Quality assessment of the Portuguese public hospitals: A multicriteria approach

António Maria Rocha
antonio.m.rocha@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2019

Abstract

The Portuguese National Health Service (SNS) was created to provide universal, equal and tendentiously free care. There are different levels of care (primary, secondary, continued and palliative), however all should deliver quality care services. Quality in healthcare is composed by several criteria such as patient safety, care appropriateness or access, it should also be efficient non-compromising the other criteria. However, during the last years political and economic events had impact in the SNS. Hence, structural reforms occurred and new health care policies were implemented, mostly focused in improving efficiency and reducing costs. This allied to divestment can increment barriers to access, compromise infrastructures and equipment and above all quality of the service provided. Thus, this dissertation aims to assess quality of the Portuguese public hospitals (secondary care providers). Here is where the ELECTRE TRI-nC comes in, enabling a multicriteria approach by formulation of a model. The formulation of a criteria tree to represent quality arrives from the literature review and the criteria weight through interactions with a decision maker. The hospital's data was collected and processed, plus a set of parameters were chosen in order to execute the ELECTRE TRI-nC. Finally, each hospital was classified and attributed to a category of existing five respectively ordered. The robustness of the model was tested through a sensitivity analysis. The result was a robust model with some space to improve. The potential application of this dissertation in future research in healthcare policy and hospital funding is high, in an SNS whose sustainability is a permanent challenge.
Keywords: Quality, Hospitals, National Health Service, ELECTRE TRI-nC, Multicriteria approach

1. Introduction

Portugal is a sovereign country whose Republic is based on the human dignity and popular will, and committed on building a free, just and supportive society. In many countries around the world this is far from being a reality. The Portuguese Constitution, that grounds and governs the principles and organization of the Portuguese State, declares in the article 64, "Everyone has the right to health protection and the duty to defend and promote it". The right to health is ensured through a National Health Service (SNS, from the Portuguese abbreviation *Serviço Nacional de Saúde*) created in 1976, which aims to promote a suitable and equitable care, tendentiously free to its citizens. The SNS is one of the oldest in the world, but nowadays is not meeting the needs of the its population [1].

In the year of 2018, 9% of Portugal's gross domestic product (GDP) was devoted to health, more than the average of the Organization for Economic Co-operation and Development (OECD) with 8.8%. Even though this attempt to maintain the investment in health, the system is struggling to accomplish its goals in a sector full of new and complex challenges, such as, aging, demographic increase, chronic diseases and the introduction of innovation responsible for generating health gains (however not eliminating costs). To face this policy-makers and healthcare managers are applying measures to improve efficiency, mostly through cost containment, that must never compromise the quality of the services provided, if possible, maximizing it. However, the Portuguese hospitals are funded based on contracts that do not consider the quality of the services provided. As for that, it is essential to assess the quality of the Portuguese public hospitals, making room to uncover effects of scarifying it and possible setting benchmarks later to be inserted on funding.

In this context, this research places emphasis on the application of a multicriteria approach, the ELECTRE TRI-nC, capable of assessing hospital's quality. Once understood the problem, an extensive literature review was performed aiming to define quality in healthcare, so that afterwards one could create a model to capable of measuring it.

2. Problem context

2.1. The health sector

The health sector is highly important position among other sectors, for both social and economic reasons [14]. This combination makes it a highly sensitive sector, as the right to health is one of the fundamental rights of any citizen. However, to assure this right the health expenditure has been increasing during the last decades. Some of the reasons for this rise were up for mentioned and concern demography shifts, increasing life expectancy of population, complex and chronic diseases, expansion of coverage by public health service and technology improvements. Furthermore, inefficient management of resources by governments, healthcare institutions and even from their workforce are also related this increase [3]. The resources to allocate to healthcare are scarce, although the needs are virtually unlimited. Therefore, the increasing of health expenditure raises issues related to the sustainability of the system, compromising quality of care and equity in access, both vital in this SNS.

All in all, this represents an optimization problem where is necessary to reduce costs without compromising quality, towards guaranteeing the sustainability of the SNS and quality care to its patients. Hospitals are the main care providers, thus it is essential to assess the quality of care they provide, so the best practices can be identified and adopted.

2.2. The Portuguese NHS and its hospitals

The SNS was implemented in 1979 after the transition between political regimes. The system is mainly based on the Beveridge model, where the Portuguese Government is responsible for managing the primary and secondary healthcare providers both public entities sustained predominantly by public taxes distributed by different ministries including the Ministry of Health. The SNS is an instrument of the state to ensure the right to health protection under the Constitution, it provides a suitable and equitable care to its beneficiaries and it is tendentiously free in order to provide access to all the citizens. In the last decades, there has been an increase of pressure in the system through different factors upwards mentioned. Due to these events the health expenditure in Portugal is rising, being the value of the health expenditure *per capita* approximately €1500, one of the highest in the E.U. compromising the existence of the SNS [9].

