Satisfaction of healthcare consumers

Analysis and comparison of different methodologies

Inês de Carvalho Vieira

Thesis to obtain the Master of Science Degree in

Industrial Engineering and Management

Supervisors: Prof. Diogo Filipe da Cunha Ferreira
Prof. Maria Isabel Craveiro Pedro

Examination Committee

Chairperson: Prof. Susana Isabel Carvalho Relvas
Supervisor: Prof. Diogo Filipe da Cunha Ferreira
Member of the Comittee: Prof. Alexandre Manuel Martins Morais Nunes

January 2021

Declaration

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

Resumo

Atualmente, a satisfação do utente e os fatores que a influenciam tornam-se numa das maiores preocupações, não só das organizações de saúde, mas também dos próprios utentes. A garantia da qualidade dos serviços oferecidos é essencial para a satisfação das expectativas e das necessidades dos utentes. Por sua vez, o sistema de saúde está em constante desenvolvimento, o que faz com que os fatores que necessitam de aperfeiçoamento se alterem ao longo do tempo. Esse é o principal foco da satisfação do utente, indicar que fatores necessitam de ser melhorados, de forma a proporcionar o melhor serviço possível.

Serve a presente dissertação para avaliar a satisfação dos utentes de uma unidade de cuidados de saúde secundários, mais precisamente, o serviço de internamento. Propõe-se determinar quais os fatores que mais influenciam a satisfação, juntamente com o nível de satisfação percecionado pelo paciente. Para tal, é realizada uma comparação metodológica (com recurso a análise fatorial, modelação de equações estruturais, regressão ordinal logística, e modelo multicritério para análise de satisfação) com o objetivo de contrapor diferentes resultados que possam advir das diferentes metodologias. Após a implementação das mesmas, conclui-se que dos onze fatores analisados, sete influenciam a satisfação do paciente, sendo estes: classificação das instalações, exames e tratamentos no hospital, pessoal auxiliar, pessoal médico, alimentação, voluntariado, e informação recebida. É esperado que com esta informação os gestores hospitalares consigam alocar os recursos disponíveis de forma mais eficiente, com o intuito de proporcionar uma melhor experiência hospitalar aos utentes Portugueses.

Palavras-chave: Satisfação do utente; Qualidade dos cuidados de saúde; Análise fatorial; Modelação de equações estruturais; Regressão ordinal logística; Modelo multicritério para análise de satisfação;

Abstract

In the current world, patient satisfaction and the factors influencing it are becoming one of the principal concerns, for health organisations and patients themselves. Assuring the quality of the provided services is essential for the fulfilment of patients' expectations and needs. Thus, with a continually evolving health system, factors that need improvement keep changing over time. That is the main focus of patient satisfaction, to point out factors and areas that have to be reviewed in order to provide the best possible service.

This dissertation evaluates patients' satisfaction regarding a secondary healthcare unit, in particular, the internment service. The main purpose is to identify which factors influence satisfaction, along with the satisfaction levels perceived by patients. To accomplish this, a methodological comparison (using factor analysis, structural equation modeling, ordinal logistic regression, and multicriteria satisfaction analysis) is performed with the goal to contrast the different results that may arise from the usage of different methodologies. After implementing the methods, it is concluded that, out of the eleven analysed factors, seven influence satisfaction, being them: accommodations, exams and treatments, auxiliary staff, medical staff, food quality, volunteering staff, and obtained information. It is expected that with this information, hospital managers are able to allocate the available resources in a more efficient manner, improving the healthcare experiences of Portuguese patients.

Keywords: Patient satisfaction; Quality of healthcare; Factor Analysis; Structural Equation Modeling; Ordinal Logistic Regression; Multicriteria Satisfaction Analysis;

Acknowledgments

The work presented in this thesis was performed at the Instituto de Engenharia de Sistemas e Computadores Investigação e Desenvolvimento (INESC-ID), Civil Engineering Research and Innovation for Sustainability (CERIS) and Centro de Estudos de Gestão (CEG-IST) of Instituto Superior Técnico (Lisbon, Portugal). This work is part of the hSNS FCT – Research Project (PTDC/EGE-OGE/30546/2017).

Throughout the writing of this dissertation I received a great deal of assistance and encouragement that cannot be disregarded. In the first place, I would like to thank my parents, Ivone and Paulo, and my siblings, Lara and Rodrigo, for their unconditional support, motivation and endless perseverance that is crucial in every single part of my life.

Hereinafter, I would like to express my deepest gratitude to my supervisors, Prof. Diogo Ferreira and Prof. Isabel Pedro, without whom I would not have been able to complete this dissertation. Thank you for always providing me with the assistance I needed, while being extremely patient at the same time. I would also like to leave a word of appreciation for the sympathy demonstrated throughout this entire journey.

To my close family and friends, thank you for always enduring with me, through good times and bad.

Thank you all.

Table of Contents

List of figures

List of tables

List of abbreviations

ANOVA	Analysis of Variance
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CPCA	Categorical Principal Components Analysis
EFA	Exploratory Factor Analysis
GFI	Goodness of Fit Index
GLS	Generalized Least Square
HPSM	Hierarchical Patient Satisfaction Model
HRA	Health Regulatory Agency
ICC	Intra-class Correlation Coefficient
KMO	Kaiser–Meyer–Olkin
MAR	Missing at Random
MCAR	Missing Completely at Random
ML	Maximum Likelihood
MNAR	Missing Not at Random
MUSA	Multicriteria Satisfaction Analysis
NFI	Normed Fit Index
NHS	National Health Service
OLR	Ordinal Logistic Regression
OR	Odds Ratio
PCA	Principal Components Analysis
PCC	Pearson’s Correlation Coefficient
PCFI	Parsimony Comparative Fit Index
PGFI	Parsimony Goodness of Fit Index
PLS	Partial Least Squares
PLUM	Polytomous Universal Model

PNFI	Parsimony Normed Fit Index
PSM	Patient Satisfaction Model
RHA	Regional Health Administrations
RMSEA	Root Mean Squared Error Approximation
SCC	Spearman's Correlation Coefficient
SEM	Structural Equations Modeling
SERVQUAL	Service Quality
SPSS	Statistical Package for the Social Sciences
ULS	Unweighted Least Square
WLS	Weighted Least Square

1. Introduction

In this chapter, a brief introduction is made about the topics and the concepts discussed in this dissertation. It gives a framework to the problem and the motivation behind it. In addition, the dissertation structure is also presented.

1.1. Problem contextualization

Healthcare systems are continually changing and improving, and so it is necessary to find a way to evaluate outputs while measuring the satisfaction of the service receiver, in this case, the patient. *Patient satisfaction* can be defined as a patient's response to several aspects of their service experience. Assessing patient satisfaction may provide valuable and unique insights about daily hospital care and quality. It is considered an indicator of healthcare quality, though the connection between the two is far from clear (Ng and Luk, 2019). Patient satisfaction is a concept that has long been neglected and cast aside but is becoming gradually more important. Inclusively, Donabedian (1988) includes it as an outcome of healthcare services; hence, it is of extreme importance to evaluate satisfaction with care. Multiple authors argue that satisfaction and the result in terms of the patient's health status are related terms (Elixhauser et al., 2003; Rogers et al., 2004; Levin-Scherz et al., 2006; Needleman et al., 2006). Measuring healthcare quality and satisfaction does not only constitute an indispensable element for appropriate management of resources but also allows focusing o n its users' preferences, giving them a chance to participate in the construction of a customized health service, better fitted to their needs and expectations (Abrantes, 2012). With this information, managers can more efficiently allocate resources to improve patients' experience and satisfaction (Al-Abri and Al-Balushi, 2014). When talking about public hospitals, there may not be a financial interest in performing these studies since they are not particularly interested in profit. However, with the increase of market competitivity, for private companies, it is vital to meet patients' needs, satisfying them, so they become loyal to the organization (Lovelock and Wright, 2001). The purpose of health care, whether provided by a public or private entity, is to guarantee the best service level. Nevertheless, due to resources' restrictions, it is not always possible to provide the best care in every service dimension. That is why the present study can be a useful tool for health care improvement. The evaluation of which dimensions influence patient satisfaction provides valuable insights on how to allocate available resources in an effective manner.

1.2. Dissertation objectives

Despite the existence of a strong legal and political commitment to the well-being of society, health inequalities are an issue in Portugal. Hence, adjustments need to be made to increase the efficiency and quality of health services. On the one hand, intensive research has been done by diverse authors, discussing the many dimensions of perceived service quality (Ferreira, 2019). On the other hand, there is still a lack of studies comparing the results of different methodologies for the same purpose. In Portugal, specifically, satisfaction studies are held on a national level. This keeps the results from translating the actual reality of each health unit. Therefore, considering the various options present in literature, and the uttermost importance of patient satisfaction, reviewing the outcomes of different methods will help fill a significant research gap. This dissertation, thus, runs four widely used methods (Factor analysis, Structural Equation Modeling (SEM), Ordinal Logistic Regression (OLR), and Multicriteria Satisfaction Analysis (MUSA)) to study patients' satisfaction and its predictors in the internment service of a public Portuguese hospital. The main objectives are to understand if the methods return similar results and study patients' satisfaction in that specific hospital service.

1.3. Dissertation structure

The structure of this dissertation is as follows:

Chapter 1 – Introduction: A contextualisation of the problem, along with the objectives, are presented so that the reader can have a better understanding of the dissertation.

Chapter 2 – Satisfaction in health: In the second chapter, the evolution of customer satisfaction through the years is explored, along with the definition of patient satisfaction and its relationship with healthcare quality.

Chapter 3 – Healthcare in Portugal: Firstly, the structure and organisation of the Portuguese healthcare system are described, followed by satisfaction studies regarding the Portuguese National Health System (NHS).

Chapter 4 – Literature review: This chapter provides a theoretical analysis resulting in an overview of key findings. The information retrieved from previous researches is presented in a table that contains the authors, country of study, sample, quality dimensions and drivers, methodology, dependent variables, and main factors affecting satisfaction. Hereinafter, the utilization and influence of each criterion are assessed.

Chapter 5 – Methodology: The theory behind the methods that will be used is presented. This chapter is divided into five sections. The first one portrays the methodology present in the literature review, and each one of the remaining sections explores a different method.

Chapter 6 – Case study: The experience of Portuguese citizens with a public hospital's service. The case study is presented describing the criteria and subcriteria used to measure satisfaction within the internment service, how data was handled, and the reckoning of the final sample.

Chapter 7 – Results discussion and implications: The main results of each method are provided, together with managerial decisions that stem from those results.

Chapter 8 – Concluding remarks, limitations, and directions for future work: An overview of the main findings and work carried out is made. Limitations inherent to this study are explored and guidelines to mitigate future similar shortcomings are delineated.

2. Satisfaction in health

This chapter is divided into two sections. Section 2.1 presents the definition of satisfaction along with its evolution throughout the years. Section 2.2 conceptualizes patient satisfaction and quality in healthcare while trying to delineate a possible relationship between them.

2.1. Definition of satisfaction

Over the past few decades, given the customer-orientation philosophy and continuous improvement principles of modern enterprises, consumer satisfaction has gained recognition as a measure for quality in many public sector services (Ferreira, 2019). Customer satisfaction had its origins in the decade of 1960 but was only given the proper attention in the '80s. This concept is an ever-evolving one that keeps changing and improving, becoming more accurate each time. The first definitions of customer satisfaction emerged around 1960. At this time, there was still a lot to be unveiled about this concept. Cardozo (1965) stated that customer satisfaction with a product is influenced by the effort spent to acquire it and the expectations concerning it. Satisfaction with the product may be higher when customers spend considerable effort to obtain the product than when they use only modest effort.

In the '70s, the interest in customer satisfaction had its focus on developing previous theories. Conceptualizations of customer satisfaction started to appear, making room for different approaches and ideas. Lundstrom and Hunt (1978) mentioned that customer satisfaction corresponded to the judgement of the experience between cognitive processes and affective elements. Haines et al. (1970) defined satisfaction as '*the buyer's cognitive state of being either adequately rewarded or not for the sacrifice he has undergone'*. Customer satisfaction was still product-centred, and the definitions were not as complex as they should be since this concept was relatively recent and was not the main focus of researches.

In the '80s, satisfaction became one of the most important factors when studying consumer behaviour. Researchers were concerned with the contextualization of the satisfaction processes, while the companies aimed to know how to measure it. These studies focused on the operationalization of customer satisfaction and the respective precedents, being an example, the expectation-disconfirmation theory (Oliver, 1980). This theory suggests that expectations set a performance standard that will influence the actual judgement of the customer. Satisfaction is defined as, roughly speaking, the effect of expectations modified by the perceived disconfirmation. On the one hand, when the product exceeds the expected value for money, expectations are positively disconfirmed. On the other hand, when the product is inferior to the expected value for money, expectations are negatively disconfirmed. Finally, individual expectations of a consumer are confirmed when the perceived performance is at the same level as the expectations. This paradigm is based upon four pillars: expectations, performance, disconfirmation and satisfaction (Churchill et al.,1982).

In the mid '80s, the focus of investigation transferred to the importance of services' quality and to the implementation of strategies that would bring to life customer expectations (Zeithaml et al., 1996). The customer satisfaction definition used to be based on quality or expectations of a product but was now evolving into services as well. The evaluation of customer satisfaction concerned not only products sold in stores but also services surrounding it (Parasuraman et al., 1985). Aligned with these theories is Grönroos (1982), stating that when a service provider knows how the service will be evaluated by the consumer, a suggestion on how to influence these evaluations in the desired direction might be given. Other authors from this decade, including Westbrook and Reilly (1983), viewed satisfaction as the pleasant state of mind that occurs when a product, service or consumer action leads to the fulfilment of personal values.

In the ’90s, two different conceptualizations of customer satisfaction were developed: cumulative satisfaction and transaction-specific satisfaction. According to Johnson et al. (1995), service and satisfaction research had grown to include an emphasis on cumulative satisfaction, defined as a customer’s overall evaluation of a product or service provider to date. The second concept defined satisfaction as a customer’s evaluation of her or his experience with (and reactions to) a particular product transaction, episode, or service encounter. Cumulative evaluations leave the period of evaluation open and recognize that customers rely on their entire experience when forming intentions and making repurchase decisions. Thus, one advantage of cumulative evaluations is that predictions about customers’ intentions and behaviour are more accurate (Olsen and Johnson, 2003). One advantage of transaction-specific measures is that companies track, more efficiently, changes in performance that result from internal changes and/or quality improvements. In contrast, it takes time for quality changes to affect more cumulative evaluations (Johnson et al., 1995). Simpler approaches were also developed. According to Rust and Oliver (1994), customer satisfaction is an extent to which a person believes that an experience creates positive feelings. Oliver (1997), also defined customer satisfaction as the consumer’s fulfilment response. A judgment that a product or service feature, or the product or service, itself, provided (or is providing) a pleasurable level of consumption-related fulfilment, including levels of under-or over-fulfilment.

With the turn of the century, different conceptualizations emerged. For Evans et al. (2006), satisfaction/dissatisfaction is more than how well a product or service performs because it also involves the consumers’ attitude and feelings. For instance, anticipatory satisfaction may not be related to the actual performance of a product/service but rather to consumer’s imagined ideas about how the product is going to function.

2.2. Patient satisfaction and quality in healthcare

The incessant demand for improved results and quality of health services offered is of extreme importance in the development of a more effective organizational policy, adjusted to the needs of the patients. Health organizations recognize that service quality is especially pertinent regarding promotion and public image of the healthcare market (Qin and Prybutok, 2013). Hence, patient satisfaction surfaces as a conductor for promoting health organizations’ quality, allowing an assessment and identification of the most relevant dimensions for patients, as well as their satisfaction level (Castro and Portela, 2016). Patient satisfaction helps to measure the quality of healthcare, thus becoming an essential and frequently used indicator. It affects clinical outcomes, and medical malpractice claims, as well as timely, efficient, and patient-centred delivery of healthcare (Prakash, 2010). One difference between the characteristics of the consumer in the health sector or any other sector is that patients only consult health services in situations of mal-being and so their tolerance to errors is reduced (Duggirala et al., 2008). Maintaining the patients satisfied is essential so they will continue to be loyal consumers, comply with treatment, and have a stable relationship with providers (Hekkert et al., 2009). Patient satisfaction and quality of health services are a priority for the services industry due to the increasing consumption, and are critical elements for the long-term success of health institutions (Ramsaran, 2005; Kaya et al., 2020).

Even though satisfaction is an essential aspect of quality, the relationship between these two concepts is not linear. On the one hand, the results of satisfaction studies can be ambiguous and may not always be impartial. Given that patients evaluate physician's performance, for which most of them lack the necessary abilities, results can be based on affinity and not on the health professional technical skills. On the other hand, providers may have to face a trade-off between providing satisfaction to their patients or better treatment outcomes (Ferreira et al., 2018). Patient satisfaction is complex to assess, given its multidimensionality. It is composed of diverse aspects that may not be related to the actual quality of the service experienced by the patient.

To transcend the current lack of clarity in the literature on how satisfaction is defined and measured, it is valuable to consider the highly cited Donabedian's framework on how to examine health services and evaluate the quality of medical care using three components (Donabedian, 2005; Ayanian and Markel, 2016; Ferreira et al., 2018):

- Structure: environment, provider's skills and administrative systems where healthcare occurs;
- Process: the constituents of the received care (measures doctors and medical staff considered to deliver proper service);
- Outcome: the result of the care provided such as recovery, avoidable readmission, and survival;

The conceptualization of patient satisfaction regarding expectations and perceptions is related to Donabedian's triad. For instance, the patient will be satisfied with hospital attributes if his/her expectations are met (Ferreira et al., 2018). However, one of the leading critics of patient satisfaction ratings is the incapacity to rationalize medical care expectations, that can be affected by previous healthcare experiences (Schoenfelder et al., 2011). The same happens with the other two components. The patient will be satisfied with the process if symptoms are reduced, and the outcome will be favourable if there is a recovery, demonstrating that perception of received care has met the prior expectations. Throughout his framework, Donabedian regarded "outcome" as the most crucial aspect, defining it as a change in a patient's current and future health status that can be confidently attributed to antecedent care (Ferreira et al., 2018).

3. Healthcare in Portugal

This chapter is divided into two sections. Section 3.1. describes the structure and organisation of healthcare in Portugal. Section 3.2. provides examples of satisfaction studies that have been performed in the NHS to give a better understanding of the satisfaction Portuguese patients experience.

3.1. Structure and organisation

Over the past few years, the structure of healthcare in Portugal has suffered a reorganisation that aims to improve the care provided. Since 1971, the Portuguese health system has been defined by a public parcel. Movements towards centralization took place, giving the State more responsibility for managing and providing a hierarchical network of services. In 1976, the new Constitution created the NHS that has been at the leading edge of health protection for all Portuguese citizens (Nunes and Ferreira, 2019).

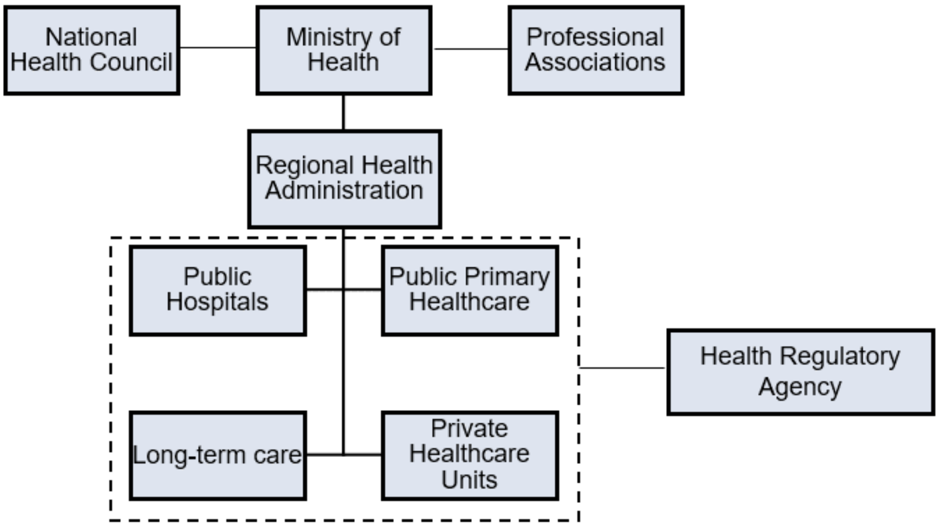


Figure 1. Overview of the health system.
Source: "Portugal: Health system review" by Simões et al. (2017)

Figure 1 shows how the healthcare system in Portugal consists of a network of public and private healthcare providers, where each one of them is connected to the Ministry of Health and patients (Simões et al., 2017). The three principal stakeholders in the Portuguese health system are: the State, that is the regulator and manager of the health system, while also acting as a provider for the NHS; the public (non-profit) sector, that has a pivotal role on society; and the private sector that focuses on providing a differentiated type of care, such as medical exams, outpatient consultations and hospitalisations. (Simões et al., 2020).

Constitutionally, the NHS control is centralized, and management is decentralized. It is managed at the regional level, being responsible for the health status of populations, provision of health services and appropriation of financial resources, depending on the population needs. Responsibility for planning and resource allocation in the Portuguese health system, at all levels, is centralized. With a centralized structure, the NHS is dominated by public orientation both in prevention and hospitalization areas (Pinho, 2009; Simões et al., 2017).

On the one hand, some of the central services that are under the government's direct administration are:

- Secretariat-general: reinforces the connection between society and the provided services, assuring, as well, an institutional articulation amongst the organisms. Offers juridical support, manages internal resources, human resources and public relations (Ministério da Saúde, 2019);
- Directorate-general of health: plans, regulates and supervises all health promotion, disease prevention and healthcare activities, institutions and services. It secures a level of quality and equal access to care providers by ensuring the execution of health policies (Ministério da Saúde, 2016);
- Inspectorate-general of health-related activities: performs audits, supervises, and implements disciplinary functions in the health sector (Ministério da Saúde, 2016);

On the other hand, some of the central services that are under the government's indirect administration are:

- Regional Health Administration (RHA): primary and secondary health care management is decentralized through five (North, Center, Lisbon and Tagus Valley, Alentejo, and Algarve) RHAs, which were introduced in 2003. Each RHA has a board of managers that supervises the delivered health care services, assures cooperation and collaboration with the private, handles secondary health care providers, and ensures the achievement of the annual financing process regarding every primary health care centre (Nunes et al., 2019);
- Central administration of the health system: it is in charge of assuring management of financial and human resources, facilities and equipment of the NHS. In addition to this, is also responsible for implementing policies, regulations and health planning, together with RHAs (Ministério da Saúde, 2020);

As already mentioned, Portugal integrates a group of countries where public (primary and hospital care) and private (pharmaceuticals, complementary diagnostics and treatments, and medical consultations) providers coexist. Essentially, there are two kinds of relations: one of complementary nature, where citizens resort to private providers upon pre-established agreements with public services, and one of replacement nature, where citizens visit private providers instead of public entities and costs are either entailed by the patient or reimbursed through a health subsystem or private insurance (Fernandes and Nunes, 2016; Nunes and Ferreira, 2019a).

Portugal's health system is characterized by a universal (served population) and general (offered specialities) coverage, mainly constituted by public entities, integrating funding and provision divided through three fundamental segments: primary healthcare, hospital care, and integrated long-term care. The primary health care segment is distributed by the five RHAs, focusing on health promotion, disease prevention and treatment of severe health conditions. Hospital care offers a more specialized secondary health care service dispersed through the entire country, depending on populational needs and number of available medical professionals. Long-term care is subsidized by the public sector and provided by the social sector, aimed at people who have loss autonomy, require palliative care, or are victims of incapacitating conditions (Fernandes and Nunes, 2016; Nunes and Ferreira, 2019b).

In line with this, the Health Regulatory Agency (HRA) was created in 2003, as an independent public entity, whose mission is to regulate the activities of healthcare providers. Englobed in this concept are public care, public primary healthcare, long-term care and private healthcare units. Its functions include the defence of patients' rights, inspections of healthcare facilities, and supervision of studies regarding the organisation of the NHS, health satisfaction and claim management. HRA's goal is to secure competition between health care providers while protecting citizens' right to complete health care coverage (Ministério da Saúde, 2016; Simões et al., 2017).

Concerning funding, the Portuguese health system did not suffer major modifications after the publication of the Health Bases Law, in 1990. Financial resources are both public and private. However, the State is the main funder, responsible for financing 70% of the total health expenditures in Portugal, according to the Beveridge model (Aarrevaara et al., 2015; Nunes and Ferreira, 2019b).

3.2. Satisfaction in the Portuguese NHS

The assessment of patients' satisfaction is done regularly through NHS's around the world allowing the measurement of health service standards. In Portugal, since the 2005 reform, patient's observations are an important aspect in NHS's evaluation (Santos, 2018). In March 2015, a study was carried out by the directorate-general of health, specifically the department of quality in health, to estimate patient's satisfaction.

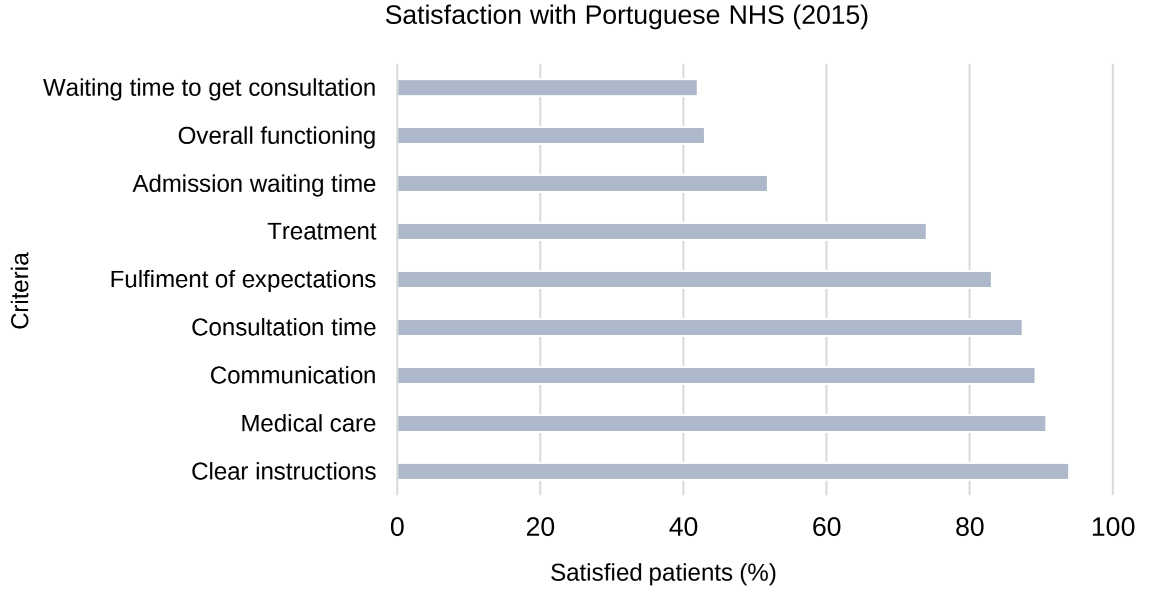


Figure 2. Satisfaction with Portuguese NHS. Source: "Estudo de satisfação dos utentes do Sistema De Saúde Português" by Department of quality in health (2015).

The results from Figure 2 show that only 43% of the patients surveyed are satisfied with the service's overall functioning. However, there are some criteria where the NHS performance keeps most patients satisfied, being them, *clear instructions, medical care,* and *communication*. The criteria that need higher improvements are *waiting time to get a consultation* and *admission waiting time*.

Other values were also assessed to evaluate NHS performance. 38.6% of the patients surveyed consider necessary a change in the health system, both in public and private sectors. These values have been increasing year after year, which can be interpreted as a lack of negative factors regarding the perceptions of the provided service (Departamento de Qualidade na Saúde [DGS], 2015).

In 2018, a study was performed using data collected by the official source of the Ministry of Health, concerning 62 different hospitals. Seven different criteria (*hospital image, admission process, facilities quality, doctors, nurses, diagnosis and treatment* and *waiting time after admission*), divided into multiple subcriteria were assessed (Ferreira et al., 2018).

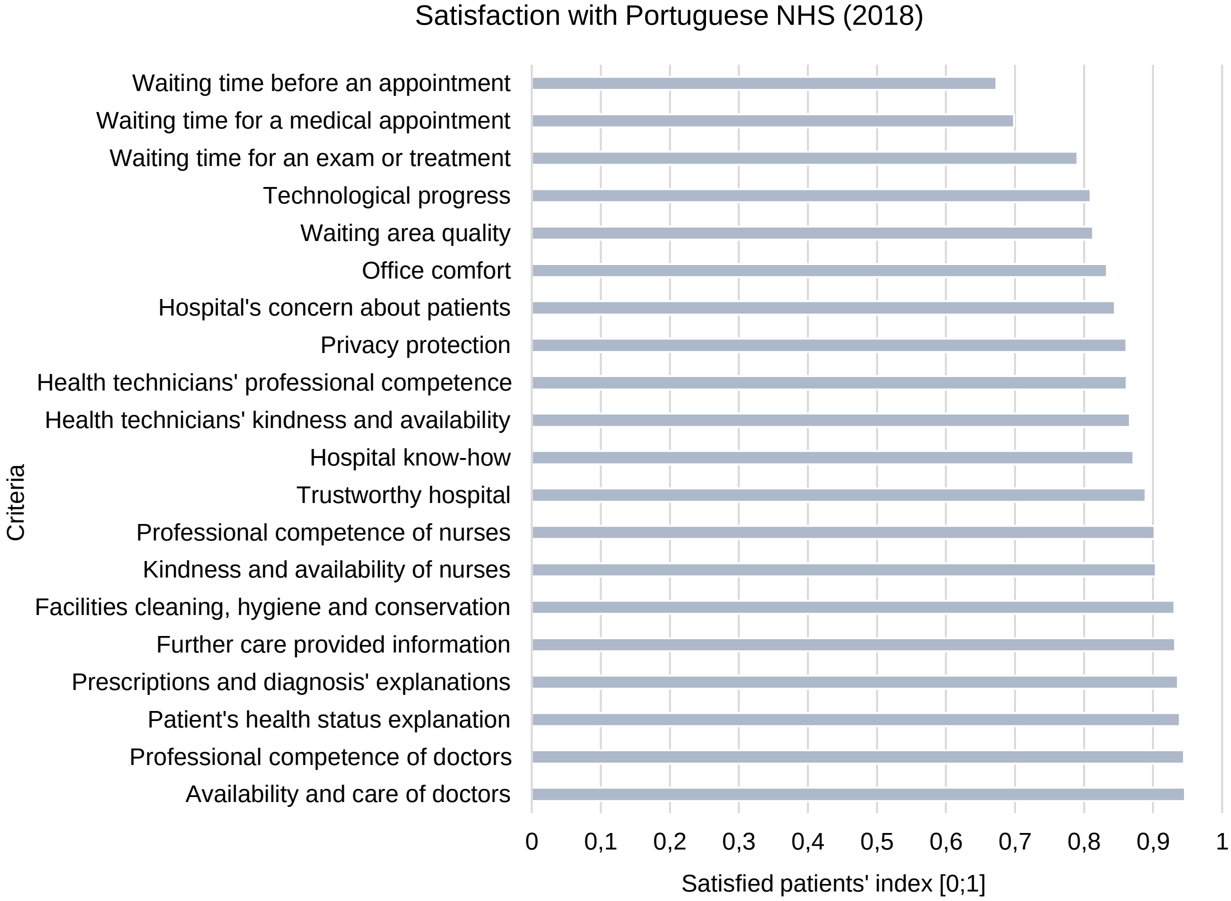


Figure 3. Satisfaction with Portuguese NHS. Source: "Patients' satisfaction: The medical appointments valence in Portuguese public hospitals" by Ferreira et al. (2018)

When observing Figure 3, it can be concluded that *doctor's availability, care and competence* are the criteria in which patients are most satisfied. *Waiting time* is shown, once again, to be the biggest problem, whether *to get a consultation* or *an exam*. Overall the level of patient satisfaction can be considered high since the criterion with the lowest index has a value of 0.6733. When comparing results from both studies, it can be observed that *waiting time* is the main reason for dissatisfaction with the Portuguese NHS. *Medical care* is present, in both studies, as one of the criteria with the highest value. From 2015 to 2018 there is an increase in patient satisfaction since the lowest value (in each study) goes from 42% (*waiting time to get a consultation*) to 67.33% (*waiting time before an appointment*), respectively. The same happens with the highest value criteria. In 2015, the value is 93.9 % (*clear instructions*) and in 2018, the value is 94.6% (*availability and care of doctors).* It is necessary to keep in mind that, until this point, Portuguese patients' satisfaction studies have only been held on a national level. This means that the results presented are calculated averages that may not translate the reality and priorities of each health entity.

