria and white cards considering Model 1 and the efficiency-oriented goal scenario.

Rank	Cards and White Cards
7	g_9
	1
6	g_6, g_7
	0
5	g_4
	0
4	g_5
	0
3	g_2, g_3, g_{10}
	1
2	g_1
	2
1	g_8

Table 6.6: Ranking of criteria and white cards considering Model 2.

Rank	Cards and White Cards
6	g_6, g_7
	0
5	g_4
	0
4	g_5
	1
3	g_2, g_3
	1
2	g_1
	2
1	g_8

The results obtained applying this method to both scenarios of the first model are represented on Tables 6.7 and 6.8 and to the second model on Table 6.9.

Table 6.7: Weights of the criteria considering all criteria and the social-oriented goal scenario (Model 1).

Criteria	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}
Weights(%)	8.6	10	10	12.1	11.4	12.8	12.8	5.1	7.2	10

Table 6.8: Weights of the criteria considering all criteria and the efficiency-oriented goal scenario (Model 1).

Criteria	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}
Weights(%)	7.6	9.1	9.1	11.3	10.6	12.1	12.1	5.4	13.6	9.1

Table 6.9: Weights of the criteria excluding the efficiency-related ones (Model 2).

Criteria	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
Weights(%)	9.6	11.7	11.7	14.9	13.9	15.9	15.9	6.4

6.6.3 Thresholds

As previously noted on Chapter 5, thresholds exist to acknowledge and attenuate the imperfect knowledge of data and the underlying arbitrariness of the criteria definition. Hence, each criterion is associated with a threshold of preference (p_j) and a threshold of indifference (q_j). The veto threshold (v_j) is also important to diminish compensatory effects.

Taking into account that all criteria are described in quantitative scales of levels and that it is possible to assign different reference actions per category, the thresholds presented on Table 6.10 were defined.

Table 6.10: Preference, indifference and veto threshold values for each criterion.

Thresholds	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}
q_j (Indifference)	0.5	2	1.5	1	0.05	0.1	0.2	2	1000	0.1
p_j (Preference)	1	4	3	2	0.1	0.2	0.3	4	3000	0.2
v_j (Veto)	10	15	15	6	0.6	0.5	2	15	20000	3

As for the credibility level, λ, it was seen on Chapter 5 that it represents the minimum level of credibility needed for the DM to validate, or not, an outranking relation, considering all criteria (Tervonen et al., 2009). In this study, an interval of values, from 0.55 to 0.65, was tried out.

6.6.4 Performance Tables

For each year, 24 performance tables were created, 12 for each model, one for each month. On Tables 6.11 and 6.12, performance tables of January are displayed, with and without the efficiency-related criteria, as an example.

Table 6.11: Performance table for January, considering all criteria (Model 1).

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}
a_1	67.2	16.40	82.4	7.14	0.00	0.00	0.00	28.1	2441	5.4
a_2	52.6	15.46	82.0	4.23	0.00	0.00	0.84	25.0	3005	5.2
a_3	96.8	6.99	77.1	7.09	0.00	0.00	0.00	32.3	2777	5.6
a_4	79.9	8.32	77.7	8.05	0.20	0.00	0.00	37.3	3142	6.0
a_5	49.3	7.42	84.8	10.70	0.00	0.00	0.00	38.4	1742	10.6
a_6	73.8	4.65	84.3	8.63	0.00	0.00	0.61	27.6	2645	7.5
a_7	63.1	8.08	83.8	7.57	0.14	0.00	0.00	37.1	3047	5.6
a_8	65.5	10.61	90.9	6.27	0.00	0.32	0.00	24.7	3006	7.2
a_9	84.6	17.73	84.3	9.95	0.00	0.00	0.00	30.2	2917	7.9
a_{10}	42.0	20.37	88.6	5.64	0.00	0.00	0.00	25.9	2476	8.1
a_{11}	65.0	1.77	73.8	9.53	0.00	0.59	1.19	29.3	3496	6.8
a_{12}	57.1	37.43	84.0	8.03	0.00	0.00	0.51	29.7	2975	6.4
a_{13}	69.1	4.53	78.9	10.76	0.00	0.45	0.00	32.3	3868	6.1
a_{14}	73.2	9.44	95.2	8.66	0.10	0.45	0.43	27.3	2538	5.1

a_{15}	61.8	6.36	85.6	10.73	0.10	0.00	0.93	45.5	2289	7.6
a_{16}	71.8	11.59	82.7	6.65	0.00	0.00	0.69	23.2	2940	6.0
a_{17}	50.7	10.93	85.5	7.41	0.13	0.64	0.00	35.4	2848	6.5
a_{18}	73.9	12.14	86.9	7.03	0.24	0.64	0.62	27.1	2519	8.7
a_{19}	56.3	4.08	73.7	5.04	0.00	0.00	0.89	34.3	3028	6.6
a_{20}	61.1	9.65	80.7	7.00	0.00	0.57	1.95	32.0	2715	6.8
a_{21}	90.4	7.18	89.4	7.65	0.13	0.38	1.13	23.5	2611	9.3
a_{22}	66.7	0.52	78.7	8.14	0.00	0.81	0.26	24.3	2787	7.7
a_{23}	63.3	1.10	76.6	9.12	0.05	0.00	0.27	41.6	2792	7.3
a_{24}	79.4	9.39	81.4	7.37	0.07	0.23	1.48	26.6	3424	6.0
a_{25}	53.3	7.46	83.3	8.08	0.00	0.48	0.32	24.6	2556	8.2
a_{26}	71.0	16.98	81.9	7.93	0.00	0.77	0.24	29.5	2565	7.9
a_{27}	65.0	8.89	81.1	10.53	0.00	0.32	1.47	42.1	3192	7.5
a_{28}	76.8	17.33	85.7	8.11	0.00	0.00	1.11	27.2	2589	6.8

Table 6.12: Performance table for January, excluding efficiency-related criteria (Model 2).

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
a_1	67.2	16.40	82.4	7.14	0.00	0.00	0.00	28.1
a_2	52.6	15.46	82.0	4.23	0.00	0.00	0.84	25.0
a_3	96.8	6.99	77.1	7.09	0.00	0.00	0.00	32.3
a_4	79.9	8.32	77.7	8.05	0.20	0.00	0.00	37.3
a_5	49.3	7.42	84.8	10.70	0.00	0.00	0.00	38.4
a_6	73.8	4.65	84.3	8.63	0.00	0.00	0.61	27.6
a_7	63.1	8.08	83.8	7.57	0.14	0.00	0.00	37.1
a_8	65.5	10.61	90.9	6.27	0.00	0.32	0.00	24.7
a_9	84.6	17.73	84.3	9.95	0.00	0.00	0.00	30.2
a_{10}	42.0	20.37	88.6	5.64	0.00	0.00	0.00	25.9
a_{11}	65.0	1.77	73.8	9.53	0.00	0.59	1.19	29.3
a_{12}	57.1	37.43	84.0	8.03	0.00	0.00	0.51	29.7
a_{13}	69.1	4.53	78.9	10.76	0.00	0.45	0.00	32.3
a_{14}	73.2	9.44	95.2	8.66	0.10	0.45	0.43	27.3
a_{15}	61.8	6.36	85.6	10.73	0.10	0.00	0.93	45.5
a_{16}	71.8	11.59	82.7	6.65	0.00	0.00	0.69	23.2
a_{17}	50.7	10.93	85.5	7.41	0.13	0.64	0.00	35.4
a_{18}	73.9	12.14	86.9	7.03	0.24	0.64	0.62	27.1
a_{19}	56.3	4.08	73.7	5.04	0.00	0.00	0.89	34.3
a_{20}	61.1	9.65	80.7	7.00	0.00	0.57	1.95	32.0
a_{21}	90.4	7.18	89.4	7.65	0.13	0.38	1.13	23.5
a_{22}	66.7	0.52	78.7	8.14	0.00	0.81	0.26	24.3
a_{23}	63.3	1.10	76.6	9.12	0.05	0.00	0.27	41.6
a_{24}	79.4	9.39	81.4	7.37	0.07	0.23	1.48	26.6
a_{25}	53.3	7.46	83.3	8.08	0.00	0.48	0.32	24.6
a_{26}	71.0	16.98	81.9	7.93	0.00	0.77	0.24	29.5
a_{27}	65.0	8.89	81.1	10.53	0.00	0.32	1.47	42.1

a_{28}	76.8	17.33	85.7	8.11	0.00	0.00	1.11	27.2
a_{29}	64.3	7.57	85.6	4.70	0.97	0.33	1.80	25.6
a_{30}	73.3	12.05	85.3	7.67	0.00	0.00	3.79	17.9

6.7 Summary

In this chapter, major elements of the case study were introduced, such as the DM, the database source (ACSS benchmarking website) and the sample chosen. 30 hospitals, HCs and PPPs (27 publicly-managed hospitals/HCs and three PPPs) were selected, from which data was analyzed through the year of 2018.

A criteria tree was built, according to the literature review, comprising five points of view - Access, Care Appropriateness, Safety, Caesarean Appropriateness, and Efficiency -, each of them with one or more criteria under it. This resulted in a family of ten criteria, measured by indicators present in the chosen database.

Then, the variables of the model were defined, necessary to carry out the method presented in the previous chapter.

Firstly, five categories were defined *a priori*, from C_1 to C_5. Then, characterizing each of them, and considering every criterion, reference actions were delineated. The criteria weights were assigned resorting to the SRF procedure, done along with the DM through the DecSpace software. In this chapter, the steps followed to perform this action and its final results are exposed.

Besides this, to mitigate imperfect knowledge, indifference, preference and veto thresholds were established by the DM for every criterion.

Finally, the last input needed was displayed, which were the performance tables. 12 performance tables were built for each model and another 12 for the second scenario of the first model, but in this chapter the only ones portrayed are the ones representing the performance values per action of the month of January, considering both models.

Having all the variables and inputs needed to execute the method, one can move on to its implementation and the results obtained. Both steps are described on the next chapter.

Chapter 7

Model Execution and Results

This chapter comprises, firstly, the steps followed to execute the model whose elements were defined on the last chapter, through the implementation of an example of a project.

Besides this, the results of the 36 implementations needed to achieve the goal of this dissertation are presented and interpreted. It comprises a range of results that include two models (with and without efficiency criteria) and two scenarios within the first model, differentiated by the criteria weight.

A Robustness Analysis is also processed using these two scenarios and different values of the credibility level. Conclusions are drawn regarding the results obtained and the sensitivity of the model.

7.1 Execution of the ELECTRE TRI-nC method

For each month of 2018, the method had to be applied three times - two times for Model 1 (with the social-oriented and efficiency-oriented goals scenarios), and one time for Model 2 -, which implied the creation of three projects per month. Since this leads to 36 individual projects, only an example of this implementation will be given in this chapter, corresponding to the Model 1, with the social-oriented goal scenario weights, for January. Note that what changes between them are the values of the weights and sizes of the action set, performance table and reference actions table.

The inputs needed to execute a project are the alternative set (actions), the criteria set, the performance table and the decision configurations, where weights, thresholds, categories, reference actions and the index of credibility are comprised. All of these elements were defined and exposed in the previous chapter.

Firstly, the list of actions is inserted, and as the Model 1 is the one being exemplified, this corresponds to a set of 28 health institutions, illustrated in Figure 7.1. For Model 2, two extra actions are included (Cascais and Loures Hospitals).

Then, criteria are added to the criterion set, shown in Figure 7.2. In this case, all ten criteria are included, and for each of them, the type of measure had to be defined (ordinal or cardinal). Note that, if Model 2 was being considered, criteria g_9 and g_{10} wouldn't be included and, besides this, in this dissertation, all criteria are cardinal measures.

The third step is the definition of the performance table, which was already prepared in an excel file and was imported in a csv format. The performance table used for this implementation is displayed in the previous chapter on Table 6.11 and corresponds to the performance of each action of the alternative set, on every criterion, during the first month of 2018. This step is presented in Figure 7.3.