In order to preserve it and its beneficiaries, policy reforms have been implemented to reduce costs and waste of financial resources, towards making healthcare entities more efficient and effective. Some of the last reforms, evolved hospital mergers, entities' corporatisation and the creation of public-private partnerships. Merges originated hospital centers, as a result of horizontal hospital mergers and local health care units as, a result of vertical mergers of both primary and secondary health care centers. Before corporatisation in 2003, financing was retrospective based on hospitals' latest year expenses, considering fixed amounts according to medical diagnosis, regardless of the costs incurred [14]. Although the payment model predicted the compensation for the real costs of care, the institutions did not have concrete and accurate ways of calculating the actual costs, resulting in over and under estimated values [6]. Thus, it is clear that the retrospective model was not efficient, as providers with higher costs would obtain more resources without added responsibilities to units or managers, adding on to the existence of failures in the control of services and costs actually incurred [4]. In case there were budgetary constraints, hospitals had less funding than expected.

The hospitals' corporatisation produced changes in funding, changing to a prospective model, where hospitals were financed according to their activity and levels of production[6]. The new model included a contract between hospitals and the State ,named *contrato-programa*, negotiated between each hospital's administration and the Ministry of Health. The objective was based on the promotion of efficiency gains, given the attribution of a value considered enough for each intervention. In this model, the type, volume and prices of services to be executed are fixed prior to their realization and independent of the actual cost. Thus, the financial risk of the hospital units depends on a sustainable management requiring the good use of their resources. This creates an incentive for a better management through the accountability of the various actors to promote efficiency, however it should not discourage quality.

Nevertheless, the payments made by the Ministry of Health to hospitals are determined by the most efficient hospital in a cluster, through averaging the unitary costs of that hospital. It is assumed that hospitals in a same cluster have analogous production technologies. The problem arrives when the concept of efficiency is not totally clear, neither the criteria used for the hospitals clustering correctly reflects both hospitals' environment, quality of care provided and management of the institutions [10]. This means that hospital financing's process is likely to produce inefficient payments. It is critical that hospital clustering (thereafter funding) includes not only the number of services and prices, but also reflecting the quality of the services and management provided, which is where this dissertation emerges on. Therefore, it is requested a detailed study about the quality of the SNS entities what is possible through building a Multiple Criteria Decision Aiding (MCDA) model.

3. Quality in healthcare services

Quality is hard to define as it assumes a subjective nature, depending on perspectives and varying within the context considered. Generic quality definitions are hard to formulate, even more for healthcare services. Thus, some authors follow disaggregated approaches based on multidimensionality and selected criteria, whereas each dimension provides an evaluation of quality on its own view, however when combining all the dimensions it results in a complex and more specific evaluation[5]. An example this approach is Donabedian's definition of quality in healthcare, as a product of three interrelated categories in his Structure-Process-Outcome model [7]: (1) Structure category denotes the attributes in which care occurs, including the material resources (*e.g.* infrastructures and equipment); (2) Process category denotes the actions performed by the staff to deliver care services (*e.g.* interactions between clinicians and patients), where the best process measures are the ones that follow clinical evidence to achieve better outcomes; (3) Outcome category denotes the effects of care services delivered on patients' quality of life. The best outcome measures are the ones which are related to the care provided.

Another multidimensional definition of quality in healthcare arises from the Institute of Medicine (IOM), defining it as a product of six dimensions [13]: (1) Safety, avoiding harm to patients from the care that is intended to help them; (2) Effectiveness, providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding overuse and underuse, respectively); (3) Person-centeredness, care that is respectful and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions; (4) Accessibility, Timeliness and Affordability, reducing waits, harmful delays, access barriers and financial risk for patients, families, and communities; (5) Efficiency, avoiding waste, including waste of equipment, supplies, ideas, and energy; (6) Equity, providing care that does not vary in characteristics such as gender, ethnicity, geographic location, and socioeconomic status.

Both multidimensional definitions, provide an accurate and correlated vision of quality in healthcare that can be used to create a unifying conceptual framework capable of representing it. Furthermore, both definitions meet the principles of medical ethics: beneficence, non-maleficence, autonomy and justice. Therefore, the dimensions are embodied in the highest guidelines of care. High quality care needs to be accessible, effective, safe, centred on patient's needs and given in a timely manner.

Once defined a unifying conceptual framework, one needs to choose metrics that better represent and measure the performance of each quality dimension. This framework will be built according to both Donabedian and the IOM multidimensional quality definitions, using quality indicators from the American Agency for Healthcare Research and Quality (AHRQ) and the IOM, who have developed indicators to measure the performance of quality care provided by the healthcare entities [12].

4. Case Study

This paper's section, defines the method used and its inputs, such as sample, scales, weights, thresholds and variables used to build the criteria tree.

4.1. Methods

In section 3, quality was defined as being multidimensional. Some authors believe MCDA approaches are highly suitable in multidimensional frameworks, when aggregating single indicators into a composite one, as they involve making decisions when it comes to combine criteria of different nature [8]. Therefore, one decided to build a model towards assessing quality in the Portuguese public hospitals, considered a sorting problem in MCDA.

The ELECTRE TRI-nC method deals with sorting problems in real-world situations and just like other MCDA models, it presents two characters who cooperate to build models: the decision maker (DM), in whose name this decision aiding (DA) is to be given; and the analyst, responsible for giving the DA through developing the model [11]. The method assigns a set of actions, $A = \{a_1, ..., a_m\}$, to a set of ordered and *a priori* defined categories, $C = \{C_1, ..., C_h, ..., C_q\}$, according to the performance of each action in a set of criteria. Criteria are denoted by g ,and $g(a)$ denotes the performance of an action a according to a given criterion, whereas a family of criteria is denoted by $G = \{g_1, g_2, ..., g_j, ..., g_n\}$. Categories are characterized by a set of reference actions, denoted $B = \{B_1, ..., B_h, ..., B_q\}$, that are further composed by a subset of reference actions giving a particular degree of freedom to characterize a category, denoted $B_h = \{b_h^r, r = 1, ..., m_h\}$.