4. Literature review

This chapter presents a comparative study of the existing literature, intending to evaluate which factors are more commonly acknowledged as the ones affecting patient satisfaction. It is an overview of an article submitted to an ISI peer-review journal that is awaiting revision.

4.1. Data collection and extraction

This research was performed respecting the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement. A checklist of 27 parameters, including the title, abstract, methods, results, discussion, and funding, was taken into consideration to ensure complete reporting of systematic reviews. Table A.1. (Appendix A) contains the full list of parameters considered in that checklist. PRISMA assures that authors prepare a transparent and complete reporting of systematic reviews and meta-analyses (Liberati et al., 2009). The PRISMA statement starts with the identification of possible studies to include in the revision, after searching in multiple databases. Papers were searched in the Scopus, Web of Science, and PubMed databases during June 2020. After testing several keywords, the search strategy used the term “patient satisfaction” to extend the number of results. Reference lists from the collected articles were also searched for additional articles. Overall, one thousand six hundred fifty-three studies composed the list of our first search. Firstly, the duplicates (241) were removed, and the remaining documents (1412) were analysed under the inclusion and exclusion criteria. Inclusion criteria included: articles from peer-review journals; written in English; published from January 2000 to June 2020; assessed which factors affect patient satisfaction (or a proxy of it); evaluated overall patient satisfaction with healthcare; quantitative studies; reviews; international studies to provide a more comprehensive analysis. Reports, books or book chapters, conference proceedings, dissertations, thesis, expert opinion, commentaries, editorials, and letters were not included. In total, 1197 studies were excluded from the list after removing duplicates because they fail all inclusion criteria or meet at least one exclusion criterion. The full-text analysis was conducted to assess the eligibility of the remaining 215 papers. Disease centred studies that did not evaluate the general aspects of patient satisfaction were excluded. Papers with unclear data collection methods, papers with unclear results, and qualitative papers were also discarded. A total of 62 papers were rejected in this step. Figure 4 outlines the PRISMA diagram detailing the study selection process. Following such a statement, 153 studies met the inclusion criteria.

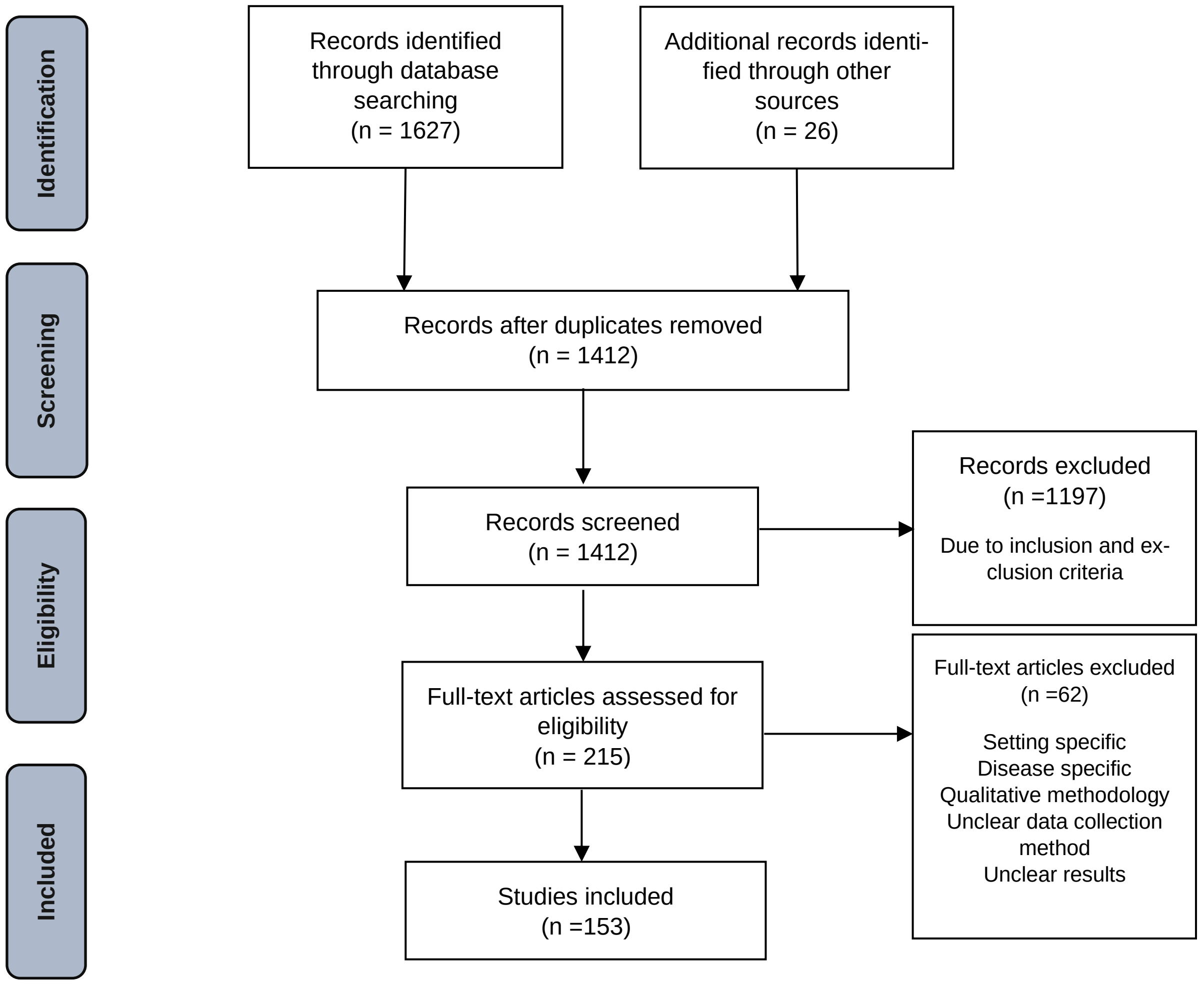


Figure 4.PRISMA statement.

4.2. A summary of studies passing the PRISMA sieve

Authors, country of study, sample, quality dimensions and drivers, methodology, dependent variable, and main factors affecting the satisfaction of each study were assessed to reach a conclusion. A dependent variable field was considered because, besides overall patient satisfaction, some articles studied proxies of satisfaction, such as willingness to recommend the hospital or willingness to return. To simplify the graphic display of information, a complete table is included in an online appendix (Appendix B[1]), whereas Table 1 presents an excerpt containing the most impactful articles (with more than 100 citations). Some past systematic reviews have revealed that interpersonal or social skills (like medical/nursing care and attitudes), technical skills, infrastructure and amenities, accommodations, environment, accessibility, continuity of care, and the outcome are the satisfaction criteria present in the majority studies related to satisfaction in healthcare (Naidu, 2009; Almeida et al., 2015; Farzianpour et al., 2015; Batbaatar et al., 2017; Ng and Luk, 2019).

[1] *Appendix B is inserted in a Google Driver folder with the following link: https://drive.google.com/file/d/13csUIwdIjXDwoElzSXvtd0YIeXcgsCia/view?usp=sharing*

Table 1. Review of remarkable studies.

Authors	Country of study	Sample	Quality dimensions & drivers	Methodology	Dependent variable	Main factors affecting satisfaction
(Aiken et al., 2012)	13 European countries and the USA;	61,168 surveys from nurses in 488 European hospitals and 617 USA hospitals, and 131,318 surveys from patients in 210 Europeans hospitals and 430 USA hospitals;	Nursing care; environment; burnout; dissatisfaction; intention to leave the job; patient safety; nursing care;	OLR; robust logistic regression with cluster; p-value;	Overall patient satisfaction with nursing care; Willingness to recommend hospital;	− nursing care; − environment;
(Bleich et al., 2009)	21 European countries;	33,734 surveys from 21 European countries;	Service experience; fulfilment of expectations; perceived health status; patient's personality;	Linear regression; p-value; r-square;	Overall patient satisfaction;	− fulfilment of expectations; − provider type; − insurance;
(Jackson et al., 2001)	USA	500 surveys from patients of 38 different physicians;	Fulfilment of expectations; information about symptoms duration; information about symptom resolution; patient's age; patient's autonomy;	Kruskal–Wallis test; p-value; maximum likelihood ratio; chi-square; multivariate regression analysis;	Overall patient satisfaction;	− fulfilment of expectations; − information provided; − patient's autonomy;
(Aldana et al., 2001)	Bangladesh	1913 surveys from a public hospital;	Working hours; waiting time; medical care; doctor's attitudes; appointment's duration; privacy; physical examination; information provided; advice given by doctor;	Multivariate regression analysis; p-value;	Overall patient satisfaction;	− appointment's duration; − privacy; − physical examination; − information provided; − the advice given by the doctor;
(Jenkinson et al., 2002)	Scotland	2249 surveys from five public hospitals;	Patient's age; patient's gender; health status; patient's education; coordination of care; comfort; emotional support; respect for patient's preferences; involvement of family; continuity of care;	Spearmen correlation coefficient (SCC); p-value; multiple linear regression; r-square;	Overall patient satisfaction;	− comfort; − emotional support; − respect for patient's preferences;

Authors	Country of study	Sample	Quality dimensions & drivers	Methodology	Dependent variable	Main factors affecting satisfaction
(Andaleeb, 2001)	Bangla-desh	216 surveys from 57 hospitals and clinics;	Ability to answer questions; doctor's assurance; nurse's assurance; staff's assurance; communication with the patient; baksheesh; doctor's attitudes;	Factor analysis; varimax rotation; Cronbach's alpha; p-value; multiple regression analysis; r-square;	Overall patient satisfaction;	− doctor's attitudes; − doctor's assurance; − nurse's assurance; − staff's assurance; − ability to answer questions; − communication with the patient;
(Westaway et al., 2003)	South Africa	263 surveys from diabetic outpatients in two public hospitals;	Doctor's kindness; doctor's encouragement; doctor's attitude; doctor's ability to listen; doctors are supportive; doctors ability to answer questions; information provided; medical skills; information provided; maintenance of contact; follow up care; fair treatment; waiting time; availability of seat on waiting area; cleanliness; privacy;	Factor analysis; varimax rotation; Cronbach's alpha; t-tests; p-value; Pearson's correlation coefficient (PCC); analysis of variance (ANOVA); Kaiser-Meyer-Olkin (KMO);	Overall patient satisfaction;	− medical care; − doctor's attitudes; − doctor's kindness; − medical skills; − information provided; − doctor's ability to answer questions; − cleanliness;
(Nguyen Thi et al., 2002)	France	533 surveys from 12 medical services at a university hospital;	Admission process; nursing care; medical care; information; hospital environment; overall quality of care; recommendations; patient's age; patient's gender; distance to the hospital; community size; patient's BMI index; patient's Karnofsky index; assistance needed at the hospital; patient's autonomy; length of stay; attitude towards the length of stay; privacy;	OLR; p-value; t-tests; PCC; chi-square;	Overall patient satisfaction;	− patient's age; − perceived health status; − admission process; − patient's autonomy; − privacy; − length of stay;
(Kitapci et al., 2014)	Turkey	369 surveys from one public hospital;	Accommodations; communication with the patient; empathy; skills; ability to answer questions;	Service quality (SERVQUAL); Cronbach's alpha; p-value; average	Overall patient satisfaction;	− kindness; − skills;

Authors	Country of study	Sample	Quality dimensions & drivers	Methodology	Dependent variable	Main factors affecting satisfaction
				variance extracted; composite assurance; factor analysis; varimax rotation; SEM;		
(Sun et al., 2000)	USA	5232 surveys from the emergency department;	Communication with family; waiting time; received help when needed; identification of health professionals; discharge process; information about return to the emergency department; signs of being aware regarding illness; side effects; provision of medication; take medication as advised; results of medical exams; follow-up appointment; information provided; doctor's attitudes;	Cronbach's alpha; SCC; p-value; OLR;	Overall patient satisfaction; Willingness to return;	− doctor's attitudes; − waiting time; − information provided; − results of medical exams;
(Fan et al., 2005)	USA	21,689 surveys from seven Veterans Affairs (VA) medical centres;	Patient's age; patient's gender; marital status; patient's education; income; occupation; health status; received care outside VA; primary care visit in previous 12 months; distance from clinic; clinic site; provider type; provider's gender; continuity of care;	t-test; Wilcoxon rank-sum test; chi-square; multivariate linear regression; Cronbach's alpha; r-square; p-value;	Overall patient satisfaction;	− continuity of care;
(Hekkert et al., 2009)	Netherlands	66,611 surveys from 8 university hospitals and 14 general hospitals;	Patient's gender; patient's age; patient's education; health status; hospital type; hospital size; population density; admission process; nursing care; medical care; communication with the patient; patient autonomy; discharge process;	Multilevel analysis; intra-class correlation coefficient (ICC); chi-square;	Overall patient satisfaction;	− patient's age; − self-perceived health status; − patient's education; − admission process; − nursing care; − medical care; − communication with the patient; − patient autonomy; − discharge process;

Authors	Country of study	Sample	Quality dimensions & drivers	Methodology	Dependent variable	Main factors affecting satisfaction
(Renzi et al., 2001)	Italy	396 surveys from the dermatology department;	Patient's age; patient's gender; patient's education level; region of residence; duration of disease; illness impact; quality of life, regarding emotions; quality of life, regarding symptoms; quality of life, regarding functioning; medical care; the accuracy of dermatological visit; doctor's ability to listen; concern for questions; appointment duration; information provided;	Principal components analysis (PCA); p-value; multiple logistic regression;	Overall patient satisfaction;	− information provided; − doctor's attitudes; − patient's age; − illness impact;
(Boudreaux et al., 2000)	USA	437 surveys from the emergency department of a municipal hospital;	Patient's age; patient's gender; patient's race; insurance; priority code; visit-time of the day; day of the week; disposition; reception courtesy; reception helpfulness; privacy; nursing care; information about treatment provided by nurses; nurses' skills; information about condition provided by doctors; medical exams explanation provided by doctors; next steps explained by doctors; follow-up instructions; discharge instructions; X-ray staff courtesy; staff care; communication with the family;	t-tests; chi-square; Mann-Whitney U tests; p-value; univariate and multivariate analysis;	Overall patient satisfaction; Willingness to recommend hospital;	− staff care; − safety; − follow-up instructions; − nurse's skills; − waiting time; − patient's age (solely for willingness to recommend hospital); − insurance (solely for willingness to recommend hospital);
(Shilling et al., 2003)	United Kingdom	1816 surveys from the oncology department;	Patient's age; physician's age; patient's gender; physician's gender; patient's physiological morbidity; waiting time; tumour site; type of treatment;	PCA; varimax rotation; Mann-Whitney U test; Kruskal-Wallis; ANOVA; Bonferroni coefficient; p-value;	Overall patient satisfaction;	− waiting time; − patient's age; − the patient's physiological morbidity;

Authors	Country of study	Sample	Quality dimensions & drivers	Methodology	Dependent variable	Main factors affecting satisfaction
(Rahmqvist and Bara, 2010)	Sweden	7245 surveys;	Patient's age; patient's gender; self-perceived health status; the origin of birth; patient's education; living area; living condition; fulfilment of expectations; medical care; waiting time; patients' participation in making decisions about treatment;	Chi-square; PCC; Fisher's exact probability test;	Overall patient satisfaction;	− patient's age; − patient's education; − self-perceived health status; − patient's nationality; − fulfilment of expectations; − medical care; − waiting time; − patients' participation in making decisions about treatment;
(Bjertnaes et al., 2012)	Norway	10,912 surveys from 63 hospitals;	Fulfilment of expectations; nursing care; medical care; incorrect treatment; health personnel in general; organization; waiting time; pain relief; communication with the patient; next of kind – handling; medical equipment; patient demographics;	Test-retest assurance; ICC; Cronbach's alpha; PCC; multivariate linear regression analysis; multilevel linear regression analysis; p-value;	Overall patient satisfaction;	− nursing care; − fulfilment of expectations; − medical care; − perceived incorrect treatment;
(Beattie et al., 2002)	USA	1868 surveys from private outpatient physical therapy clinics;	Therapist's ability to answer questions; therapist's ability to listen; therapist's kindness; appointment's duration; information provided; staff's kindness; cleanliness; medical equipment; working hours; the complexity of registration; waiting area; parking; waiting time; location;	Inter-item correlation; p-value; multiple regression analysis; r-square; Cronbach's alpha; chi-square; PCA; oblimin rotation;	Overall patient satisfaction;	− appointment's duration; − information provided;
(Schoenfelder et al.,2011b)	Germany	8428 surveys from 39 hospitals;	Fulfilment of expectations; outcome; the kindness of the nurses; the kindness of the doctors; organization of procedures and operations; quality of food; accommodation; medical care; discharge process; physician's	PCA; Cronbach's alpha; non-parametric Kruskal–Wallis test; p-value; chi-square; Fisher's exact test;	Overall patient satisfaction;	− outcome; − the kindness of nurses; − the kindness of doctors; − the organisation of procedures and operations; − quality of food;

Authors	Country of study	Sample	Quality dimensions & drivers	Methodology	Dependent variable	Main factors affecting satisfaction
			knowledge of patient anamnesis; admission process; communication with the patient; cleanliness;	logistic regression analysis;		− accommodations; − medical care; − discharge process; − physician's knowledge of patient anamnesis; − admission process;

Regarding the dependent variables of the collected articles, in general, studies about patient satisfaction try to unveil factors associated with his/her overall satisfaction with one or more services (96% of the collected studies) or willingness to recommend the hospital/clinic (9%), instead. A smaller percentage of studies (7%) included both dependent variables (Boudreaux et al., 2000; Cheng et al., 2003; Otani and Kurz, 2004; Brown et al., 2005; Haase et al., 2006; Chahal, 2010; Otani et al., 2011; Aiken et al., 2012; Otani et al., 2012). It means that there is one dependent variable (typically the overall satisfaction) explained by a series of criteria and other external factors. However, one can also use other dependent dimensions as proxies for such overall satisfaction. Examples include the willingness to return (Sun et al., 2000; Brown et al., 2005; Singh et al., 2016), medical services satisfaction, accommodations services satisfaction, and nursing services satisfaction (Matis et al., 2009), satisfaction with the quality of medical information (Soufi et al., 2010), and healthcare quality (Widayati et al., 2017). The statistical analysis applied to these data focuses on the global scenario, as every dependent variable is accounted for. To provide a more unambiguous graphic representation of the analysis, some of the patient satisfaction related factors were grouped into a single factor. These are some examples: (*i*) concern (from the doctor, the nurse, or other staff, either clinical or not); (*ii*) clinical staff social characteristics (assurance, attention, attitudes, kindness, skills, and speciality); (*iii*) hospital characteristics (image, location, quality, size, and type); and (*iv*) patient's social characteristics (autonomy, dignity, emotional support, income, life expectancy, marital status, nationality, occupation, race, residence, satisfaction with life, and stress level).

4.3. A statistical overview of data

To provide a brief insight into the entire literature review, Table 2 contains statistical measures applied to the collection of 153 studies.

Table 2. Statistical measures applied to the data collected.

	Sample size	No. Methods	No. Criteria	No. Explanatory variables	No. Critical factors
Mean	18640	1	8	3	4
Median	728	1	6	3	4
Mode	200	1	6	0	3
Standard deviation	84507	0,39	6	3	2
Coefficient of variation	453%	33%	73%	100%	59%
Minimum	37	1	0	0	1
Maximum	934800	2	26	14	13

The sample size, the first object of analysis, shows a significant coefficient of variance due to the dispersion of the values, as can be seen through the minimum and maximum values. Regarding the methodology used, most studies applied only one method. However, some studies used two methods in a complementary way. The factors influencing satisfaction were divided into two groups. Criteria and explanatory variables. The number of criteria used to assess patient satisfaction has a low variance. Researchers give more importance to criteria than to explanatory variables translated through a higher mean, median, and mode of criteria, meaning that researchers tend to disregard the vital aspect of satisfaction drivers. The number of both criteria and explanatory variables have a minimum value of zero, meaning that there are studies that assess only the importance of one of these factors. The number of critical factors has low variance, and the minimum is equal to one since each study seeks to find out the determinants of patient satisfaction.

4.4. Global analysis over the utilization and influence of satisfaction criteria and explanatory variables

From Table B.1. (Appendix B), the utilization and influence of each factor related to patient satisfaction were analysed. The percentage of utilization is the ratio between the number of studies using it and the total number of evaluated studies. The influence rate of a factor measures the relative number of papers concluding that this factor is critical for patient satisfaction.

All factors related to patient satisfaction were analysed and clustered in terms of satisfaction criteria, and explanatory variables, regardless of the dependent variable used by researchers. The “quality dimensions and drivers” section of Table B.1. was reviewed to analyse the most utilized factors. The fifteen most utilized factors were divided into criteria and explanatory variables; Figure 5 and Figure 6 represent them, respectively. These factors are the ones that most researchers use to study patient satisfaction and may not correspond to the most important and influential factors of patient satisfaction.

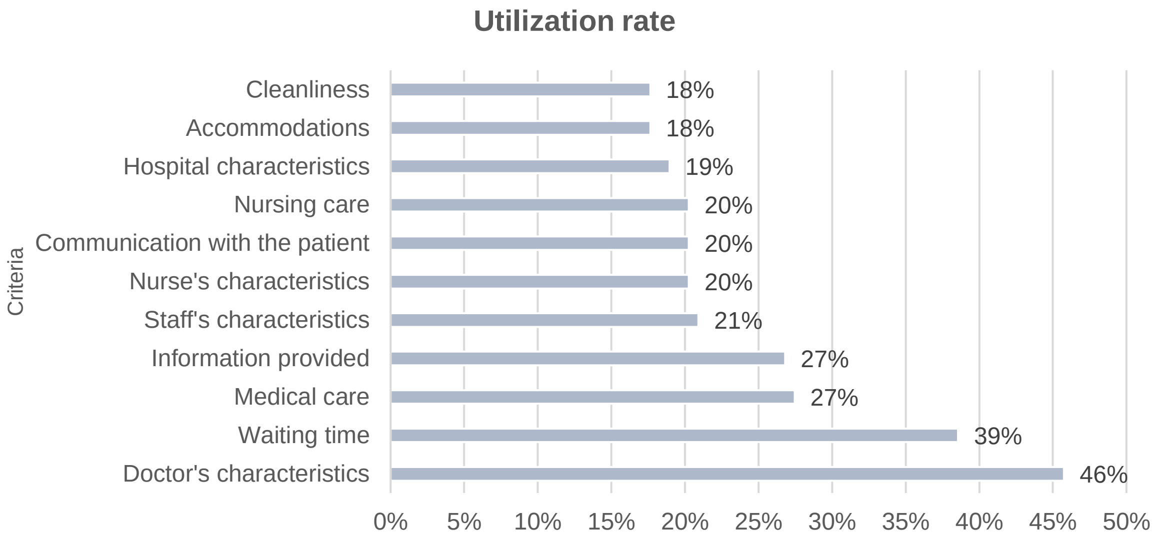


Figure 5.Analysis of the most utilized criteria in literature. Source: The author.

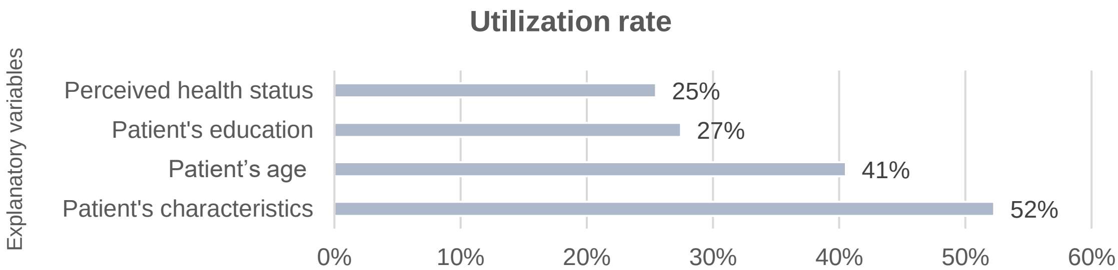


Figure 6. Analysis of the most utilized explanatory variables in literature. Source: The author.

From the fifteen most used factors, eleven are criteria, and four are explanatory variables. On the one hand, doctor’s characteristics, waiting time, medical care, and information provided have the highest utilization rates within the criteria. On the other hand, patient’s social characteristics, patient’s age, patient’s education, and perceived health status also have the highest utilization rates within explanatory variables.

Figure 7 ranks criteria deemed as the most influential of patient satisfaction. In contrast, Figure 8 presents the most influential explanatory nondiscretionary dimensions. This analysis resulted in fifty-six factors, divided into forty-seven criteria and nine explanatory variables.

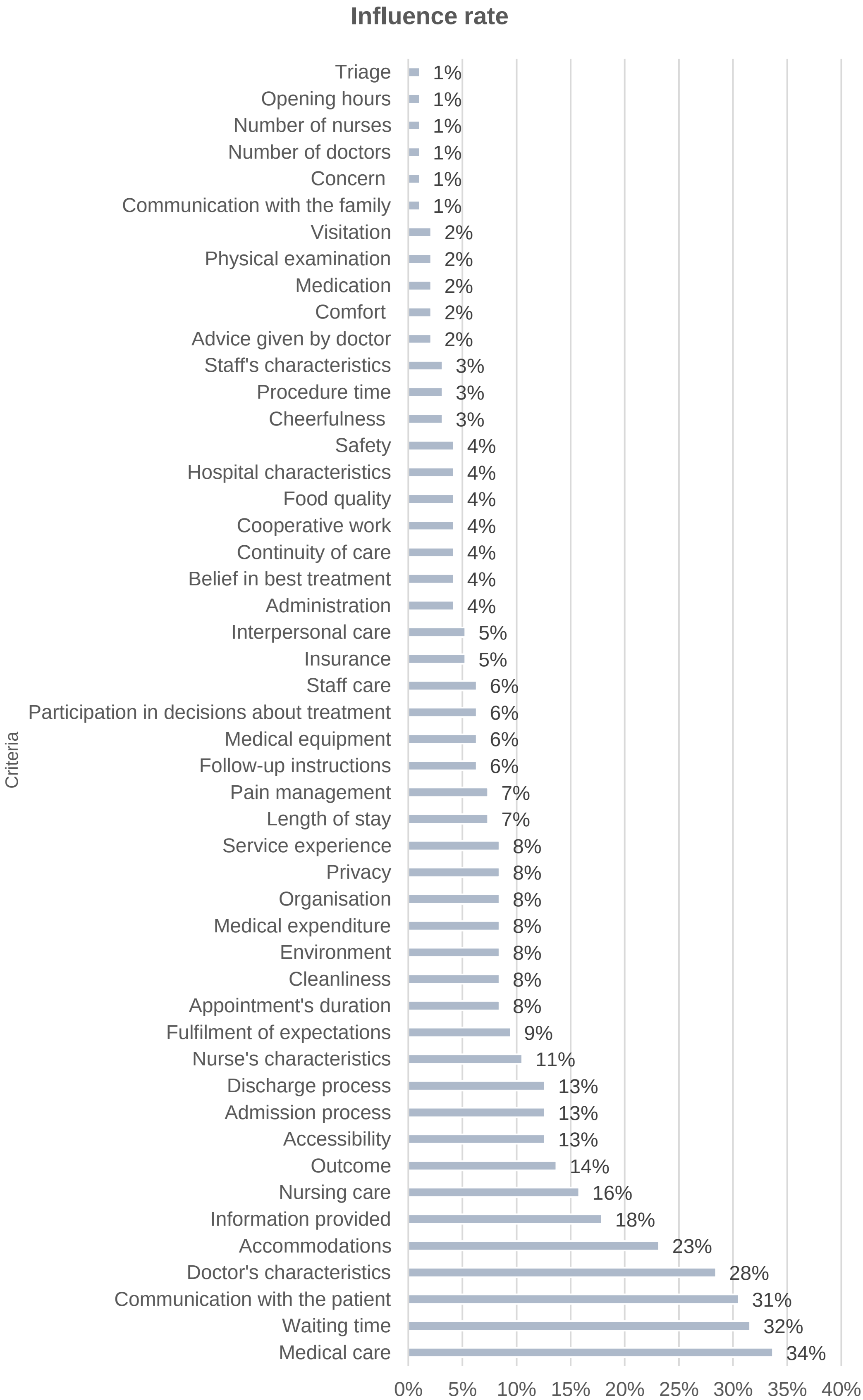


Figure 7. Analysis of criteria deemed as the most critical in literature. Source: The author.