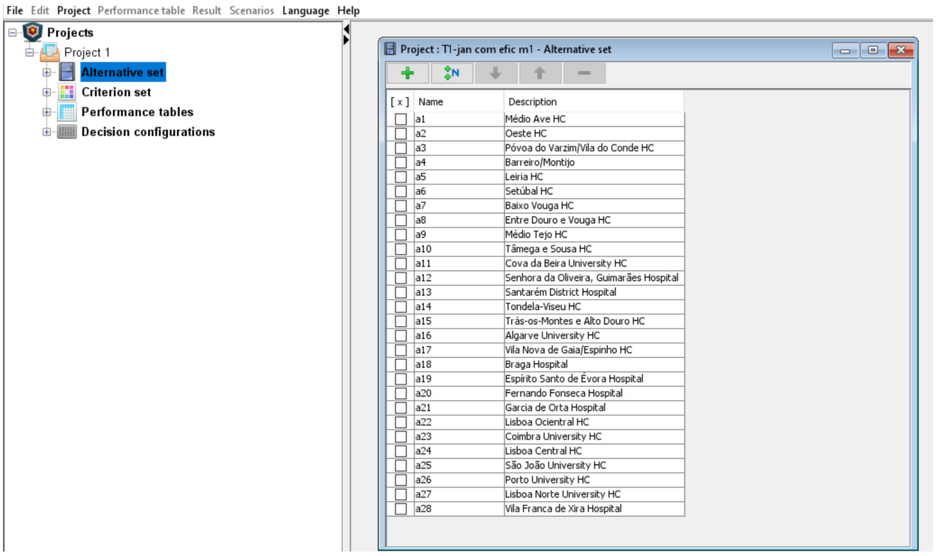

Figure 7.1: Introduction of the alternative set (actions).

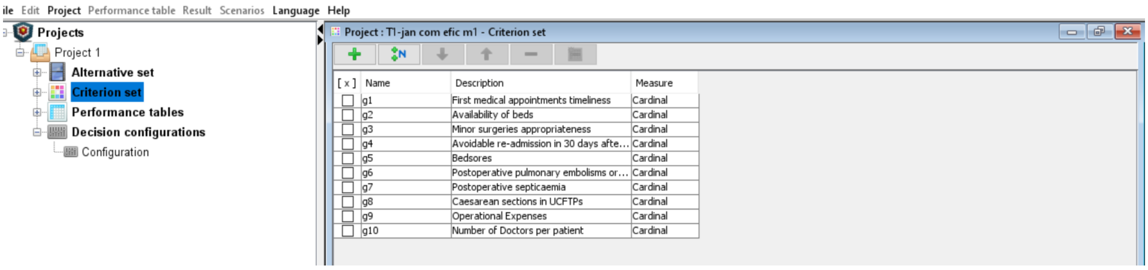

Figure 7.2: Introduction of the criterion set.

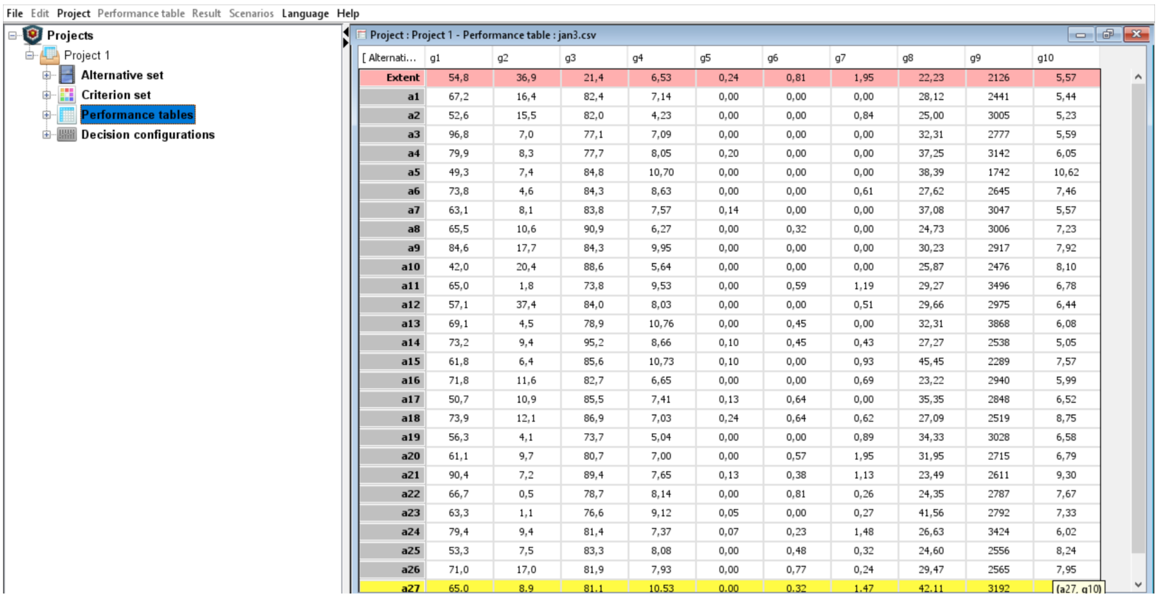

Figure 7.3: Introduction of the performance table of January, for Model 1.

Next, one must choose the method to be executed by adding a "decision configuration" tab and selecting the option intended. In this analysis it was the ELECTRE Tri-nC method, as shown in Figure 7.4.

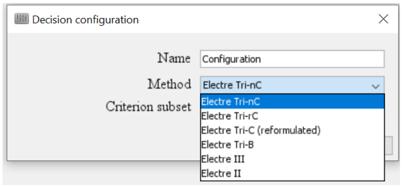

Figure 7.4: Choosing the method.

Following the choice of the method, the introduction of parameters within the "decision configuration" tab is the next step and is presented in Figures 7.5 and 7.6. This includes the weights, thresholds and credibility level, the categories and the reference actions, all previously described on Chapter 6. The latter is presented as another performance table, which was also imported as a csv file. Note that the best category ("Very Good") must be at the top, which one considered C_5, and the worst ("Very Weak") must be at the bottom, which corresponds to C_1. This order of numbers also applies for the reference actions.

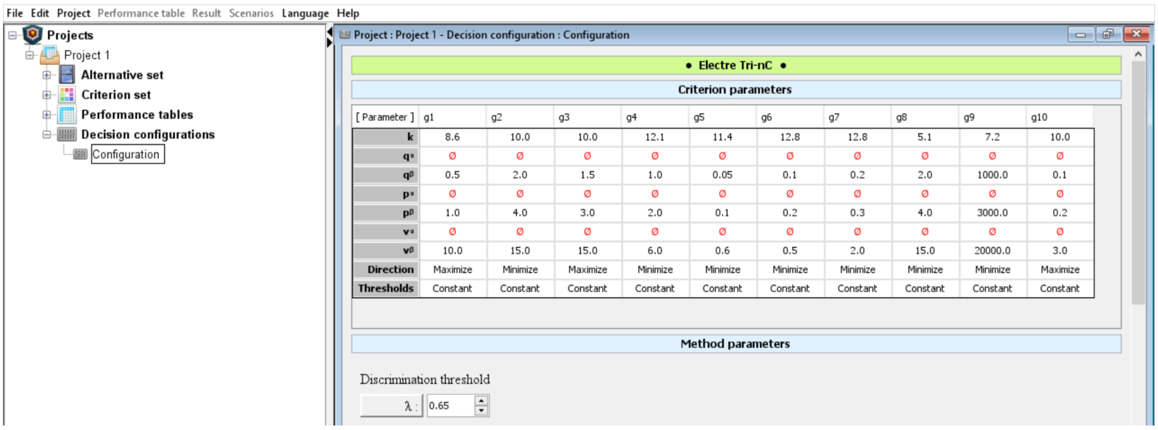

Figure 7.5: Introduction of the weights, thresholds and credibility level.

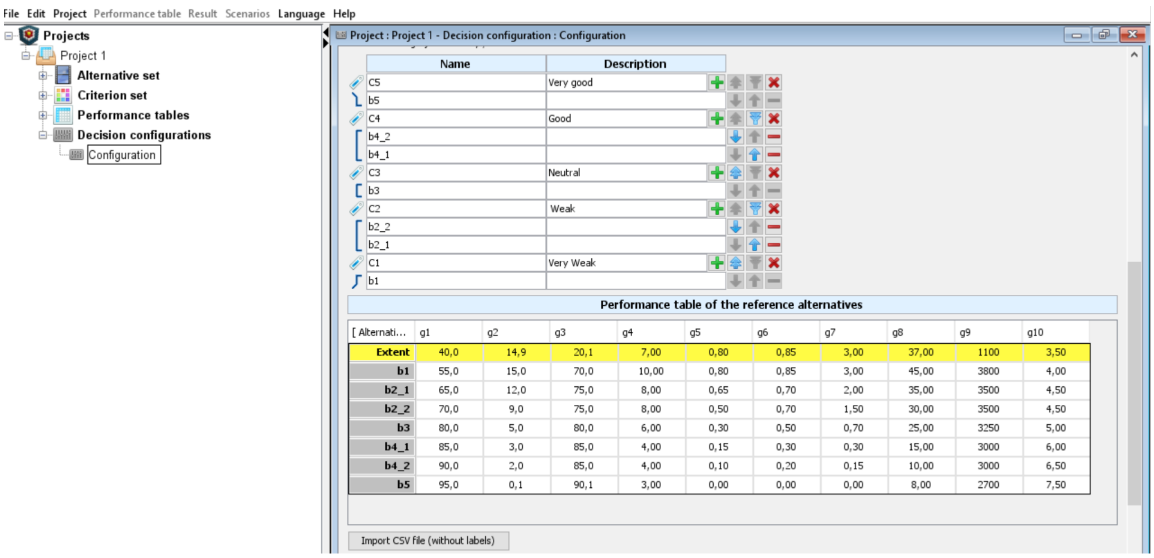

Figure 7.6: Introduction of the Categories and Reference Actions.

Finally, the project can be executed, which is accomplished following the example of Figure 7.7, which leads to the window in Figure 7.8, and allows the display of the results. In the case of spotted errors in any of the parameters, the software exhibits a warning message.

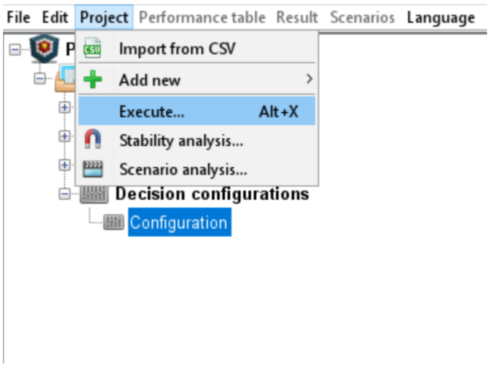

Figure 7.7: Execution of the project.

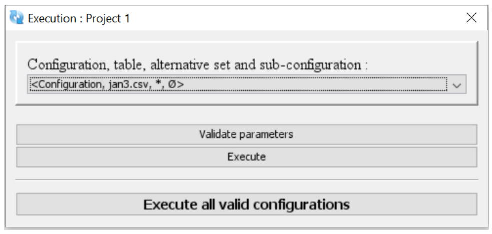

Figure 7.8: Execution window.

7.2 Results

As seen on Chapter 5, the output of the ELECTRE TRI-nC method consists on a range of possible categories that an entity can be assigned to or, if the two categories are the same, the one category the entity is assigned to. For limitations of space and very extensive results, it was chosen to present only an example by selecting the results for the month of January in both models, which can be seen on Table 7.1. The cells marked in grey represent the modifications, from one model to the other, between assignments of an action. Note that the Model 1 results presented (which considers efficiency criteria) only take into account the social-oriented goal scenario, since the weights are coherent with Model 2, making a comparison between the two possible. Throughout this chapter, there will only be portrayed tables concerning the results for the month of January for the reasons mentioned above. However, the remaining tables can be consulted in Appendix A.

Table 7.1: Results of January 2018, for Model 1 and Model 2.