The method has two main steps: (1) The construction of outranking relationships, represented by aS_ja', which means "an action a is at least as good as a', according to a criterion g_j". For this to be valid there are two conditions to be fulfilled: *Concordance*, where the majority of criteria should be in favour of this relationship, measured by the global concordance index (that considers the weight of criteria to validate an outranking relation); and *non-discordance*, when none of the opposing criteria exercises its veto power to this assertion, measured by the non-discordance index (that attributes a power veto to criteria). Then, the credibility index is responsible to justify, aS_ja', considering all the

criteria denoted $\sigma(a, a')$. Finally, towards converting fuzzy relations into a crisp outranking relationship it is used the level of credibility, λ. Through comparing the level of credibility with the categorical credibility index, calculated between each action and the set of reference actions of each category, it is possible to identify four binary relationships for credibility level: λ-outranking, λ-preference, λ-indifference, λ-incomparability. These four relationships are the ones that validate or not $aS_j a'$. (2) The exploitation of the outranking relationships, that aim to assign an action to a category. The procedure uses two joint rules together, descending and ascending rules that originate a function $\rho(\{a\}, B_h)$ (see equation 1), that compare an action, a, to subsets of reference actions, B_h, considering a defined credibility level. Each rule selects a possible category for an action, in case the category is the same the action is assigned to a unique category, otherwise intervals of categories are created.

$$\rho(\{a\}, B_h) = min\{\sigma(\{a\}, B_h), \sigma(B_h, \{a\})\} \quad (1)$$

The ELECTRE TRI-nC presents highly valuable features for this study: categories can be defined by one or more reference actions (providing freedom to the DM when characterizing a category), criteria with different weights, indifference and preference thresholds (to tackle imperfect knowledge and arbitrariness), qualitative and quantitative scales (direct use of data without recoding), as well as the possibility of being heterogeneous, and method's non-compensatory character (as in health, the worst performances on certain criteria can't be systematically compensated by better performances on others [15]), reinforced with the presence of a veto threshold.

4.2. Data and sample

The data used in this study needed to be reliable and integrable in our model according to the literature review. Therefore, one used the data provided by the Portuguese Central Health System Administration (ACSS), available via http://benchmarking.acss.min-saude.pt. This data belongs to a benchmarking developed among hospitals of the SNS, providing a set of indicators and respective performance values. The data processing resulted in a data set with twenty-five hospitals/actions, a_m for m=1,...,25, operating in 2017 and 2018 and according to twenty-four indicators.

4.3. Criteria tree

This subsection presents the approach and variables used to build our model's criteria tree, displayed in table 1. The tree formulation was limited by the indicators provided by our source, which was mainly related to process in Donabedian's definition. Thus, we firstly collected indicators as a start for our study and only afterwards build the tree by defining criteria and subcriteria respectively. However, the selection of the indicators considered the literature review and the quality indicators present in the IOM and AHRQ. Subsequently, the criteria were chosen considering the multidimensional quality definitions of the Donabedian and IOM. The subcriteria selection was based on indicators, since they are operationalized them. Finally, the subcriteria were grouped on families and included in the respective criteria

4.4. Criteria scales

Scales are a set of levels fully ordered. When comparing two actions according to a criterion g, one compares both scores in a scale for evaluating their respective performance towards assessing the performance of each action. It is necessary to define scales and directions of preference for each criterion, to assure valid comparisons between actions.

For subcriteria, one used quantitative continuous scales and reference levels, as the DM agreed and as indicators assumed the same nature. However, for criteria there was a problem with their multidimensionality, one of the major problems to tackle in this dissertation. None of the literature review approaches was applicable in our case to build scales combining multiple dimensions, since each criterion aggregates several subcriteria and each of them have several levels, generating boundless combinations. So we opt for an innovative approach, to define the levels for the criteria set: (1) Define the levels for all the subcriteria; (2) Execute the ELECTRE TRI-C in subcriteria families to assess the categories to actions; (3) Convert the categories assessed to each action to a level between 1 to 5 in an ordinal scale, where 1 is the minimum and 5 the maximum. As we defined five categories this conversion is direct (a C_1 represents a level 1, a C_2 represents a level 2, until C_5 that represents a level 5), unless the ELECTRE method had assessed an interval of categories. In case of that, we opt for creating a descendant view and an ascendant view. In case an action is defined by an interval, the best category assigned is included in the descendant view and the worst category assigned is included in the ascendant view. Actions assigned to only one category, present that same category in both views. This conversion was performed for the five families of subcriteria, an example is illustrated in table 2.

4.5. Criteria weighting

The interaction between criteria in ELECTRE methods is modelled through the weights of the interaction coefficients and the modifications in the concordance index. The criteria weights were established applying the SRF procedure, a software developed by Roy & Figueira (1998) into DecSpace

Table 1: Criteria, subcriteria, and corresponding indicators.