From Figure 7, it is possible to conclude that the three most influential criteria are medical care, waiting time, and communication with the patient. Despite not being on the top three, criteria related to doctor's social skills exhibit a high importance rate and should be noticed as well. It is interesting to note that researchers tend to conclude that criteria related to social skills of staff, such as communication, are more important than others, for instance, food quality and comfort. Also, criteria associated with the technical skills of staff appear to be less critical. It seems to be in line with some authors claiming that patients are usually unable to judge health professionals in those terms (Elixhauser et al., 2003; Rogers et al., 2004; Levin-Scherz et al., 2006; Needleman et al., 2006). Additionally, waiting time is one of the most critical criteria to study patient satisfaction. For instance, Ferreira et al. (2018) classified this criterion as a critical must-be requirement. It means that patients take it for granted and neither get satisfied nor dissatisfied if the waiting time is null. However, their dissatisfaction intensely increases when waiting time becomes more substantial. The authors also verified that waiting time was the most crucial criterion for patients in medical appointment services (Vieira et al., 2020).

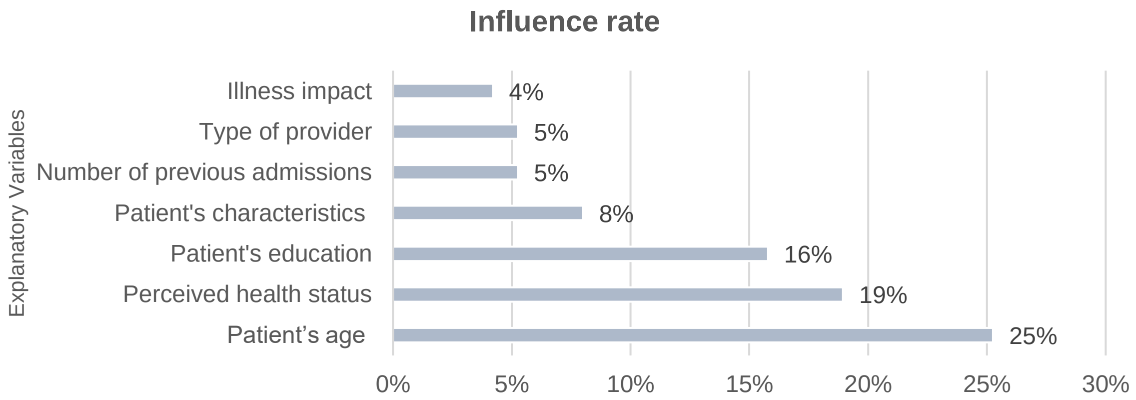


Figure 8. Analysis of explanatory variables deemed as the most critical in literature. Source: The author.

From Figure 8, patient's age, perceived health status, and patient's education are the variables that researchers tend to consider as the most influential. The conclusions from previous studies saying that age, education, and self-reported health status have an evident and significant influence on the satisfaction outcomes were confirmed (Hekkert et al., 2009). Older patients or the ones with better self-perceived health status are typically more satisfied, while highly educated people are less satisfied with the healthcare services provided (Nguyen Thi et al., 2002; Rahmqvist and Bara, 2010).

Comparing results from the utilization analysis and the influence analysis, differences arise. Figure 5 and Figure 7, both portraying criteria, resonate differences in the ranking positions. Doctor's characteristics, the most utilized criterion, was placed fourth on the importance-related ranking. Communication with the patient also occupies different positions in the analysis. Figure 5 shows this criterion in the seventh position, while in Figure 7, it is the third criterion with the highest importance rate.

Regarding the explanatory variables, Figure 6 and Figure 8 also display disparities. Patient's social characteristics are the most used explanatory variables, but it occupies the fourth position on the influence analysis. Patient's age has the second-highest utilization rate and the highest influence rate. Patient's education occupies the third position in both analyses. At last, perceived health status is ranked fourth in Figure 6, but secondly in Figure 8.

As stated on the introduction of this section, past reviews (Naidu, 2009; Almeida et al., 2015; Farzianpour et al., 2015; Batbaatar et al., 2017) acknowledge interpersonal or social skills (like medical/nursing care and attitudes), technical skills, infrastructure and amenities, accommodations, environment, accessibility, continuity of care, and the outcome as the most important factors. In terms of explanatory variables, these reviews also point out the frequent use of variables like the patient's gender, age, education, and marital status. Despite the similarity of results between previous studies and this bibliographic review, some factors seem to occupy a place of relevancy not seen before. Waiting time and information provided are not present on previous reviews. On the one hand, waiting time is a determinant of dissatisfaction in healthcare, regardless of the stage in which the inpatient is. Waiting time and waiting lists are frequently seen as barriers to access. Meanwhile, efficient hospitals usually have short waiting times (Sofaer and Firminger, 2005). The longer the waiting time, the more dissatisfied the customer is (Davis and Maggard, 1990). However, the converse is not necessarily true. If the waiting time is very short or even null, the customer may take it for granted because he/she needs the medical/nursing act and get neither satisfied nor dissatisfied. It means that waiting time is usually pointed out as a must-be requirement (Ferreira et al., 2018). On the other hand, the criterion information provided may refer to any process of care since the patient enters the system until he/she leaves it. For instance, the inadequate post-discharge care and lack of patients' preparedness are two potential determinants of readmissions for further care (Benbassat and Taragin, 2000). Missing or confusing information provided by the clinical staff contributes to the lack of preparedness and, by consequence, to customer dissatisfaction. The fact that this criterion does not appear in previous reviews is perhaps the result of a merging of some criteria related to it. However, the need for high discrimination of criteria during a satisfaction survey is pointed out (Vieira et al., 2020).

5. Methodology

In this chapter, the methodology in which the project is developed is presented. A description of the methods used in this dissertation is provided to give a better understanding of the mathematical assumptions that underlie each of them.

5.1. Methodology present in the literature review

Figure 9 provides a chart comparison of the different methods used in the literature devoted to the patient's satisfaction analysis. Four main methods were identified: regression analysis, factor analysis, SEM, and MUSA.

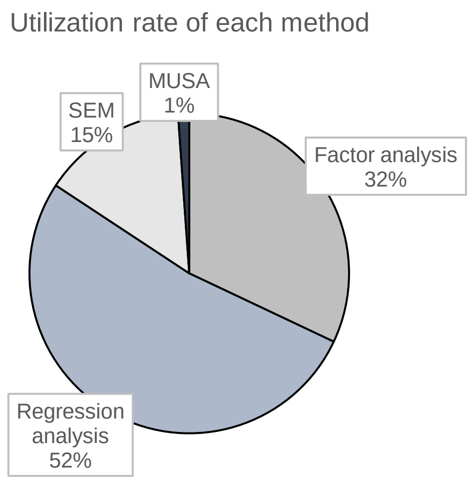


Figure 9. Analysis of methods utilized in the literature. Source: The author.

It can be observed that regression analysis is the methodology chosen by most researchers (52%). It is important to mention that, to simplify this comparison analysis, six different types of regression analysis were clustered into one big group. Out of the 52% of studies that apply regression analysis, 31% use multivariate regression analysis, 29% employ OLR, 24% utilize linear regression, 11% use multiple regression analysis, 3% use multilevel analysis, and 2% utilize stepwise regression. As can be seen, there are many regression methods available and present in the literature. However, due to the characteristics of our sample, something that is further discussed, OLR is the method that will be implemented. Factor analysis comes in second place, with a 32% utilization rate, followed by SEM (15%), and at last, MUSA (1%). From the 153 collected articles, 27 (18%) combined different methods in a complementary nature: factor analysis with regression analysis (16 of the 27 articles, or 59%), and factor analysis with SEM (11 of the 27 articles, or 41%). The difference in the level of utilization of each method can be due to the difficulty of implementation. SEM and MUSA are more complex than the other two, and thus harder to implement. Logistic regression and factor analysis are simpler and easier to implement, becoming more attractive to the researcher. Each of the four methods has different advantages and disadvantages that are explained in the following sections. These differences must be considered when pondering which method to use since different methods can deliver different results. For this specific project, the combination of the four methods is used both on a complementary and comparative nature. Factor analysis is the first method to be applied to the entire sample. It is employed in a complementary nature to eliminate redundancy that might exist. Subsequently, SEM, OLR, and MUSA are implemented, and the results of each method are compared, as seen in Figure 10.

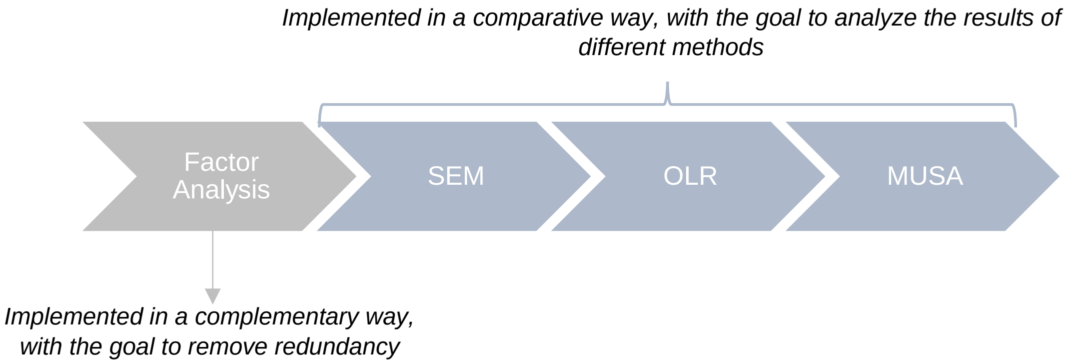


Figure 10. Sequencing of the different methods.

5.2. Factor Analysis

Factor analysis is a mathematical model that explains the correlation between a large set of variables in terms of a small number of underlying factors (Mardia et al.,1994), using procedures that summarize information included in a data matrix and replacing original variables by a small number of composite variables or factors. The usage of factor analysis has the following objectives (Vilares and Coelho, 2011):

- To identify and interpret subjacent dimensions that explain the correlations between the original set of variables;
- To identify a new and reduced set of non-correlated variables that replace the original variables in subsequent multivariate analysis;
- To select a small set of variables, from a greater set, to use in subsequent multivariate analysis;

The goal of factor analysis is to diminish the dimensionality of the original space and to interpret the resulting space, covered by a reduced number of variables that dominate the previous ones (Rietveld and Van Hout, 1993). The realistic goal, however, is to obtain a parsimonious solution that grants a close approximation to reality. Thus, the hypothesis of perfect fit is not empirically interesting (Fabrigar et al.,1999). Factor analysis removes redundancy and should be executed, for instance, before implementing SEM. When performing a satisfaction study, it is common to have a great number of variables that are correlated, being therefore desirable to reduce the number of variables into dimensions that are easier to interpret (Vilares and Coelho, 2011).

Two models differ in purpose and computation: PCA and common factor analysis (Fabrigar et al.,1999). PCA transforms original variables, that can be or not correlated, into a smaller set of non-correlated variables. The principal components result from linear combinations of original variables that aim to reduce original data without losing information. PCA retains those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components (the ones that explain a large part of the variance present in the data) and ignoring higher-order ones. In common factor analysis, the main goal is to explain covariance structure amongst original variables through a hypothetical set of common factors (unobservable) (Vilares and Coelho, 2011). The main difference between the two models relies on the way communalities are used. In PCA communalities are initially one. On the one hand, PCA considers that the total variance of variables can be explained by its components, and so there is no error variance. On the other hand, common factor analysis assumes error variance.

This is shown by the fact that communalities have to be estimated, making it more complicated than PCA (Finch, 2013). Based on the common factor model, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) have been presented (Thurstone, 1947). EFA attempts to discover the nature of the constructs influencing a set of responses and is used when a researcher wants to find the number of factors affecting variables. CFA tests whether a specified set of constructs is influencing responses in a predicted way. (Decoster and Hall, 1998). Factor analysis is exploratory if the researcher does not have a hypothesis about the number of factors measured by tests, and confirmatory if the researcher has such hypotheses and conducts statistical tests of them (Fienberg and Junker, 2010).

Determining the number of factors to extract can be a difficult task. Extracting a large number of factors can result in variance errors but extracting a small number can leave out valuable common variance (Yong and Pearce, 2013). Specifying too few factors in a model can originate more errors than specifying too many factors. When too few factors are included in a model, a substantial error is likely. Research suggests that over factoring originates less error to factor loading estimates than under factoring (Fabrigar et al., 1999).

The most common criteria to determine the number of factors to retain are:

- Retaining all factors with eigenvalues higher than one (Guttman-Kaiser rule);
- The “scree test”, that consists of eigenvalues and factors. Eigenvalues of the reduced correlation matrix are computed and represented in descending order of values. Factors above the “break” are the ones to retain (Costello and Osborne, 2005). The “scree test” is only dependable for a sample size of at least 200 (Yong and Pearce, 2013);
- Retaining all factors that explain, at least, 70%-80% of the variance (Finch, 2013);
- Parallel analysis: simulation of a set of random data with the same number of variables as the real data. That random data is submitted to PCA and eigenvalues are saved. This process is repeated at least 100 times and the resulting set of eigenvalues averaged and compared with real data. The eigenvalues extracted from real data that exceed those extracted from random data indicate the number of factors to retain (Watkins, 2018);

To get a better interpretation of the results, the rotation has to take place due to ambiguity present in unrotated factors (Yong and Pearce, 2013). Varimax, quartimax, and equamax are the most used orthogonal methods of rotation. Orthogonal rotation is when factors are rotated 90° from each other. Direct oblimin, quartimin, and promax are the most used oblique methods of rotation. Oblique rotation is when factors are not rotated 90° from each other. Orthogonal rotation produces uncorrelated factors, alternatively, oblique methods allow correlation between factors. It is advised for researchers to use orthogonal rotation because it produces more easily interpretable results. However, in social sciences correlation between factors is expected (Osborne et al., 2011).

Cluster analysis also groups data in order to reduce the number of factors. However, the organisation is different from factor analysis. While factor analysis combines variables that measure similar phenomena, cluster analysis groups individuals that have similar perceptions (Vilares and Coelho, 2011). A set of objects is divided into several groups or clusters so that objects within the same group are more similar to each other than objects in different groups. Clustering techniques belong in different categories: hierarchical, optimization-partitioning, density-seeking and clumping methods. (Bratchell, 1989). In satisfaction studies, cluster analysis is usually applied to identify client segments with similar attitudes and perceptions towards attributes of service quality.

To assure the validity and adequacy of the results provided by the analysis, multiple coefficients are considered.

- **ANOVA** allows discovering if the results from a survey are significant. There are multiple types of this analysis of variance, but only one-way ANOVA is discussed since it is the one used later on. One-way ANOVA is characterized as a "fixed effects" model where the goal is to assess the differences among the subject groups, in this present case study, the difference between two genders. A significant result translates inequalities between the two groups (Armstrong et al., 2002).
- **Bartlett's test of sphericity** examines the hypothesis of the correlation matrix being the identity matrix. If significant, the null hypothesis, that says that variables are not correlated, is rejected and different rules are employed to identify the number of components to retain (Peres-Neto et al., 2005).
- **Cronbach's alpha** is one of the most celebrated measures of reliability on the social sciences. It is considered a measure of scale reliability and internal consistency since it evaluates how closely related a set of variables are as a group. The higher the value of the coefficient, the strongest is the consistency. A value above 0,700 is considered an acceptable indicator of internal consistency (Bonett and Wright, 2015).
- **ICC** refers to correlations among data, as opposed to correlations between two groups or classes of data. It is calculated through the ratio: (variance of interest) / (total variance) = (variance of interest) / (variance of interest + unwanted variance). If the unwanted variance is higher than the variance of interest, the reliability of the model is poor, with a value lower than 0,500. Values above 0,800 are considered signs of good reliability (Liljequist et al., 2019).
- **PCC** and **SCC** are similar coefficients that measure the intensity of the relationship among variables of interest. The main difference is that PCC is a linear measure that indicates the direction of the relationship, as opposed to SCC that is calculated from vectors of ranks (Hauke and Kossowski, 2011).
- **KMO** test indicates how fitting the data is for factor analysis. This test measures sampling adequacy for every variable and at last, for the total model. Values above 0,700 are considered a good fit (Glen, 2016).
- **Mann–Whitney U test** examines differences between two groups on a single variable. The objective of this test is to assess if two groups come from the same population being that the null hypothesis dictates that both samples come from the same population. To examine the null hypothesis, observations from the two groups are combined into a single group and are ranked.

When the result is presented with a non-significant level, it is possible to conclude that the two groups do not differ in a significant manner (McKnight and Najab, 2009).

- **Independent t-test** is a statistical test that investigates if there are statistically significant differences between the means of two independent groups. The null hypothesis of the independent t-test states that there are no significant differences regarding the means of the two sets (Kim, 2015).

5.3. Structural Equation Modeling

SEM is a general modeling technique, used to test the validity of theoretical models that define causal, and hypothetical relations between variables. These relationships are represented by parameters that indicate the magnitude of the effect that some variables (independent variables) have on others (dependent variables). SEM is an extension of general linear models that considers measurement errors associated with variables under study (Marôco, 2014).

SEM is a statistical methodology that takes hypothesis-testing to the multivariate analysis. Multivariate procedures commonly used in market research are essentially descriptive or exploratory, so hypothesis testing is difficult, if not impossible. SEM generally involves the specification of an underpinning linear regression type model (incorporating the structural relationships or equations between unobserved or latent variables) together with several observed or measured indicator variables (Byrne, 1994). The SEM technique is widely used to explore and test causal relationships in the social sciences, specifically in health. SEM is a combination of factor analysis, multiple correlations, regression and path analysis. Compared with other multivariate analysis methods, SEM can estimate dependence relationships, represent unobserved concepts in these relationships, consider measurement errors in estimation, and define a model explaining an entire set of relationships (Cho et al., 2009; Kline, 2011; Xiong et al., 2014).

In exact sciences, it is common to work with observable variables. However, when dealing with social and health sciences, researchers often encounter variables that are not directly observable (latent variables), thus, the traditional analysis does not always allow the evaluation of flexible theoretical references without excessive contamination of statistical errors (Marôco, 2014). Latent variables, unlike observable variables, are not directly observed but are rather assessed from other directly observed and measured variables through a mathematical model (Cepeda-Carrion et al., 2019). Exogenous latent variables are synonymous of independent variables and endogenous latent variables are synonymous of dependent variables. Endogenous variables are influenced by exogenous variables either directly or indirectly (Byrne,1994). The general SEM can be also known as Linear Structural Relationships, a linear model that establishes relations between variables. This model can be organised into two sub-models: measurement and structural model. The structural model shows the relationship between latent variables and is synchronously estimated with the measurement model. When a model only contains observable variables, the structural model is reduced to path analysis. The measurement model shows the relationship between observed and latent variables. Its goal is to illustrate how well the observable variables measure the latent variables. Measurement variables are assessed by CFA. This method determines relations between observed and latent variables and tests them in order to confirm the suggested structure (Wang and Wang, 2012; Marôco, 2014). Regarding the measurement model, two distinct concepts have to be acknowledged. Reflective and formative models. Reflective models are the most common under the SEM methodology and assume that indicators are caused by the latent variable, whereas in formative models, it is considered that indicators cause the latent variable (Freeze et al., 2007). This differentiation is further discussed in section 7.2.

For this case study, SEM is going to be executed in the AMOS software where criteria and subcriteria are graphically represented whether as observable or latent variables. Each observable variable is paired together with a measurement error term that represents the existent amount of variation in the observed variable that is due to measurement errors. These measurement error terms can have covariances between themselves, also known as modification indices (Murti, 2016). It is important to give a distinction between these two concepts: covariance and correlation. Although very similar, covariance refers to the direction of the linear relationship of both variables follows. Correlation, in its turn, is a value, from -1 to 1, that measures the strength and direction of the linear relationship of the two variables (Saha, 2018; Tripathi, 2019). When these covariances happen for errors of variables that belong to the same component, its interpretation and explanation are easy to fundament, given that information sharing between the variables is justified. Correlated errors are also possible among variables using identical vocabulary or placed near each other on a questionnaire (Bollen and Lennox, 1991; Marôco, 2014). However, when covariances are linked to errors from variables of different components, its acceptance is more debatable, because it means that variables are not the only source of contribution for the definition of the respective components. If it is decided that a model contains covariances between measurement errors, a justification must be presented (Gerbing and Anderson, 1984). Latent constructs can be associated with each other, suggesting that covariance is allowed between latent components. The influence of each component on the dependent variable is represented by the component loading (Murti, 2016; Ramlall, 2016). The dependent variable is, for this study, patients' satisfaction with the service provided. Figure 11 graphically demonstrates the SEM path diagram of a reflective model where causality flows from latent constructs to observable variables.

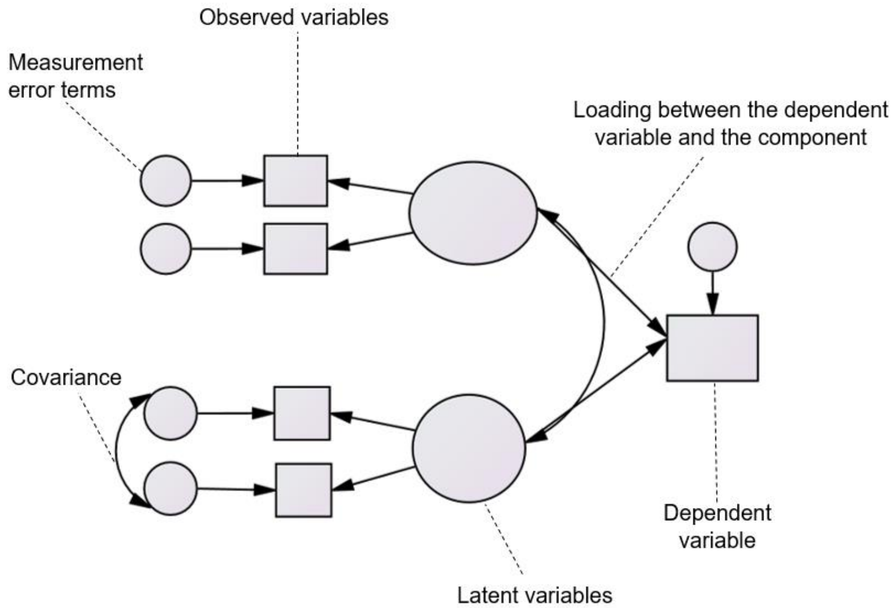


Figure 11. Structure of a SEM path diagram.

When applying the SEM method, an estimation model has to be selected. This estimation consists of obtaining parameters estimates that represent, in the best way possible, data of the analysis. It is executed through covariance matrixes of observable variables (Marôco, 2014). The estimation process can be executed through multiple techniques accordingly to the used data. The most common are maximum likelihood (ML), unweighted least square (ULS), weighted least square (WLS), and generalized least square (GLS).

ML is the most used estimation method. It searches for the value of the parameter vector that maximizes the likelihood function. The bigger the sample, the more reliable the results. However, it is important to note that this method can only be applied when data presents multivariate normal distribution. In general, this is a robust method, that provides unbiased results when assumptions are met (Marôco, 2014). ULS estimates the model parameters that minimize the sum of the squared errors. This method does not have assumptions that need to be met but is not asymptotically efficient. It is specific for data that do not follow a normal distribution (Marôco, 2014). WLS weights the observations concerning the error variance of that observation, thus, conquering the question of non-constant variance. It is applied to data that do not follow a multivariate normal distribution and delivers best results when applied to a large sample (Kantar, 2015). GLS estimation is accomplished though the weighting of estimated errors of the residual matrix with the weights correspondent to the inverse of the covariance matrix. This method is best suited for large samples, with more than 500 observations (Markus et al., 2003). It is possible to observe that these alternatives have different characteristics and might be best suited for different types of data. However, there are no specific guidelines on which method is best to follow. When searching through the literature, it is noticeable that authors agree that ML is the method that delivers the most robust results. For instance, Olsson et al. (2000) compared the performance of ML, WLS, and GLS, and concluded that ML presents more realistic indices and less biased parameter estimates.

Another aspect to account for is the presence of outliers. These are data points that are strongly deviated from other values. Outliers can lead to significant fluctuations in parameters estimates and must not be ignored. The diagnosis of outliers can be done in two different alternatives: univariate measures and multivariate measures (Cohen et al., 2003; Hunter and Schmidt, 2004; Kutner et al., 2004). With univariate measures, an observation is considered an outlier if it lies above or below the interval of equation 1. P_{25} and P_{75} represent, respectively, the 25 and 75 distribution percentiles. A value that is not declared an outlier through this alternative might still be a multivariate outlier (Marôco, 2014).

$$P_{75} \pm 1{,}5x(P_{75}\text{-}P_{25}) \qquad (1)$$

Within multivariate measures, *Mahalanobis distance* is the most used technique. It considers a centroid as the junction of the means of the predictor variables and calculates the distance between it and a data point. A large distance is an indicator of an outlier (Aguinis et al., 2013).

It is recognized that all SEMs are simplified approximations to reality, not hypotheses that might be true. Accordingly, various indices have been developed as measures of goodness of approximation to the distribution from which the sample was originated. Absolute and relative fit indices can be differentiated. Absolute indices are functions of discrepancies. Relative indices compare a function of discrepancies from the fitted model to a function of discrepancies from a null model. In the last-mentioned type, all variables are usually uncorrelated (McDonald and Ho, 2002). Absolute fit indices are:

- **Chi-Square value:** traditional measure for evaluating the overall model fit and assess the magnitude of discrepancy between the sample and fitted covariances matrices (Hooper et al., 2008). Large sample sizes represent problems for significance tests based on chi-square statistics. The issue is that the larger the sample, the greater the power, and so even smaller differences are reported as indicating a statistically significant misfit between the data and the model. Very large sample sizes have a higher probability of dealing with type I errors. These errors reject the hypothesis saying the model is well adjusted when the adjustment is in fact good (Byrne, 1994).
- **Root Mean Squared Error Approximation (RMSEA):** is used to provide a mechanism to adjust the sample size where chi-square statistics are used. It tells us how well the model, with unknown but optimally chosen parameter estimates, would fit the population's covariance matrix (Byrne, 1998);
- **GFI:** created by Jöreskog and Sorbom (1984) as an alternative to the Chi-Square test and calculates the proportion of variance that is accounted for by the estimated population covariance (Tabachnick and Fidell, 2012). GFI explains covariances' proportions, observed amongst manifested variables, explained by the adjusted model. It tends to increase with sample size and the number of variables on the model (Marôco, 2014);

Relative fit indices can be acknowledged as:

- **Normed fit index (NFI):** This statistic assesses the model by comparing the chi-square value of the model to the chi-square of the null model. The null model is the worst-case scenario as it specifies that all measured variables are uncorrelated. A major disadvantage to this index is that it is sensitive to sample size, underestimating fit for samples less than 200 (Hooper et al., 2008);
- **Comparative fit index (CFI):** This coefficient is a revised form of NFI that is less sensitive to sample size**.** It captures the fit of one's hypothesized model as an empirical increment above a simpler model. CFI attempts to adjust for model complexity or parsimony. It does so by including the degrees of freedom used in the model directly into the computation (Iacobucci, 2010);
- **Parsimonious fit indices:** Proposed by Mulaik et al., (1989), these indices are adjustments of the relative indices mentioned above. The adjustments made prioritize simplicity of the model and penalize models that are less parsimonious, with the goal to favour the simpler alternatives. The more complex a model is the poorer adjustment of the fit index. Parsimony Goodness-of-Fit Index (PGFI), Parsimony Comparative Fit Index (PCFI) and Parsimony Normed Fit Index (PNFI) are examples of theses indices (Hooper et al., 2008).

Many other adjustment indices are available to researchers. However, due to critics and apparent obsolete nature, only the indices above mentioned are used in this project.

5.4. Ordinal Logistic Regression

Regression analysis is commonly used to model the association between a response and several potential explanatory variables, with each association estimated in terms of an odds ratio. As previously mentioned, there are many types of regression analyses, each presenting different characteristics and being best suited for different types of variables and samples. OLR, that can be referred to as the proportional odds model, is a special type of multinomial regression, which can be advantageous when the response variable is ordinal (Koletsi and Pandis, 2018). Since that is the case with our sample, OLR was the regression method chosen.

Important features of OLR are (Kumar and Sankar, 2008):

- It provides a single regression coefficient estimate of covariates for each response category;
- It follows stochastic ordering;
- Is easy and simple to apply;
- Needs few parameters to estimate;
- Odds are proportional across the response variable;

OLR is a statistical method where one variable is explained or understood based on one or more variables. The variable that is being explained is called the dependent, or response variable. The other variables used to explain or predict the response are called independent variables. In an OLR model, the outcome variable has more than two levels. It estimates the probability of being at or below a specific outcome level given a collection of explanatory variables (Xing Liu and Koirala, 2012). Opposed to what happens in linear regression models, for regression models it is not possible to calculate a single R^2, thus, approximations are measured. Since there are no specific guidelines on how to use or interpret these measures, values above 0,5 are considered indicators of good adjustment (Lomax and Hahs-Vaughn, 2012; Pituch, 2016).

Statistical Package for the Social Sciences (SPSS), the software that is used to implement this method, returns three pseudo-R-squared coefficients (UCLA: Statistical Consulting Group, 2011):

- **Cox and Snell's R^2** is a transformation of the statistic used to dictate the convergence of logistic regression. It is based on the log-likelihood for the estimated model in comparison to the log-likelihood for a baseline model. The maximum of this value is less than one.
- **Nagelkerke's R^2** is an adjustment of the Cox and Snell's R^2 with values extended to one.
- **McFadden's R^2** is the third measure and is based on the log-likelihood kernels for the baseline model and the estimated model.

Limitations are, however, also a part of this model (Kumar and Sankar, 2008):

- Large samples are required since the coefficients are estimated by maximum likelihood estimate;
- Proportional odds assumption should be satisfied, meaning that the odds ratio is constant across the cut-off point for each of the covariate in the model. If this assumption is not truthful, the estimate of the parameters obtained is not valid;

When performing this method on SPSS, some measures are evaluated. The goodness of fit test contains the Deviance and Pearson chi-square tests, that are used to determine if a model exhibits good fit to the data. Non-significant test results are indicators that the model fits the data well (Petrucci, 2009; Field, 2018). In OLR models, there is an assumption regarding ordinal odds. According to this assumption, parameters cannot change for different categories. This means that the correlation between predictor variables and the dependent variable cannot change for the dependent variable's categories. The null hypothesis of this test says that the slope coefficients in the predicted model are equal in all categories, thus the significance level must be sig>0,05 (Ari and Yildiz, 2014).

When a logistic regression is calculated, the regression coefficient (β) portraits the estimated increase in the log odds of the *dependent variable* for every unit *increase* in the value of the *predictor.* The OR for each unit increase in the predictor, is its turn, the exponential function of the regression coefficient (e^{β}). An OR > *1* is associated with an increased probability of being in a higher category on the dependent variable along with increasing values on an independent variable. An OR < 1 leads to a decreasing probability with increasing values on an independent variable. An OR = 1 implies that there is no predicted change in the likelihood of a predictor being on a higher category (Szumilas, 2010).