	Model 1 (social)		**Model 2**	
	Min	**Max**	**Min**	**Max**
a_1	C_3	C_3	C_3	C_3
a_2	C_2	C_3	C_2	C_3
a_3	C_3	C_4	C_4	C_4
a_4	C_3	C_3	C_3	C_3
a_5	C_2	C_5	C_2	C_3
a_6	C_3	C_4	C_3	C_4
a_7	C_3	C_3	C_3	C_3
a_8	C_3	C_3	C_3	C_3
a_9	C_3	C_3	C_3	C_3
a_{10}	C_1	C_4	C_1	C_4
a_{11}	C_2	C_2	C_2	C_2
a_{12}	C_1	C_3	C_1	C_3
a_{13}	C_2	C_2	C_2	C_2
a_{14}	C_3	C_4	C_3	C_3
a_{15}	C_3	C_3	C_3	C_3
a_{16}	C_3	C_4	C_3	C_4
a_{17}	C_2	C_3	C_2	C_3

a_{18}	C_3	C_4	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3
a_{20}	C_2	C_2	C_2	C_2
a_{21}	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_2	C_3
a_{23}	C_3	C_3	C_3	C_3
a_{24}	C_3	C_3	C_3	C_3
a_{25}	C_2	C_3	C_2	C_3
a_{26}	C_3	C_3	C_2	C_2
a_{27}	C_2	C_3	C_2	C_2
a_{28}	C_3	C_4	C_3	C_4
a_{29}			C_2	C_2
a_{30}			C_3	C_3

Looking at every month's results, it is possible to draw conclusions concerning the hospitals with best and worst performances. The considered ”best” hospitals were the ones that, in the majority of months, were assigned to the categories ”Very Good” or ”Good”. The ”worst” hospitals were the ones assigned to the categories ”Very Bad”, ”Bad” and within ”Very Bad” and ”Neutral”, in the majority of the months. Note that every entity is analyzed individually each month, thus an analysis per group is not included in this discussion.

Hence, for a credibility level of λ=0.65, the ”best” hospitals, for Model 1, were a_4, a_{14} and a_{21}, which correspond to Barreiro/Montijo HC, Tondela-Viseu HC and Garcia de Orta Hospital. When considering Model 2, Póvoa do Varzim/Vila do Conde HC, a_3, also figures in this set. This happens because this hospital showed performance values within the categories ”Bad” and ”Neutral” for the efficiency criteria, g_9 and g_{10}, thus, when removing these parameters, these values are not accounted for and the categories to which this action is assigned to change for the better.

For the same data, the considered ”worst” hospitals for Model 1 were a_2, a_{25} and a_{26}, corresponding to Oeste HC, São João University HC and Porto University HC, respectively. However, when looking at the results for Model 2, a_{25} and a_{26} are no longer joined by a_2, but a_{27} becomes part in this set (Lisboa Norte University HC), as well as a PPP, Cascais Hospital, a_{29}. These alterations can be justified by the fact that a_2 shows worse results in the efficiency criteria, so when they are removed, so is this action from the worst performing hospitals. The opposite happens with a_{27}, which showed Neutral/Good performance in the efficiency criteria, so when these are not accounted for, this hospital is assigned to worse categories.

It is also worth noticing that, even hospitals labeled as the ”best” ones, present months where their performance was worse, and vice-versa. For instance, a_4 in August was assigned to the interval of categories between C_1 and C_3. Besides this, both ”best” and ”worst” hospitals show ”Neutral” performance in some of the months of the year.

Focusing on PPPs - Vila Franca de Xira Hospital, Cascais Hospital and Loures Hospital -, the only hospital standing out - negatively - in this study is Cascais Hospital. Comparing to publicly-managed hospitals, none of the other PPPs is considered one of the ”best” or one of the ”worst” hospitals, falling in the middle classification. However, in a monthly analysis, more changes can be seen.

When considering Model 1, only Vila Franca de Xira Hospital can be analyzed, and it performed well (Neutral/Good performance) in a couple of months. Likewise, it fell into the worst performing hospitals in one month, September.

Without efficiency criteria, for Model 2, the other two PPPs can also be scrutinized. Cascais Hospital was in the worst performing hospitals in a few months, only reaching a Neutral/Good performance in September. Finally, Loures Hospital was considered one of the best ones (Good performance) in only one month and one of the worst ones (Bad performance) in two.

In conclusion, it is hard to make a clear comparison to a_{29} and a_{30} since they do not provide efficiency data, which as we have seen has the power to put or take an hospital from a classification, being significant in the whole picture. Besides this, the way hospitals perform in certain months individually is not an indicator of its overall performance. There is, then, no clear evidence that one group outperforms the other.

However, the profile of the entities considered the best performing hospitals can be used for benchmarking purposes, allowing other hospitals to seek improvement (Augusto et al., 2008).

7.2.1 Robustness Analysis

In MCDA, one of the main concerns is the robustness of its methods (Rangel-Valdez et al., 2018).

According to Rosenhead and Mingers (2001), Robustness Analysis is "*a method for evaluating initial decision commitments under conditions of uncertainty, where subsequent decisions will be implemented over time.*" The robustness of an initial commitment measures the flexibility that decision will leave for useful future decision choice, thus an analysis allows measuring the capacity of a model for resisting vague approximations and/or zones of ignorance derived from its initial representation (Rosenhead and Mingers, 2001; Rangel-Valdez et al., 2018).

In short, this kind of analysis is important to verify the stability or sensitivity of the results, by changing preference parameters and seeing how the results behave. For this to be accomplished, one can vary the credibility level and the weights assigned to the criteria. As previously seen, this dissertation takes into account two scenarios under Model 1, distinguished solely for their weights. This already allows the robustness to be tested but, in addition, one also varied the credibility level in both models to test the evolution of the assignments, considering λ=0.55, 0.60, 0.65.

7.2.1.1 Changing the credibility level

This analysis was made for three different credibility levels, λ=0.55, 0.60, and 0.65. It is likely that, when increasing the credibility level, both ascending and descending rules converge in terms of the assigned category, possibly originating an unique category.

An example of the changes in assignments for different credibility levels is shown on Table 7.2, which displays the results for January, considering Model 1. The cells that represent modifications are marked in grey. The analysis for the remaining months can be consulted in Appendix B.

Table 7.2: Results of Model 1 for January 2018, considering different λ values.

	Model 1 (social)						**Model 1 (efficiency)**					
	λ=0.55		**λ=0.60**		**λ=0.65**		**λ=0.55**		**λ=0.60**		**λ=0.65**	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_2	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_3	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_3
a_5	C_2	C_5	C_2	C_5	C_2	C_5	C_2	C_5	C_2	C_5	C_2	C_5
a_6	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4

a_7	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_8	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_9	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_3	C_3	C_3
a_{10}	C_1	C_3	C_1	C_4	C_1	C_4	C_1	C_3	C_1	C_4	C_1	C_4
a_{11}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{12}	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3
a_{13}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{14}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_{15}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{16}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_{17}	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3
a_{18}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{28}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4

The total percentages of assignments to each category were calculated and shown on Tables 7.3 and 7.4.

Table 7.3: Comparison of the total percentages of assigned categories for different λ values, considering the social-oriented goal scenario of Model 1.

Min	**Max**	**λ=0.55**	**λ=0.60**	**λ=0.65**
C_1	C_2	0.89%	0.89%	0.60%
C_1	C_3	2.68%	2.38%	2.68%
C_1	C_4	0%	0.30%	0,30%
C_1	C_5	0.30%	0.30%	0.30%
C_2	C_2	14.58%	13.10%	10.71%
C_2	C_3	14.29%	14.88%	16.96%
C_2	C_4	0.30%	0.60%	0.60%
C_2	C_5	0.60%	0.60%	0.60%
C_3	C_3	43.45%	43.45%	44.35%
C_3	C_4	11.01%	11.01%	10.31%
C_3	C_5	0,30%	0,30%	0.30%
C_4	C_4	11.31%	11.90%	11.01%
C_4	C_5	0.30%	0%	0%
C_5	C_5	0%	0.30%	0.30%

Table 7.4: Comparison of the total percentages of assigned categories for different λ values, considering the efficiency-oriented goal scenario of Model 1.

Min	**Max**	**λ=0.55**	**λ=0.60**	**λ=0.65**
C_1	C_2	1.79%	1.79%	1.19%
C_1	C_3	3.27%	2.98%	3.57%
C_1	C_4	0%	0.30%	0.30%
C_1	C_5	0.60%	0.60%	0.60%
C_2	C_2	16.67%	14.58%	12.80%
C_2	C_3	15.77%	15.77%	16.67%
C_2	C_4	2.08%	2.68%	2.68%
C_2	C_5	0.60%	0.60%	0.60%
C_3	C_3	38.39%	39.29%	40.77%
C_3	C_4	9.82%	9.82%	9.82%
C_3	C_5	0.60%	0.60%	0.60%
C_4	C_4	10.42%	11.01%	10.42%

It can be verified that, for smaller credibility levels, there are stronger assignments, i.e., a bigger percentage of actions in an unique category, such as C_2 and C_4.

However, in cases of actions assigned to C_2, they go to the interval of categories immediately above, being within C_2 and C_3. The same is verified for some actions that were assigned only to C_3, but in higher credibility levels change to an interval of categories, between C_3 and C_4.

In cases of actions in C_4, there are three cases (in 336) in the first scenario and one in the second scenario, in which the actions lower their category, being assigned to an interval of categories, C_3 and C_4.

The percentages of actions changing to stronger assignments (from intervals of categories to an unique one) were 35.3% and 39% in the first and second scenarios, respectively.

Hence, although higher credibility levels are linked to weaker assignments in this study, they are also more accurate. The ”best” and ”worst” hospitals, in general, are maintained.

Overall, for the first case, comparing λ=0.55 and λ=0.60, only 4.76% of the assignments have changed, while for λ=0.65, 5.36% of the assignments have changed. For the second case, 7.14% of the assignments have changed when changing from λ=0.55 to λ=0.60, and 5,06% to λ=0.65.

The same process was applied to Model 2, which is represented on Table 7.5.

Table 7.5: Comparison of the total percentages of assigned categories for different λ values, considering Model 2.

Min	**Max**	**λ=0.55**	**λ=0.60**	**λ=0.65**
C_1	C_2	0.83%	0.83%	0.83%
C_1	C_3	2.50%	1.94%	1.67%
C_1	C_4	0%	0.55%	0.83%
C_2	C_2	20.56%	16.67%	14.44%
C_2	C_3	12.22%	15.45%	15.56%
C_2	C_4	0%	1.11%	1.94%
C_3	C_3	46.39%	44.72%	44.17%
C_3	C_4	6.94%	10.00%	10.56%
C_4	C_4	10.28%	9.72%	9.72%
C_4	C_5	0.30%	0%	0%
C_5	C_5	0%	0.30%	0.30%

Overall, comparing λ=0.55 and λ=0.60, 12.5% of the assignments have changed, while for λ=0.65, only 8% of the assignments have changed.

In general, the changes between credibility levels are the same as previously demonstrated on Table 7.3. However, it can be noticed that, in this model, less categories are covered. For instance, although with a low percentage, in Model 1, actions were assigned to very broad intervals of categories, such as between C_1 and C_5 and C_3 and C_5, which does not happen in this model, suggesting the existence of more oscillations in the efficiency criteria.

With both examples it is possible to conclude that the model is robust, since the percentages of changes are not high enough for it to have a significant impact in the rough results.

7.2.1.2 Changing the weights of the criteria

For this part of the robustness analysis, it is enough to carry a comparison between the two considered scenarios of Model 1, differentiated by their goals (social goal or efficiency goal), to which weights were assigned resorting to the SRF procedure, as seen on the last chapter. Once again, an example of the results is given only for one month, January, and is displayed on Table 7.6. For this analysis, λ=0.65.

Table 7.6: Comparison of the assigned categories for different scenarios, using January 2018 and λ=0.65.