Criteria	Subcriteria	Indicator
g1, Access	g1,1: First medical appointments timeliness	Number of non-urgent first medical appointments performed in adequate time per 100 first medical appointments
	g1,2: Enrolled patients for surgery	Number of enrolled in the surgical waiting list within the mean guaranteed response time
	g1,3: Availability of beds	Occupancy rate
	g1,4: Availability of doctors	Doctors per 1000 inhabitants
	g1,5: Availability of nurses	Nurses per 1000 inhabitants
g2, Care Appropriateness	g2,1: Minor surgeries appropriateness	Number of outpatient surgeries per 100 potential outpatient procedure
	g2,2: Avoidable re-admission prior 30 days after discharge	Number of readmissions in 30 days after discharge per 100 inpatients
	g2,3: Excessive staying delay	Number of long-stay inpatients per 100 admissions
	g2,4: Hip surgery timeliness	Number of hip surgeries performed in the first 48 hours per 100 hip surgeries
	g2,5: Delay before surgery	Average waiting time before surgery
g3, Patient Safety	g3,1: Bedsores	Number of bedsores per 100 inpatients
	g3,2: Bloodstream infections related to CVC	Bloodstream infection rate related to CVC per 100 inpatients
	g3,3: Postoperative pulmonary embolisms or thrombosis	Postoperative pulmonary embolism/deep venous thrombosis cases per 100 surgical procedures
	g3,4: Postopertative septicaemia	Postoperative septicemia cases per 100 inpatients
	g3,5: Non-instrumental vaginal deliveries with severe laceration	Cases of trauma on vaginal delivery (third and fourth degree lacerations), without instrumentation, per 100 assisted deliveries
	g3,6: Assisted vaginal deliveries with severe laceration	Cases of trauma on vaginal delivery (third and fourth degree lacerations), with instrumentation, per 100 assisted deliveries
g4, Efficiency	g4,1: Expenses with staff	Expenses with staff per standard patient
	g4,2: Expenses with drugs, pharmaceutical products and clinical consumables	Expenses with drugs, pharmaceutical products and clinical consumables per standard patient
	g4,3: Expenses with supplies and external services	Expenses with supplies and external services per standard patient
	g4,4: Expenses with overtime	Expenses with overtime per total expenses with staff
	g4,5: Expenses with outsourcing	Expenses with outsourcing per total expenses with staff
g5, Caesarean Appropriateness	g5,1: Volume of caesarean sections	Number of caesarean sections per 100 deliveries
	g5,2: Caesarean sections in UCFTPs	Number of caesarean sections in UCFTPs per 100 sections in UCFTPs
	g5,3: First caesarean sections in UCFTPs	Number of first caesarean sections in UCFTPs per 100 deliveries in UCFTPs without caesarean section before

Table 2: Scales conversion from subcriteria to criteria

Action	Categories		Viewpoint	
	Min	Max	Ascendant View	Descendant View
a_1	C_3	C_4	3	4
a_3	C_1	C_2	1	2
a_4	C_4	C_4	4	4

Table 3: Weight attributed by the DM to each criterion.

Criteria	Normalized weight
g_1	23.60
g_2	28.69
g_3	31.25
g_4	3.12
g_5	13.34

(a web software). The SRF procedure allows the DM to hierarchize the different criteria of a family in a given context, using his know-how and expertise in the sector, to provide the analyst the information needed to attribute a numerical weight to each of the criteria, in table 3 it is described the weight of the criteria family, for space reasons not possible to show subcriteria weights. The SRF includes two variables, later to be used in the sensitivity analysis: the blank cards that set differences bigger than one unit between two criteria, and ratio-z a numerical value representing how many times the best criterion is more important than the worst. This procedure was repeated six times, one for each family of subcriteria and another for the criteria family.

4.6. Categories and reference actions

Hospitals can use different approaches to provide the best services, having different production technology and consume of resources. In line with the DM, one classified the hospitals in a set of five predefined categories ordered by performance regarding quality: C_1 Very poor, C_2 poor, C_3 neutral, C_4 good, and C_5 very good.

Later, the DM defined the reference value of performance per criterion and subcriterion in a given category. For the criteria the DM was able to identify different reference actions per category, allowing to apply the ELECTRE TRI-nC method. The criteria reference actions are presented in table 4. For the subcriteria, the DM identified one reference action per category, since in his opinion already represent accurately the categories. Thus, for subcriteria it was only applicable the ELECTRE TRI-C.

Table 4: Reference actions for criteria family

Category	Ref. Action	Criteria				
		g_1	g_2	g_3	g_4	g_5
C_5	b_5^1	5	5	5	5	5
	b_5^2	5	4	5	4	5
	b_5^3	5	4	5	4	4
C_4	b_4^1	4	4	5	4	5
	b_4^2	4	4	5	4	4
	b_4^3	4	4	4	4	4
C_3	b_3^1	4	4	4	3	4
C_2	b_2^1	3	3	4	3	4
	b_2^2	3	3	3	3	3
C_1	b_1^1	3	2	3	2	3
	b_1^2	2	2	2	2	3
	b_1^3	2	2	2	1	3

4.7. Thresholds and credibility level

ELECTRE methods use preference (p_j) and indifference (q_j) thresholds, to tackle imperfect knowledge. The family of criteria was described in ordinal scales, not being possible for an action to be defined by two different levels, thus no indifference and preference thresholds were applicable. However considering the veto threshold, (v_j), the DM felt relevant to include it for g_2 (($v_j = 2$)) and g_3 (($v_j = 3$) regarding their importance in quality.