5.5. Multicriteria Satisfaction Analysis

The MUSA model was developed to measure and analyse customer's satisfaction from a specific product or service, but the same principles can be used to measure global satisfaction of a group of individuals regarding a specific service or operation that they interact with (Muhtaseb et al.,2012). The basic principle of MUSA is the aggregation of individual judgments into a collective value function, assuming that customers' global satisfaction depends on a set of criteria representing service characteristic dimensions. This preference disaggregation methodology is implemented through an ordinal regression-based approach used for the assessment of a set of marginal satisfaction functions in such a way, that the global satisfaction criterion becomes consistent with customer's judgments (Drosos et al., 2015).

5.5.1.Notation

Take into account the following notation (Ferreira et al., 2018):

- $G = \{g_1, \ldots, g_j, \ldots, g_n\}$ portrays a set of family criteria;
- G_j, with $j = 1, \ldots, n$, is the j^{th} criterion of set G (n stands for the number of criteria);
- E_j represents the discrete scale of criterion g_j, $(j = 1, \ldots, n)$;
- g^l_j, $j = 1, \ldots, n$ and $l = 1, \ldots, L$, represents the l^{th} dissatisfaction/satisfaction level (hereinafter named satisfaction levels), i.e., $E_j = \{g^1_j, \ldots, g^l_j, \ldots, g^{L_j}_j\}$;
- $g^1_j \preceq g^l_j \preceq g^{L_j}_j$ denotes a total order for g^l_j; symbols $\prec$ and _ mean "strictly less preferred than" and "as preferable as", respectively; e.g., the totally satisfied level $(l = L_j)$ is strictly more preferred than the totally dissatisfied level $(l = 1)$;
- $E = \{g^1, \ldots, g^l, \ldots, g^L\}$ is a discrete scale associated with the overall satisfaction; as before, $g^1 \preceq g^l \preceq \ldots \preceq g^L$ denotes a total order for g^l, $l = 1, \ldots, L$;
- $P = \{1, \ldots, q, \ldots, p\}$ represents a set of patients whose satisfaction respecting a hospital is being assessed; each patient $q \in \{1, \ldots, p\}$, characterizes the hospital according to a single level of each scale E_j, for $j = 1, \ldots, n$ and E.
- $x_j^{(q)} \in E_j$ represents the satisfaction level assigned by patient q regarding the j^{th} criterion, g_j;

- $x^{(q)} \in E$ stands for the overall satisfaction level assigned by patient q with respect to the entire hospital;
- $\hat{x}^{(q)} \in E$ represents the overall satisfaction level;
- $v(x^{(q)})$: $E \rightarrow [0, 1]$ is a monotone non-decreasing value function of its argument $x^{(q)} \in E$; $v(x^{(q)})$ is the value function related with each overall satisfaction score, and $v(g^1) = 0 \leq \ldots \leq v(g^l) \leq \ldots \leq v(g^L) = 1$;
- $v_j(x^{(q)}_j)$: $E \rightarrow [0, 1]$ is a monotone non-decreasing value function related with the partial satisfaction score j, with $v_j(g^l_j) = 0 \leq \ldots \leq v_j(g^l_j) \leq \ldots \leq v_j(g^L_j) = 1$;
- $\alpha^{(q)}$ is a free error variable associated with patient $q \in \{1, \ldots, p\}$ that can be decomposed into two non-negative error variables, $\alpha^{(q)+}$ (overestimation error) and $\alpha^{(q)-}$ (underestimation error), such that $\alpha^{(q)} = \alpha^{(q)-} - \alpha^{(q)+}$;

Given that $x^{(q)} \in E$ is the level assigned by patient q to define the overall satisfaction of the hospital, the value of $\hat{x}^{(q)}$ is described by $v(\hat{x}^{(q)})$. If an additive model can be applied with the partial values' aggregating purposes, the following is verified:

$$v(x^{(q)}) = \sum_{j=1}^{n} v_j(x_j^{(q)}) \quad (2)$$

where $x_j^{(q)} \in E_j$ is the level of satisfaction patient q elected to describe the hospital according to criterion g_j. The overall satisfaction level of a patient q should be the same as the aggregating results, i.e., there should be an indifference between $\hat{x}^{(q)}$ and $x^{(q)}$ represented by $\hat{x}^{(q)} \sim x^{(q)}$, suggesting:

$$v(\hat{x}^{(q)}) = v(x^{(q)}) \quad (3)$$

On some occasions, the latter is not verified, and some errors can be present. The free variable $\alpha^{(q)}$ is introduced to equation (3):

$$v(\hat{x}^{(q)}) = \sum_{j=1}^{n} v_j(x_j^{(q)}) + \alpha^{(q)} \quad (4)$$

Given that $\alpha^{(q)}$ is free, it can be rewritten as two non-negative variables, $\alpha^{(q)+}$ and $\alpha^{(q)-}$.

$$v(\hat{x}^{(q)}) = \sum_{j=1}^{n} v_j(x_j^{(q)}) + \alpha^{(q)+} + \alpha^{(q)-} \quad (5)$$

5.5.2. The basic MUSA

The method finds an additive utility function representing the satisfaction level of a set of customers based on their expressed preferences collected in a satisfaction survey's data. Customers are asked to give a satisfaction level for a service or a product, as well as a marginal satisfaction level for each one of its characteristics (Grigoroudis and Siskos, 2003). Previously, integral concepts necessary to formulate the mathematical problem were introduced. From there, the Patient Satisfaction Model (PSM) can be defined as (Ferreira et al., 2018):

In the objective function, there is a minimization of the sum of the non-negative error variables, $\alpha^{(q)+}, \alpha^{(q)-}$, for all $q = 1, \ldots, p$, (Grigoroudis and Siskos, 2002). This function focuses on minimizing the deviation between patients' overall and partial judgements. It is acknowledged from linear programming that both deviations cannot have a positive value simultaneously. If the value of the objective function is zero, it means that the information provided by the patients at the comprehensive level and the per-criterion levels is consistent. Contrarily, some inconsistencies may appear. The value of the objective function is a form of measuring the number of inconsistencies present in the model.

$$minimize\ z = \sum_{q=1}^{p} (\alpha^{(q)+} + \alpha^{(q)-}) \tag{6.1.}$$

subject to:

$$v(g^L) - v(x^{(q)}) = (\sum_{j=1}^{n} v_j(g_j^{(Lj)}) - v_j(x_j^{(q)})) + \alpha^{(q)+} - \alpha^{(q)-}, q = 1, \ldots, p \tag{6.2.}$$

Equation (6.2.) portraits the indifference relation between the overall satisfaction and the conjoint aggregation of partial satisfaction. The difference between the values of the highest satisfaction level and the overall qth judgement must be equal to the difference between aggregating results, including an error term.

$$v(g^l) - v(g^{l-1}) \geq 0, \qquad l = 2, \ldots, L \tag{6.3.}$$

$$v_j(g_j^l) - v_j(g_j^{l-1}) \geq 0, \qquad j = 1, \ldots, n, \quad l = 2, \ldots, L_j \tag{6.4.}$$

These two equations state that the respective value functions are non-decreasing monotone functions.

$$v(g^L) = 1 \tag{6.5.}$$

Equation (6.5.) denotes that the value of the best performance is unitary, meaning that there is no preferable satisfaction level than the highest one.

$$v(g^l) = 0 \tag{6.6.}$$

Equation (6.6.) states that the value of the worst satisfaction level is zero because there is no satisfaction level worst then the lowest one. With this, is concluded that satisfaction levels are bounded within the range [0;1].

$$\sum_{j=1}^{n} v_j(g_j^{Lj}) = 1 \tag{6.7.}$$

Equation (6.7.) proves that the cumulative value of the best performance in all criteria must be equal to the best performance's value in overall judgements.

$$v_j(g_j^1) = 0, \qquad j = 1, \dots, n, \tag{6.8.}$$

Equation (6.8.) establishes that the partial value of the worst performance in each subcriterion has to be zero. Furthermore, the cumulative value of the lowest satisfaction level in all subcriteria has to also be zero.

$$v(g^l) \geq 0, \qquad l = 2, \dots, L-1, \tag{6.9.}$$

$$v_j(g_j^l) \geq 0, \quad j = 1, \dots, n, \quad l = 2, \dots, L_j \tag{6.10.}$$

$$\alpha^{(q)-}, \alpha^{(q)+} \geq 0, \quad q = 1, \dots, p. \tag{6.11.}$$

These three equations establish the non-negativity of the variables that are optimized.

5.5.3. The hierarchical MUSA

In several applications of the MUSA model, it seems rather useful to consider a hierarchical, instead of flat, structure of criteria. Aligned with this, a new formulation of the MUSA method was developed and named as hierarchical MUSA. In this new model, for each criterion g_j, there is a set of subcriteria $G_j = \{g_{jk}^1, \dots, g_{jk}^l, \dots, g_{jk}^{Lkj}\}$, for $j = 1, \dots, n$. Each subcriterion g_{jk} has its own level scale $E_{jk} = \{g_{jk}^1, \dots, g_{jk}^l, \dots, g_{jk}^{Lkj}\}$, where $v_{jk}(g_{jk}^1) = 0$ for $j = 1, \dots, n$. The new constraints can be interpreted similarly as the ones from the PSM model, previously explained, The Hierarchical Patient Satisfaction Model (HPSM) is as follows:

$$minimize\ z = \sum_{q=1}^{p} (\alpha^{(q)+} + \alpha^{(q)-}) + \sum_{q=1}^{p} \sum_{j=1}^{n} (\alpha^{(q)+} + \alpha^{(q)-}) \tag{7.1.}$$

subject to:

$$v_j(g_j^{Lj}) - v_j(x_j^{(q)}) = (\sum_{k=1}^{nj} v_{jk}(g_{jk}^{(Lkj)}) - v_{jk}(x_{jk}^{(q)})) + \alpha_j^{(q)+} - \alpha_j^{(q)-} \tag{7.2.}$$

$$q = 1, \dots, p$$

$$j = 1, \dots, n$$

$$v_{jk}(g_{jk}^l) - v_{jk}(g_{jk}^{l-1}) \geq 0, \qquad j = 1, \dots, n, \tag{7.3.}$$

$$k = 1, \dots, n_{j,}$$

$$l = 2, \dots, L_{kj}$$

$$\sum_{j=1}^{n} v_{jk}(g_{jk}^{Lkj}) = v_j(g_j^{Lj}), \qquad j = 1, \dots, n \tag{7.4.}$$

$$v_{jk}(g_{jk}^{1}) = 0, \qquad j = 1, \dots, n, \quad k = 1, \dots, n_j \tag{7.5.}$$

$$v_{jk}(g_{jk}^{l}) \geq 0, \qquad j = 1, \dots, n, \quad k = 1, \dots, n_j, \quad l = 2, \dots, L_{kj} \tag{7.6.}$$

$$\alpha^{(q)-}, \alpha^{(q)+} \geq 0, \quad p = 1, \dots, q, \quad j = 1, \dots, n. \tag{7.7.}$$

These constraints follow the same principles as the basic model previously presented. For this specific case study, the model HPSM is applied through the MATLAB software.

5.5.4. MUSA and the Kano's model

Kano's model, proposed by the Japanese professor Noriaki Kan, is a useful tool to understand customer needs and their impact on customer satisfaction. The Kano diagram specifies three types of relationships between the degree of customer satisfaction and the fulfilment of expectations (Kano et al., 1984; Wang and Ji, 2010):

- Must-be Attributes: customers take must-be characteristics for granted. If these requirements are not sufficiently met, customers will be dissatisfied. However, their presence does not contribute to customer satisfaction.
- One-dimensional attributes*:* their fulfilment is positively and linearly related to the level of customer satisfaction. The higher the level of fulfilment, the higher the degree of customer satisfaction, and vice versa;
- Attractive attributes: fulfilment of attractive attributes will lead to greater satisfaction. However, since customers are not expecting these requirements, they will not be dissatisfied in the case of absence. These requirements are seen as *'pleasant surprises'.*

Yang (2005) refined Kano's model by acknowledging the degree of importance that each attribute holds. The three original categories are subdivided into two different groups (Yang and Yang, 2011):

- Critical or necessary (must-be attributes): Critical attributes are essential to customers and their fulfilment must be a priority. Necessary attributes should be met at a required level to not dissatisfy customers.
- High value-added or low value-added (one-dimensional attributes): high value-added quality attributes have a high impact on customer satisfaction and can lead to increased revenue. Thus, an effort to ensure that these attributes are provided is recommended. Low value-added quality attributes are not as influential but cannot be ignored. Providing too little of these requirements can leave customers feeling dissatisfied.

– Highly or less (attractive attributes): Highly attractive attributes are an effective way of attracting new customers. Less attractive attributes are not as attractive and can be reduced if there are not enough funds.

Determining which criteria are the most important can be a big help for managers and decision-makers who want to improve customers, or in the specific case study, patients' satisfaction. The allocation of each criterion/subcriterion to an attribute category can be easily executed through the linkage of MUSA and the Kano's model. Firstly, patients are divided into globally satisfied and globally dissatisfied groups. Satisfied patients are those who present a satisfaction score above the neutral level. Dissatisfied patients, in their turn, reveal a satisfaction score below the neutral level.

The categorization according to the Kano's model is accomplished by comparing the weights that satisfied and dissatisfied patients assign to each requirement. For instance, dissatisfied patients attach a higher weight to must-be requirements than satisfied patients. When the assigned weights are similar to both satisfied and dissatisfied patients, the requirement in question is said to be one dimensional. Finally, requirements are considered attractive when the weight assigned to them is higher for satisfied patients (Ferreira et al., 2018). For the subdivision into refined Kano's model categories, the weights of each criterion/subcriterion and the weight's centroid are compared. Table 3 provides a graphical representation of the division into different categories and specifications that need to be evaluated.

Table 3. Kano's model and refined version applied to MUSA. Subcriterion g_{jk}, j=1,..., n; k=1,..., n_j; Weight associated with dissatisfied patients: ($w_{jk}^{d} = v_{jk}^{d}(g_{jk}^{Lkj})$); Weight associated with satisfied patients: $w_{jk}^{s} = v_{jk}^{s}(g_{jk}^{Lkj})$); Weight's centroid: $\overline{w_{jk}}$.

Subcriterion g_k			
	Must-be attribute $w_{jk}^{d} > w_{jk}^{s}$	Critical $w_{jk} > \overline{w_{jk}}$	Necessary $w_{jk} < \overline{w_{jk}}$
	One dimension attribute $w_{jk}^{d} \approx w_{jk}^{s}$	High value-added $w_{jk} > \overline{w_{jk}}$	Low value-added $w_{jk} < \overline{w_{jk}}$
	Attractive attribute $w_{jk}^{d} < w_{jk}^{s}$	Highly $w_{jk} > \overline{w_{jk}}$	Less $w_{jk} < \overline{w_{jk}}$

5.5.5. Important indexes

<u>Satisfaction index</u>

The evaluation of a performance norm, through satisfaction indexes, can be useful for customer satisfaction analysis. According to Matsatsinis and Siskos (2003), it is one of the characteristics that differentiates the MUSA method from the remaining approaches. The satisfaction indexes for each criterion/subcriterion are as follows (João et al., 2007):

$$S_j = \sum_{l=2}^{L_j} P_{lg_j}.\, v_j(g_j^l) \tag{8}$$

$$S_{jk} = \sum_{l=2}^{L_{kj}} P_{lg_{jk}}.\, v_{jk}(g_{jk}^{lkj}) \tag{9}$$

The first term of the sums represents the frequency of patients rating the j^th criterion and k^th subcriterion, available in Table 4.

Average improvement index

When deciding which areas to improve and enhance it is first necessary to assess the room for improvement of each dimension. Every criterion/subcriterion is associated with an average improvement index that depends on the importance that each criterion/subcriterion has on patients and how dissatisfied they feel towards it. These indices are normalised in the interval [0;1] and are calculated through the following equations (Grigoroudis and Siskos, 2002; Grigoroudis and Siskos, 2010; Ferreira et al., 2018):

$$\Delta_j = v_j(g_j^{Lj}) \,.\, (1 - \sum_{l=1}^{L_j} P_{lgj} \,.\, v_j(g_j^l)), \qquad j = 1, \dots, n$$

$$\Delta_{jk} = v_{jk}(g_{jk}^{Lkj}) \,.\, (1 - \sum_{l=1}^{L_{kj}} P_{lgjk} \,.\, v_{jk}(g_{jk}^{lkj})), \qquad j = 1, \dots, n \text{ and } K = 1, \dots, n_j \tag{10}$$

The first term of both equations corresponds to the importance of each j^th criterion or k^th subcriterion and the second term represented the dissatisfaction level. The higher the dissatisfaction, the larger room for improvement of that criterion/subcriterion. These indices are null when patients attribute no weight to the criterion/subcriterion or are completely satisfied with it.

Demanding index

The demanding nature of patients varies according to the criterion/subcriterion in question. Grigoroudis and Siskos (2002, 2010) identified three different demanding levels:

- *Non-demanding patients:* Patients say they are satisfied, but only a small part of their expectations are met. The value function has a concave form, meaning that the main relative improvement of the criterion's value function happens on the first satisfaction levels (the lowest levels).
- *Neutral patients:* The percentage of fulfilled expectations is directly related to how satisfied patients express they are. The value function presents a linear form due to the linear relationship explained.
- *Demanding patients:* Patients are only satisfied when they get the best quality. The value function is represented by a convex form since the main relative improving occurs on the highest satisfaction levels.

Demanding indexes are specified under the interval [-1;1]. Demanding patients are characterized by positive values, while non-demanding patients manifest negative values. Neutral patients reveal values close to zero. The final expression that allows the computation of this index is (Ferreira et al., 2021):

$$\mathrm{D_j} = 1 - \frac{1}{v_j(g_j^{Lj}) \,.\, g_j^{Lj}} \,.\, \sum_{l=2}^{L_j} (v_j(g_j^l) + v_j(g_j^{l-1})) \tag{11}$$

5.5.6. Strategic priorities

When considering the improvement and demanding indexes, improvement diagrams can be created, according to demand (high/low) and room for improvement (high/low), as seen on Figure 12 (Grigoroudis and Siskos 2002, 2010).

Third priority *High demand, and low room for improvement*	**Second Priority** *High demand, and high room for improvement*
Second Priority *Low demand, and low room for improvement*	**First Priority** *Low demand, and high room for improvement*

Figure 12.Improvement diagram of strategic priorities. Source: Adapted from Grigoroudis and Siskos (2002, 2010).

- *1st priority* – Highly effective dimensions with non-demanding patients are considered direct improvement actions.
- *2nd priority* – Dimensions that have either high or low demand and room for improvement.
- *3rd priority* – Satisfaction dimensions that have a small room for improvement and might require considerable effort.

To differentiate between high or low demand/room for improvement, the distribution centroid has to be assessed. High demand/room for improvement values are located above the centroid, and low values are below.

5.5.7. Market opportunities

The combination of weights and satisfaction indices can generate action diagrams as the one in Figure 13. These diagrams inform us about the strong and weak points of patient satisfaction, acknowledging the dimensions that require an increased improvement effort (Grigoroudis and Siskos 2002, 2010).

Transfer Resources *High satisfaction, low importance, low priority, reallocate resources elsewhere*	**Leverage Opportunities** *High satisfaction, high importance, and medium priority*
Status Quo *Low satisfaction, low importance, and no action required*	**Action Opportunity** *Low satisfaction, high importance, and high priority*

Figure 13. Action diagram of market opportunities. Source: Adapted from Grigoroudis and Siskos (2002, 2010).

- *Action opportunity*: First priority dimensions where improvement efforts should be centralized.
- *Leverage opportunity:* These dimensions can be advantageous when used against the competition but only require medium priority.
- *Transfer resources:* Low priority with a recommendation for resources to be allocated elsewhere.
- *Status Quo:* No action is necessary given the low levels of satisfaction and low weights.

As previously explained, the distinction between high or low weights/satisfaction values is focused on the location of the centroid.

6. Case study: The experience of Portuguese citizens with a public hospital's service

6.1. Sample and validation

There has been an increase in interest to perform surveys assessing patient satisfaction in the Portuguese NHS throughout the years. Surveys' results can help to understand how patient perceive their care and treatment. They also can be used to check for progress and improvement, to promote an improvement in the quality of the NHS. For this study, a patient satisfaction survey composed of 65 questions was delivered, in 2018, to patients in the internment service of a secondary healthcare unit, that due to privacy reasons cannot be named. The survey is comprised of one question regarding the partial satisfaction for each of the 53 subcriteria, one question about the global satisfaction for each of the eleven criteria, and finally one question concerning the global satisfaction with the service provided. A total number of 251 responses were gathered from patients between 20 and 92 years old (with an average of 42 years old), both male (27%) and female (73%), out of six different medical internment specialities: paediatrics (63%), gastroenterology (14%), nephrology (8%), urology (6%), orthopaedics (5%) and internal medicine (4%).

The survey followed the Portuguese NHS's official template, allowing an understanding of patients' partial and global satisfaction, translated through a seven-point Likert-type scale, where 1 means *very dissatisfied*, and 7 means *very satisfied*. A seven-point scale is considered to have the most common number of response alternatives such as stated by Cox III (1980) *'if the number of alternative responses was to be established democratically, seven would probably be selected'.* However, there are different scale alternatives present in literature, such as the three-point scale, five-point scale, or nine-point scale. When comparing the options, there are noticeable distinctions between them. Scale reliability, for instance, is the focus of many studies, and there is a consensus that seven-point Likert scales are the most reliable (Nunnally, 1967; Finn, 1972; Ramsay, 1973; Oaster, 1989). Cicchetti et al (1985), proved through Monte Carlo simulations that there is an increase in reliability when comparing seven-point to two-point Likert scales, while also concluding that eight, nine or ten-point Likert scales are not more reliable than seven-point Likert scales. Preston and Colman (2000) established that scales with seven, eight, nine or ten points are more reliable than scales with two, three or four points. Accordingly, Joshi et al. (2015) also demonstrated that seven-point scales have a better reliability performance than five-point scales given their more comprehensive range of options, increasing the probability of translating the respondent's perceptions into a number. Recent studies state that utilizing a seven-point Likert-type scale increases reliability and validity and facilitates results' interpretation (Jung et al., 2011; Maeng et al., 2012; Gupta et al., 2013). A different issue with scale alternatives is that with three-point or five-point scales, uncertain or neutral responses are used more frequently than seven-point or nine-point scales (Matell and Jacoby, 1971). When using numbers to classify health care aspects, a psychological bias can be associated with the given responses, since the patient may not reasonably differentiate each level. A solution to this problem might be presenting a description for each scale level, as stated by Dickinson and Zellinger (1980), Krosnick and Berent (1993), and Krosnick and Presser (2009).

Going into further detail about the evaluated parameters, the eleven criteria present on the survey are: *obtained information, accommodation's quality, visits, food quality, medical staff, nursing staff, auxiliary staff, administrative staff, volunteering staff, exams and treatments, and discharge process.* Comparing these criteria to the results gathered from the utilization statistical analysis, it is possible to conclude that every criterion has also been included on the surveys of previous studies. This is an indicator that the parameters assessed in this case study are reliable and aligned with past researches.

Table 4 provides information regarding the relative frequency of each satisfaction level. The percentage of patients that evaluated each subcriterion with a determined satisfaction level is calculated to identify the service dimensions where patients seem mostly satisfied or dissatisfied. In all subcriteria and criteria, level seven is the one with the highest frequency, indicating that patients are very satisfied with the service provided. It is important to note that since the scale used on the questionnaires is a seven-point Likert scale, the expected level of satisfaction is based on the sum of the answers of the fifth, sixth, and seventh levels. From a simple observation of Table 4, it is possible to conclude that *nursing staff* is the dimension that mostly satisfies patients. It is followed by *auxiliary staff* and *food quality*. When assessing the areas where patients are mostly dissatisfied, *volunteering staff* is the criteria that provides the worst service. O*btained information* and *visits* also leave patients with a feeling of dissatisfaction. There are divergences when comparing these results to the previous NHS studies displayed in Figure 2 and Figure 3. As already mentioned, those studies were held on a national level, not portraying the reality of each health entity. The major difference between these studies is the criteria where patients are most (dis)satisfied. *Wating time*, the main source of dissatisfaction on the first two studies is replaced by *volunteering staff*. As for the criterion where patients are more satisfied, *medical care* was considered the best service attribute for the national-level studies, while *nursing staff* is the criterion that leaves patients feeling more satisfied in this specific study. Considering the level of overall satisfaction, the lowest satisfaction level is of 85.1% (*volunteering staff*), a significant increase from the 67.33% (*waiting time to get an appointment*) and 42% (*waiting time to get a consultation*) satisfaction levels experienced on the previous studies. The highest satisfaction level is of 97.6% *(nursing staff),* showing, as well, an increase when compared to the 94.6% (*availability and care of doctors*) and 93.9% (*clear instructions*) levels of the former studies. These findings corroborate the notion that national-level studies do not translate the reality of each health unit, highlighting the gap in the literature that needs to be tackled. A comparative analysis to assess patient satisfaction throughout the different specialities would also be of interest. However, due to a small sample size, it is not possible to achieve reliable conclusions.

Table 4. Criteria G_j, subcriteria g_{jk}, $j=1,\ldots,11$ and $k=1,\ldots, n_j$, and relative frequencies (%) of different satisfaction levels. Source: the author.

Criteria	Subcriteria	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Obtained information [G_1]	Patient's guide [g_{11}]	2.6	0.4	1.7	7.3	8.6	18.2	61.2
	Patient's rights and duties [g_{12}]	2.6	0.4	1.7	6.0	8.2	19.8	61.3
	Complaint means [g_{13}]	2.8	1.4	0.9	10.1	7.4	17.5	59.9
	Substitution in decision making [g_{14}]	1.9	1.4	1.9	10.5	7.2	18.1	59.0

Criteria	Subcriteria	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
	Anticipated vital will [g_{15}]	4.1	1.8	1.2	12.5	10.7	18.5	51.2
	Global	1.7	1.3	1.7	6.6	7.4	21.3	60.0
Accommo-dations' quality [G_2]	Cleanliness [g_{21}]	1.2	0.0	0.8	1.2	6.8	23.5	66.5
	Comfort and commodity [g_{22}]	1.2	0.0	1.2	2.4	4.0	19.2	72.0
	Privacy [g_{23}]	2.0	0.0	1.2	4.4	8.1	23.8	60.5
	Furniture [g_{24}]	1.2	0.4	1.2	5.6	9.6	25.3	56.6
	Noise [g_{25}]	3.2	0.8	4.4	10.0	12.0	23.7	45.8
	Temperature [g_{26}]	2.0	0.4	1.6	2.4	11.2	25.6	56.8
	Entertainment [g_{27}]	1.6	1.2	2.4	9.3	8.1	18.6	58.7
	Global	1.6	0.8	1.2	2.8	8.4	26.8	58.4
Visits [G_3]	Visitation hours [g_{31}]	2.1	2.1	4.2	7.9	11.7	16.3	55.6
	Visit duration [g_{32}]	2.1	0.8	4.2	8.3	12.5	17.5	54.6
	Number of visits [g_{33}]	1.7	1.3	3.8	8.4	9.3	18.1	57.4
	Easy access for close relatives [g_{34}]	1.7	0.8	1.7	5.4	7.9	15.9	66.5
	Global	1.7	0.4	2.5	5.9	8.4	19.2	61.9
Food Quality [G_4]	Preparation, *etc.* [g_{41}]	1.6	0.4	2.0	6.5	15.4	26.3	47.8
	Variety [g_{42}]	1.2	0.4	2.0	8.6	16.7	24.1	46.9
	Quantity [g_{43}]	1.6	0.8	1.2	4.5	11.7	23.5	56.7
	Meal support [g_{44}]	2.1	0.4	0.8	4.9	9.1	22.2	60.5
	Global	2.0	0.0	0.4	4.9	13.8	26.3	52.6
Medical staff [G_5]	Availability [g_{51}]	1.6	0.0	1.6	2.4	8.8	14.9	70.7
	Attention [g_{52}]	1.2	0.4	2.0	2.4	6.8	16.9	70.3
	Kindness [g_{53}]	1.2	0.0	1.6	3.6	5.6	17.3	70.7
	Information of patient's health state [g_{54}]	2.4	0.8	2.0	1.6	5.6	22.1	65.5
	Information of medical treatment [g_{55}]	1.6	0.8	1.2	3.2	5.6	22.0	65.6
	Information of medical exams [g_{56}]	2.0	0.4	2.4	2.8	6.9	20.7	64.6
	Health advising and teaching [g_{57}]	1.6	0.8	1.2	5.3	7.7	19.5	63.8
	Global	2.6	0.4	0.9	2.6	4.8	20.8	68.0
Nursing staff [G_6]	Availability [g_{61}]	1.2	0.0	0.0	1.2	3.6	15.6	78.4
	Attention [g_{62}]	1.2	0.0	0.4	1.2	3.2	15.9	78.1
	Kindness [g_{63}]	1.2	0.0	0.4	0.0	4.4	13.7	80.3
	Information of patient's health state [g_{64}]	0.8	0.4	0.4	1.2	4.4	19.5	73.3
	Information of nursing treatment [g_{65}]	0.8	0.8	0.0	0.8	4.8	18.4	74.4
	Health advising and teaching [g_{66}]	0.8	0.8	0.4	2.8	4.0	20.2	70.9

Criteria	Subcriteria	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
	Global	0.8	0.4	0.4	0.8	2.8	15.9	78.9
Auxiliary staff [G_7]	Availability [g_{71}]	1.2	0.0	0.0	2.0	3.6	16.2	76.9
	Attention [g_{72}]	1.2	0.0	0.0	2.0	3.2	16.6	76.9
	Kindness [g_{73}]	1.2	0.0	0.4	2.0	2.8	14.1	79.4
	Performance efficiency [g_{74}]	0.8	0.4	0.0	2.4	3.6	16.5	76.2
	Global	0.8	0.4	0.0	2.0	2.8	15.8	78.1
Adminis-trative staff [G_8]	Availability [g_{81}]	1.0	0.0	1.0	8.7	3.4	21.2	64.9
	Attention [g_{82}]	1.0	0.0	1.0	8.2	3.9	19.8	66.2
	Kindness [g_{83}]	1.0	0.5	0.5	7.7	5.3	18.7	66.5
	Performance efficiency [g_{84}]	1.4	0.0	0.5	8.2	3.8	18.8	67.3
	Global	1.0	1.0	0.5	7.7	4.3	19.7	65.9
Volunteering staff [G_9]	Availability [g_{91}]	2.1	0.0	2.8	9.7	5.5	13.1	66.9
	Attention [g_{92}]	2.1	0.0	2.8	11.0	3.4	14.5	66.2
	Kindness [g_{93}]	2.1	0.0	2.8	9.7	3.4	12.4	69.7
	Global	2.0	0.0	2.7	10.1	2.0	14.2	68.9
Information provided [G_{10}]	Availability [g_{101}]	0.8	1.3	0.8	6.8	9.7	19.4	61.2
	Attention [g_{102}]	1.3	0.4	0.4	5.5	7.6	19.3	65.5
	Kindness [g_{103}]	1.3	0.0	0.4	3.8	8.9	19.0	66.7
	Information of patient's health state [g_{104}]	1.7	0.4	0.0	6.6	9.6	19.2	62.4
	Information of medical treatment [g_{105}]	1.8	0.4	0.9	5.3	10.5	18.9	62.3
	Information of medical exams [g_{106}]	1.7	0.9	0.9	5.2	10.8	19.0	61.6
	Health advising and teaching [g_{107}]	1.8	0.9	0.4	6.2	9.7	18.5	62.6
	Global	0.8	0.8	0.4	4.2	9.7	16.5	67.5
Discharge process [G_{11}]	Homecare provided information [g_{111}]	1.8	0.5	0.5	4.5	4.5	14.9	73.3
	Waiting time after discharge [g_{112}]	2.3	0.5	0.5	5.9	4.5	17.2	69.2
	Global	1.8	0.4	0.0	4.0	5.8	16.1	71.7

6.2. Handling missing data

Research centred on consumer surveys has the goal to capture behaviour, attitudes, preferences and characteristics from participants or key informants. However, there is one major drawback that hinders the accomplishment of such objectives. That is missing values, either at the participant or criterion level (Streiner, 2002; Acock, 2005). Whether the participant does not wish to partake in the survey, and does not respond to the total questionnaire, or assigns the NR/blank option to determining items in the questionnaire, these are the two scenarios where missing values emerge. The latter scenario is the one that happens in this present study. When surveys present data with missing values, the use of statistical methods becomes controversial and complicated, being associated with reduced sample size, loss of statistical power, and inflation of standard errors (Anderson et al., 1983; Peng et al., 2006; Karanja et al., 2013).