	Social-oriented goal scenario		**Efficiency-oriented goal scenario**	
	Min	**Max**	**Min**	**Max**
a_1	C_3	C_3	C_3	C_3
a_2	C_2	C_3	C_2	C_3
a_3	C_3	C_4	C_3	C_4
a_4	C_3	C_3	C_3	C_3
a_5	C_2	C_5	C_2	C_5
a_6	C_3	C_4	C_3	C_4
a_7	C_3	C_3	C_3	C_3
a_8	C_3	C_3	C_3	C_3
a_9	C_3	C_3	C_3	C_3
a_{10}	C_1	C_4	C_1	C_4
a_{11}	C_2	C_2	C_2	C_2
a_{12}	C_1	C_3	C_1	C_3
a_{13}	C_2	C_2	C_2	C_2
a_{14}	C_3	C_4	C_3	C_4
a_{15}	C_3	C_3	C_3	C_3
a_{16}	C_3	C_4	C_3	C_4
a_{17}	C_2	C_3	C_2	C_3
a_{18}	C_3	C_4	C_4	C_4
a_{19}	C_3	C_3	C_3	C_3
a_{20}	C_2	C_2	C_2	C_2
a_{21}	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_3	C_3
a_{23}	C_3	C_3	C_3	C_3
a_{24}	C_3	C_3	C_3	C_3
a_{25}	C_2	C_3	C_2	C_3
a_{26}	C_3	C_3	C_3	C_3
a_{27}	C_2	C_3	C_2	C_3
a_{28}	C_3	C_4	C_3	C_4

Analyzing every month, of a total of 336 assignments, 50 actions change categories (14.88%), in which 13 - 26% - changed to stronger assignments (from an interval of categories to an unique one), up or down, 14 started to include intervals with higher categories (28%), and 21 started to include intervals with lower categories (42%) and two included a lower and higher category (4%). The changes for lower or higher categories are expected since "Operational Expenses", g_9, criterion's weight increased, which means the actions that had bad values for this criterion are likely to change to a worse category, and vice-versa.

It is also noteworthy that almost half of the identified changes (48%) happened in only one month, April. This was a month where a lot of fluctuations on the categories were spotted, with the change in weights making a lot of difference.

Overall, taking into consideration the percentage of changes obtained, changing the importance of this one criterion does not have a major impact in the assignments, reinforcing the robustness of the model.

Additionally, one made small changes on the reference actions, making them less strict, and the only difference spotted was that there were more actions going from C_3 to reaching category C_4.

Comparing the results to the ones obtained by Jorge (2019), whose analysis was made for the same year and using most of the same actions, one of the best (Póvoa de Varzim/Vila do Conde HC) and two of the worst (Oeste HC and Lisboa Norte University HC) hospitals are coherent, but not the others. However, it is important to note that, even though the criteria are structured in the same way, unlike Rocha (2019) who uses a hierarchical way that can't be compared, the criteria considered by Jorge (2019) are different than the ones in this study, as well as the weights assigned to each of them and reference actions considered to each category. Besides this, his analysis considers yearly data, while this one is done monthly, making the comparison rather insignificant.

7.3 Summary

In this chapter, the model was implemented and an example of the steps required to execute a project was given, using the month of January. These steps were repeated three times for each month: the first using a model with efficiency criteria and a first set of weights, the second using the same model with a different set of weights and finally using a model without efficiency criteria, including all three PPP hospitals.

The results of all these implementations were presented for the same month, but an analysis was carried out taking into account every month. The considered best entities were the ones who were assigned to categories "Very Good" and "Good" in the majority of the months, and the worst entities were the ones who were assigned to the categories "Very Bad", "Bad" or within "Very Bad" and "Neutral" in the majority of the months.

For Model 1, the best ones were Barreiro/Montijo HC, Tondela-Viseu HC and Garcia de Orta Hospital, and the worst ones Oeste HC, São João University HC and Porto University HC.

For Model 2, the best ones were Barreiro/Montijo HC, Tondela-Viseu HC, Garcia de Orta Hospital and Póvoa do Varzim/Vila do Conde HC, and the worst ones São João University HC, Porto University HC and Lisboa Norte University HC.

The differences are due to the fact that when efficiency criteria are excluded, the hospitals who perform better in these two criteria compared to others worsen their assigned categories and vice-versa.

A Robustness Analysis was done by changing two parameters, the credibility level, λ, and the criteria weights, which had already been changed for the general analysis.

It was noticed that those variations did not provoke major alterations in the rough results.

In fact, when changing the credibility level, it was verified that, for the first scenario of Model 1, comparing λ=0.55 and λ=0.60, only 4.76% of the assignments changed, while for λ=0.65, 5.36% of the assignments changed. For the second scenario, 7.14% of the assignments changed when substituting λ=0.55 for λ=0.60, and 5.06% for λ=0.65. For Model 2, more modifications were found, with 12.5% of different assignments between λ=0.55 and λ=0.60, and 8% for λ=0.65.

When analyzing both scenarios, where one had an efficiency-related goal, with the "Operational Expenses" criterion as the one with the highest weight, 14.88% of the initial assignments shifted, and almost half of these changes happened in only one month, April.

Overall, taking into consideration the percentage of changes obtained in both adjustments, changing the values of λ was not significant, suggesting that the model is robust. Likewise, the importance of this criterion, "Operational Expenses", did not have a major impact in the assignments, reinforcing the robust characterization of the model.

Chapter 8

Conclusion, Limitations and Future Work

In this last chapter, the main conclusions drawn from the results obtained are exposed, as well as the limitations faced while conducting it.

8.1 Conclusions

Recently, Portuguese healthcare services are not meeting the needs of the population when it comes to the services provided. They are considered rivalrous due to the limited availability of staff, beds and other hospital resources, being linked to geographical disparities in the availability of services, long waiting times and long waiting lists. Besides this, factors such as the increase of average life expectancy, the aging population, and the higher incidence of chronic diseases contribute to the lack of responsiveness and overall challenges faced by the SNS.

In response to some of these issues and also to try to reduce the high expenditures and waste of public funds in the healthcare sector, while improving the efficiency and the effectiveness of health-care providers, health reforms were implemented in Portugal, such as the attempt of employing private management tools in the public sector, with the first wave of PPPs emerging.

However, it is common for measures based on cost containment and efficiency improvement to jeopardize the quality of the provided services.

The main goal of this dissertation was to evaluate and compare the quality of publicly-managed Portuguese hospitals to the PPPs currently operating (Vila Franca de Xira Hospital, Cascais Hospital and Loures Hospital), while trying to find out if one of the groups outperformed the other and whether PPPs are a better alternative or not. This was done following on previous work and using decision support techniques through the application of the ELECTRE TRI-nC multi-criteria model. Using this method, one could incorporate various indicators regarding health, attribute different weights to criteria, use several reference actions, among other characteristics that proved to be very useful when assessing quality in the health sector.

To draw significant conclusions, the methodology used in this study is characterized as being robust, which was confirmed through the performance of tests.

As of the main focus of this work, the main conclusion was that both groups, publicly-managed hospitals and PPPs, present similar performances regarding the criteria chosen to represent the access, care appropriateness, safety, caesarean appropriateness and efficiency points of view. In terms of ”best” and ”worst” performers, both classifications belonged mainly to EPEs, although Cascais Hospital also figured in the ”worst” set in one of the models.

Despite the fact that there is not a significant difference between both groups’ performance, this does not mean they fail and/or succeed in the same areas or levels, thus both groups need to improve their delivered services in general.

Other authors have reached similar conclusions, such as Ferreira and Marques (2019) and Entidade Reguladora da Saúde (2016), using different methodologies.

Both ERS and Ferreira and Marques (2019) relied on DEA, as well as Nunes and Matos (2017), to compare EPE hospitals to those under the PPP regime.

In the first case, Entidade Reguladora da Saúde (2016) considered four components in their study, which were technical efficiency, effectiveness, clinical quality and regulatory costs. For instance, regarding quality, which takes into account safety, infrastructures and equipment, and patient satisfaction, PPP hospitals showed, on average, more satisfying results than publicly-managed hospitals. But, in general, the results were mixed for different indicators and no obvious conclusion could be taken concerning the comparison between both groups’ performance.

Ferreira and Marques (2019) used data from 2012 to 2017 and indicators of access and quality from the ACSS benchmarking website as well, through DEA and a Benefit of the Doubt approach. The results from this study showed that publicly-managed hospitals and PPPs exhibit similar social performances, none of them being better social performers (delivering timelier, safer or more appropriate healthcare services) than the other.

Nunes and Matos (2017), in their turn, used the DEA methodology for data of the years 2013, 2014 and 2015, in which four PPP hospitals were analyzed. Three of them showed good efficiency levels during the period under analysis, and one of them was considered inefficient. Overall, most of the PPP hospitals were considered efficient when compared both within their group and with the other EPE hospitals, but not all of them and not always, as expected.

In their study, Ferreira and Marques (2019) expose how their results can implicate policymakers, regulators, citizens, hospital managers, clinical staff, academics, project managers, operational researchers, and others, and it can also be applied in this case, since both studies are, overall, similar.

For instance, as one mentioned at the beginning of this dissertation, one of the biggest arguments in the public opinion against PPPs is that these entities are not capable of delivering the same level of quality and access as publicly-managed hospitals, since being managed by private partners would likely lead them to an ultimate goal: maximizing profit.

The findings of this study do not support this hypothesis, since PPPs seem to be capable of providing health services as good as publicly-managed hospitals, no matter the profit.

In general, there is no evidence that one group outperforms the other in terms of access, safety or appropriate care. However, it is important to notice that the major categories Portuguese hospitals were assigned to were between ”Bad” and ”Neutral” levels, which suggests that Portuguese hospitals, overall, show substantial performance problems and plenty of room for improvement.

The results from this research can help hospitals and their clinical staff to improve their performance through benchmarking, comparing their practices to others better classified and employing new techniques.

When it comes to the political and management points of view, it can also be useful. Since, based on this study’s findings, PPPs seem to exhibit a quality of performance at least equal to publicly-managed hospitals, policymakers, whose responsibilities include deciding if new contracts should be created, if an ending contract should be renewed or if the management of a PPP hospital should be assigned fully to the public party, can use it as support to make this kind of decisions.

Particularly, the results obtained through this dissertation would not suggest changing the private/public management of these entities, which would only be a costly and unworthy process, considering the absence of differences in their performances.

Moreover, the aforementioned applications go in line with the goals of the hSNS Project, which aimed for these results to be useful to improve the quality of the delivered Portuguese healthcare services, support management by monitoring performance indicators and improve hospitals' financing according to their performance.

The main objectives of this dissertation were achieved, with a successful application of a MCDA approach in assessing the quality of the Portuguese hospitals. The fact that the results meet what had already been found in the literature regarding other methods, validates the implementation of the model chosen, ELECTRE TRI-nC, to reach this dissertation's goals.

8.2 Limitations and future work

It is important to outline the limitations of this study.

To achieve the goal of analysing and comparing EPE and PPP hospitals, it was necessary to circumvent some obstacles, such as missing data. In some cases, approximations resorting to correlations and linear regressions had to be done to cover the lacking information. Because the PPPs do not provide information regarding efficiency and productivity dimensions, two distinct models had to be created and implemented, where one did not comprise all PPP hospitals but included these criteria, and other where three PPP hospitals were under scrutiny but two criteria were not considered. These issues suggest that efforts should be made for these entities to become more attentively regulated, with reports including the information that is currently not provided becoming available, which would facilitate these types of comparisons and studies that intend to analyze both groups of hospitals accurately in a benchmarking perspective.

Besides this, only a few criteria were considered, even regarding all the indicators present in the ACSS benchmarking website. A more complete research could be carried out if more criteria and indicators were to be analyzed.

Even though the criteria employed in this study were considered important under the chosen points of view, these results should be compared with future results that take into account other quality and access related criteria. In fact, according to the literature review conducted prior to the definition of the model, other types of variables, that were not comprehended in this dissertation, were pointed out as relevant. It would be interesting to include information concerning facilities and infrastructures, patient satisfaction, and other outcomes.

If the sample evaluated comprised more years than an unique one (2018), the results could be more complete and robust.

The thresholds (indifference, preference and veto) have been considered as constants throughout this analysis, but they can vary, so it could also make sense to modify this in future studies.

Finally, this process was done from a subjective point of view, since the parameters and variables were decided between the Analyst and the single DM. This includes, for instance, the selection of criteria and the definition of reference actions, thresholds and weights. If more than one DM were involved, it would be possible to see the differences on opinions and considerations and the way this would reflect on the results.

The framework detailed through this study can have multiple applications in different areas and scenarios, especially based on ranking projects using sets of indicators. This can be used not only on a social point of view but also financial, environmental, etc.

It is really important to consider quality and access to make sure patients are being properly cared for, which involves timely, equitable, safe and appropriate healthcare. Thus, when assessing and investing in the SNS and health sector in general, studies of this nature could be very useful.

Bibliography

Akdag, H., Kalayci, T., Karagöz, S., Zülfikar, H., and Giz, D. (2014). The evaluation of hospital service quality by fuzzy MCDM. *Applied Soft Computing Journal*, 23:239–248.