For subcriteria families, being all described by quantitative scales, it is possible that an action is defined by different values that result in a same level of a scale. Thus, in line with DM's opinion the subcriteria thresholds are shown in table 5.

The credibility level is the minimum degree, λ, which is necessarily considered by the DM for validating or not an outranking statement taking in account all criteria. The credibility level is a cutting level, since it converts a fuzzy relation into a crisp outranking relation [11]. Typically, λ, takes

Table 5: Preference, indifference thresholds of the subcriteria.

Thresholds	Subcriteria					
	$g_{1,1}$	$g_{1,2}$	$g_{1,3}$	$g_{1,4}$	$g_{1,5}$	
q	2.0	2.0	2.0	0.4	0.4	
p	5.0	5.0	3.0	0.5	0.5	
	$g_{2,1}$	$g_{2,2}$	$g_{2,3}$	$g_{2,4}$	$g_{2,5}$	
q	3.0	1.0	0.3	3.0	0.2	
p	5.0	2.0	0.5	5.0	0.3	
	$g_{3,1}$	$g_{3,2}$	$g_{3,3}$	$g_{3,4}$	$g_{3,5}$	$g_{3,6}$
q	0.01	0.01	0.01	0.01	0.02	0.02
p	0.01	0.01	0.01	0.01	0.02	0.02
	$g_{4,1}$	$g_{4,2}$	$g_{4,3}$	$g_{4,4}$	$g_{4,5}$	
q	50.0	50.0	25.0	0.5	0.5	
p	100.0	100.0	50.0	1.0	1.0	
	$g_{5,1}$	$g_{5,2}$	$g_{5,3}$			
q	3.0	3.0	1.0			
p	5.0	5.0	3.0			

a value within the range $[0.5, 1[$, from a discussion with the DM it was decided to use $\lambda = 0.6$.

5. Results and discussion

Once implemented the model, it was possible to establish outranking relations, and afterwards finally assigns hospitals to categories through the assignment procedure. Note that this model was executed considering the ascendant and descendant view point, as the study was performed for 2017 and 2018, it resulted in four executions. The executions were done using the MCDA-ULaval, that provides all method's functionalities in a free software.

5.1. Outranking relation

As stated in subsection 4.1, the ELECTRE methods allow to build one or more outranking relations considering the performances of each action on each criterion. These relations enable to state whether an action a, is preferred to an action a' according to a criterion g_j. The credibility of these relations is measured by the credibility indices, σ [2]. The calculated categorical credibility indices over the potential actions and *vice versa*, for the descendant view of 2018 are presented in table 6 (not possible to display all four executions), where one can notice the presence of outranking relations justified by the presence of categorical credibility indices equal to 1. It is also observable, that most of the potential actions denote a categorical credibility equal or close to one for $\sigma(a, B_1)$, *i.e.*, the majority of the potential actions demonstrated an outranking relation over B_1, therefore that set of potential actions are at least as good as the B_1 reference action set. This observation sustains the fact that C_1 is the worst category from the existing five. The opposite happened for C_5 as expected, where the category demonstrated an outranking relation over all the potential actions, $\sigma(B_5, a) = 1, \forall a$. This sustains

the fact that C_5 is the best category from the existing five.

5.2. Exploitation of an outranking relation
Once constructed the outranking relations, it was possible to assign the actions to the corresponding categories. The assignment procedure uses the joint rules described in subsection 4.1, generating a range of possible categories $\Gamma(a)$ that follows the ELECTRE TRI-nC properties [2]: (1) Happens when an action a is neither λ-indifferent nor λ-incomparable to B_h, $h = 1, ..., q$: the result is $\Gamma(a)$ being composed of one or two consecutive categories; Happens when a is λ-indifferent to at least one subset of reference actions B_h: $\Gamma(a)$ is composed of the subset of consecutive categories defined by such λ-indifference, and, possibly, by including one or two of the adjacent categories to them; Happens when a is λ-incomparable to at least one subset of reference actions B_h: $\Gamma(a)$ is composed of the subset of consecutive categories defined by such λ-incomparability, and, possibly, by including one or two of the adjacent categories to them. The results of the assignment procedure for all the four executions are represented in table 7.

5.3. Findings
The actions in the ascendant views were assigned to equal or worst categories comparing to the descendant views, corroborating the expected from the approach created in subsection 4.4 regarding scales.

In the year of 2017 only one interval of categories, $[C_1, C_2]$, maintained the same number of actions assigned for both the views with 5 actions; and also three categories with 0 actions assigned C_3, C_4 and C_5. In the ascendant view the most represented category was the worst category, C_1, with 11 actions and 92% of the actions were assigned to a category equal or lower than C_2. In respect to the highest category assigned it was $[C_3, C_4]$ with only one action, a_2. In the descendant view the most represented category was C_2 with 10 actions, outperforming C_1 with 4 actions representing a reduction of 60% comparing to the ascendant view and 76% of the actions were assigned to a category equal or lower than C_2 representing a reduction of 17% comparing to the ascendant view. Furthermore, in the descendant view the highest category assigned to an action was in the interval of $[C_4, C_5]$ with one action assigned, again a_2.