There are three main categories where patterns of missing data occur. The fundamental distinction between the three categories is the relationship between the collected criteria in the dataset and the probability of existing missing data (Baraldi and Enders, 2010).

- Missing Completely at Random (MCAR): Happens when the probability of a criterion having a missing value is not related to the criterion itself nor any other model measure. When this is the case, any treatment can be employed without risking incorporating bias into the model (Heitjan, 1997; Bennett, 2001; Patrician, 2002; Batista and Monard, 2003; Karanja et al., 2013).
- Missing at Random (MAR): Occurs when the probability of a criterion having a missing value is not related to the criterion itself but to other model measures. This missing data scenario is labelled as ignorable because it can be ignored by the researcher without inducing additional bias. (Bennett, 2001; Dong and Peng, 2013; Karanja et al., 2013).
- Not Missing at Random (MNAR): When the probability of a criterion having missing data is directly related to the criterion itself. This scenario is mentioned as "non-ignorable" because the mechanism to deal with it must be specified by the researcher and incorporated into the analysis in order to reduce bias. (Bennett, 2001; Dong and Peng, 2013; Karanja et al., 2013).

To countermeasure the problems risen by missing data, researchers have to be aware of its patterns and proportion due to the effects it has on the validity of estimates. Thus, an understanding of all these characteristics is necessary to address deficiencies associated with missing data and to be able to employ the most adequate methods to mitigate said deficiencies (Peng et al., 2006; Karanja et al., 2013; Tabachnick and Fidell, 2012). With this in mind, a complete and introspective analysis was performed in SPSS where it is possible to conclude in which data missing category the dataset is included in, along with its proportion and patterns.

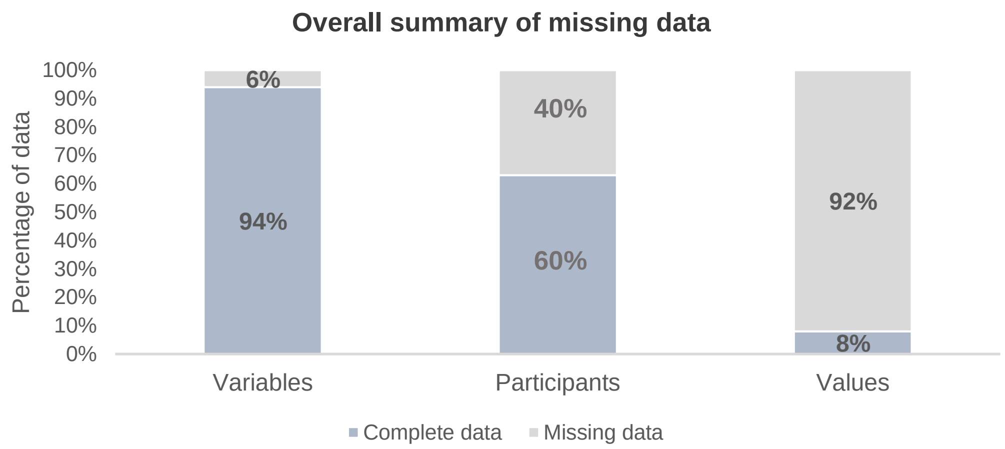


Figure 14. The output of an overall summary of missing data analysis using SPSS software. Source: The author.

Figure 14 provides an insight into the missing data present in the sample. Each of the three columns represents different attributes and is divided into sections according to its percentage of complete and missing data. Column "variables" (containing criteria and subcriteria) informs us the number of variables that have at least one missing value, being them 94%. Only 6% (corresponding to four variables, *i.e.: G_{21}: cleanliness; G_{62}: attention of nursing staff; G_{64}: information regarding patient's health state* *provided by nursing staff; G_6: global satisfaction with nursing staff service)* do not have any missing value. Column "participants" indicates the number of survey participants who did and did not respond to every question. On the one hand, a total of 40% (100 patients) answered to the entire survey with a valid Likert scale level. On the other hand, 60% (151 patients) either left a blank space or used the N/A option when answering part of the questions. Lastly, the column "values" indicates the total number of values that are missing from the sample. In this specific case, there are 65 survey questions and 251 survey participants, attaining to a total of 16,315 through the entire dataset. It is possible to observe that only 8% (1251) of the total values are missing from the sample.

To develop a more profound knowledge regarding the patterns of missing data, namely, to associate which missing data category is present in the sample, further analysis was performed and is displayed in Figure 15 and Figure 16.

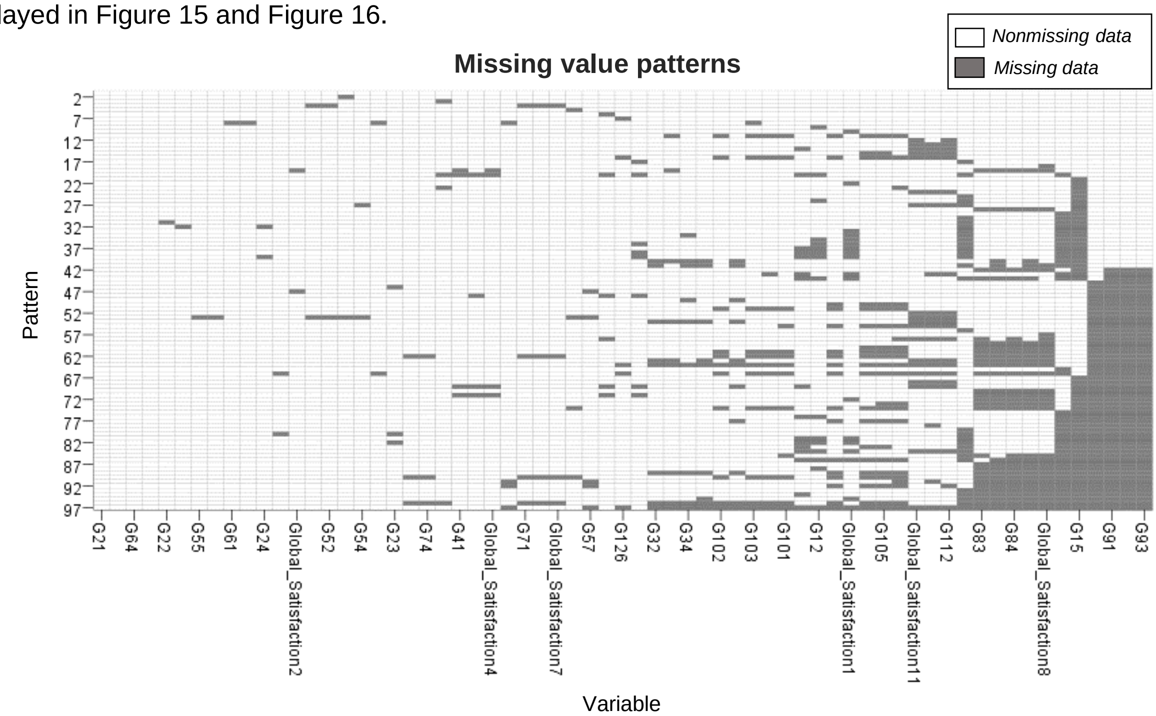


Figure 15. Missing value patterns. Source: SPSS software.

Figure 15 shows the missing value patterns where each pattern corresponds to a set of cases with the same pattern of incomplete or complete data. A total of 97 patterns are demonstrated on the graph. For example, Pattern 1 displays cases that have no missing values. Pattern 32, in turn, represents cases that have missing values on variable G_{22} (*comfort and commodity*) and G_{24} (*furniture)*. The graph is ordered from left to right regarding the number of missing values, with the goal to disclose monotonicity. Monotone missing data occurs when a participant stops responding at one point during the survey and does not complete any further questions. A way to conclude if data are monotone is to verify if all missing and nonmissing cells are adjacent, meaning that there are no isolated missing or nonmissing cells. As can be seen in Figure 15, there are isolated cells, therefore monotonicity is not present, which in turn, is a sign of randomness in the missing values.

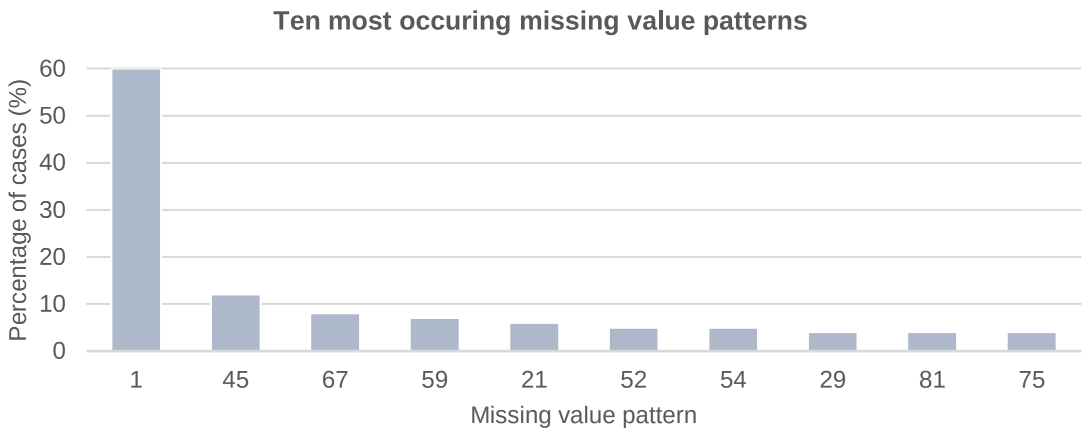


Figure 16.The output of most occurring missing value patterns analysis. Source: SPSS software.

Figure 16 presents the ten most occurring missing value patterns. It can be seen that Pattern 1, in which no missing values are present across all variables, is the most prominent. The nine remaining patterns have a reduced and similar percentage throughout them all. Coupling the results from Figure 15 and Figure 16, it is possible to conclude that there is randomness in terms of missing data, meaning that missing data does not follow a determined pattern, and the odds of having a bias in the sample is reduced. Comparing these conclusions to the three missing value categories earlier mentioned, Missing Completely at Random appears to be the category for which the dataset is best suited for. Following these assumptions, the literature preferred methods to treat MCAR data are analysed and discussed, in order to implement the most fitted. It is important to note that there are no stipulated guidelines on which is the best alternative to treat missing data in survey-based research, and so the findings are subjective and may be dependable of interpretation (Karanja et al., 2013).

Several methods can be used to handle missing data. However, the three most common approaches relay on deletion techniques, single imputation, and multiple imputation. Given the fact that the missing values on our dataset are considered MCAR, multiple imputation techniques are not going to be further discussed or implemented, since they only provide valid results under the MAR condition (Little and Rubin, 2002; Patrician, 2002).

Deletion techniques

Deletion techniques are the most traditional missing data techniques. The most common one, *listwise deletion*, removes every respondent with missing data, restricting the analysis only to complete cases. The advantage of using the approach is that it constructs a complete data set, which allows the use of standard analysis techniques. However, it presents multiple disadvantages. Removing all incomplete cases causes a serious sample reduction, that can lead to a lack of power of stability tests, and misleading results when a large portion of data is dismissed (Acock, 2005; Baraldi and Enders, 2010).

Schafer (1997) states that this method is only valid for datasets with less than 5% missing cases. Bennett (2001) advocates for a 10% cut off point, while Peng et al. (2006) suggests that a 20% cut off point is not unusual but any values beyond this can be problematic. When applying this method to our dataset, the sample size is reduced from 251 patients to 93 patients, translating into a 37% cut off point, which is not supported by the literature. Furthermore, regarding the minimum sample size necessary to obtain reliable results, Kline (2011) proposes a ratio of 20 participants for each parameter. Schreiber et al. (2006) recommend a ratio of 10 participations per parameter, and Bentler and Chou (1987) suggest a ratio of five to one. Given that our model contains 52 criteria, it can be concluded that even the original sample might not have the ideal size according to the literature, so, implementing listwise deletion and removing more than 100 participants is not recommended.

The second deletion technique is *pairwise deletion*. It consists of using the largest set of available cases to estimate the parameter of interest. Incomplete cases are used on an analysis-by-analysis basis, in a way that one case can be utilized in one analysis but not on others. On the one hand, this approach has an advantage over listwise deletion because it minimizes the reduction of the sample size (Bennett, 2001). On the other hand, interpretation difficulties may arise since different sample sizes are applied to each static, what can lead to inconsistent correlations and a non-positive definite covariance matrix (Kim and Curry, 1977; Graham, 2009; Karanja et al., 2013). The latter error occurred when implementing this technique in our sample in the SPSS software. Because of this, when dealing with factor analysis, SEM and OLR (methods performed with SPSS software) pairwise deletion is not possible to implement and a different deletion technique has to be applied. However, since MUSA is performed in MATLAB, it is possible to apply pairwise deletion to this specific method. Instead of employing a single imputation technique, it is preferable to use pairwise deletion in this method. Due to its sensitivity, imputing fictitious numbers to fill the gaps, increases the bias present in the dataset and is not recommended.

Single imputation techniques

Single imputation techniques replace the missing data with suitable estimations. Mean imputation substitutes missing values with the arithmetic mean of the collected data. This technique can only be applied when less than 10% of vales are missing, what is verified in our data sample (Patrician, 2002; Karanja et al., 2013). It was concluded that mean substitution was the best-suited method to handle missing data since it does not further reduce the sample size and is adequate given the proportion of the missing values. For factor analysis, SEM, and OLR this is the chosen imputation technique.

Table 5 shows the number of valid answers for each criterion/subcriterion after the implementation of pairwise deletion, that is, the MUSA database. As one can conclude *volunteering staff* is the criterion that suffered the biggest reduction, with the removal of more than 100 answers. Despite this, a sample size of 145 is still considered robust enough to proceed with the analysis.

Table 5. Number of valid answers for each criteria G_j, and subcriteria g_{jk}, $j=1,...,11$ and $k=1,..., n_j$.
Source: the author.

Criteria	Subcriteria	Number of valid answers
Obtained information [G_1]	Patient's guide [g_{11}]	232
	Patient's rights and duties [g_{12}]	232
	Complaint means [g_{13}]	217
	Substitution in decision making [g_{14}]	200
	Anticipated vital will [g_{15}]	168
	Global	*160*
Accommodations' quality [G_2]	Cleanliness [g_{21}]	251
	Comfort and commodity [g_{22}]	250
	Privacy [g_{23}]	248
	Furniture [g_{24}]	249
	Noise [g_{25}]	249
	Temperature [g_{26}]	250
	Entertainment [g_{27}]	247
	Global	*238*
Visits [G_3]	Visitation hours [g_{31}]	239
	Visit duration [g_{32}]	240
	Number of visits [g_{33}]	237
	Easy access for close relatives [g_{34}]	239
	Global	*233*
Food Quality [G_4]	Preparation, etc. [g_{41}]	247
	Variety [g_{42}]	245
	Quantity [g_{43}]	247
	Meal support [g_{44}]	243
	Global	240

Criteria	Subcriteria	Number of valid answers
Medical staff [G_5]	Availability [g_{51}]	249
	Attention [g_{52}]	249
	Kindness [g_{53}]	249
	Information regarding patient's health state [g_{54}]	249
	Information regarding medical treatment [g_{55}]	250
	Information regarding medical exams [g_{56}]	246
	Health advising and teaching [g_{57}]	246
	Global	*238*
Nursing staff [G_6]	Availability [g_{61}]	250
	Attention [g_{62}]	251
	Kindness [g_{63}]	249
	Information of patient's health state [g_{64}]	251
	Information of nursing treatment [g_{65}]	250
	Health advising and teaching [g_{66}]	247
	Global	*246*
Auxiliary staff [G_7]	Availability [g_{71}]	247
	Attention [g_{72}]	247
	Kindness [g_{73}]	248
	Performance efficiency [g_{74}]	248
	Global	*247*

Criteria	Subcriteria	Number of valid answers
Administrative staff [G_8]	Availability [g_{81}]	208
	Attention [g_{82}]	207
	Kindness [g_{83}]	209
	Performance efficiency [g_{84}]	208
	Global	*206*
Volunteering staff [G_9]	Availability [g_{91}]	145
	Attention [g_{92}]	145
	Kindness [g_{93}]	145
	Global	*145*
Exams and treatments [G_{10}]	Availability [g_{101}]	237
	Attention [g_{102}]	238
	Kindness [g_{103}]	237
	Information regarding patient's health state [g_{104}]	229
	Information regarding medical treatment [g_{105}]	228
	Information regarding medical exams [g_{106}]	232
	Health advising and teaching [g_{107}]	227
	Global	222
Discharge process [G_{11}]	Homecare provided information [g_{111}]	221
	Waiting time after discharge [g_{112}]	221
	Global	*219*
Global satisfaction [G12]		221

7. Results discussion and implications

7.1. Factor Analysis results

Despite already having a defined structure, exploratory factor analysis was performed to assess if a new structure could explain the relationship among data. The emergence of a new structure might be an indicator of a need for revaluation and restructuring of the survey. This problem is, however, outside of the scope of this thesis and is not further developed. Factor analysis was performed on SPSS software through the principal components' method along with a varimax rotation. Bartlett's sphericity test demonstrated a p-value<0.001, meaning that the null hypothesis was rejected and factor analysis to extract components could proceed. In order to know the number of factors to retain, conditions explained in the previous section, methodology, were verified. The Guttman-Kaiser rule was followed, meaning that only factors with eigenvalues higher than one were selected. Communalities present high values, which means that the extracted components are a suited representation of each criterion. Table 6 shows the explained variance of the nine components, 82.558%, throughout the three stages of the analysis, proving there is no significant loss of information from the original variables. Despite having the same cumulative percentage, when comparing the total variance, from the unrotated factor model to the rotated one, it can be noticed that it is more equally spread through the factors, which means that the rotated matrix is easier to interpret having a more uniform coupling of the subcriteria into components.

Table 6. Total variance explained.

	Initial eigenvalues			Extraction sum of squared loadings			Rotation sums of squared loadings		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	30.523	57.591	57.591	30.523	57.591	57.591	10.311	19.454	19.454
2	3.028	5.712	63.304	3.028	5.712	63.304	6.888	12.997	32.451
3	2.525	4.765	68.069	2.525	4.765	68.069	5.158	9.733	42.184
4	2.432	4.588	72.657	2.432	4.588	72.657	4.824	9.102	51.286
5	1.966	3.710	76.366	1.966	3.710	76.366	4.277	8.070	59.356
6	1.400	2.641	79.007	1.400	2.641	79.007	3.854	7.271	66.627
7	1.242	2.344	81.351	1.242	2.344	81.351	3.595	6.784	73.411
8	1.167	2.202	83.553	1.167	2.202	83.553	3.357	6.333	79.744
9	1.002	1.891	85.444	1.002	1.891	85.444	3.021	5.700	85.444

The rotated component matrix shows the rotated component loadings (represent how the variables are weighted per component), as well as the correlations between variables and the said component. Results presented in Table C.1. (Appendix C) show a clear grouping of subcriteria into components with meaningful loadings. According to the rotated component matrix, it is possible to conclude that subcriteria can be grouped into nine components, instead of the eleven originally presented on the questionnaire. The new components are broken down, in Table 7, showing the junction between the two sets of criteria that form component five and component six.

Table 7. Components resulted from the PCA.

Component	Criteria
G_1	Obtained information
G_2	Accommodation
G_3	Visits
G_4	Food
G_5	**Medical services** (Medical staff + Discharge process)
G_6	**Health staff** (Nursing staff + Auxiliary staff)
G_7	Administrative staff
G_8	Volunteering staff
G_9	Exams and treatments

The new coupling of *medical staff* with *discharge process* translates the way patients evaluate doctors depending on the perceived quality of their indications on the discharge process. There may be a lack of sufficient knowledge to differentiate both professionals in many situations, regarding *nursing and auxiliary staff*.

The adequacy of the analysis was evaluated by multiple coefficients, as explained in section 5.3, with the results presented in Table 8.

Table 8. Adequacy analysis.

Adequacy measure	Value	Meaning
ANOVA	0.280	There is not a significant difference between the two genders;
Cronbach's alpha	0.964	Excellent internal consistency;
ICC	0.963	Excellent reliability of measurements;
PCC	0.01	Correlations significative at the 0.01 level;
SCC	0.01	Correlations significative at the 0.01 level;
KMO	0.936	Excellent adequacy of the sample;
Mann–Whitney U test	p-value>0.050	There is not a significant difference between the two genders;
Independent t-test	Males (mean=6.634; standard deviation=0.866); Females (mean=6.629; standard deviation=0.808); t(215)=0.041; p-value=0.909;	Despite male patients being slightly more satisfied, the null hypothesis is not rejected since there is not a significant difference between the two genders;

As can be seen in Table 8, the sample demonstrates good adequacy and consistency throughout all coefficients. Regarding the possibility of analysing the sample in two distinct groups, being the patients divided by gender, it is proven by ANOVA, Mann-Whitney U and independent t-test that there are no significant differences that justify the division of data.

An alternative analysis with the initial structure of eleven criteria is also performed. In this analysis, the dataset did not undergo EFA and criteria are treated as observable variables, using patient's survey responses, as opposed to constructs. To simplify the reckoning of each analysis, analysis A is assigned to the database that resulted from EFA, containing nine criteria. Analysis B refers to the analysis of the original dataset that did not undergo EFA, including the eleven original criteria.

7.2. Structural Equations Modeling results

In this section, the results of the SEM method are presented for both analyses. Despite differences in implementation, an estimation model, common to both analyses has to be elected, and given the alternatives previously explored, ML is considered the best option. Since this method can only be applied to data with multivariate normal distribution, Skewness and Kurtosis coefficients are verified. With no value violating the stipulated intervals (|Skewness|<3 and |Kurtosis|<10), it is found that ML can be applied to the data (Marôco, 2014).

7.2.1. Analysis A

From the rotated component matrix, a SEM path diagram was created on Amos SPSS software. The initial path diagram includes nine latent variables and 54 observable variables. Their relationship flows from the construct (latent variable) to the indicator (observable variable), making this a reflective model. This type of model acknowledges that the construct causes indicators and that latent constructs exist independent of the indicators used (Freeze et al., 2007). It is possible to remove indicators without changing the construct's conceptual meaning in a reflective model (Rossiter, 2005). Each observable variable is paired with a measurement error term that represents unknown variability sources that are not considered in the model (Murti, 2016). Regarding construct validity, 'classical test theory', such as reliability and stability testing is appropriate for this type of model (Jarvis et al., 2003; MacKenzie et al., 2005). Aligned with these conclusions, a stability analysis, with results present in Table 9, was performed to assess if the initial model was adequate.

Table 9. Stability analysis from the original model (Analysis A).

Goodness of fit measure	Value	Criteria
X^2	5925.539	
X^2/df	4.774	<5
GFI	0.514	< 0.8 poor adjustment
RMSEA	0.123	>0.1 poor adjustment
CFI	0.793	<0.8 poor adjustment

NFI	0.748	<0.8 poor adjustment
PCFI	0.735	>0.6 good adjustment
PGFI	0.459	<0.6 poor adjustment
PNFI	0.698	>0.6 good adjustment

As it can be noticed from Table 9, the stability analysis demonstrates poor adjustment with only three measures displaying fair adjustment (X^2/df, PCFI, PNFI). The inflated X^2 value shows that the null hypothesis was rejected and that the covariance matrix of the sample differs from the covariance matrix estimated by the model. The number of degrees of freedom states that there are approximately four independent alternative model structures that do not disrupt any constraints. GFI has a value of 0.514 meaning that approximately 0.500 of the covariance, present in observable variables, is not explained by the model and a RMSEA of 0.123 indicating poor adjustment. CFI and NFI compare the fit of the interest model to the fit of a null model. With a value of approximately 0.800 for both coefficients, it indicates that the model of interest improves the fit by 80% in relation to the null model. PCFI is a parsimonious index that evaluates the complexity of the model. The more complex the model, the lower the fit index. With PCFI and PNFI demonstrating good adjustment it is possible to conclude that the model is presented in a simple form. However, when observing PGFI, which is also a parsimonious index that favours the simplicity of the model, poor adjustment is underlined. These two different outcomes lead to an inconclusive result regarding the simplicity of the model. With a majority of indices demonstrating poor adjustment, some alterations were made.

Firstly, the existence of outliers was evaluated through *Mahalanobis distance*. In total, fourteen observations presented high distance values (>50) and p_1 and p_2 <0.001. These observations were removed from the dataset. A stability analysis followed these modifications. However, the indices still demonstrated poor adjustment, and further alterations were necessary.

Secondly, modification indices were assessed. As a general rule, modification indices with values superior to eleven should be appraised (Marôco, 2014). Subcriteria *attention of auxiliary staff and kindness of auxiliary staff* (G_{72} and G_{73}, respectively) showed modification indices of around 200. With such high modification indices, removal of the subcriteria was recommended. Since these subcriteria were grouped into component *health staff*, still composed of eight other subcriteria, reliability of the component was maintained after their elimination. Despite these adjustments, modification indices' values were still superior to what was desired, translating the model need for changes. Covariances between measurement errors were, thus, established according to their modification indices. When a third stability analysis was performed in Amos software, results still showed poor adjustment.

From there, subcriteria *information provided by medical staff regarding patient's health state* and *kindness of professionals from exams and treatments* (G_{54} and G_{103}) were removed due to model discrepancies (inflating X^2, and decreasing GFI, CFI, PGFI and PCFI). These subcriteria belong to a large component and were redundant in the face of all available information. After this, covariances were, once more, established according to modification indices. As previously mentioned, covariances between measurement errors of subcriteria within the same component are understandable, and easy to interpret and to explain, since redundancy can naturally be present in these cases. A fourth, and final, stability analysis was performed, demonstrating reasonable adjustment throughout the majority coefficients indicating goodness of fit and adequacy of the model. Results from the four analyses are displayed in Table 10.

Table 10. Summary of stability analyses results.

Goodness of fit measure	**Values for respective stability analyses**			
	First analysis	**Second analysis**	**Third analysis**	**Fourth analysis**
X^2	5925.539	5872.269	3364.898	2885.310
X^2/df	4.774	4.445	3.093	2.567
GFI	0.514	0.521	0.637	0.690
RMSEA	0.123	0.117	0.091	0.081
CFI	0.793	0.795	0.877	0.912
NFI	0.748	0.751	0.829	0.864
PCFI	0.735	0.740	0.812	0.608
PGFI	0.459	0.468	0.605	0.837
PNFI	0.698	0.751	0.767	0.793

It is possible to infer that, on the final stability analysis, most coefficients demonstrate good or excellent adjustment. On the one hand, the reason why some coefficients still demonstrate a “not so good” adjustment can be due to the patient’s false perception regarding the service provided, which can lead to data bias. Patients’ false perception can be understood, for example, as the misinterpretation of the indications given by the doctors or the lack of empathy felt by the patient. On the other hand, evaluation of the coefficients is a subjective task since there are no empirical proves of the optimal values. When considering a GFI of 0.683, it means that approximately 0.300 of the covariance, present in observable variables, is not explained by the model, which is not an elevated value. The value of X^2 remains inflated, thus, the null hypothesis is still rejected. Nonetheless, there is a lot of criticism surrounding the effectiveness of this measure. For instance, Marôco (2014) states the X^2 test is useless because it tests a hypothesis that is not credible. Evaluating the adjustment of a model, expecting it to be perfect is a false assumption because every model has some extent of the error. After careful consideration and in alignment with the results provided by the stability analysis, outputs generated by SEM were validated. The final path diagram includes nine latent variables and 50 observable variables and is broken down by criteria in order to provide a simpler graphical visualization. In the end of this section, the total path diagram is displayed along with standardized regression weights of the most influential criteria.

Figure 17 represents the component *obtained information.* Measurement errors of the subcriteria *patient’s guide and patient’s rights and duties* (G_{11} and G_{12}, respectively), are connected through a covariance relationship that can be explained by the ambiguity of the own subcriteria since the patient’s guide can already include the rights and duties. Standardized regression weights tell us that *substitution in decision making (G_{14})* is the most important subcriterion in these criteria, whereas *anticipated vital will* (G_{15}) is the subcriterion with the smallest weight, therefore being the least substantial.

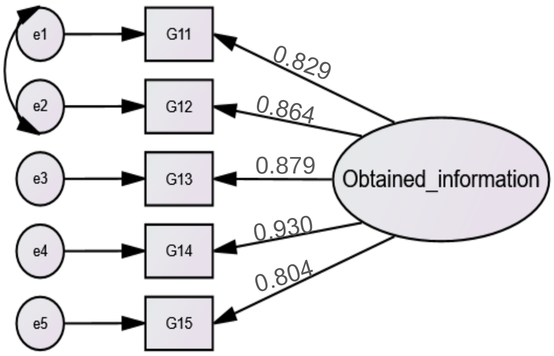


Figure 29. Path diagram of component "obtained information", with subcriteria G_{11}: patient's guide; G_{12}: Patient's rights and duties; G_{13}: complaint means; G_{14}: substitution in decision making; G_{15}: anticipated vital will. Source: SPSS software.