Almeida-Dias, J., Figueira, J. R., and Roy, B. (2012). A multiple criteria sorting method where each category is characterized by several reference actions: The Electre Tri-nC method. *European Journal of Operational Research*, 217(3):567–579.

Augusto, M., Lisboa, J., Yasin, M., and Figueira, J. R. (2008). Benchmarking in a multiple criteria performance context: An application and a conceptual framework. *European Journal of Operational Research*, 184(1):244–254.

Badri, M. A., Attia, S., and Ustadi, A. M. (2009). Healthcare quality and moderators of patient satisfaction: Testing for causality. *International Journal of Health Care Quality Assurance*, 22(4):382–410.

Baltussen, R., Marsh, K., Thokala, P., Diaby, V., Castro, H., Cleemput, I., Garau, M., Iskrov, G., Olyaee-manesh, A., Mirelman, A., Mobinizadeh, M., Morton, A., Tringali, M., van Til, J., Valentim, J., Wagner, M., Youngkong, S., Zah, V., Toll, A., Jansen, M., Bijlmakers, L., Oortwijn, W., and Broekhuizen, H. (2019). Multicriteria decision analysis to support health technology assessment agencies: Benefits, limitations, and the way forward. *Value in Health*, 22(11):1283–1288.

Barata, A. N., Tokuda, Y., and Martins, H. M. (2012). Comparing healthcare systems: Considering Japan and Portugal. *Revista Portuguesa de Saude Publica*, 30(2):180–192.

Barros, P. P. and Martinez-Giralt, X. (2009). Contractual design and PPPs for hospitals: Lessons for the Portuguese model. *European Journal of Health Economics*, 10(4):437–453.

Belton, V. and Stewart, T. (2002). *Multiple Criteria Decision Analysis: An Integrated Approach*. Springer US, first edition.

Bem, A., Ucieklak-Jeż, P., and Prędkiewicz, P. (2014). Measurement of health care system efficiency. *Management Theory and Studies for Rural Business and Infrastructure Development*, 36(1):25–33.

Benbassat, J. and Taragin, M. (2000). Hospital readmissions as a measure of quality of health care: Advantages and limitations. *Archives of Internal Medicine*, 160(8):1074–1081.

Betrán, A. P., Torloni, M. R., Zhang, J.-J., Gülmezoglu, A., WHO Working Group on Caesarean Section, Aleem, H., Althabe, F., Bergholt, T., de Bernis, L., Carroli, G., et al. (2016). WHO statement on caesarean section rates. *BJOG: An International Journal of Obstetrics & Gynaecology*, 123(5):667–670.

Blatnik, P., Bojnec, Š., and Tušak, M. (2017). Measuring efficiency of secondary healthcare providers in Slovenia. *Open Medicine (Poland)*, 12(1):214–225.

Bouyssou, D. (1990). Building Criteria: A Prerequisite for MCDA. In *Readings in Multiple Criteria Decision Aid*. Springer, Berlin, Heidelberg.

Buse, K. and Harmer, A. (2004). Power to the partners?: The politics of public-private health partnerships. *Development*, 47(2):49–56.

Campbell, S. M., Roland, M. O., and Buetow, S. A. (2000). Defining quality of care. *Social Science and Medicine*, 51(11):1611–1625.

Chen, T. Y., Chang, C. H., and Rachel Lu, J. F. (2013). The extended QUALIFLEX method for multiple criteria decision analysis based on interval type-2 fuzzy sets and applications to medical decision making. *European Journal of Operational Research*, 226(3):615–625.

Comissão para a Sustentabilidade do Financiamento do Serviço Nacionalde Saúde (2007). Relatório Final da Comissão para a Sustentabilidade do Financiamento do SNS. *Technical report 2006/2007*.

Costa, A. and Figueira, J. (2016). O método ELECTRE TRI-nC (in Portuguese). In *A Investigação Operacional em Portugal, Novos Desafios, Novas Ideias: Homenagem ao Professor Luís Valadares Tavares*, pages 203–218. IST PRESS, first edition.

Crisp, L. N., Berwick, D., Kickbusch, I., Bos, W., Antunes, J., Barros, P., and Soares, J. (2014). Um futuro para a saúde: todos temos um papel a desempenhar. *Lisboa: Fundação Calouste Gulbenkian*.

Cruz, C. O. and Marques, R. C. (2013a). Flexible contracts to cope with uncertainty in public-private partnerships. *International Journal of Project Management*, 31(3):473–483.

Cruz, C. O. and Marques, R. C. (2013b). Integrating infrastructure and clinical management in PPPs for health care. *Journal of Management in Engineering*, 29(4):471–481.

Doetsch, J., Pilot, E., Santana, P., and Krafft, T. (2017). Potential barriers in healthcare access of the elderly population influenced by the economic crisis and the troika agreement: A qualitative case study in Lisbon, Portugal. *International Journal for Equity in Health*, 16(1):1–17.

Dolan, J. G. (2010). Multi-criteria clinical decision support: A primer on the use of multiple-criteria decision-making methods to promote evidence-based, patient-centered healthcare. *Patient*, 3(4):229–248.

Donabedian, A. (2005). Evaluating the quality of medical care. *New England Journal of Medicine*, 288(25):1352–1353.

Eapen, Z. J., Fonarow, G. C., Dai, D., O'Brien, S. M., Schwamm, L. H., Cannon, C. P., Heidenreich, P. A., Bhatt, D. L., Peterson, E. D., and Hernandez, A. F. (2011). Comparison of composite measure methodologies for rewarding quality of care: An analysis from the american heart association's get with the guidelines program. *Circulation: Cardiovascular Quality and Outcomes*, 4(6):610–618.

El Gibari, S., Gómez, T., and Ruiz, F. (2019). Building composite indicators using multicriteria methods: a review. *Journal of Business Economics*, 89(1):1–24.

Emanuel, L., Berwick, D., Conway, J., Combes, J., Hatlie, M., Leape, L., Reason, J., Schyve, P., Vincent, C., and Walton, M. (2009). What exactly is patient safety? *Journal of Medical Regulation*, 95(1):13–24.

Entidade Reguladora da Saúde (2015). Estudo sobre o Desempenho das Unidades Locais de Saúde. Technical report, ERS.

Entidade Reguladora da Saúde (2016). Estudo De Avaliação Das Parcerias Público-Privadas Na Saúde. Technical report, ERS.

European Commission (2014). Expert Panel on Effective Ways of Investing in Health (EXPH): Health and economic analysis for an evaluation of the public-private partnerships in health care delivery across Europe. Technical report, European Comission.

Ferreira, D. and Marques, R. C. (2015). Did the corporatization of Portuguese hospitals significantly change their productivity? *European Journal of Health Economics*, 16(3):289–303.

Ferreira, D. and Marques, R. C. (2018). Identifying congestion levels, sources and determinants on intensive care units: the Portuguese case. *Health Care Management Science*, 21(3):348–375.

Ferreira, D. C. and Marques, R. C. (2019). Do quality and access to hospital services impact on their technical efficiency? *Omega (United Kingdom)*, 86:218–236.

Ferreira, D. C. and Marques, R. C. (2020). Public-private partnerships in health care services: Do they outperform public hospitals regarding quality and access? Evidence from Portugal. *Socio-Economic Planning Sciences*, page 100798.

Ferreira, D. C., Marques, R. C., and Nunes, A. M. (2018a). Economies of scope in the health sector: The case of Portuguese hospitals. *European Journal of Operational Research*, 266(2):716–735.

Ferreira, D. C., Marques, R. C., and Nunes, A. M. (2019). Pay for performance in health care: a new best practice tariff-based tool using a log-linear piecewise frontier function and a dual–primal approach for unique solutions. *Operational Research*, pages 1–46.

Ferreira, D. C., Marques, R. C., Nunes, A. M., and Figueira, J. R. (2018b). Patients' satisfaction: The medical appointments valence in Portuguese public hospitals. *Omega (United Kingdom)*, 80:58–76.

Ferreira, D. C. and Nunes, A. M. (2019). Technical efficiency of Portuguese public hospitals: A comparative analysis across the five regions of Portugal. *International Journal of Health Planning and Management*, 34(1):e411–e422.

Ferreira, D. C., Nunes, A. M., and Marques, R. C. (2020). Optimizing payments based on efficiency, quality, complexity, and heterogeneity: the case of hospital funding. *International Transactions in Operational Research*, 27(4):1930–1961.

Figueira, J. and Roy, B. (2002). Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. *European Journal of Operational Research*, 139(2):317–326.

Figueira, J. R., Almeida-Dias, J., Matias, S., Roy, B., Carvalho, M. J., and Plancha, C. E. (2011). Electre Tri-C, a multiple criteria decision aiding sorting model applied to assisted reproduction. *International Journal of Medical Informatics*, 80(4):262–273.

Figueira, J. R., Greco, S., Roy, B., and SŁOWIŃSKI, R. (2012). An overview of ELECTRE methods and their recent extensions. *Journal of MultiCriteria Decision Analysis*, 20:61–85.

Galea, G. and McKee, M. (2014). Public-private partnerships with large corporations: Setting the ground rules for better health. *Health Policy*, 115(2-3):138–140.

Greco, S., Figueira, J. R., and Ehrgott, M. (2005). *Multicriteria Decision Analysis: State of the art surveys*. Springer, second edition.

Greco, S., Ishizaka, A., Tasiou, M., and Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141(1):61–94.

Grigoroudis, E., Orfanoudaki, E., and Zopounidis, C. (2012). Strategic performance measurement in a healthcare organisation: A multiple criteria approach based on balanced scorecard. *Omega*, 40(1):104–119.

Gulliford, M., Figueroa-Munoz, J., Morgan, M., Hughes, D., Gibson, B., Beech, R., and Hudson, M. (2002). What does 'access to health care' mean? *Journal of Health Services Research and Policy*, 7(3):186–188.

Guven, P. (2007). Measuring efficiency in health care: Analytic techniques and health policy. *International Journal of Pharmaceutical and Healthcare Marketing*, 1:264–265.

Handler, A., Issel, M., and Turnock, B. (2001). A Conceptual Framework to Measure Performance of the Public Health System. *American Journal of Public Health*, 91(8):1235–1239.

Hibbard, J. and Pawlson, L. G. (2004). Why not give consumers a framework for understanding quality? *Joint Commission Journal on Quality and Safety*, 30(6):347–351.

Hughes, R. (2008). *Patient safety and quality: An evidence-based handbook for nurses*, volume 3. Rockville, MD: Agency for Healthcare Research and Quality.

Institute of Medicine (1993). *Access to Health Care in America*. Washington, DC: The National Academies Press (US).

Jacobs, R., Goddard, M., and Smith, P. C. (2005). How robust are hospital ranks based on composite performance measures? *Medical Care*, 43(12):1177–1184.

Jacobs, R., Smith, P. C., and Goddard, M. K. (2004). *Measuring performance: an examination of composite performance indicators: a report for the Department of Health*, volume 29. Centre of Health Economics, University of York.

Jacobs, R., Smith, P. C., and Street, A. (2006). *Measuring efficiency in health care: analytic techniques and health policy*. Cambridge University Press.

Johnston, J. and Gudergan, S. P. (2007). Governance of public-private partnerships: Lessons learnt from an Australian case? *International Review of Administrative Sciences*, 73(4):569–582.

Jorge, R. (2019). A multiple criteria sorting method in a web-based platform. Master's thesis, Instituto Superior Técnico, Universidade de Lisboa.

Kisliakovskii, I., Balakhontceva, M., Kovalchuk, S., Zvartau, N., and Konradi, A. (2017). Towards a simulation-based framework for decision support in healthcare quality assessment. *Procedia Computer Science*, 119:207–214.

Marques, R. C. and Silva, D. (2008). As parcerias público-privadas em Portugal. Lições e recomendações. *Tékhne-Revista de Estudos Politécnicos*, 6(10):33–50.

Marsh, K., Goetghebeur, M., Thokala, P., and Baltussen, R. (2017). *Multi-criteria decision analysis to support healthcare decisions*, volume 10. Springer.

Marsh, K., Lanitis, T., Neasham, D., Orfanos, P., and Caro, J. (2014). Assessing the value of healthcare interventions using multi-criteria decision analysis: A review of the literature. *PharmacoEconomics*, 32(4):345–365.