In the year of 2018 only two categories, C_4 and C_5, maintained the same number of actions assigned for both the views with 0 actions. In the ascendant view the most represented category was the worst category, C_1, as well as category C_2 both with 7 actions assigned and 80% of the actions were assigned to a category equal or lower than C_2, in respect to the highest category assigned it was C_3 with 3 actions assigned. In the descendant view the most represented category was C_2 with 8 actions, outperforming C_1 with 4 actions assigned, a reduction of 43% comparing to the ascendant view and 60% of the actions were assigned to a category equal or lower than C_2 representing a reduction of 25% comparing to the ascendant view. Furthermore, in the descendant view the highest category assigned to an action was in the interval of $[C_4, C_5]$ with one action assigned ,a_1.

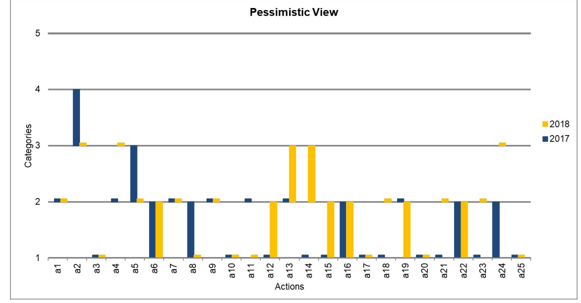

Figure 1: Assignment procedure for ascendant viewpoint in both years.

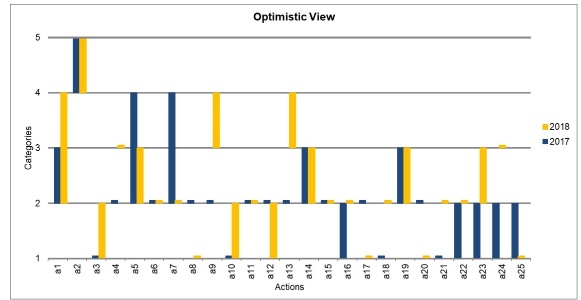

Figure 2: Assignment procedure for descendant viewpoint in both years.

Globally, there was a maintenance or an improvement of the quality in the healthcare entities between the years of 2017 and 2018, expressed when comparing both ascendant and descendant views. For a better visualization figure 1 presents a plot for the ascendant view in both years and 2, presents a plot for the descendant view in both years. Concerning all the four assignment procedures, on one hand no actions were assigned to the best category C_5 what allows us to conclude that there was no under evaluation of the reference actions that define that category; on the other hand, regarding the worst category, C_1, there were several actions assigned to it.

Looking to the actions with better categories assigned, a_2, was considered the best action for both years in the descendant view, the best in the ascendant view in 2017 and together with a_{24} the best in the ascendant view for 2018. These results show that the action a_2 is consistently assigned to the

Table 6: Categorical credibility indices calculated between the potential actions and the set of reference actions of each category for the descendant view of 2018

Actions	$\sigma(a, B_h)$					$\sigma(B_h, a)$				
	C_1	C_2	C_3	C_4	C_5	C_1	C_2	C_3	C_4	C_5
a_1	1.00	1.00	0.87	0.63	0.31	0.00	0.68	0.97	1.00	1.00
a_2	1.00	0.97	0.83	0.84	0.84	0.00	0.08	0.45	0.76	1.00
a_3	1.00	0.71	0.00	0.00	0.00	0.66	1.00	1.00	1.00	1.00
a_4	0.87	0.63	0.63	0.60	0.60	0.00	0.34	0.69	1.00	1.00
a_5	1.00	1.00	0.87	0.60	0.27	0.00	0.71	1.00	1.00	1.00
a_6	1.00	0.76	0.31	0.27	0.13	0.00	0.71	1.00	1.00	1.00
a_7	1.00	0.76	0.45	0.42	0.24	0.00	0.713	1.00	1.00	1.00
a_8	1.00	0.40	0.00	0.00	0.00	0.69	0.69	0.69	1.00	1.00
a_9	1.00	0.76	0.76	0.73	0.42	0.00	0.58	0.87	0.87	1.00
a_{10}	0.87	0.84	0.14	0.00	0.00	0.71	1.00	1.00	1.00	1.00
a_{11}	0.76	0.76	0.26	0.23	0.00	0.29	1.00	1.00	1.00	1.00
a_{12}	1.00	0.73	0.00	0.00	0.00	0.71	1.00	1.00	1.00	1.00
a_{13}	1.00	0.69	0.69	0.27	0.00	0.00	0.00	0.55	0.58	1.00
a_{14}	0.87	0.63	0.63	0.60	0.27	0.00	0.71	1.00	1.00	1.00
a_{15}	1.00	0.97	0.29	0.06	0.03	0.58	1.00	1.00	1.00	1.00
a_{16}	1.00	1.00	0.55	0.27	0.13	0.00	0.71	1.00	1.00	1.00
a_{17}	0.87	0.27	0.00	0.00	0.00	0.71	1.00	1.00	1.00	1.00
a_{18}	1.00	0.76	0.45	0.42	0.24	0.00	0.71	1.00	1.00	1.00
a_{19}	1.00	1.00	0.69	0.45	0.27	0.00	0.68	0.97	1.00	1.00
a_{20}	1.00	0.45	0.01	0.01	0.00	0.68	0.97	0.97	1.00	1.00
a_{21}	1.00	0.76	0.26	0.26	0.01	0.29	0.97	0.97	1.00	1.00
a_{22}	1.00	0.69	0.41	0.13	0.00	0.00	0.71	1.00	1.00	1.00
a_{23}	1.00	1.00	0.69	0.45	0.27	0.00	0.68	0.97	1.00	1.00
a_{24}	1.00	1.00	0.55	0.31	0.16	0.00	0.00	0.68	0.68	1.00
a_{25}	1.00	0.33	0.00	0.00	0.00	0.84	0.97	0.97	1.00	1.00