Component *accommodations* is represented in Figure 18. Covariance is present between measurement errors of subcriteria *comfort and commodity*, and *privacy* (G_{22} and G_{23}, respectively), highlighting the fact that the more privacy provided to the patient, the more comfortable he/she feels. The standardized regression weights show that *comfort and commodity* (G_{22}) is the subcriterion with the highest weight, as opposed to *entertainment* (G_{27}) that appears to be the least impactful subcriterion within this cluster.

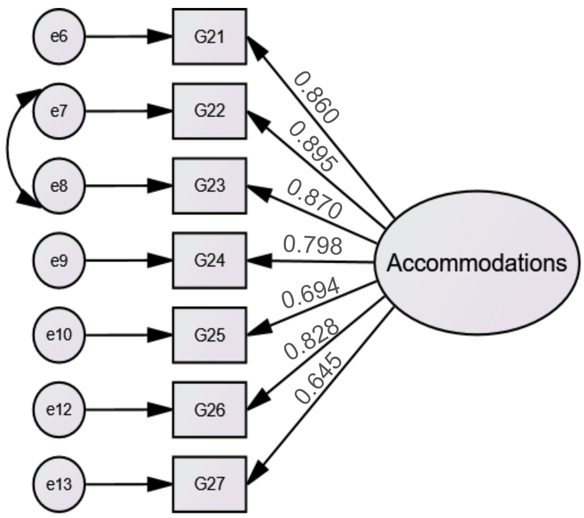


Figure 49. Path diagram of component "accommodations", with subcriteria G_{21}: cleanliness, G_{22}: comfort and commodity, G_{23}: privacy, G_{24}: furniture, G_{25}: noise, G_{26}: temperature, G_{27}: entertainment. Source: SPSS software.

In figure 19, component *visits* is presented with the four subcriteria that integrate it, along with the respective measurement errors. There are no covariances between measurement errors. *Visit duration* (G_{32}) is the most important subcriterion, having the highest standardized regression weight, and *easy access for close relatives* (G_{33}) is the least influential subcriterion.

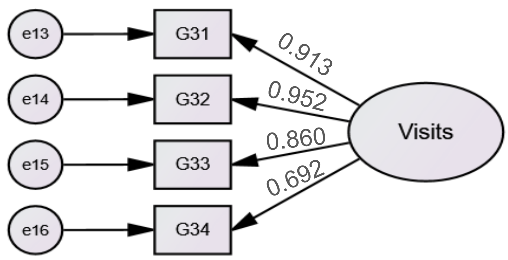


Figure 69. Path diagram of component "visits", with subcriteria G_{31}: visitation hours; G_{32}: visit duration; G_{33}: number of visits; G_{34}: easy access for close relatives. Source: SPSS software.

Figure 20 represents the component *food quality.* Measurement errors of subcriteria *preparation* and *variety of food (*G_{41} and $G_{42,}$ respectively) have a covariance that can be explained by the fact that when there is a wider variety of food from each food group, the preparation of the meal took longer and was well planned. According to standardized regression weights, *quantity of food* (G_{43}) is the paramount subcriterion, contrarily *to preparation, appearance, temperature, and taste* (G_{41}).

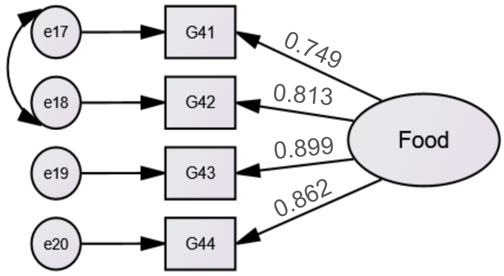


Figure 89. Path diagram of component "food", with subcriteria G_{41}: Preparation, appearance, temperature, taste; G_{42}: variety; G_{43}: quantity; G_{44}: meal support. Source: SPSS software.

In Figure 21*,* it is possible to observe two covariances. The first is between the measurement errors of *information regarding medical exams provided from medical staff* and *health advising and teaching from medical staff* (G_{56} and G_{57}, respectively). These two are correlated because health advising is a direct consequence of the results of medical exams, thus when the patient is gathering information provided by the doctor, these two variables can be intertwined. The second pair of measurement errors presenting a covariance relationship is *homecare provided information* and *waiting time after discharge* (G_{11} and G_{12}, respectively), both comprehended under *discharge process*. *Attention of medical staff* (G_{52}) is the most relevant subcriterion, having the highest regression weight. Differently, *waiting time after discharge* (G_{112}) is the least relevant subcriterion, with the lowest regression weight.

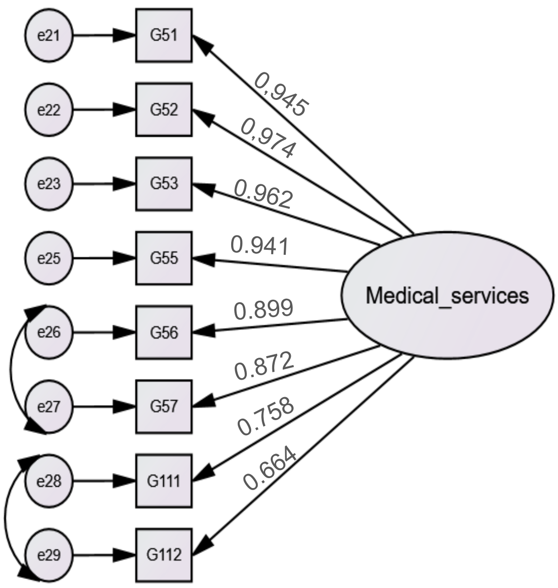


Figure 109. Path diagram of component "medical services", with subcriteria G_{51}: availability of medical staff; G_{52}: attention of medical staff; G_{53}: kindness of medical staff; G_{55}: information regarding medical treatment; G_{56}: information regarding medical exams; G_{57}: health advising and teaching from medical staff; G_{111}: homecare provided information; G_{112}: waiting time after discharge. Source: SPSS software.

Figure 22 displays component *health staff*, in which there are two pairs of covariance relationships present. *Information regarding patient's health state provided from nursing staff* and *information regarding nursing treatment* (G_{64} and G_{55}, respectively) is the first pair. Once again, nursing treatment is a direct consequence of the patient's health state. Provided information regarding these two variables may not have a clear and direct separation point, making it difficult for the patient to make a distinction when evaluating each variable. *Availability* and *performance efficiency of auxiliary staff* (G_{71} and G_{74}*, respectively)* form the second pair of covariance relationships. Chores of an auxiliary employee are not highly specialized, which means that availability has a big impact on the evaluation of the service provided by these professionals, nonetheless on the performance efficiency, because being available means being able to provide more help to patients. Respecting the standardized regression weights, *availability of nursing staff* (G_{61}) is the most critical subcriterion, whereas *health advising and teaching provided by nursing staff* (G_{66}) has the least importance.

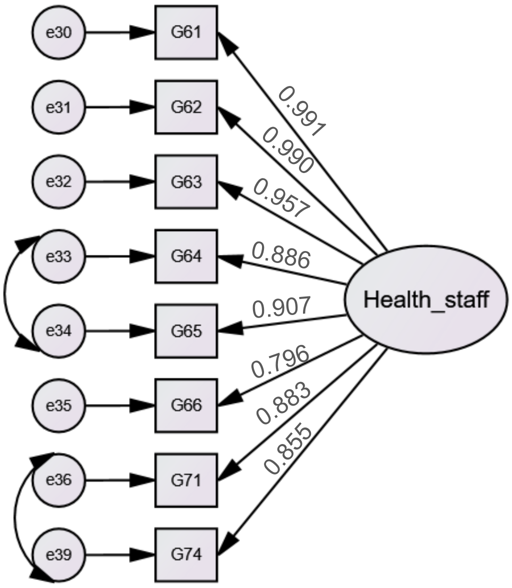


Figure 129. Path diagram of component "health staff", with subcriteria G_{61}: availability of nursing staff; G_{62}: attention of nursing staff; G_{63}: kindness of nursing staff; G_{64}: information regarding patient's health state; G_{65}: information regarding nursing treatment; G_{66}: information regarding nursing exams; G_{67}: health advising and teaching; G_{71}: availability of auxiliary staff; G_{74}: performance efficiency. Source: SPSS software.

On Figure 23, component *administrative staff* is presented. There are no covariances between measurement errors. *Attention of administrative staff* (G_{82}) appears as the most significant subcriterion contrasting with *availability of administrative staff* (G_{81}).

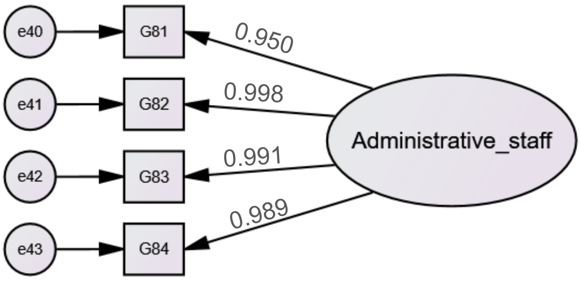
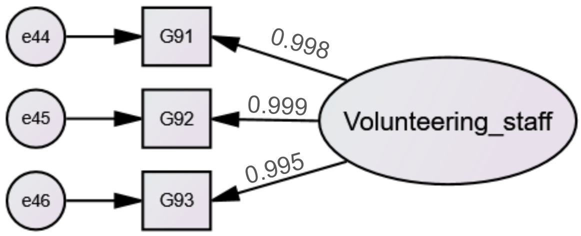


Figure 149. Path diagram of component "administrative staff" with subcriteria G_{81}: availability of administrative staff; G_{82}: attention of administrative staff; G_{83}: kindness of administrative staff; G_{84}: performance efficiency of administrative staff. Source: SPSS software.

On Figure 24, component volunteering staff is displayed. There are no covariances between measurement errors, and the standardized weight regressions demonstrate that *attention of volunteering staff* (G_{92}) has the highest relevance.



On Figure 25, component *exams and treatments* is presented with the six subcriteria that form it, as well as their respective measurement errors. There are no covariances between measurement errors. Subcriterion G_{103}, as previously mentioned, has been removed and does not appear in the representation of the path diagram. *Information regarding medical treatment (G_{105})* appears to be the most important subcriterion*.* *Availability of exams and treatments* (G_{101}), in its turn, is the least influential subcriterion.

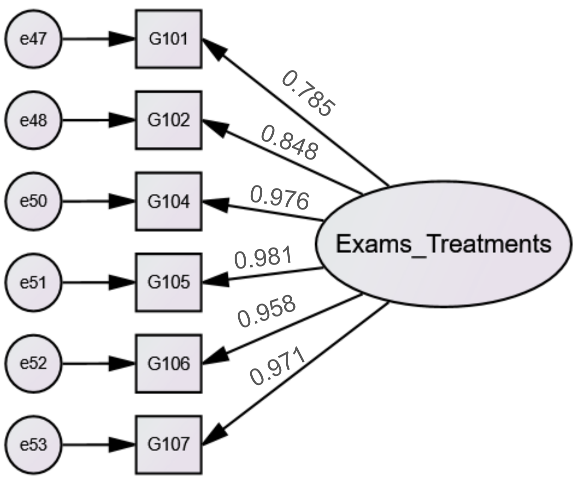


Figure 181. Path diagram of component "exams and treatments", with subcriteria G_{101}: availability of exams and treatments; G_{102}: attention of exams and treatment; G_{104}: information regarding patient's health state; G_{105}: information regarding medical treatment; G_{106}: information regarding medical exams; G_{107}: health advising and teaching. Source: SPSS software.

After the establishment and explanation of these relationships, the model's structure was finalised and is displayed in Figure 26. It is concluded that four criteria influence patient satisfaction, given their positive standardized regression weights and statistically significant p-values.

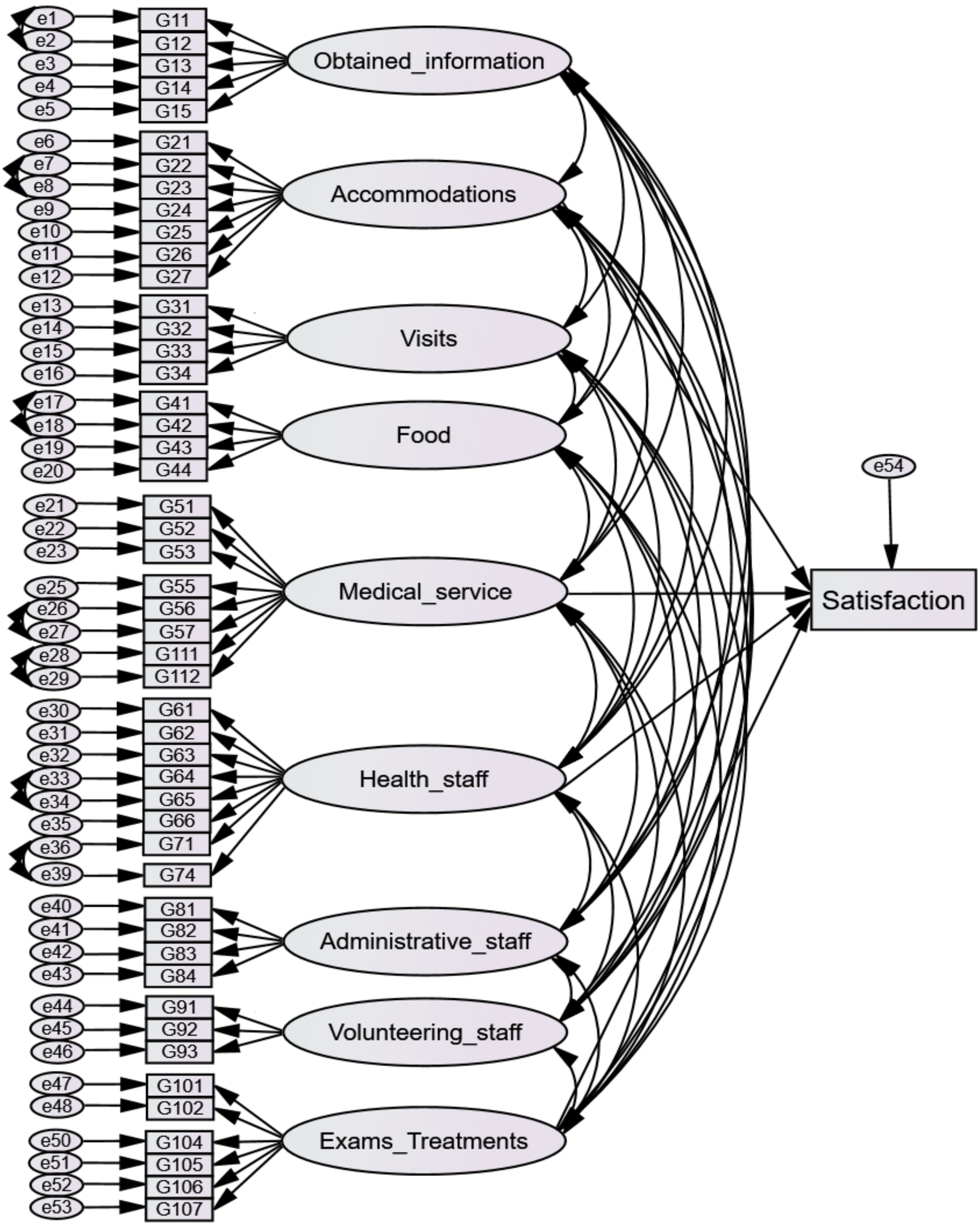


Figure 201. Complete path diagram (Analysis A). Source: SPSS software.

Accomodations is the component with the highest loading, thus, being the one that most influences patient satisfaction. It is followed by *exams and treatments*, *medical services* and lastly, *health staff.* Detailed information concerning the components is provided in Table 11.

Table 11. Detailed information about components (Analysis A).

Component	Component loading (standardized estimates)	Significance level
Accommodation	0.329	<0.001
Exams and treatments	0.277	<0.001
Medical services	0.202	<0.001
Health staff	0.192	0.004

7.2.1. Analysis B

For this analysis, an alternative SEM model with 65 observable variables was designed on AMOS SPSS software. Since this model only contains observable variables, it is considered a path analysis model where an observable construct (endogenous variable) has a linear relationship with two or more observable indicators (exogenous variables). This type of relationship between variables is a characteristic of a formative model. Opposed to a reflective model, as seen on the previous analysis, a formative model views criteria as endogenous variables that are determined by their respective subcriteria, exogenous variables (Freeze et al., 2007). As a direct consequence, removal of an indicator might alter the construct itself, therefore it is not recommended (Diamantopoulos, 1999). In formative models, error measurements are not associated with indicators. Errors are associated with the construct. A problem related to this type of model is the measure of reliability and consistency. There is no universally valid measure to assess the reliability of formative variables (Coltman et al., 2008). Marôco (2014) states that the reliability of the model should be assessed based on the R^2 coefficient. This coefficient estimates the fraction of variability of the dependent variable that is explained by the model. Despite considering criteria as dependent variables, the assessment of R^2 is done through the dependent variable *patient satisfaction*. Firstly, the existence of outliers is verified through *Mahalanobis distance*. In total, thirteen observations presented high distance values (>50) and p_1 and p_2 <0,001. These observations were removed, and the analysis proceeded with a total of 237 observations. Since no stability analysis can be performed in this model, R^2 is verified to assure the model is valid. The dependent variable *satisfaction* has a R^2 value of 0.645, which indicates that the model is adequate. Therefore, no further measures to increase reliability were explored. The final path diagram is presented in Figure 27 with the respective standardized regression weights.

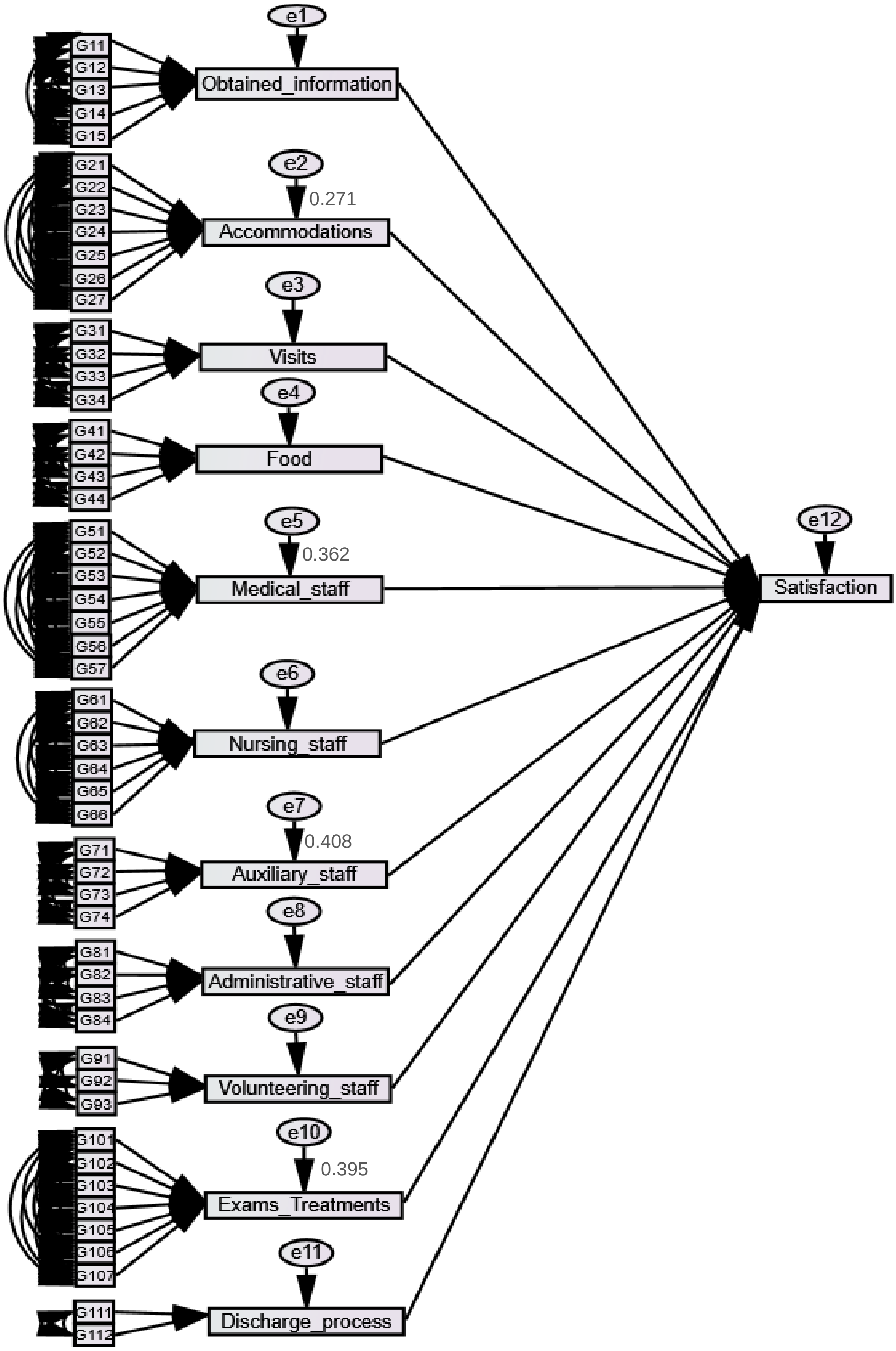


Figure 221. Complete path diagram (Analysis B). Source: SPSS software.

From this analysis, it is concluded that four criteria influence patient satisfaction. *Auxiliary staff* is the most influential criteria, followed by *exams and treatments, medical staff*, and *accommodations.* These criteria have positive standardized regression weights and statistically significant p-values, as seen in Table 12.

Table 30. Detailed information about components (Analysis B).

Component	Component loading (standardized estimates)	Significance level
Auxiliary staff	0.408	<0.001
Exams and treatments	0.395	<0.001
Medical staff	0.362	<0.001
Accommodations	0.271	<0.001

7.2.2. Results comparison

When comparing results from both analyses, some differences arise. From analysis A, *medical services (medical staff + discharge process)* is considered influential. However*, discharge process* is not seen as influential on analysis B. The same happens for *nursing staff*. In analysis A*, health staff (nursing staff + auxiliary staff)* is treated as a determinant of patient satisfaction, yet in the results of analysis B, *nursing staff* is not mentioned. The remaining criteria are common to both analyses, but some have different ranking positions. *Exams and treatment* and *medical staff* are on the second and third position, respectively, on both scenarios. *Accommodations* is the most important criteria in analysis A but is ranked last on analysis B. Contrarily, *auxiliary staff* is named as the most influential factor on analysis B, yet on analysis A it is the least influential of the four. Since in analysis A criteria were treated as latent variables, their values were created by SEM and do not correspond to the real values used in analysis B. The differences that emerge might be due to misjudgements attributed, by SEM, to latent constructs. In conclusion, both analyses demonstrated rather similar results. Thus, criteria deemed as influential according to SEM, and common to the two analyses, are *accommodations, auxiliary staff, exams and treatments, and medical staff.*

7.3. Ordinal Logistic Regression results

This section contains the results of OLR executed through the Polytomous Universal Model (PLUM) procedure on the SPSS software. The output of the regression is displayed for both analyses, including adequacy coefficients and parameter estimates, such as log odds and significance level, along with the odds ratio of each variable.

7.3.1. Analysis A

Results of an OLR between the dependent variable (*patient satisfaction*) and the nine latent constructs, provided by AMOS through regression imputation, are presented below. From the likelihood ratio chi-square test, it is possible to conclude that the final model is a significant improvement in fit when compared to the null model with a significance level below 0.001. The Pearson and deviance tests show non-significant levels (1.000 each), indicating good model fit. Pseudo R-squared values were also evaluated and can be observed in Table 13.

Table 31. Pseudo R-squared coefficients (Analysis A).

Pseudo R^2	Value
Cox and Snell's R^2	0.563
Nagelkerke's R^2	0.692
McFadden's R^2	0.503

Since there are no specific guidelines on how to handle these coefficients, values above 0.500 are assumed to be indicators of good model fit. Despite the disparities of the values, all three demonstrate good model adequacy. To finish this analysis, parallel lines test was performed, showing a non-significant result of 0,983, indicating that the proportional odds assumption is verified. Once the assessment of model fit measures is done, parameter estimates, on Table 14, present the results of the ordinal logistic regression.

Table 32. Parameter estimates (Analysis A).

Location	Estimate (β)	Exp (β)	Standardized error	Significance level	95% confidence interval Lower bound	Upper bound
Obtained information	-0.170	0.844	0.220	0.440	-0.601	0.261
Accommodations	1.159	4.937	0.460	0.001	0.696	2.497
Visits	-0.061	0.941	0.185	0.743	-0.424	0.302
Food	0.660	1.934	0.358	0.065	-0.042	1.361
Medical service	0.789	2.201	0.313	0.012	0.175	1.402
Health staff	0.426	1.532	0.399	0.285	-0.355	1.208
Administrative staff	-0.145	0.865	0.231	0.532	-0.598	0.308
Volunteering staff	-0.110	0.896	0.196	0.570	-0.495	0.274
Exams and treatments	0.983	2.673	0.244	0.000	0.505	1.461

Accommodations appears as the most influential predictor of patient satisfaction. The proportional odds model shows the positive effect β=1.159 which is statistically significant (sig=0.001). The OR is exp(β)= 4.937, meaning that the odds of a patient being more satisfied increase by 4.937 for every unit increase on *accommodations*. The 95% confidence interval for the cumulative odds ratio shows that this increase is between $e^{0.696}$ and $e^{2.497}$. Since this interval (2.005; 12.460) does not contain the value 1, corresponding to the null hypothesis of independence, it indicates that *accommodations* influences *patient satisfaction.*

Exams and treatments is the second most influential predictor of the dependent variable, with an OR of 2.673 and statistical significance (sig=0.000). For every unit increase on *exams and treatments*, there is a predicted increase of 2.673 on the odds of *patient satisfaction.* From the 95% confidence interval ($e^{0.505}$ = 1.657; $e^{1.461}$= 4.310), it is concluded that the null hypothesis is rejected since the value 1 is not included on the interval, thus suggesting that this criterion is influential.

Medical service is the last predictor of patient satisfaction, with an OR of 1.532 and a significance level of 0.012. OR shows that for every unit increase in *medical service*, there is a predicted increase of 1.532 on the odds of *patient satisfaction.* The 95% confidence interval ($e^{0.175}$=1.191; $e^{1.402}$=4.063) also signals that this criterion is influential since it does not include the value 1. The remaining criteria are not statistically significant (sig>0.005), thus, are not considered influential regardless of their log-odds and OR.

7.3.2. Analysis B

An OLR for the dependent variable and the eleven initial criteria was also performed. The likelihood ratio chi-square test, with a significance level of 0.000, shows that the final model is a significant improvement in fit when compared to the null model. The Pearson test is statistically significant (sig<0.005), demonstrating a bad model fit. The deviance test is non-significant (sig>0.050), indicating good model fit. Because of these differences, no conclusions regarding the model fit can be retrieved from these coefficients. As previously mentioned, pseudo R-squared values have to be interpreted with caution, but values above 0.500, as seen in Table 15, signal good model fit.

Table 40. Pseudo R-squared coefficients (Analysis B).

Pseudo R^2	Value
Cox and Snell's R^2	0.561
Nagelkerke's R^2	0.689
McFadden's R^2	0.501

The final step of the adequacy analysis is the parallel lines test. This test returns a non-significant value of 0.994, thus, the proportional odds assumption is verified. With the adequacy analysis finalised, the outputs of the ordinal logistic regression are disclosed in Table 16.

Table 41. Parameter estimates (Analysis B).

					95% confidence interval	
Location	**Estimate (β)**	**Exp (β)**	**Standardized error**	**Significance level**	**Lower bound**	**Upper bound**
Obtained information	-0.037	0.964	0.180	0.835	-0.390	0.315
Accommodations	0.876	2.401	0.213	0.000	0.459	1.293
Visits	-0.078	0.925	0.180	0.664	-0.432	0.275
Food	0.044	1.045	0.171	0.799	-0.291	0.378
Medical staff	0.239	1.270	0.192	0.214	-0.138	0.615
Nursing staff	0.271	1.311	0.355	0.445	-0.425	0.968
Auxiliary staff	1.276	3.582	0.339	0.000	0.613	1.940
Administrative staff	0.095	1.100	0.262	0.716	-0.418	0.609
Volunteering staff	-0.161	0.851	0.201	0.424	-0.556	0.234
Exams and treatments	0.973	2.646	0.238	0.000	0.507	1.439
Discharge process	0.363	1.438	0.199	0.068	-0.027	0.753

According to Table 16, *auxiliary staff* is the criteria that most determines *patient satisfaction*. Exp(β) has a value of 3.582, meaning that for each unitary increase there is a predicted increase of 3.582 on the odds of *patient satisfaction*. From the 95% confidence interval it is concluded that the predicted increase is between $e^{0.613}$ (1.846) and $e^{1.940}$ (6.960). The value 1 is not comprehended on the confidence interval, rejecting the null hypothesis of independence, and assuring that *auxiliary staff* influences *patient satisfaction.*

E*xams and treatments* also influences *patient satisfaction*. With an OR of 2.646, for every unitary increase on *exams and treatments* there is a predicted increase of 2.646 on the odds of *patient satisfaction.* The 95% confidence interval for this cumulative odds ratio says that the increase is between $e^{0.507}$ and $e^{1.439.}$ This interval (1.660; 4.216) does not include the value 1, attesting for the influence of *exams and treatment*s on *patient satisfaction.*

The last influential criterion is *accommodations*. There is a predicted increase of 2.401 on the odds of *patient satisfaction* for each unit increase on accommodations. With a 95% confidence interval ($e^{0.459}$ = 1.582; $e^{1.293}$= 3.644), not containing the value 1, the null hypothesis of independence is rejected, and it is verified that *accommodations* influences *patient satisfaction.*

Given the statistically non-significant levels of the remaining criteria, they are treated as non-influential, despite their log-odds and OR.

7.3.3. Results comparison

On the one hand, analysis A states that *accommodations, exams and treatments*, and *medical service* are the predictors of patient satisfaction. On the other hand, analysis B finds *auxiliary staff, exams and treatments* and *accommodations* as the criteria that influence patient satisfaction. The main distinction between the two analyses is the presence of *auxiliary staff as* the most influential criterion for analysis B, despite not being considered influential in analysis A. This difference is also present on the results of the SEM method, and the same justification applies. *Accommodations* and *exams and treatments* are found influential on both analyses despite being in different positions. Once more, results from both analyses do not completely converge but can be seen as approximate. In lines with this, *accommodations* and *exams and treatments* are the criteria seen as satisfaction predictors by OLR.