McKee, M., Edwards, N., and Atunc, R. (2006). Public-Private partnerships for hospitals. *Bulletin of the World Health Organization*, 84(11):890–896.

Meleddu, M., Pulina, M., and Scuderi, R. (2019). Public and private healthcare services: What drives the choice? *Socio-Economic Planning Sciences*, page 100739.

Mitropoulos, P., Mitropoulos, I., and Sissouras, A. (2013). Managing for efficiency in health care: The case of Greek public hospitals. *European Journal of Health Economics*, 14(6):929–938.

Moore, L., Lavoie, A., Sirois, M. J., Belcaid, A., Bourgeois, G., Lapointe, J., Sampalis, J. S., Le Sage, N., and Émond, M. (2013). A comparison of Methods to obtain a composite performance indicator for evaluating clinical processes in trauma care. *Journal of Trauma and Acute Care Surgery*, 74(5):1344–1350.

Mosadeghrad, A. M. (2014). Factors influencing healthcare service quality. *International Journal of Health Policy and Management*, 3(2):77–89.

Nardo, M., Saisana, M., Saltelli, A., and Tarantola, S. (2005). Tools for composite indicators building. *European Comission, Ispra*, 15(1):19–20.

National Academies of Sciences, Engineering, and Medicine (2018). *Crossing the global quality chasm: Improving health care worldwide*. National Academies Press.

Naylor, C. D. (1998). What is appropriate care? *New England Journal of Medicine*, pages 1915–1921.

Nunes, A. and Matos, A. (2017). Análise do desempenho dos hospitais em parceria público privada no Serviço Nacional de Saúde português. *Jornal Brasileiro de Economia da Saúde*, 9(3):304–309.

Nunes, A. M. (2018). A Recuperação Do Serviço Nacional De Saúde Português: Novas Perspectivas Para A Política De Saúde. *Revista de Gestão em Sistemas de Saúde*, 7(1):16–28.

Nunes, A. M. and Ferreira, D. C. (2019a). Reforms in the Portuguese health care sector: Challenges and proposals. *International Journal of Health Planning and Management*, 34(1):e21–e33.

Nunes, A. M. and Ferreira, D. C. (2019b). The health care reform in Portugal: Outcomes from both the New Public Management and the economic crisis. *International Journal of Health Planning and Management*, 34(1):196–215.

Nunes, A. M., Ferreira, D. C., and Fernandes, A. C. (2019). Financial crisis in Portugal: Effects in the health care sector. *International Journal of Health Services*, 49(2):237–259.

OECD (2001). *Health at a Glance 2001*. OECD Publishing.

OECD (2004). The OECD-JRC Handbook on Practices for Developing Composite Indicators. *OECD Committee on Statistics*, pages 7–8.

OECD and JRC (2008). *Handbook on constructing composite indicators: methodology and user guide*. OECD Publishing.

OECD/European Observatory on Health Systems and Policies (2017). *Portugal: Country Health Profile 2017, State of Health in the EU*. OCDE Publishing.

Oliveira, M. D. and Pinto, C. G. (2005). Health care reform in Portugal: An evaluation of the NHS experience. *Health Economics*, 14(S1):203–220.

Olmsted, M. G., Powell, R., Murphy, J., Bell, D., Stanley, M., Mellissa, M., Marshica, S., and Geisen, E. (2018). U.S. News & World Report 2018-19 Best Hospitals: Specialty Rankings. Technical report, Research Triangle Institute International.

Park, G. W., Kim, Y., Park, K., and Agarwal, A. (2016). Patient-centric quality assessment framework for healthcare services. *Technological Forecasting and Social Change*, 113:468–474.

Parker, L. A., Zaragoza, G. A., and Hernández-Aguado, I. (2019). Promoting population health with public-private partnerships: Where's the evidence? *BMC Public Health*, 19(1):1–8.

Pereira, M., Ferreira, D., and Marques, R. (2020). A critical look at the portuguese public–private partnerships in healthcare. *International Journal of Health Planning and Management*.

Perelman, J., Felix, S., and Santana, R. (2015). The great recession in Portugal: Impact on hospital care use. *Health Policy*, 119(3):307–315.

Podinovskii, V. V. (1994). Criteria importance theory. *Mathematical Social Sciences*, 27(3):237–252.

Pronovost, P. J., Thompson, D. A., Holzmueller, C. G., Lubomski, L. H., and Morlock, L. L. (2005). Defining and measuring patient safety. *Critical Care Clinics*, 21(1):1–19.

Queenan, J. T. (2011). How to stop the relentless rise in cesarean deliveries. *Obstetrics and Gynecology*, 118(6):1421–1422.

Ramos, F. V. (2006). Despacho nº 721/2006. *Diário da república nº 8, série ii*, pages 468–471.

Rangel-Valdez, N., Fernandez, E., Cruz-Reyes, L., Gomez-Santillan, C., Rivera, G., and Florencia, R. (2018). Robustness analysis of an outranking model parameters' elicitation method in the presence of noisy examples. *Mathematical Problems in Engineering*, 2018.

Reeves, D., Campbell, S. M., Adams, J., Shekelle, P. G., Kontopantelis, E., and Roland, M. O. (2007). Combining multiple indicators of clinical quality: An evaluation of different analytic approaches. *Medical Care*, 45(6):489–496.

Rego, G., Nunes, R., and Costa, J. (2010). The challenge of corporatisation: The experience of Portuguese public hospitals. *European Journal of Health Economics*, 11(4):367–381.

Robertson-Preidler, J., Biller-Andorno, N., and Johnson, T. J. (2017). What is appropriate care? An integrative review of emerging themes in the literature. *BMC Health Services Research*, 17(1):1–17.

Robson, M., Hartigan, L., and Murphy, M. (2013). Methods of achieving and maintaining an appropriate caesarean section rate. *Best Practice and Research: Clinical Obstetrics and Gynaecology*, 27(2):297–308.

Robson, M. S. (2018). Use of indications to identify appropriate caesarean section rates. *The Lancet Global Health*, 6(8):e820–e821.

Rocha, A. (2019). Quality assessment of the Portuguese public hospitals: A multicriteria approach. Master's thesis, Instituto Superior Técnico, Universidade de Lisboa.

Rosenhead, J. and Mingers, J. (2001). Robustness analysis: Keeping your options open. In *Rational analysis for a problematic world revisited*, volume 32, pages 181–207. Wiley, Chichester.

Rosenthal, M. B., Landrum, M. B., Meara, E., Huskamp, H. A., Conti, R. M., and Keating, N. L. (2007). Using performance data to identify preferred hospitals. *Health Services Research*, 42(6p1):2109–2119.

Roy, B. (1991). The outranking approach and the foundations of ELECTRE methods. *Theory and Decision*, 31:49–73.

Roy, B. (2016). Paradigms and challenges. In *Multiple criteria decision analysis*, pages 19–39. Springer.

Saisana, M. and Tarantola, S. (2002). *State-of-the-art report on current methodologies and practices for composite indicator development*, volume 214. Citeseer.

Sakellarides, C., Castelo-Branco, L., Barbosa, P., and Azevedo, H. (2014). The impact of the financial crisis on the health system and health in portugal. *Copenhagen: World Health Organization*.

Schang, L., Hynninen, Y., Morton, A., and Salo, A. (2016). Developing robust composite measures of healthcare quality: Ranking intervals and dominance relations for Scottish Health Boards. *Social Science and Medicine*, 162:59–67.

Shwartz, M., Burgess, J. F., and Berlowitz, D. (2009). Benefit-of-the-doubt approaches for calculating a composite measure of quality. *Health Services and Outcomes Research Methodology*, 9(4):234–251.

Shwartz, M., Burgess, J. F., and Zhu, J. (2016). A DEA based composite measure of quality and its associated data uncertainty interval for health care provider profiling and pay-for-performance. *European Journal of Operational Research*, 253(2):489–502.

Shwartz, M., Cohen, A. B., Restuccia, J. D., Ren, Z. J., Labonte, A., Theokary, C., Kang, R., and Horwitt, J. (2011). How well can we identify the high-performing hospital? *Medical Care Research and Review*, 68(3):290–310.

Shwartz, M., Ren, J., Peköz, E. A., Wang, X., Cohen, A. B., and Restuccia, J. D. (2008). Estimating a composite measure of hospital quality from the hospital compare database. *Medical Care*, 46(8):778–785.

Shwartz, M., Restuccia, J. D., and Rosen, A. K. (2015). Composite measures of health care provider performance: A description of approaches. *Milbank Quarterly*, 93(4):788–825.

Simões, J., Augusto, G. F., Fronteira, I., and Hernández-Quevedo, C. (2017). Portugal: Health system review. *Health Systems in Transition*, 19(2):1–184.

Talib, F., Azam, M., and Rahman, Z. (2015). Service quality in healthcare establishments: a literature review. *International Journal of Behavioural and Healthcare Research*, 5(1–2):1–24.

Tànfani, E. and Testi, A. (2012). *Advanced decision making methods applied to health care*, volume 173. Springer Science & Business Media.

Tervonen, T., Figueira, J. R., Lahdelma, R., Dias, J. A., and Salminen, P. (2009). A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research*, 192(1):236–242.

Thokala, P. and Duenas, A. (2012). Multiple criteria decision analysis for health technology assessment. *Value in Health*, 15(8):1172–1181.

Torchia, M., Calabrò, A., and Morner, M. (2015). Public–private partnerships in the health care sector: A systematic review of the literature. *Public Management Review*, 17(2):236–261.

Veillard, J., Garcia-Armesto, S., Kadandale, S., and Klazinga, N. (2009). International health system comparisons: from measurement challenge to management tool. *Performance measurement for health system improvement: experiences, challenges and prospects*, pages 641–672.

Wind, Y. and Green, P. E. (2013). *Marketing research and modeling, Progress and prospects: A tribute to Paul E. Green*, volume 14. Springer Science & Business Media.

Wood, D. J. (1991). Corporate social performance revisited. *The Academy of Management Review*, 16(4):691–718.

World Health Organization (2010). *Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies*. World Health Organization.

World Health Organization (2016). *Health system efficiency: how to make measurement matter for policy and management*. World Health Organization. Regional Office for Europe.

Yang, O., Chan, M. K., Cheng, T. C., and Yong, J. (2020). Cream skimming: Theory and evidence from hospital transfers and capacity utilization. *Journal of Economic Behavior and Organization*, 173:68–87.

Zineldin, M. (2006). The quality of health care and patient satisfaction: An exploratory investigation of the 5Qs model at some Egyptian and Jordanian medical clinics. *International Journal of Health Care Quality Assurance*, 19(1):60–92.