best categories in both viewpoints and years inclusively it achieves an interval of categories $[C_4, C_5]$ which is almost the maximum of the scale. Concerning action a_{24} () together with a_2 it is assigned with the best category for the ascendant view of 2018 with a C_3, having the same evaluation for the descendant view. This suggests that a_{24} is the one presenting better quality healthcare among the largest hospital centres of the NHS. Hospitals a_9 and a_{13} for the year of 2018 in the descendant view were assigned with an interval of categories $[C_3, C_4]$, still far from the results presented by a_2, however amongst the best categories attributed in this case study.

Although, it is visible that there were actions continuously being assigned to the worst categories C_1 and C_2, where the worst ones were a_{10}, a_3 and a_{17},and from the largest hospital centres a_{25} and a_{20}. Those represent the hospitals with lowest quality healthcare performances according to the case study. In overall there are others where the categories assigned are not much better than the ones assigned to the previous actions, which suggests that the quality provided by the hospitals of the SNS is low, in almost all the cases between the categories C_1 and C_2.

5.4. Sensitivity analysis

One performed an extensive sensitivity analysis, in the family of criteria assignment results, to test the robustness of the model. This analysis focused firstly on changes in the credibility index, testing the model with $\lambda = 0.55$ and $\lambda = 0.65$. There were less alterations for $\lambda = 0.65$, than for $\lambda = 0.55$, suggesting that the credibility level validated by the DM is a consistent value.

Secondly, one created three alternative scenarios to test changes in criteria's weight by inducing changes in the SRF procedure, namely in the cards order and in ratio-z. The scenarios are next described, as for ratio-z for the criteria family was $Z = 10$, so one tested the scenarios together with two variations of Z, $Z = 9$ and $Z = 11$. In scenario A one reduced to two units the distance between g_1 and g_5, comparing to the three in the original scenario. In scenario E, one added one unit of distance between g_4 and g_5, comparing to the three in the original scenario. In scenario I one added two units of distance between g_1 and g_5, comparing to the three in the original scenario and one added

Table 7: Assignment procedure for the years of 2017 and 2018 and respective viewpoints.

Actions	2017				2018			
	Asc. view		Des. view		Asc. view		Des. view	
	Min	Max.	Min	Max.	Min	Max.	Min	Max
a_1	C_2	C_2	C_2	C_3	C_2	C_2	C_2	C_4
a_2	C_3	C_4	C_4	C_5	C_3	C_3	C_4	C_5
a_3	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_2
a_4	C_2	C_2	C_2	C_2	C_3	C_3	C_3	C_3
a_5	C_2	C_3	C_2	C_4	C_2	C_2	C_2	C_3
a_6	C_1	C_2	C_2	C_2	C_1	C_2	C_2	C_2
a_7	C_2	C_2	C_2	C_4	C_2	C_2	C_2	C_2
a_8	C_1	C_2	C_2	C_2	C_1	C_1	C_1	C_1
a_9	C_2	C_2	C_2	C_2	C_2	C_2	C_3	C_4
a_{10}	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_2
a_{11}	C_2	C_2	C_2	C_2	C_1	C_1	C_2	C_2
a_{12}	C_1	C_1	C_2	C_2	C_1	C_2	C_1	C_2
a_{13}	C_2	C_2	C_2	C_2	C_2	C_3	C_3	C_4
a_{14}	C_1	C_1	C_2	C_3	C_2	C_3	C_2	C_3
a_{15}	C_1	C_1	C_2	C_2	C_1	C_2	C_2	C_2
a_{16}	C_1	C_2	C_1	C_2	C_1	C_2	C_2	C_2
a_{17}	C_1	C_1	C_2	C_2	C_1	C_1	C_1	C_1
a_{18}	C_1	C_1	C_1	C_1	C_2	C_2	C_2	C_2
a_{19}	C_2	C_2	C_2	C_3	C_1	C_2	C_2	C_3
a_{20}	C_1	C_1	C_2	C_2	C_1	C_1	C_1	C_1
a_{21}	C_1	C_1	C_1	C_1	C_2	C_2	C_2	C_2
a_{22}	C_1	C_2	C_1	C_2	C_1	C_2	C_2	C_2
a_{23}	C_1	C_1	C_1	C_2	C_2	C_2	C_2	C_3
a_{24}	C_1	C_2	C_1	C_2	C_3	C_3	C_3	C_3
a_{25}	C_1	C_1	C_1	C_2	C_1	C_1	C_1	C_1

one unit of distance between g_4 and g_5, comparing to the three in the original scenario. No alterations were registered for the alternative scenarios, however for Z=11 in the original scenario 3.5% of the assignments change suggesting the importance ratio-z assumes in attributing weight to criteria.