7.4. Multicriteria Satisfaction Analysis results

The results from the MUSA method performed on MATLAB are presented in Table 17 and Table 18. Along with every subcriterion and criterion's weight, multiple indices are assessed to provide a deep insight into patients' preferences. Oppositely to what happened with SEM and OLR, for MUSA, results are not developed in two separate analysis (analysis A and analysis B). To put it simply, MUSA is only applied to the original database because this method cannot work with data returned by factor analysis. There is an exception to this rule, that is using of a categorical principal components analysis (CPCA), as seen in studies associated with satisfaction, such as Valle et al. (2011) and Vuković et al. (2012).

Table 49. MUSA main results.

Criteria	Subcriteria	Subcriteria weight	Criteria weight [0-1]	Overall subcriteria weight [0-1]	Satisfaction index [0-100%]	Demanding index [-1;1]	Average improvement index [0-100%]	Strategic improvement	Market Opportunities
G_1	g_{11}	0.1648		0.0340	3.1100	-0.0800	15.9675	2nd priority	Leverage Opportunity
	g_{12}	0.2383		0.0491	4.6500	-0.0280	22.7219	1st priority	Leverage Opportunity
	g_{13}	0.1521		0.0313	2.7900	-0.1200	14.7856	2nd priority	Leverage Opportunity
	global		0.1144		9.4400	0.0200	10.3601	1st priority	Leverage Opportunity
G_2	g_{21}	0.1036		0.0058	0.5300	-0.0500	10.3051	2nd priority	Status Quo
	g_{22}	0.1774		0.0099	0.8100	0.3100	17.5963	3rd priority	Status Quo
	g_{23}	0.1471		0.0082	0.6500	0.1900	14.6144	3rd priority	Status Quo
	g_{24}	0.2047		0.0114	0.9000	0.1700	20.2858	2nd priority	Status Quo
	g_{25}	0.1031		0.0057	0.4000	0.2200	10.2688	3rd priority	Status Quo
	g_{26}	0.1134		0.0063	0.5300	0.0100	11.2799	2nd priority	Status Quo
	g_{27}	0.1506		0.0084	0.7000	0.0100	14.9546	2nd priority	Status Quo
	global		0.0557		4.1600	0.2100	5.3383	3rd priority	Status Quo
G_3	g_{31}	0.3294		0.0297	2.1000	0.1900	32.2483	2nd priority	Leverage Opportunity
	g_{32}	0.1436		0.0130	0.9400	0.1400	14.2250	2nd priority	Status Quo
	g_{33}	0.3142		0.0283	2.3300	-0.1200	30.6879	1st priority	Leverage Opportunity
	g_{34}	0.2128		0.0191	1.6100	0.1000	20.9374	1st priority	Status Quo
	global		0.0902		7.7900	-0.0600	8.3173	1st priority	Transfer Resources
G_4	g_{41}	0.2501		0.0316	1.9500	0.3500	24.5223	2nd priority	Leverage Opportunity
	g_{42}	0.3534		0.0446	3.4900	0.0400	34.1066	1st priority	Leverage Opportunity
	g_{43}	0.2051		0.0259	1.7900	0.3800	20.1429	3rd priority	Leverage Opportunity
	g_{44}	0.1915		0.0242	1.9500	0.2700	18.7766	3rd priority	Leverage Opportunity
	global		0.1262		9.2500	0.3400	11.4527	2nd priority	Leverage Opportunity
G_5	g_{51}	0.0896		0.0072	0.5800	0.3300	8.9080	3rd priority	Status Quo
	g_{52}	0.1079		0.0087	0.6700	0.3900	10.7177	3rd priority	Status Quo
	g_{53}	0.1027		0.0083	0.6900	0.1800	10.1991	3rd priority	Status Quo
	g_{54}	0.1849		0.0149	1.0300	0.5300	18.2996	3rd priority	Status Quo
	g_{55}	0.1587		0.0128	0.9700	0.3900	15.7161	3rd priority	Status Quo
	g_{56}	0.2499		0.0201	1.4900	0.4400	24.6176	2nd priority	Action Opportunity
	g_{57}	0.1063		0.0086	0.6300	0.4300	10.5630	3rd priority	Status Quo
	global		0.0804		6.2600	0.4400	7.5367	2nd priority	Status Quo
G_6	g_{61}	0.1123		0.0100	0.8700	0.3900	11.1323	3rd priority	Status Quo

	g_{62}	0.1510		0.0135	1.2200	0.2300	14.9158	3rd priority	Status Quo
	g_{63}	0.1524		0.0136	1.2300	0.2000	15.0525	3rd priority	Status Quo
	g_{64}	0.2778		0.0248	2.1400	0.1900	27.1855	2nd priority	Leverage Opportunity
	g_{65}	0.1281		0.0114	1.0500	-0.1100	12.6755	2nd priority	Status Quo
	g_{66}	0.1783		0.0159	1.3000	0.2600	17.5982	3rd priority	Status Quo
	global		0.0892		7.9000	0.2500	8.2153	2nd priority	Transfer Resources
G_7	g_{71}	0.2439		0.0204	1.8200	0.1800	23.9461	2nd priority	Leverage Opportunity
	g_{72}	0.2640		0.0221	1.9400	0.2200	25.8878	2nd priority	Leverage Opportunity
	g_{73}	0.1843		0.0154	1.4400	-0.0900	18.1646	2nd priority	Status Quo
	g_{74}	0.3078		0.0257	2.4200	-0.2600	30.0351	1st priority	Leverage Opportunity
	global		0.0836		7.8100	-0.1500	7.7071	1st priority	Transfer Resources
G_8	g_{81}	0.2942		0.0257	2.0800	0.1700	28.8081	2nd priority	Leverage Opportunity
	g_{82}	0.3777		0.0330	2.7600	0.1400	36.7275	1st priority	Leverage Opportunity
	g_{83}	0.1495		0.0130	1.0900	0.0700	14.7870	2nd priority	Status Quo
	g_{84}	0.1785		0.0156	1.2600	0.2000	17.6251	3rd priority	Status Quo
	global		0.0873		7.5200	-0.1000	8.0735	1st priority	Transfer Resources
G_9	g_{91}	0.2913		0.0339	2.7300	0.2400	28.3348	2nd priority	Leverage Opportunity
	g_{92}	0.3457		0.0402	3.4300	0.0100	33.3842	1st priority	Leverage Opportunity
	g_{93}	0.3630		0.0422	3.6100	0.1000	34.9896	1st priority	Leverage Opportunity
	global		0.1163		9.0400	0.1100	10.5786	1st priority	Leverage Opportunity
G_{10}	g_{101}	0.1405		0.0105	0.8500	0.0800	13.9306	2nd priority	Status Quo
	g_{102}	0.1314		0.0098	0.8700	-0.0800	13.0257	2nd priority	Status Quo
	g_{103}	0.0912		0.0068	0.5600	0.2600	9.0689	3rd priority	Status Quo
	g_{104}	0.1343		0.0100	0.8000	0.2200	13.3226	3rd priority	Status Quo
	g_{105}	0.2052		0.0154	1.3300	0.0500	20.2471	1st priority	Status Quo
	g_{106}	0.1149		0.0086	0.7900	-0.2400	11.3992	2nd priority	Status Quo
	g_{107}	0.1825		0.0137	1.1500	0.1300	18.0401	2nd priority	Status Quo
	global		0.0748		6.1400	0.0800	0.0000	2nd priority	Status Quo
G_{11}	g_{111}	0.5221		0.0428	8.8600	0.0100	47.5842	1st priority	Leverage Opportunity
	g_{112}	0.4779		0.0391	3.0500	0.2800	46.3324	2nd priority	Leverage Opportunity
	global		0.0819		6.5700	0.3300	7.6519	2nd priority	Transfer Resources
Subcriteria centroid				0.0192	1.6900	0.1475	20.2284		
Criteria centroid			0.0909		6.2900	0.1130	6.9621		

Table 50. Kano's model applied to MUSA.

Criteria	Subcriteria	Dissatisfied patients	Satisfied patients	Kano's model category
G_1	g_{11}	0.0379	0.0336	Must-be, critical
	g_{12}	0.0379	0.0707	Highly attractive
	g_{13}	0.0347	0.0334	Must-be, critical
	global	0.1106	0.1377	Highly attractive
G_2	g_{21}	0.0068	0.0060	Must-be, necessary
	g_{22}	0.0120	0.0041	Must-be, necessary
	g_{23}	0.0125	0.0034	Must-be, necessary
	g_{24}	0.0157	0.0065	Must-be, necessary
	g_{25}	0.0147	0.0009	Must-be, necessary
	g_{26}	0.0100	0.0017	Must-be, necessary
	g_{27}	0.0100	0.0067	Must-be, necessary
	global	0.8168	0.2944	Must-be, necessary
G_3	g_{31}	0.0365	0.0168	Must-be, critical
	g_{32}	0.0181	0.0113	Must-be, necessary
	g_{33}	0.0238	0.0607	Highly attractive
	g_{34}	0.0199	0.0229	Less attractive
	global	0.9837	0.1117	Less attractive
G_4	g_{41}	0.0422	0.0163	Must-be, critical
	g_{42}	0.0379	0.0485	Must-be, critical
	g_{43}	0.0366	0.0106	Must-be, critical
	g_{44}	0.0310	0.0112	Must-be, critical
	global	0.1477	0.8661	Must-be, critical
G_5	g_{51}	0.0071	0.0022	Must-be, necessary
	g_{52}	0.0077	0.0031	Must-be, necessary
	g_{53}	0.0068	0.0036	Must-be, necessary
	g_{54}	0.0177	0.0023	Must-be, necessary
	g_{55}	0.0113	0.0049	Must-be, necessary
	g_{56}	0.0204	0.0045	Must-be, critical
	g_{57}	0.0101	0.0033	Must-be, necessary
	global	0.0811	0.0239	Must-be, necessary
G_6	g_{61}	0.0069	0.0131	Less attractive
	g_{62}	0.0085	0.0181	Less attractive
	g_{63}	0.0089	0.0134	Less attractive
	g_{64}	0.0126	0.0321	Highly attractive
	g_{65}	0.0067	0.0257	Less attractive
	g_{66}	0.0125	0.0241	Less attractive
	global	0.0561	0.1264	Less attractive
G_7	g_{71}	0.0173	0.0146	Highly attractive
	g_{72}	0.0189	0.0196	Highly attractive
	g_{73}	0.0126	0.0243	Less attractive
	g_{74}	0.0138	0.0450	Highly attractive
	global	0.0626	0.1035	Less attractive
G_8	g_{81}	0.0246	0.0266	One dimensional, high valued added
	g_{82}	0.0302	0.0415	Highly attractive
	g_{83}	0.0152	0.0238	Less attractive
	g_{84}	0.0176	0.0179	One dimensional, low value added
	global	0.0877	0.1099	Less attractive
G_9	g_{91}	0.0407	0.0330	Must-be, critical
	g_{92}	0.0423	0.0430	One dimensional, high value added
	g_{93}	0.0474	0.0364	Must-be, critical
	global	0.1305	0.1124	Must-be, critical
G_{10}	g_{101}	0.0114	0.0142	Less attractive
	g_{102}	0.0092	0.0134	Less attractive
	g_{103}	0.0082	0.0071	Must-be, necessary
	g_{104}	0.0116	0.0117	One dimensional, low value added
	g_{105}	0.0125	0.0149	Less attractive
	g_{106}	0.0083	0.0142	Less attractive
	g_{107}	0.0127	0.0180	Less attractive
	global	0.0737	0.0936	Less attractive
G_{11}	g_{111}	0.0333	0.0394	Highly attractive
	g_{112}	0.0366	0.0257	Must-be, critical
	global	0.0699	0.0650	Must-be, necessary

Obtained information: This criterion is composed of five subcriteria that translate the impact of providing valuable information regarding the internment service so that the patient feels at ease with what is expecting him. The results of MUSA revealed that *substitution in decision making*, and *anticipated vital will* (g_{14} and g_{15}, respectively) have weights equivalent to zero, and were thus, removed from the analysis. This criterion is seen as the third and last satisfaction predictor, weighting 11.44% (located above the criteria centroid). Having the highest satisfaction index out of all criteria (9.44%), it is safe to assume that patients are satisfied with this criterion. Regarding the remaining subcriteria, *patient's rights and duties*, g_{12}, has the highest overall weight (4.91%), but there is not a substantial difference between the weights of the three subcriteria. *Patient's guide* and *complaint means* (g_{11} and g_{13}, respectively) are critical must-be requirements, as a result of the weights associated with dissatisfied patients being higher than those associated with satisfied patients. The critical nature of the requirements is due to their overall weights (3.40% and 3.13%) being above the subcriteria centroid (1.92%). The fact that these subcriteria are considered as must-be requirements shows that patients view instructions on how to proceed during the hospital stay and the existence of channels to press a complaint as basic and vital service characteristics, being dissatisfied when their expectations are not met. *Patient's rights and duties* and *global obtained information* (g_{12} and G_1, respectively) are highly attractive attributes that share a considerable room for improvement and should be treated as leverage opportunities of first priority. Given that patients are not overly demanding when it comes to these subcriteria and criterion, it can be of great interest to develop strategies that enhance the benefits of providing valuable information, with a special concern for *patient's rights and duties* that is somehow seen as surprise element that leaves patients feeling even more satisfied.

Accommodations' quality: Comprised of seven subcriteria regarding hospital's infrastructures, this criterion is considered to be the least influential due to having the lowest criterion weight (5.57%) (this discovery goes against what has been found with the other methods, something that is further discussed) and simultaneously, the lowest satisfaction score (4.16%). From the Kano's model, all subcriteria and the criterion itself are necessary must-be attributes, meaning that patients view these as innate service characteristics that should exist independently of the circumstances. Patients are dissatisfied when these attributes do not exist but are not additionally satisfied when their expectations are met. Given their low satisfaction indexes (a fact that is in line with the conclusions retrieved from Table 4) and patient's demanding nature, some changes might be implemented to improve patient satisfaction. However, due to the low weighs of each subcriterion/criterion and the fact that there is little room for improvement, according to MUSA, it is not advantageous from a business point of view to implement alterations on these attributes since the outcome is likely to bring little to no benefits for the provider.

Visits: When considering the internment service, it is essential to ponder the importance visitations have on the well-being of the patient himself. Thus, evaluating patient satisfaction with this criterion is imperative to provide the best possible service. With a satisfaction index of 7.79% (above the criteria centroid), patients seem to be satisfied with the visitation's service. *Visitation hours* and *number of visits* (g_{31} and g_{33}, respectively) are the most significant subcriteria, having the highest overall subcriteria weights (2.97% and 2.83%, respectively), and are seen as leverage opportunities. These opportunities do not need to be assessed immediately but can be advantageous in a competitive scenario. On the one hand, the Kano's model tells us that *number of visits* and *easy access for close relatives* (g_{33} and g_{34}, respectively) are attractive requirements that generate patient satisfaction but are not an extreme necessity, given that their absence does not dictate patient dissatisfaction. They are seen as a bonus, especially *easy access for close relatives* that is highly attractive, and even more valued by patients. On the other hand, *visitation hours* and *visitation duration* (g_{31} and g_{32}, respectively) are seen as must-be requirements, posing as basic service characteristics that patients assume are inherent to the service provided. Patients do not feature a demanding nature for any of the subcriteria/criterion except for *visitation hours* that is, coincidentally, the subcriterion with the highest room for improvement in this set, while also being a considered a leverage opportunity. This demonstrates that a modification in the visitation hours, most likely an extension, would be deeply valued by the patients. Globally, this criterion has a low weight (9.02%), thus, not being considered as a predictor of patient satisfaction. It is also viewed as an opportunity to transfer resources to some other dimensions that might be more important and influential. With this said, it is prudent to only perform alterations to this set of attributes once resources have been applied to more important dimensions.

Food quality: When treating inpatients, the hospital is responsible for every aspect of care, and food quality cannot be dismissed. Hospital food is not usually associated with excellent quality or taste. However, when treating diseased patients, it is of paramount importance to provide food that can nourish the body in the best possible manner, customized according to each patient's needs. Contrarily to what was found with SEM and OLR, *food quality* is the most influential criterion according to MUSA. Patients are delighted with this criterion, presenting the second highest satisfaction index (9.25%). Since this entire set of attributes is composed of critical must-be requirements and given the high satisfaction indexes (not only of the criterion but also of the subcriteria), modifications to this specific area do not seem to be a priority. However, there is still some room for improvement given the high demanding nature that patients have towards these attributes, turning them into leverage opportunities that should be evaluated to optimize their benefits. Special attention should be given to *food variety* (g_{42}), the most influential subcriterion that is seen as first priority improvement. Despite displaying the highest satisfaction index (3.49%) out of the four subcriteria, the increased patient's demanding nature makes up a significant room for improvement that shall not be disregarded.

Medical staff: Despite the importance that medical staff has on overall patient satisfaction, it is no surprise that when taking into account the internment service (where doctors are not as present as the remaining health professionals), physicians are not the main source of satisfaction/dissatisfaction, reaching a weight of only 8.40% (below the criteria centroid). Patients appear rather dissatisfied with this criterion (and corresponding subcriteria), while also being extremely demanding and considering them as must-be requirements. This could be an indicator that a revision might be required, however, due to the low room for improvement, these attributes are classified as status quo opportunities that do not generate any benefits. *Information regarding medical exams* (g_{56}) is the subcriterion with the largest weight (2.01%) that presents a low satisfaction index with substantial room for improvement, coming up as a second priority action opportunity, where some alterations might be beneficial for the provider. Overall, patients view the service rendered by doctors as a granted attribute, becoming extremely dissatisfied when their prospects are not met. An increase in the efficiency of medical staff might be suitable to provide the patient with a better stay. Nonetheless, from a managerial perspective, alterations to this set are seen as low-valued added.

Nursing staff: Being one of the professionals who keep close contact with the patient, it would be expected of the criterion to have a high weight. However, with a weight of 8.92% (located below the centroid), it is concluded that this is not an influential criterion. The results for *nursing staff* are similar to the ones obtained for *medical staff*, with a slight difference in the satisfaction index. Overall, patients are satisfied with the work of the *nursing staff*, what is aligned with the results from Table 4. However, when looking into the subcriteria, solely *information regarding patient's health state* (g_{64}) has a satisfaction index that is placed above the centroid. Demanding indexes show that patients are not as critical when evaluating *nursing staff* as compared to *medical staff*, whilst still presenting a demanding nature for the majority of the attributes*. Information regarding patient's health state* (g_{64}), the most influential subcriterion in this set, presents a large room for improvement and is consequently viewed as a leverage opportunity of second priority, where modifications might be seen as beneficial. The Kano model labels this set of attributes as highly/less attractive because they display a large weight for satisfied patients and a small weight for dissatisfied patients. This type of result is not expected since *nursing staff* (and its respective subcriteria) should be characterized as one-dimensional or must-be requirements. This outcome might be a result of the mathematics that are inherent to the MUSA method. Since it optimizes weights and value functions by minimizing errors that can emerge from patient’s evaluations, it is possible to obtain weights that have no meaning (Ferreira et al., 2018).

Auxiliary staff: Auxiliaries are the personnel with which patients maintain the closest proximity, thus, extreme importance would be expected as was pointed out by SEM and OLR. However, using the MUSA method, patients seem to not allocate big importance to interpersonal relationships. The results of this set of attributes are identical to the results of *nursing staff* and *medical staff*, what once again, proves the devaluation of interpersonal care from part of the patients. With a weight of 8.36% (below the centroid), patients seem reasonably satisfied with this criterion (satisfaction index of 7.81 %). Given the low importance of this criterion, it might not be advisable to spend resources enhancing it. However, looking at the non-demanding nature of patients and the available room for improvement, it is suggested that, once all major alterations are performed, if there are still resources available, this set of attributes be treated as a first priority. Congruently to what was observed with nursing staff, the Kano model categorizes these attributes as highly/less attractive requirements. This type of result is not expected, since auxiliary staff is a cornerstone of the internment service, going against the definition of ‘pleasant surprise’ that is given to attractive requirements.

Administrative staff: Being responsible for the legal and financial aspects of the admission and discharge process, while having little to no contact with the patient himself, the influence of this criterion on patient satisfaction is predicted to be reduced. The latter is verified through the criterion weight of 8.73% (below the centroid). Patients seem to be satisfied with the service provided by *administrative staff*, although *kindness* and *performance efficiency* (g_{83} and g_{84}, respectively) did not completely satisfy patients. Given the non-demanding patient nature towards *attention of administrative staff* (g_{82}) and the increased room for improvement, the enhancement of this subcriterion is viewed as a first priority leverage opportunity. Adding the fact that *attention of administrative staff* (g_{82}) is a highly attractive requirement that patients perceive as a *bonus*, an upgrade in this area might be profitable. Globally, this criterion poses as a non-influential dimension that leaves patients feeling satisfied. Yet, at the same time, presents a reasonable room for improvement, that when coupled with patient's non-demanding nature and the fact that it is considered an attractive requirement that has the potential to leave patients feeling more satisfied, it might be perceived as an interesting option to create some changes if resources are available.

Volunteering staff: Albeit not having specific training or education, *volunteering staff* is recognized as a crucial part of the internment service, since they are the people who help, guide, and communicate with patients. The latter is verified with a criterion weight of 11.63%, making this the second most influential patient satisfaction predictor. Patients are satisfied with the service provided by these personnel. However, there is still an increased room for improvement, so these items are viewed as leverage opportunities that should be assessed with the respective priority. The Kano model classifies them as critical must-be requirements due to having a larger weight attributed by dissatisfied patients than by satisfied patients. *Attention of volunteering staff* (g_{92}), however, is classified as a high valued-added one-dimension requirement because satisfied patients allocate the same weight to this requirement as dissatisfied patients. This type of requirement leaves patients feeling satisfied when their expectations are met and dissatisfied, otherwise.

Exams and treatments: Exams and treatments are necessary tools to improve patients' health status. However, when asking an inpatient to evaluate his satisfaction with this criterion, complications may arise. This is because medical exams can be evasive procedures that leave no satisfaction in the patient, despite being completely necessary, making it hard for the patient to provide an unbiased review. Medical treatments' evaluations can also be biased because patients may not have the scientific knowledge to assess if a prescribed treatment is the best option or not. With this said, *exams and treatments* was not found to be a predictor of patient satisfaction, having a criterion weight of 8.19% (below the centroid). Despite being considered low priority status quo opportunities, the dissatisfaction felt by patients is an indicator that some changes might be in order. This criterion and respective subcriteria are viewed as less attractive requirements, with exception of *kindness of staff* and *information regarding patient's health state* (g_{103} and g_{104}, respectively) that are treated as a necessary must-be requirement and a low value-added one-dimensional requirement, respectively. A reason for this distinction lays on the fact that patients identify *kindness of staff* and *information regarding patient's health state* as implicit service characteristics.

Discharge process: *The discharge process* is usually viewed as a positive event through the eyes of the patient. With only two subcriteria, of similar weights, *homecare provided information* and *waiting time after discharge* (g_{111} and g_{112}, respectively) leave patients extremely satisfied, while also having the highest improvement index. Because of this, they are leverage opportunities that shall be tackled with critical priority. According to the Kano model, they belong to different classes of requirements. On the one hand, *homecare provided information* (g_{111}) is a highly attractive requirement, meaning that patients are surprisingly satisfied when a thorough and detailed explanation is provided. On the other hand, *waiting time after discharge* (g_{112}) is a critical must-be requirement. The longer the waiting time, the more dissatisfied the patient gets. However, if the waiting time is reduced, there is no feeling of satisfaction because patients are highly demanding, expecting waiting times approximate to zero. Overall, this criterion is a necessary must-be requirement with little influence on patient satisfaction (8.19%).

7.5. Managerial implications

Once all methodologies have been applied to the dataset, a final comparison needs to be performed. Since multiple methods, each one relying on distinct mathematical assumptions as explained in section 5, were used to study patient satisfaction, divergent results are expected. However, since validity, reliability, and sample adequacy for each method was verified, different results are not directly related to errors.

When comparing the results from SEM and OLR, both methods consider *accommodations, and exams and treatments* as influential. The disparity is that SEM also deems *medical staff and auxiliary staff* as influential. These criteria were also seen as influential on analysis A of OLR, but since they were not common to both analyses (A and B), were discarded as non-influential. These two methods return relatively similar results, and concise conclusions can be retrieved. However, when observing results from MUSA, discrepancies arise*. Food quality, volunteering staff and obtained information* are the predictors of patient satisfaction according to MUSA. With this said, there are multiple reasons for the existent discrepancy, but the validity of results remains unimpaired.

Comparing these results with the results yielded from the literature review, more specifically, Figure 7, it is possible to conclude that these outcomes are aligned with previous findings since all seven predictors are present on the influence analysis. As referred to earlier, healthcare's focus is to provide the best care across all service dimensions. However, due to capital and resource restrictions, it may not always possible to implement successful changes, and compromises must be made. Evaluating how patients perceive the service provided, and how satisfied they are about it is a practical approach that can help managers decide on how to allocate the available resources. It is suggested that health care managers implement periodic patient satisfaction surveys to supervise the impact that modifications might have on patient satisfaction. In line with the results introduced above, there are seven dimensions (*accommodations, exams and treatments, auxiliary staff, medical staff, food quality, volunteering staff, and obtained information*) that deserve special attention when implementing health policies. Combining the seven predictors of patient satisfaction with dissatisfaction issues, there are several modification suggestions that can be presented. From the survey's results (Table 3), it is possible to conclude that *volunteering staff and obtained information* are the two dimensions where patients are less satisfied. Since these dimensions are predictors of satisfaction, hospital managers should pay close attention to possible improvements. Considering that volunteering staff are not specialized personnel, improvements in this dimension should be focused on the humanization of health care, such as educating volunteering staff on how to interact with inpatients. Regarding obtained information, the creation of different communication channels, such as a digital platform where patients can have access to the necessary information while also being able to post a complaint if desired. Improvements in the remaining predictors should also be assessed, such as:

- Assuring the comfort of inpatients through the enhancement of accommodations and updated technology;
- Developing techniques to improve auxiliary and medical staff productivity, through either incentives that increase their motivation, penalties to their performance evaluation, or training to surpass difficulties;
- Assuring cutting edge technology is available for medical exams and treatments. However, if there are not sufficient financial resources, some different areas can be improved. For instance, managers should guarantee that waiting time and wating lists for medical exams and treatments are not too long, and that there is always available equipment in case of emergency;
- The presence of penalties on the contracts of food operators to guarantee the quality of every meal.

It is important to remember that this is a case study in a specific internment setting, thus, the final results and eventual implications are only valid for this particular scenario.

8. Concluding remarks, limitations, and directions for future research

The focus of this thesis was the evaluation of patient satisfaction through the implementation and comparison of different methodologies, which originated two scientific articles. One being a review article, and the other an empirical article with the methods' results.

Several steps were taken before reaching the final list of satisfaction predictors. An extensive literature review that allowed for an initial insight into which service dimensions patients value the most was the first step of this thesis. An analysis into which methods were used in the collected articles was performed to assure the methodology to be implemented in this project was aligned with literature. After data treatment and validation, factor analysis was conducted in a complementary nature to the remaining methods. From there, two separate analyses (depending on which dataset was going to be examined) emerged. SEM was applied to both analyses and conclusions were achieved considering common predictors. Nevertheless, bias is closely related to all aspects of SEM. For an estimator, a large amount of bias can turn a real positive value into a negative one, and vice versa. As explained, when performing SEM, the dataset has to follow a multivariate normal distribution to allow the ML estimator's usage, otherwise, results will be biased. This assumption was verified for our dataset, yet comparing results using other estimation methods might be of interest. If an analysis is based on biased estimators, the results will be misleading. A way to assess the bias present in estimates is the size of standard errors. Smaller standard errors are indicators of better efficiency. Standardized errors can be estimated through Monte Carlo simulations. This is a recommendation for future work, in order to eliminate bias that might be present in the results. (Zhong and Yuan, 2011). Furthermore, some multivariate models are not suitable for dealing with ordinal scales, as is the case of SEM and factor analysis, methods that are still being used by some researchers (47% of the collected articles). They undertake mathematical operations that are not consistent with Stevens' theory and data categorization (Stevens, 1946), and should not be applied to this kind of data (Vieira et al., 2020).

OLR was the following method to be implemented in the two analyses. As in the case of SEM, final results considered predictors common to both analyses. Evidently, this method also has limitations that might add bias to results. OLR assumes that error variances are homoscedastic, meaning that they do not vary even if the value of the predictor changes. Depending on the sample, this assumption can hold or not, but its inadequate usage might cause less precise results. Weighted regression, a method that allocates a weight (based on variance) to each data point might be the best way to replace heteroscedasticity with homoscedasticity (Frost, 2019). There is no indication that heteroscedasticity is present in the dataset, but its evaluation should be taken into account for future research.

Lastly, MUSA was applied solely to the dataset of analysis B. The justification for this fact is based on the fact that MUSA can solely work with results from CPCA. This is a type of factor analysis that deals with categorical variables and can be applied to ordinal categorical data, such as the case of our dataset. Thus, for further research, it would be advantageous to implement CPCA, compare the outputs with the ones from PCA, and use the new dataset in a complementary nature with MUSA.

The outputs of the MUSA method proved to be different from the other two methods, and such difference can rely on MUSA's assumption that criteria/subcriteria are independent of each other. In some cases, this assumption may not be met, given the presence of cognitive biasing effects, such as the Halo effect (a cognitive bias where people form an opinion about an attribute based on their impression on another attribute (Costa and Remedios, 2014). From the survey's answers, there is no way of knowing if patients' judgements towards an attribute were influenced by external factors, such as past experiences or other service attributes. Thus, the Hallo effect cannot be confirmed nor denied. To correctly assess the interactions between criteria/subcriteria and diminish the Halo effect, MUSA-INT, which considers positive and negative synergies between attributes, could be applied (Angilella et al., 2014; Ferreira et al., 2018). This new method should be considered for future research. Another issue that might arise from the MUSA method is its sensitivity to the number of constraints and variables, which means that as a linear programming model, the higher the number of satisfaction levels (in this case, seven), the higher the probability of introducing instability and, ultimately, returning an infeasible model. However, as explained in past sections, a seven point-Likert scale seems ideal, so a compromise between the two conditions might be necessary.

The importance of satisfaction predictors can be assessed through several methods, as previously seen. Given the easiness of handling and computation, factor and regression analyses are the elected methods when it comes to healthcare management. Nevertheless, despite the low usage rate of MUSA, this is an effective method with several advantages over the traditional customer satisfaction models, since it considers customers' judgments in the way they are expressed in questionnaires (Angilella et al., 2014; Ferreira et al., 2018). The low utilization of MUSA in healthcare and its potential when compared with other alternatives generates opportunities for a broader diffusion of studies using this method (Vieira et al., 2020).

Once the methods' implementation was completed, a conclusion that *accommodations, auxiliary staff, exams and treatments, medical staff, food quality, volunteering staff, and obtained information* are satisfaction predictors was achieved. It was not possible to verify if the results of this study are aligned with the Andalusian Agency for Healthcare Quality (ACSA) programme, the official accreditation standard, according to the Portuguese Health Ministry, due to difficulties on the interpretation of the programme. This should be assessed and further explored in a future research.