Appendix A

Tables of Results

Table A.1: Results of February 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	**Max**	**Min**	**Max**	**Min**	**Max**
a_1	C_3	C_3	C_3	C_3	C_3	C_3
a_2	C_2	C_3	C_2	C_3	C_2	C_3
a_3	C_3	C_3	C_3	C_3	C_3	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_2	C_3	C_2	C_4	C_2	C_4
a_6	C_3	C_4	C_3	C_3	C_3	C_3
a_7	C_3	C_3	C_3	C_4	C_3	C_4
a_8	C_3	C_3	C_3	C_3	C_3	C_4
a_9	C_3	C_4	C_3	C_4	C_3	C_4
a_{10}	C_1	C_3	C_1	C_3	C_1	C_3
a_{11}	C_1	C_3	C_1	C_3	C_1	C_3
a_{12}	C_2	C_3	C_2	C_3	C_2	C_3
a_{13}	C_3	C_3	C_3	C_3	C_2	C_3
a_{14}	C_3	C_3	C_3	C_3	C_3	C_3
a_{15}	C_3	C_3	C_3	C_3	C_3	C_3
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_3	C_3	C_3	C_3
a_{18}	C_3	C_4	C_3	C_4	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_2	C_2	C_2	C_2	C_2	C_2
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_2	C_3	C_2	C_2	C_2	C_2
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_3	C_2	C_3	C_2	C_3
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_2	C_3	C_2	C_3	C_2	C_3
a_{28}	C_4	C_4	C_4	C_4	C_3	C_4
a_{29}					C_3	C_3
a_{30}					C_3	C_3

Table A.2: Results of March 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	**Max**	**Min**	**Max**	**Min**	**Max**
a_1	C_2	C_2	C_2	C_2	C_2	C_3
a_2	C_2	C_2	C_2	C_2	C_2	C_3
a_3	C_3	C_4	C_3	C_4	C_4	C_4
a_4	C_4	C_4	C_3	C_3	C_4	C_4
a_5	C_3	C_3	C_2	C_3	C_3	C_3
a_6	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_3	C_3	C_2	C_3	C_3	C_3
a_8	C_3	C_4	C_2	C_4	C_3	C_4
a_9	C_3	C_3	C_3	C_3	C_3	C_3
a_{10}	C_2	C_4	C_2	C_4	C_2	C_4
a_{11}	C_3	C_3	C_2	C_3	C_3	C_3
a_{12}	C_3	C_3	C_2	C_3	C_3	C_3
a_{13}	C_3	C_4	C_3	C_3	C_3	C_3
a_{14}	C_3	C_3	C_3	C_3	C_3	C_3
a_{15}	C_3	C_3	C_2	C_3	C_3	C_3
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_2	C_2	C_2	C_2	C_2
a_{18}	C_3	C_4	C_3	C_4	C_3	C_3
a_{19}	C_3	C_3	C_2	C_2	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_3	C_3
a_{22}	C_2	C_2	C_2	C_2	C_2	C_2
a_{23}	C_3	C_3	C_2	C_3	C_3	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_2	C_3	C_2	C_2	C_2	C_3
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}					C_3	C_3
a_{30}					C_3	C_3

Table A.3: Results of April 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	Max	Min	Max	Min	Max
a_1	C_3	C_3	C_3	C_3	C_3	C_3
a_2	C_2	C_2	C_2	C_3	C_2	C_2
a_3	C_4	C_4	C_2	C_3	C_4	C_4
a_4	C_3	C_3	C_2	C_4	C_3	C_3
a_5	C_3	C_4	C_2	C_4	C_3	C_3
a_6	C_1	C_3	C_1	C_5	C_1	C_3
a_7	C_2	C_3	C_2	C_2	C_2	C_3
a_8	C_2	C_3	C_3	C_3	C_2	C_3
a_9	C_3	C_4	C_1	C_3	C_3	C_4
a_{10}	C_2	C_3	C_2	C_3	C_2	C_3
a_{11}	C_3	C_3	C_1	C_2	C_3	C_3
a_{12}	C_2	C_3	C_2	C_4	C_2	C_3
a_{13}	C_3	C_3	C_1	C_2	C_3	C_3
a_{14}	C_4	C_4	C_2	C_4	C_4	C_4
a_{15}	C_2	C_3	C_1	C_3	C_2	C_3
a_{16}	C_3	C_3	C_2	C_2	C_3	C_3
a_{17}	C_3	C_3	C_3	C_3	C_3	C_3
a_{18}	C_3	C_3	C_3	C_5	C_3	C_3
a_{19}	C_3	C_3	C_2	C_2	C_3	C_3
a_{20}	C_3	C_3	C_3	C_4	C_3	C_3
a_{21}	C_5	C_5	C_4	C_4	C_5	C_5
a_{22}	C_3	C_3	C_2	C_3	C_3	C_3
a_{23}	C_3	C_3	C_2	C_4	C_3	C_3
a_{24}	C_3	C_3	C_1	C_3	C_3	C_3
a_{25}	C_2	C_2	C_2	C_4	C_2	C_2
a_{26}	C_2	C_2	C_3	C_3	C_2	C_2
a_{27}	C_3	C_3	C_1	C_3	C_3	C_3
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}					C_2	C_2
a_{30}					C_4	C_4

Table A.4: Results of May 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	Max	Min	Max	Min	Max
a_1	C_4	C_4	C_4	C_4	C_4	C_4
a_2	C_3	C_3	C_3	C_3	C_3	C_3
a_3	C_4	C_4	C_4	C_4	C_4	C_4
a_4	C_3	C_3	C_3	C_3	C_3	C_3
a_5	C_3	C_4	C_3	C_4	C_3	C_4
a_6	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_2	C_2	C_2	C_2	C_2	C_2
a_8	C_2	C_3	C_2	C_3	C_2	C_3
a_9	C_3	C_3	C_3	C_3	C_3	C_3
a_{10}	C_3	C_4	C_3	C_4	C_3	C_4
a_{11}	C_2	C_2	C_2	C_2	C_2	C_2
a_{12}	C_3	C_3	C_3	C_3	C_3	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_3	C_3	C_3	C_3	C_3	C_3
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_2	C_2	C_2	C_2	C_3
a_{18}	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_3	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_2	C_3	C_2	C_3	C_2	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_3	C_3	C_3	C_3	C_3	C_3
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}					C_3	C_3
a_{30}					C_3	C_3

Table A.11: Results of December 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	Max	Min	Max	Min	Max
a_1	C_3	C_4	C_3	C_4	C_3	C_4
a_2	C_1	C_3	C_1	C_3	C_1	C_3
a_3	C_2	C_3	C_2	C_3	C_2	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_3	C_3	C_3	C_3	C_3	C_3
a_6	C_3	C_4	C_3	C_4	C_3	C_4
a_7	C_3	C_3	C_3	C_3	C_3	C_3
a_8	C_2	C_5	C_2	C_5	C_2	C_3
a_9	C_4	C_4	C_4	C_4	C_4	C_4

a_{10}	C_3	C_3	C_3	C_3	C_3	C_3
a_{11}	C_2	C_3	C_2	C_3	C_2	C_3
a_{12}	C_2	C_3	C_2	C_3	C_2	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_2	C_3	C_2	C_3	C_2	C_3
a_{16}	C_1	C_2	C_1	C_2	C_1	C_2
a_{17}	C_3	C_5	C_3	C_5	C_3	C_3
a_{18}	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_2	C_3	C_2	C_3	C_2	C_2
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_1	C_3	C_1	C_3	C_2	C_3
a_{24}	C_2	C_2	C_2	C_2	C_2	C_2
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_1	C_5	C_1	C_5	C_1	C_2
a_{28}	C_2	C_3	C_2	C_3	C_2	C_3
a_{29}					C_2	C_2
a_{30}					C_2	C_2

Table A.5: Results of June 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	**Max**	**Min**	**Max**	**Min**	**Max**
a_1	C_4	C_4	C_4	C_4	C_4	C_4
a_2	C_2	C_2	C_2	C_2	C_2	C_2
a_3	C_4	C_4	C_4	C_4	C_4	C_4
a_4	C_3	C_3	C_3	C_3	C_3	C_3
a_5	C_3	C_3	C_3	C_3	C_3	C_4
a_6	C_4	C_4	C_4	C_4	C_4	C_4
a_7	C_3	C_3	C_3	C_3	C_2	C_3
a_8	C_3	C_3	C_3	C_3	C_3	C_3
a_9	C_3	C_3	C_3	C_3	C_3	C_3
a_{10}	C_3	C_3	C_3	C_3	C_3	C_3
a_{11}	C_3	C_3	C_3	C_3	C_3	C_3
a_{12}	C_3	C_3	C_3	C_3	C_3	C_3
a_{13}	C_2	C_2	C_2	C_2	C_2	C_2
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_2	C_3	C_2	C_3	C_2	C_4
a_{16}	C_3	C_4	C_3	C_4	C_3	C_4
a_{17}	C_3	C_3	C_3	C_3	C_3	C_3
a_{18}	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_4	C_4	C_4	C_4	C_3	C_4
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_3	C_3	C_3	C_3	C_3	C_3
a_{28}	C_3	C_3	C_3	C_4	C_3	C_3
a_{29}					C_2	C_2
a_{30}					C_3	C_3

Table A.6: Results of July 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	**Max**	**Min**	**Max**	**Min**	**Max**
a_1	C_3	C_3	C_3	C_3	C_3	C_3
a_2	C_2	C_3	C_2	C_3	C_2	C_3
a_3	C_2	C_3	C_2	C_3	C_2	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_2	C_3	C_2	C_3	C_2	C_3
a_6	C_3	C_4	C_3	C_4	C_3	C_4
a_7	C_3	C_3	C_3	C_3	C_3	C_3
a_8	C_2	C_3	C_2	C_3	C_2	C_3
a_9	C_3	C_3	C_3	C_3	C_3	C_4
a_{10}	C_2	C_2	C_2	C_3	C_2	C_3
a_{11}	C_3	C_4	C_3	C_4	C_3	C_4
a_{12}	C_3	C_3	C_3	C_3	C_3	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_3	C_3	C_3	C_3	C_3	C_4
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_2	C_3	C_2	C_3
a_{18}	C_3	C_3	C_3	C_3	C_2	C_2
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_3	C_3	C_3	C_3	C_3	C_3
a_{22}	C_3	C_4	C_3	C_4	C_3	C_4
a_{23}	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_1	C_2	C_1	C_2	C_1	C_2
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_2	C_3	C_2	C_3	C_2	C_3
a_{28}	C_3	C_3	C_3	C_4	C_3	C_3
a_{29}					C_2	C_2
a_{30}					C_3	C_3

Table A.7: Results of August 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	Max	Min	Max	Min	Max
a_1	C_3	C_3	C_3	C_3	C_3	C_4
a_2	C_3	C_3	C_3	C_3	C_3	C_32
a_3	C_1	C_3	C_1	C_3	C_1	C_4
a_4	C_1	C_3	C_1	C_3	C_1	C_3
a_5	C_1	C_3	C_1	C_3	C_1	C_4
a_6	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_3	C_3	C_3	C_3	C_3	C_3
a_8	C_2	C_3	C_2	C_2	C_2	C_2
a_9	C_4	C_4	C_4	C_4	C_4	C_4
a_{10}	C_2	C_3	C_2	C_3	C_2	C_3
a_{11}	C_3	C_4	C_3	C_4	C_3	C_4
a_{12}	C_3	C_3	C_3	C_3	C_3	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_3	C_4	C_3	C_4	C_3	C_4
a_{15}	C_2	C_3	C_2	C_2	C_2	C_2
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_3	C_3	C_3	C_3	C_3	C_3
a_{18}	C_4	C_4	C_4	C_4	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_3	C_3
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_2	C_3	C_2	C_3	C_2	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_3	C_3	C_3	C_3	C_3	C_3
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}					C_3	C_3
a_{30}					C_3	C_4

Table A.8: Results of September 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	Max	Min	Max	Min	Max
a_1	C_3	C_4	C_3	C_3	C_3	C_3
a_2	C_2	C_3	C_2	C_3	C_2	C_3
a_3	C_4	C_4	C_3	C_3	C_3	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_2	C_3	C_2	C_3	C_2	C_4
a_6	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_2	C_2	C_2	C_3	C_2	C_3
a_8	C_2	C_3	C_2	C_3	C_2	C_3
a_9	C_3	C_4	C_3	C_4	C_3	C_4
a_{10}	C_2	C_3	C_2	C_3	C_2	C_3
a_{11}	C_3	C_4	C_3	C_4	C_3	C_4
a_{12}	C_2	C_2	C_2	C_2	C_2	C_2
a_{13}	C_3	C_3	C_3	C_3	C_2	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_3	C_3	C_3	C_3	C_3	C_4
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_2	C_3	C_2	C_3
a_{18}	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_2	C_2	C_2	C_3	C_2	C_3
a_{20}	C_2	C_2	C_2	C_2	C_2	C_2
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_3	C_2	C_2	C_2	C_3
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_2	C_2	C_2	C_2	C_2	C_2
a_{28}	C_2	C_2	C_2	C_2	C_2	C_2
a_{29}					C_3	C_4
a_{30}					C_2	C_2

Table A.9: Results of October 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	Max	Min	Max	Min	Max
a_1	C_3	C_3	C_3	C_3	C_3	C_3
a_2	C_2	C_2	C_2	C_2	C_2	C_3
a_3	C_3	C_4	C_3	C_4	C_4	C_4
a_4	C_3	C_3	C_3	C_3	C_3	C_3
a_5	C_3	C_3	C_3	C_3	C_3	C_3
a_6	C_3	C_3	C_3	C_3	C_3	C_4
a_7	C_2	C_3	C_2	C_3	C_2	C_3
a_8	C_2	C_3	C_2	C_3	C_2	C_3
a_9	C_3	C_3	C_3	C_3	C_3	C_3
a_{10}	C_3	C_3	C_3	C_3	C_3	C_3
a_{11}	C_3	C_3	C_3	C_3	C_3	C_3
a_{12}	C_2	C_3	C_2	C_3	C_2	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_3	C_4	C_3	C_4	C_3	C_4
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_2	C_3	C_3	C_4
a_{18}	C_4	C_4	C_4	C_4	C_3	C_3
a_{19}	C_2	C_2	C_2	C_2	C_2	C_2
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_2	C_2	C_2	C_2	C_2	C_2
a_{23}	C_3	C_4	C_3	C_4	C_3	C_3
a_{24}	C_2	C_2	C_2	C_2	C_2	C_2
a_{25}	C_2	C_3	C_2	C_3	C_2	C_2
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_2	C_3	C_2	C_3	C_2	C_2
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}					C_2	C_2
a_{30}					C_3	C_3

Table A.10: Results of November 2018, for Model 1 and Model 2.