It would also be interesting to target different patient groups, such as age, gender, comorbidities, and specifically for the internment service, medical speciality, and length of stay. For instance, patients with a longer stay, due to a more serious condition, base their judgement on different experiences than patients with a one-night stay. Such segmentation was not possible to execute with our dataset due to a reduced size sample. However, since different patient groups value different service dimensions, data segmentation can provide more reliable results.

References

Aarrevaara, T., Carapinheiro, G., Carvalho, T., Correia, T., Dent, M., Hirvonen, H., Ribeiro, J., Saks, M., Santiago, R., & Serra, H. (2015). *Professionalism, managerialism and reform in higher education and the health services: The European welfare state and the rise of the knowledge society* (T. Carvalho & R. Santiago (eds.); 1st ed.). Palgrave Macmillan, London.

Abrantes, M. J. A. (2012). *Qualidade e satisfação: Opinião dos utilizadores de serviços de saúde hospitalares*. Universidade de Coimbra.

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family*, *67*(4), 1012–1028.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270–301.

Aiken, L. H., Sermeus, W., Van Den Heede, K., Sloane, D. M., Busse, R., McKee, M., Bruyneel, L., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., Tishelman, C., Scott, A., Brzostek, T., Kinnunen, J., Schwendimann, R., Heinen, M., Zikos, D., Sjetne, I. S., Smith, H. L., & Kutney-Lee, A. (2012). Patient safety, satisfaction, and quality of hospital care: Cross sectional surveys of nurses and patients in 12 countries in Europe and the United States. *BMJ*, *344*(7851), 1–14.

Al-Abri, R., & Al-Balushi, A. (2014). Patient satisfaction survey as a tool towards quality improvement. *Oman Medical Journal*, *29*(1), 3–7.

Aldana, J. M., Piechulek, H., & Al-Sabir, A. (2001). Client satisfaction and quality of health care in rural Bangladesh. *Bulletin of the World Health Organization*, *79*(6), 512–517.

Almeida, R. S., Bourliataux-Lajoinie, S., & Martins, M. (2015). Satisfaction measurement instruments for healthcare service users: A systematic review. *Cadernos de Saúde Pública*, *31*(1), 11–25.

AMA. (2020). *Administração central do sistema de saúde*.

Andaleeb, S. S. (2001). Service quality perceptions and patient satisfaction: A study of hospitals in a developing country. *Social Science and Medicine*, *52*(9), 1359–1370.

Anderson, A. B., Basilevsky, A., & Hum, D. P. J. (1983). Missing data: A review of the literature. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (1st ed., pp. 415–494). Academic Press.

Angilella, S., Corrente, S., Greco, S., & Słowiński, R. (2014). MUSA-INT: Multicriteria customer satisfaction analysis with interacting criteria. *Omega*, *42*(1), 189–200.

Ari, E., & Yildiz, Z. (2014). Parallel lines assumption in ordinal logistic regression and analysis approaches. *International Interdisciplinary Journal of Scientific Research*, *1*(3), 8–23.

Armstrong, R. A., Eperjesi, F., & Gilmartin, B. (2002). The application of analysis of variance (ANOVA) to different experimental designs in optometry. *Ophthalmic and Physiological Optics*, *22*(3), 248–256.

Ayanian, J. Z., & Markel, H. (2016). Donabedian's lasting framework for health care quality. *New England Journal of Medicine*, *375*(3), 205–207.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*(1), 5–37.

Beattie, P. F., Pinto, M. B., Nelson, M. K., & Nelson, R. (2002). Patient satisfaction with outpatient physical therapy: Instrument validation. *Physical Therapy*, *82*(6), 557–565.

Benbassat, J., & Taragin, M. (2000). Hospital readmissions as a measure of quality of health care: Advantages and limitations. *Archives of Internal Medicine*, *160*(8), 1074–1081. https://doi.org/10.1001/archinte.160.8.1074

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand*

Journal of Public Health, *25*(5), 464–469.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*(1), 78–117.

Bjertnaes, O. A., Sjetne, I. S., & Iversen, H. H. (2012). Overall patient satisfaction with hospitals: Effects of patient-reported experiences and fulfilment of expectations. *BMJ Quality and Safety*, *21*(1), 39–46.

Bleich, S. N., Ozaltin, E., & Murray, C. J. L. (2009). How does satisfaction with the healthcare system relate to patient experience? *Bulletin of the World Health Organization*, *87*(4), 271–278.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314.

Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior*, *36*(1), 3–15.

Boudreaux, E. D., Ary, R. D., Mandry, C. V, & McCabe, B. (2000). Determinants of patient satisfaction in a large, municipal ED: The role of demographic variables, visit characteristics, and patient perceptions. *The American Journal of Emergency Medicine*, *18*(4), 394–400.

Bratchell, N. (1989). Cluster Analysis. *Chemometrics and Intelligent Laboratory Systems*, *6*(2), 105–125.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS-Windows: Basic concepts, applications, and programming* (1st ed.). Sage Publications.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.

Cardozo, R. N. (1965). Experimental study and expectation. *Journal of Marketing Research*, *2*(3), 244–249.

Cepeda-Carrion, G., Cegarra-Navarro, J. G., & Cillo, V. (2019). Tips to use partial least squares structural equation modelling (PLS-SEM) in knowledge management. *Journal of Knowledge Management*, *23*(1), 67–89.

Cho, K. M., Hong, T. H., & Hyun, C. T. (2009). Effect of project characteristics on project performance in construction projects based on structural equation model. *Expert Systems with Applications*, *36*(7), 10461–10470.

Churchill, G. A., Surprenant, T., & Surprenant, C. (1982). Investigation into customer: The determinants of satisfaction. *Journal of Marketing Research*, *19*(4), 491–504.

Cicchetti, D. V., Showalter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. *Applied Psychological Measurement*, *9*(1), 31–36.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). LAWRENCE ERLBAUM ASSOCIATES.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250–1262.

Costa, L., & Remedios, R. (2014). Different methods, different results: Examining the implications of methodological divergence and implicit processes for achievement goal research. *Journal of Mixed Methods Research*, *8*(2), 162–179.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, *10*(7).

Davis, M., & Maggard, M. (1990). An analysis of customer satisfaction with waiting times in a two-stage service process. *Journal of Operations Management*, *9*(3), 324–334.

Decoster, J. (1998). *Overview of factor analysis*. http://stat-help.com/factor.pdf

Departamento de Qualidade na Saúde (DGS). (2015). *Estudo de satisfação dos utentes do sistema de saúde Português*. Direção-Geral Da Saúde. https://www.dgs.pt/qualidade-e-seguranca/satisfacao-do-utente-no-sns/documentos-e-publicacoes/estudo-de-satisfacao-dos-utentes-do-sistema-de-saude-portugues-2015.aspx

Diamantopoulos, A. (1999). Export performance measurement: reflective versus formative indicators. *International Marketing Review*, *16*(6), 444–457.

Dickinson, T. L., & Zellinger, P. M. (1980). A comparison of the behaviorally anchored rating mixed standard scale formats. *Journal of Applied Psychology*, *65*(2), 147–154.

Donabedian A. (1988). The quality of care: how can it be assessed? *Journal of the American Medical Association*, *260*, 1743–1748.

Donabedian, A. (2005). Evaluating the quality of medical care. *Milbank Quarterly*, *83*(4), 691–729.

Dong, Y., & Peng, C. Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*(1), 1–17.

Duggirala, M., Rajendran, C., & Anantharaman, R. N. (2008). Provider-perceived dimensions of total quality management in healthcare. *Benchmarking: An International Journal*, *15*(6), 693–722.

Elixhauser, A., Steiner, C., & Fraser, I. (2003). Volume thresholds and hospital characteristics in the United States. *Health Affairs*, *22*(2), 167–177.

Evans, M., Jamal, A., & Foxall, G. (2009). *Consumer Behaviour* (2nd ed.). John Wiley & Sons.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299.

Fan, V. S., Burman, M., McDonell, M. B., & Fihn, S. D. (2005). Continuity of care and other determinants of patient satisfaction with primary care. *Journal of General Internal Medicine*, *20*(3), 226–233.

Farzianpour, F., Byravan, R., & Amirian, S. (2015). Evaluation of patient satisfaction and factors affecting it: a review of the literature. *Health*, *7*(11), 1460.

Fernandes, A. C., & Nunes, A. M. (2016). Os hospitais e a combinação público-privado no sistema de saúde português. *Acta Medica Portuguesa*, *29*(3), 217–223.

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

Ferreira, D., & 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, S. (2019). *A multiple criteria satisfaction analysis tool for measuring patients ' satisfaction with Portuguese secondary care : The case of internment services* (Issue October). Instituto Superior Técnico.

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). SAGE Publications Ltd.

Fienberg, S. E., & Junker, B. W. (2010). An overview of statistics in education. In *International Encyclopedia of Education* (3rd ed.).

Finch, W. H. (2013). Exploratory Factor Analysis. In T. Teo (Ed.), *Handbook of Quantitative Methods for Educational Research* (pp. 167–186). SensePublishers.

Finn, R. H. (1972). Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement*, *34*, 885–892.

Freeze, R., Raschke, R. L., & Freeze, R. D. (2007). An Assessment of Formative and Reflective Contructs in IS Research. *Proceedings of the 15th European Conference on Information Systems, ECIS 2007*, 1481–1492.

Frost, J. (2019). *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models* (1st ed.). Statistics By Jim Publishing.

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research*, *11*(1), 572.

Glen, S. (2016). *Kaiser-Meyer-Olkin (KMO) test for sampling adequacy*. StatisticsHowTo.Com: Elementary Statistics for the Rest of Us! https://www.statisticshowto.com/kaiser-meyer-olkin/

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

Grigoroudis, E., & Siskos, Y. (2002a). Preference disaggregation for measuring and analysing customer satisfaction: The MUSA method. *European Journal of Operational Research*, *143*(1), 148–170.

Grigoroudis, E., & Siskos, Y. (2002b). Preference disaggregation for measuring and analyzing customer satisfaction: the MUSA method. *European Journal Operational Research*, *143*(1), 148–170.

Grigoroudis, E., & Siskos, Y. (2010). MUSA: Multicriteria Satisfaction Analysis. In F. S. Hillier (Ed.), *Customer Satisfaction Evaluation: Methods for Measuring and Implementing Service Quality* (pp. 91–121). Springer Science+Business Media, Inc.

Grönroos, C. (1982). An applied service marketing theory. *European Journal of Marketing*, *16*(7), 30–41.

Haines, G. H., Howard, J. A., & Sheth, J. N. (1970). The theory of buyer behavior. *Journal of the American Statistical Association*, *65*(331), 1406–1407.

Hauke, J., & Kossowski, T. (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, *30*(2), 87–93.

Heitjan, D. F. (1997). Annotation: what can be done about missing data? Approaches to imputation. *American Journal of Public Health*, *87*(4), 548–550.

Hekkert, K. D., Cihangir, S., Kleefstra, S. M., van den Berg, B., & Kool, R. B. (2009). Patient satisfaction revisited: A multilevel approach. *Social Science and Medicine*, *69*(1), 68–75.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Sage.

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*(1), 90–98.

III, E. P. C. (1980). The optimal number of response alternatives for a scale: A Review. *Journal of Marketing Research*, *17*(4), 407.

Jackson, J. L., Chamberlin, J., & Kroenke, K. (2001). Predictors of patient satisfaction. *Social Science & Medicine*, *52*(4), 609–620.

Jarvis, C. B., Mackenzie, S. B., Podsakoff, P. M., Giliatt, N., & Mee, J. F. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199–218.

Jenkinson, C., Coulter, A., Bruster, S., Richards, N., & Chandola, T. (2002). Patients' experiences and satisfaction with health care: Results of a questionnaire study of specific aspects of care. *Quality and Safety in Health Care*, *11*(4), 335–339.

João, I. M., Bana e Costa, C. A., & Figueira, J. F. (2007). An alternative to the MUSA method for customer satisfaction analysis. *EURO XXII - 22nd European Conference on Operational Research*.

Johnson, M. D., Anderson, E. W., & Fornell, C. (1995). Rational and adaptive performance expectations in a customer satisfaction framework. *Journal of Consumer Research*, *21*(4), 695–707.

Joshi, A., Kale, S., Chandel, S., & Pal, D. (2015). Likert scale: Explored and explained. *British Journal* *of Applied Science & Technology*, *7*(4), 396–403.

Kano, N., Seraku, N., Takahashi, F., & Tsuji, S. (1984). Attractive quality and must-be quality. *Journal of the Japanese Society for Quality Control*, *41*, 39–48.

Karanja, E., Zaveri, J., & Ahmed, A. (2013). How do MIS researchers handle missing data in survey-based research: A content analysis approach. *International Journal of Information Management*, *33*(5), 734–751.

Kaya, O., Teymourifar, A., & Ozturk, G. (2020). Analysis of different public policies through simulation to increase total social utility in a healthcare system. *Socio-Economic Planning Sciences*, *70*(December 2017), 100742.

Kim, J. O., & Curry, J. (1977). The treatment of missing data in multivariate analysis. *Sociological Methods Research*, *6*, 215–240.

Kim, T. K. (2015). T test as a parametric statistic. *Korean Journal of Anesthesiology*, *68*(6), 540–546.

Kitapci, O., Akdogan, C., & Dortyol, İ. T. (2014). The impact of service quality dimensions on patient satisfaction, repurchase intentions and word-of-mouth communication in the public healthcare industry. *Procedia-Social and Behavioral Sciences*, *148*, 161–169.

Kline, R. (2011). *Principles and Practice of Structural Equation Modeling* (T. D. Little & D. A. Kenny (eds.); 3rd ed.). The Guildfors Press.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (D. A. Kenny & D. L. Rodd (eds.); 3rd ed.). The Guildfors Press.

Koletsi, D., & Pandis, N. (2018). Ordinal logistic regression. *American Journal of Orthodontics and Dentofacial Orthopedics*, *153*(1), 157–158.

Krosnick, J. A., & Berent, M. K. (1993). Comparisons of party identification and policy preferences: The impact of survey question format. *American Journal of Political Science*, *37*(3), 941–964.

Krosnick, J. A., & Presser, S. (2009). *Handbook of Survey Research* (J. D. Wright & P. V. Marsden (eds.); 2nd ed.). Elsevier.

Kumar, R., & Sankar, R. J. (2008). *An application for ordinal logistic (proportional odds) regression model using SPSS* [University College of Medical Sciences]. https://www.researchgate.net/publication/256613909_An_application_for_ordinal_logistic_proportional_odds_regression_model

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models* (B. Gordon & R. T. Hercher Jr. (eds.); 5th ed.). McGraw-Hill / Irwin.

Levin-Scherz, J., DeVita, N., & Timbie, J. (2006). Impact of pay-for-performance contracts and network registry on diabetes and asthma HEDIS® measures in an integrated delivery network. *Medical Care Research and Review*, *63*(1), 14–28.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*(7).

Liljequist, D., Elfving, B., & Roaldsen, K. S. (2019). Intraclass correlation – A discussion and demonstration of basic features. *PLoS ONE*, *14*(7).

Little, R., & Rubin, D. (2002). *Statistical Analysis with Missing Data* (D. J. Balding, P. Bloomfield, N. A. Cressie, N. I. Fisher, I. M. Johnstone, J. B. Kadane, L. M. Ryan, D. W. Scott, A. F. Smith, & J. L. Teugels (eds.); 2nd ed.). John Wiley & Sons, Inc.

Liu, X., & Koirala, H. (2012). Ordinal regression analysis: Using generalized ordinal logistic regression models to estimate educational data. *Journal of Modern Applied Statistical Methods*, *11*(1), 242–254.

Lomax, R. G., & Hahs-Vaughn, D. L. (2012). *An Introduction to Statistical Concepts* (3rd ed.). Routledge.

Lovelock, C., & Wright, L. (2001). *Serviços: Marketing e gestão*. Saraiva.

Lundstrom, W. J., & Hunt, H. K. (1978). Conceptualization and measurement of consumer satisfaction and dissatisfaction. *Journal of Marketing Research*, *15*(1), 167.

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, *90*(4), 710–730.

Malkanthie, A. (2015). The Basic Concepts of Structural Equation Modeling. *Lap Lambert Academic Publishing*, *1*(January), 55.

Mardia, K., Kent, T., & Bibby J. (1980). *Multivariate Analysis* (Z. W. Birnbaum & E. Lukacs (eds.); 1st ed.). Academic Press.

Markus, M., Vernon Knapp, H., & Tasker, G. D. (2003). Entropy and generalized least square methods in assessment of the regional value of streamgages. *Journal of Hydrology*, *283*(1–4), 107–121.

Marôco, J. (2014). *Análise de Equações Estruturais: Fundamentos Teóricos, Softwares e Aplicações* (2nd ed.).

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for likert scale items? study 1: Reliability and validity. *Educational and Psychological Measurement*, *31*(3), 657–674.

Matsatsinis, N., & Siskos, Y. (2003). *Intelligent Support Systems for Marketing Decisions* (F. S. Hillier (ed.)). Springer Science+Business Media.

McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64–82.

McKnight, P. E., & Najab, J. (2009). Mann Whitney U Test. In *The Corsini Encyclopedia of Psychology*.

Mert Kantar, Y. (2015). Generalized least squares and weighted least squares estimation methods for distributional parameters. *Revstat Statistical Journal*, *13*(3), 263–282.

Ministério da Saúde. (2016). *Lei Orgânica*. https://www.sns.gov.pt/institucional/ministerio-da-saude/lei-organica/

Ministério da Saúde. (2019). *Secretaria-geral do Ministério da Saúde*. https://www.sns.gov.pt/entidades-de-saude/secretaria-geral-do-ministerio-da-saude/

Monteiro de Castro, J., & Lopes Portela, S. (2016). *A satisfação do cliente em contexto de consulta externa num hospital privado* [ISCTE-Business school]. https://repositorio.iscte-iul.pt/bitstream/10071/13049/1/Tese - Satisfação do Cliente num serviço de Consulta Externa.pdf

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

Muhtaseb, R., Lakiotaki, K., & Matsatsinis, N. (2012). Applying a multicriteria satisfaction analysis approach based on user preferences to rank usability attributes in E-tourism websites. *Journal of Theoretical and Applied Electronic Commerce Research*, *7*(3), 28–48.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, *105*(3), 430–445.

Murti, H. (2016). How to conduct path analysis and structural equation model for health research. *International Conference on Public Health*.

Naidu, A. (2009). Factors affecting patient satisfaction and healthcare quality. *International Journal of Health Care Quality Assurance*, *22*(4), 366–381.

Needleman, J., Buerhaus, P. I., Stewart, M., Zelevinsky, K., & Mattke, S. (2006). Nurse staffing in hospitals: Is there a business case for quality? *Health Affairs*, *25*(1), 204–211.

Ng, J. H. Y., & Luk, B. H. K. (2019). Patient satisfaction: Concept analysis in the healthcare context. *Patient Education and Counseling*, *102*(4), 790–796.

Nguyen Thi, P. L., Briançon, S., Empereur, F., & Guillemin, F. (2002). Factors determining inpatient satisfaction with care. *Social Science and Medicine*, *54*(4), 493–504.

Nunes, A. M., & Ferreira, D. C. (2019a). Reforms in the Portuguese health care sector: Challenges and proposals. *International Journal of Health Planning and Management*, *34*(1), 21–33.

Nunes, A. M., & 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.

Nunnally, J. C. (1967). *Psychometric Theory* (2nd ed.). McGraw-Hill.

Oaster, T. R. F. (1989). Number of alternatives per choice point and stability of Likert-type scales. *Perceptual and Motor Skills*, *68*(2), 549–550.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*(4), 460–469.

Oliver, R. L. (1997). *Satisfaction: A behavioral perspective on the consumer*. McGraw Hill.

Olsen, L. L., & Johnson, M. D. (2003). Service equity, satisfaction, and loyalty: From transaction-specific to cumulative evaluations. *Journal of Service Research*, *5*(3), 184–195.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling*, *7*(4), 557–595.

Osborne, J., Osborne, J. W., Costello, A. B., & Kellow, J. T. (2014). *Best practices in exploratory factora analysis*. CreateSpace.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, *49*(4), 41–50.

Patrician, P. A. (2002). Multiple imputation for missing data. *Research in Nursing and Health*, *25*(1), 76–84.

Peng, C. Y., Harwell, M. R., Liou, S. M., & Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. In S. S. Sawilowsk (Ed.), *Real data analysis* (pp. 31–78).

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics and Data Analysis*, *49*(4), 974–997.

Petrucci, C. J. (2009). A primer for social worker researchers on how to conduct a multinomial logistic regression. *Journal of Social Service Research, 35*(2), 193–205.

Pituch, K. A., & Stevens, J. P. (2016). *Applied Multivariate Statistics for the Social Sciences* (D. Riegert & A. Symons (eds.); 6th ed.). Routledge.

Prakash, B. (2010). Patient satisfaction. *Journal of Cutaneous and Aesthetic Surgery*, *3*(3), 151–155.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15.

Qin, H., & Prybutok, V. R. (2013). A quantitative model for patient behavioral decisions in the urgent care industry. *Socio-Economic Planning Sciences*, *47*(1), 50–64.

Rahmqvist, M., & Bara, A. C. (2010). Patient characteristics and quality dimensions related to patient satisfaction. *International Journal for Quality in Health Care*, *22*(2), 86–92.

Ramlall, I. (2016). *Applied sctructural equation modelling for researchers and practitioners: Using R and Stata for behavioural research* (1st ed.). Emerald Group Publishing Limited.

Ramsaran-Fowdar, R. R. (2005). Identifying health care quality attributes. *Journal of Health and Human Services Administration*, *27*(4), 428–443.

Ramsay, J. O. (1973). The effect of number of categories in rating scales on precision of estimation of scale values. *Psychometrika*, *38*, 513–533.

Renzi, C., Abeni, D., Picardi, A., Agostini, E., Melchi, C. F., Pasquini, P., Puddu, P., & Braga, M. (2001). Factors associated with patient satisfaction with care among dermatological outpatients. *British Journal of Dermatology*, *145*(4), 617–623.

Rietveld, T., & Van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour*. Mouton de Gruyter.

Rogers, A. E., Hwang, W. T., Scott, L. D., Aiken, L. H., & Dinges, D. F. (2004). The working hours of hospital staff nurses and patient safety. *Health Affairs*, *23*(4), 202–212.

Rossiter, J. R. (2005). Reminder: A horse is a horse. *International Journal of Research in Marketing*, *22*(1), 23–25.

Rust, R. T., & Oliver, R. L. (1993). *Service quality: New directions in theory and practice* (R. T. Rust & R. L. Oliver (eds.); 1st ed.). Sage Publications.

Saha, S. (2018). *Baffled by Covariance and Correlation?* Towards Data Science. https://towardsdatascience.com/let-us-understand-the-correlation-matrix-and-covariance-matrix-d42e6b643c22

Santos, E. (2018). Patients' satisfaction with the NHS: A novel indicator for Portugal. *International Journal of Healthcare Education & Medical Informatics*, *05*(01), 12–17.

Saúde, M. da. (2016). *Entidade Reguladora de Saúde (ERS)*. https://www.sns.gov.pt/entidades-de-saude/entidade-reguladora-da-saude/

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data* (1st ed.). Chapman and Hall/CRC.

Schoenfelder, T., Klewer, J., & Kugler, J. (2011). Determinants of patient satisfaction: A study among 39 hospitals in an in-patient setting in Germany. *International Journal for Quality in Health Care*, *23*(5), 503–509.

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*(6), 323–338.

Shilling, V., Jenkins, V., & Fallowfield, L. (2003). Factors affecting patient and clinician satisfaction with the clinical consultation: can communication skills training for clinicians improve satisfaction? *Psycho-Oncology*, *12*(6), 599-611.

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

Simões, J., Fronteira, I., & Augusto, G. F. (2020). The 2019 health basic law in Portugal: Political arguments from the left and right. *Health Policy*, 4–9.

Sofaer, S., & Firminger, K. (2005). Patient perceptions of the quality of health services. *Annual Review of Public Health*, *26*(1), 513–559.

Stevens, S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Streiner, D. L. (2002). The case of the missing data: Methods of dealing with dropouts and other research vagaries. *Canadian Journal of Psychiatry*, *47*(1), 68–75.

Sun, B. C., Adams, J., Orav, E. J., Rucker, D. W., Brennan, T. A., & Burstin, H. R. (2000). Determinants of patient satisfaction and willingness to return with emergency care. *Annals of Emergency Medicine*, *35*(5), 426–434.

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *19*(3), 227–229.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Pearson Education.

Thurstone, L. L. (1947). *Multiple factor analysis*. University of Chicago Press.

Tripathi, A. (2019). *What are covariance and correlation coefficients and their significance?* Towards Data Science. https://towardsdatascience.com/covariance-and-correlation-321fdacab168

UCLA: Statistical Consulting Group. (2011). *What are pseudo R-squareds?* https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/

Valle, P. O., Mendes, J., Guerreiro, M., & Silva, J. A. (2011). Can welcoming residents increase tourist satisfaction? *Anatolia*, *22*(2), 260–277.

Vieira, I., Ferreira, D. C., & Pedro, M. I. (2020). Patient satisfaction with healthcare provision services: What does the literature tell us about it? *Socio-Economic Planning Sciences*.

Vilares, M. J., & Coelho, P. S. (2011). *Satisfação e Lealdade do Cliente: Metodologias de Avaliação, Gestão e Análise* (2nd ed.). Escolar Editora.

Vuković, M., Gvozdenović, B. S., Gajić, T., Stamatović Gajić, B., Jakovljević, M., & McCormick, B. P. (2012). Validation of a patient satisfaction questionnaire in primary health care. *Public Health*, *126*(8), 710–718.

Wang, J., & Wang, X. (2012). *Structural Equation Modeling: Applications Using Mplus* (1st ed.). WILEY.

Wang, T., & Ji, P. (2010). Understanding customer needs through quantitative analysis of Kano's model. *International Journal of Quality and Reliability Management*, *27*(2), 173–184.

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*(3), 219–246.

Westaway, M. S., Rheeder, P., Van Zyl, D. G., & Seager, J. R. (2003). Interpersonal and organizational dimensions of patient satisfaction: The moderating effects of health status. *International Journal for Quality in Health Care*, *15*(4), 337–344.

Westbrook, R. A., & Reilly, M. D. (1983). Value-percept disparity: An alternative to the disconfirmation of expectations theory of consumer satisfaction. *Advances in Consumer Research*, *10*, 256–261.

Xiong, B., Skitmore, M., Xia, B., Masrom, M. A., Ye, K., & Bridge, A. (2014). Examining the influence of participant performance factors on contractor satisfaction: A structural equation model. *International Journal of Project Management*, *32*(3), 482–491.

Yang, C. C. (2005). The refined Kano's model and its application. *Total Quality Management and Business Excellence*, *16*(10), 1127–1137.

Yang, C. C., & Yang, K. J. (2011). An integrated model of value creation based on the refined Kano's model and the blue ocean strategy. *Total Quality Management and Business Excellence*, *22*(9), 925–940. https://doi.org/10.1080/14783363.2011.611358

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, *60*(2), 31–46.

Zhong, X., & Yuan, K. H. (2011). Bias and efficiency in structural equation modeling: Maximum likelihood versus robust methods. *Multivariate Behavioral Research*, *46*(2), 229–265.

Appendices

Appendix A

Table A. 1.PRISMA checklist. Source: http://prisma-statement.org/PRISMAStatement/Checklist. Accessed on 1 June 2020.

Section/topic	#	Checklist item
TITLE		
Title	1	Identify the report as a systematic review, meta-analysis, or both.
ABSTRACT		
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.
INTRODUCTION		
Rationale	3	Describe the rationale for the review in the context of what is already known.
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).
METHODS		
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each

		meta-analysis.
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.
RESULTS		
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).
DISCUSSION		
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.
FUNDING		
Funding	27	Describe sources of funding of the systematic review and other support (e.g., supply of data); role of funders for the systematic review.

Appendix C

Table C. 1. Rotated component matrix.

	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9
Patient's guide **[g_{11}]**	0.769								
Patient's rights and duties **[g_{12}]**	0.725								
Complaint means **[g_{13}]**	0.799								
Substitution in decision making **[g_{14}]**	0.800								
Anticipated vital will **[g_{15}]**	0.765								
Cleanliness **[g_{21}]**		0.577							
Comfort and commodity**[g_{22}]**		0.625							
Privacy **[g_{23}]**		0.701							
Furniture **[g_{24}]**		0.682							
Noise **[g_{25}]**		0.567							
Temperature **[g_{26}]**		0.594							
Entertainment **[g_{27}]**		0.509							
Visitation hours **[g_{31}]**			0.822						
Visit duration **[g_{32}]**			0.835						
Number of visits **[g_{33}]**			0.792						
Easy access for close relatives **[g_{34}]**			0.563						
Preparation, appearance, temperature, taste **[g_{41}]**				0.794					
Variety **[g_{42}]**				0.799					
Quantity **[g_{43}]**				0.697					
Meal support **[g_{44}]**				0.528					
Availability **[g_{51}]**					0.640				
Attention **[g_{52}]**					0.683				
Kindness **[g_{53}]**					0.673				
Information regarding patient's health state **[g_{54}]**					0.758				
Information regarding medical treatment **[g_{55}]**					0.733				
Information regarding medical exams **[g_{56}]**					0.724				
Health advising and teaching **[g_{57}]**					0.725				
Availability **[g_{61}]**						0.731			
Attention **[g_{62}]**						0.773			
Kindness **[g_{63}]**						0.799			
Information regarding patient's health state **[g_{64}]**						0.717			

Information regarding nursing treatment **[g_{65}]**						0.721			
Health advising and teaching **[g_{66}]**						0.644			
Availability **[g_{71}]**						0.809			
Attention **[g_{72}]**						0.819			
Kindness **[g_{73}]**						0.826			
Performance efficiency **[g_{74}]**						0.801			
Availability **[g_{81}]**							0.745		
Attention **[g_{82}]**							0.779		
Kindness **[g_{83}]**							0.777		
Performance efficiency **[g_{84}]**							0.789		
Availability **[g_{91}]**								0.885	
Attention **[g_{92}]**								0.884	
Kindness **[g_{93}]**								0.900	
Availability **[g_{101}]**									0.516
Attention **[g_{102}]**									0.538
Kindness **[g_{103}]**									0.495
Information regarding patient's health state **[g_{104}]**									0.768
Information regarding medical treatment **[g_{105}]**									0.752
Information regarding medical exams **[g_{106}]**									0.740
Health advising and teaching **[g_{107}]**									0.775
Homecare provided information **[g_{111}]**					0.619				
Waiting time after discharge **[g_{112}]**					0.486				