	Model 1 (social)		Model 1 (efficiency)		Model 2	
	Min	Max	Min	Max	Min	Max
a_1	C_3	C_4	C_3	C_4	C_3	C_4
a_2	C_2	C_3	C_2	C_3	C_2	C_3
a_3	C_3	C_4	C_3	C_4	C_4	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_3	C_3	C_3	C_3	C_3	C_3
a_6	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_2	C_3	C_2	C_3	C_2	C_3
a_8	C_2	C_3	C_2	C_3	C_2	C_3
a_9	C_3	C_4	C_3	C_4	C_3	C_4
a_{10}	C_3	C_3	C_3	C_3	C_3	C_3
a_{11}	C_4	C_4	C_4	C_4	C_4	C_4
a_{12}	C_2	C_3	C_2	C_3	C_2	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_3	C_4	C_3	C_4	C_3	C_4
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_2	C_3	C_2	C_3
a_{18}	C_3	C_4	C_3	C_4	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_2	C_2
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_2	C_3	C_2	C_3	C_2	C_3
a_{23}	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_2	C_3	C_2	C_2	C_2	C_2
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_2	C_3	C_2	C_2	C_2	C_2
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}					C_2	C_3
a_{30}					C_2	C_3

Appendix B

Robustness Analysis

Table B.1: Results of Model 1 and Model 2 for February 2018, considering different λ values.

	Model 1 (social)						**Model 1 (efficiency)**						**Model 2**					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_2	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_2	C_3	C_2	C_4	C_2	C_3	C_2	C_3	C_2	C_4	C_2	C_4	C_2	C_3	C_2	C_4	C_2	C_4
a_6	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_3	C_3	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4
a_8	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_4
a_9	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_{10}	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3
a_{11}	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3
a_{12}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{14}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{15}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_3	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_3	C_3	C_2	C_3	C_2	C_3	C_3	C_3
a_{18}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{28}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_3	C_4

a_{29}													C_3	C_3	C_3	C_3	C_3	C_3
a_{30}													C_3	C_3	C_3	C_3	C_3	C_3

Table B.2: Results of Model 1 and Model 2 for March 2018, considering different λ values.

	Model 1 (social)						**Model 1 (efficiency)**						**Model 2**					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_3
a_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_3
a_3	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_4	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_6	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_2	C_2	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_8	C_3	C_3	C_3	C_4	C_3	C_4	C_2	C_3	C_2	C_4	C_2	C_4	C_3	C_3	C_3	C_4	C_3	C_4
a_9	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_{10}	C_2	C_4	C_2	C_4	C_2	C_4	C_2	C_4	C_2	C_4	C_2	C_4	C_2	C_3	C_2	C_3	C_2	C_4
a_{11}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_2	C_2	C_2	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{12}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_2	C_2	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{13}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_3	C_3	C_3
a_{14}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{15}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{18}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_{22}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{23}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_3
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}													C_3	C_3	C_3	C_3	C_3	C_3
a_{30}													C_3	C_3	C_3	C_3	C_3	C_3

Table B.3: Results of Model 1 and Model 2 for April 2018, considering different λ values.

	Model 1 (social)						**Model 1 (efficiency)**						**Model 2**					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2
a_3	C_4	C_4	C_4	C_4	C_4	C_4	C_2	C_3	C_2	C_3	C_2	C_3	C_4	C_4	C_4	C_4	C_4	C_4
a_4	C_3	C_4	C_3	C_4	C_3	C_3	C_2	C_4	C_2	C_4	C_2	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_5	C_3	C_3	C_3	C_3	C_3	C_4	C_2	C_4	C_2	C_4	C_2	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_6	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_5	C_1	C_5	C_1	C_5	C_1	C_3	C_1	C_3	C_1	C_3
a_7	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_3
a_8	C_2	C_2	C_2	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_2	C_2	C_3	C_2	C_3
a_9	C_3	C_4	C_3	C_4	C_3	C_4	C_1	C_2	C_1	C_2	C_1	C_3	C_3	C_4	C_3	C_4	C_3	C_4
a_{10}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{11}	C_3	C_3	C_3	C_3	C_3	C_3	C_1	C_2	C_1	C_2	C_1	C_2	C_3	C_3	C_3	C_3	C_3	C_3
a_{12}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_4	C_2	C_4	C_2	C_4	C_2	C_3	C_2	C_3	C_2	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3	C_1	C_2	C_1	C_2	C_1	C_2	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4	C_2	C_4	C_2	C_4	C_2	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_2	C_3	C_2	C_3	C_2	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{18}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_5	C_3	C_5	C_3	C_5	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_4	C_2	C_2	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_5	C_5	C_5	C_5	C_5	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_5	C_5	C_5	C_5	C_5
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_3	C_3	C_3	C_3
a_{23}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_4	C_2	C_4	C_2	C_4	C_2	C_4	C_3	C_3	C_3	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_4	C_2	C_4	C_2	C_4	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_3	C_3	C_3	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_3	C_3	C_3	C_3	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}													C_2	C_2	C_2	C_2	C_2	C_2
a_{30}													C_4	C_4	C_4	C_4	C_4	C_4

Table B.4: Results of Model 1 and Model 2 for May 2018, considering different λ values.

	Model 1 (social)						Model 1 (efficiency)						Model 2					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_3	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_4	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_5	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4
a_6	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_8	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_3
a_9	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{10}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_4
a_{11}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{12}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_3
a_{18}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_3	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3
a_{23}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_3
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_2	C_3	C_3	C_3	C_2	C_3	C_2	C_3	C_3	C_3
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4
a_{29}													C_3	C_3	C_3	C_3	C_3	C_3
a_{30}													C_3	C_3	C_3	C_3	C_3	C_3

Table B.5: Results of Model 1 and Model 2 for June 2018, considering different λ values.

	Model 1 (social)						**Model 1 (efficiency)**						**Model 2**					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_3	C_4	C_4	C_4	C_4	C_3	C_4	C_3	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_4	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_5	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_4
a_6	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_3	C_4
a_7	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_8	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_9	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{10}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{11}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{12}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{13}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{14}	C_4	C_4	C_4	C_4	C_3	C_4	C_3	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_4	C_2	C_4
a_{16}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_{17}	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_2	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3
a_{18}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_3	C_4	C_3	C_4
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{28}	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}													C_2	C_2	C_2	C_2	C_2	C_2
a_{30}													C_3	C_3	C_3	C_3	C_3	C_3

Table B.6: Results of Model 1 and Model 2 for July 2018, considering different λ values.

	Model 1 (social)						**Model 1 (efficiency)**						**Model 2**					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_2	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_6	C_4	C_4	C_4	C_4	C_3	C_4	C_4	C_4	C_4	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_7	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_8	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_9	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_4
a_{10}	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_3
a_{11}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_{12}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_4
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{18}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_2	C_2	C_2	C_2	C_2
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{22}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_{23}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}													C_2	C_2	C_2	C_2	C_2	C_2
a_{30}													C_3	C_3	C_3	C_3	C_3	C_3

Table B.7: Results of Model 1 and Model 2 for August 2018, considering different λ values.

	Model 1 (social)						Model 1 (efficiency)						Model 2					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_4
a_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_4
a_4	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3
a_5	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_4	C_1	C_4
a_6	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_8	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_9	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{10}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{11}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_4	C_4	C_3	C_4	C_3	C_4
a_{12}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_3	C_4	C_4	C_4	C_4	C_4	C_3	C_4	C_4	C_4	C_4	C_4	C_3	C_4
a_{15}	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_2	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3
a_{18}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}													C_3	C_3	C_3	C_3	C_3	C_3
a_{30}													C_4	C_4	C_3	C_4	C_3	C_4

Table B.8: Results of Model 1 and Model 2 for September 2018, considering different λ values.

	Model 1 (social)						**Model 1 (efficiency)**						**Model 2**					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_3
a_2	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_3	C_3	C_4	C_4	C_4	C_4	C_4	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_3	C_3	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_4	C_2	C_4
a_6	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_3
a_8	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_9	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_{10}	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{11}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_{12}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_4
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{18}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3	
a_{20}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_2	C_3	C_2	C_3	C_3	C_3	C_2	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{25}	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_3
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{28}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{29}													C_3	C_3	C_3	C_4	C_3	C_4
a_{30}													C_2	C_2	C_2	C_2	C_2	C_2

Table B.9: Results of Model 1 and Model 2 for October 2018, considering different λ values.

	Model 1 (social)						**Model 1 (efficiency)**						**Model 2**					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_3
a_3	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_4	C_4
a_4	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_5	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_6	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_4
a_7	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_8	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3
a_9	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{10}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{11}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{12}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_4	C_2	C_4
a_{18}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{23}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{25}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{27}	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_3	C_3	C_3
a_{29}													C_2	C_2	C_2	C_2	C_2	C_2
a_{30}													C_3	C_4	C_3	C_3	C_3	C_3

Table B.10: Results of Model 1 and Model 2 for November 2018, considering different λ values.

	Model 1 (social)						**Model 1 (efficiency)**						**Model 2**					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_2	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_3	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_4	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_6	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_7	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_8	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_3
a_9	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_{10}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{11}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{12}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_4	C_3	C_4
a_{16}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{17}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{18}	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_2	C_2	C_2	C_2	C_2	C_2
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{22}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{23}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{24}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_2	C_2	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{28}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{29}													C_2	C_2	C_2	C_3	C_2	C_3
a_{30}													C_3	C_3	C_3	C_3	C_3	C_3

Table B.11: Results of Model 1 and Model 2 for December 2018, considering different λ values.

	Model 1 (social)						**Model 1 (efficiency)**						**Model 2**					
	λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65		λ=0.55		λ=0.60		λ=0.65	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
a_1	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_2	C_1	C_2	C_1	C_2	C_1	C_3	C_1	C_2	C_1	C_2	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3
a_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_4
a_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_5	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_6	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_3	C_3	C_3	C_3	C_4	C_3	C_4	C_3	C_4	C_3	C_4
a_7	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_8	C_2	C_5	C_2	C_5	C_2	C_5	C_2	C_5	C_2	C_5	C_2	C_5	C_2	C_3	C_2	C_3	C_2	C_3
a_9	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{10}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{11}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{12}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{13}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{14}	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4	C_4
a_{15}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_3	C_2	C_3
a_{16}	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
a_{17}	C_3	C_5	C_3	C_5	C_3	C_5	C_3	C_5	C_3	C_5	C_3	C_5	C_3	C_3	C_3	C_3	C_3	C_3
a_{18}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{19}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_2	C_2	C_2	C_2	C_2
a_{20}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{21}	C_3	C_4	C_4	C_4	C_4	C_4	C_3	C_4	C_4	C_4	C_4	C_4	C_3	C_4	C_4	C_4	C_4	C_4
a_{22}	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3	C_3
a_{23}	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_1	C_3	C_2	C_2	C_2	C_2	C_2	C_2
a_{24}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{25}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{26}	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2	C_2
a_{27}	C_1	C_5	C_1	C_5	C_1	C_5	C_2	C_2	C_2	C_2	C_2	C_2	C_1	C_2	C_1	C_2	C_1	C_2
a_{28}	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
a_{29}													C_2	C_2	C_2	C_2	C_2	C_2
a_{30}													C_2	C_2	C_2	C_2	C_2	C_2