Finally, one tested both analysis at the same time (in a total of 96 tests) presented in table 8. From 2400 assignments generated, only 192 suffered alterations (an equivalent to 8%) considering our original model, corroborating the model's robustness.

6. Conclusions

The application of a multicriteria approach, more particularly the ELECTRE TRI-nC, to perform a quality assessment of the Portuguese public hospitals proved to be a reliable tool. The only change applied in methodology was the approach to create the family of criteria scales, that was based on the method's assignments for subcriteria. Fortunately, it displayed the expected results with the descendant view over equalling or exceeding the ascendant view and allowing us to achieve our final goal. Besides that, most of the method's inherent features demonstrated great detail to represent quality in healthcare, especially due its non-compensatory nature, the presence of thresholds and veto power, different weight criteria and the possibility of a category being characterized by more than one reference action. This features not only help to achieve a more real model, but also allow the DM to express all his knowledge to be inputted in the model.

As to limitations, the data used from the ACSS benchmarking is reliable, although there is a lack of indicators concerning outcomes. As outcomes are the ultimate validators according to Donabedian, it would be interesting to test the model with outcomes' data [7]. Also, the existence of only one DM (besides is precious help), it would be enriching to have other collaborations, introducing new opinions and scenarios, to generate new confronting results.

If there is a need of reformulating and investing in the SNS, quality must be considered to assure the needs and safety of patients, workforce and the health institutions. In future, this model could work as a guideline to create in-depth assessing from hospital funding, to physicians' evaluations. Furthermore, at a time that public-private-partnership are so criticized, the application of the model concerning data of these hospitals would be interesting, to-

Table 8: Number of alterations in the assignment of categories comparing to Scenario DM and ratio-z=10 in each of the executions when varying the scenarios and ratio-z.

	2017 Asc. view			2017 Des. view			2018 Asc. view			2018 Des. view		
Scenario	Z=9	Z=10	Z=11	Z=9	Z=10	Z=11	Z=9	Z=10	Z=11	Z=9	Z=10	Z=11
DM	0		0	0		2	0		2	0		3
A	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0

wards proving data to assess these partnerships and their existence.

To end, the MCDA application to assess the quality of the Portuguese public hospitals was achieved, validating this model and dissertation.

Acknowledgements
This dissertation was included in the hSNS FCT - Research Project (PTDC/EGEOGE/30546/2017). I would also like to mention the EIT Health MSc in Technological Innovation in Health, that cooperated in this dissertation.

Furthermore, I would like to thank my supervisors for guiding me during this dissertation, Prof. José Rui Figueira and Prof. Rui Cunha Marques. A special thanks to Alexandre Nunes, Diogo Ferreira and Miguel Pereira, the first one for providing his time and knowledge as the DM of this dissertation, the last two for helping me from the simplest to the most complex question during this work.

Finally, thank you to my family and friends for giving me love and strength to be better every day and never stop following my dreams.

References

[1] National health care in Portugal : a new opportunity. *The Lancet*, 394(10206):1298, 2019.

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

[3] P. Barros. *Economia da saúde - conceitos e comportamentos.* Coimbra (Almedina), 2013.

[4] P. Barros, J. Pereira, and J. Simões. A sustentabilidade financeira do Serviço Nacional de Saúde. Technical report, 2007.

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

[6] C. Costa, R. Santana, and P. Boto. Financiamento por capitação ajustada pelo risco : conceptualização e aplicação. *Revista Portuguesa de Saúde Pública*, pages 67–102, 2008.

[7] A. Donabedian. The Quality of Care: how can it be assessed? *Jama*, 260(12):1743–1748, 1988.

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

[9] D. Ferreira, R. Marques, and A. Nunes. Economies of scope in the health sector: The case of Portuguese hospitals. *European Journal of Operational Research*, 2018.

[10] D. Ferreira and A. Nunes. Technical efficiency of Portuguese public hospitals: A comparative analysis across the five regions of Portugal. *International Journal of Health Planning and Management*, 34:411–422, 2018.

[11] S. Greco, M. Ehrgott, and J. Figueira. *Multiple criteria decision analysis: State of the art surveys.* Springer's International Series,2005, 2016.

[12] A. Handler, M. Issel, and B. Turnock. A conceptual framework to measure performance of the public health system. *American Journal of Public Health*, 91(8):1235–1239, 2001.

[13] National Academies of Sciences, Engineering and Medicine. *Crossing the global quality chasm: Improving health care worldwide.* 2018.

[14] A. Nunes. *Reformas na gestão hospitalar: Análise dos efeitos da empresarialização.* PhD thesis, Instituto Superior de Ciências Sociais e Políticas, 2016.

[15] K. Otani, B. Waterman, K. Faulkner, S. Boslaugh, and C. Dunagan. How patient reactions to hospital care attributes affect the evaluation of overall quality of care , willingness to recommend , and willingness to return. *Journal of Healthcare Management*, 55(1):25–38, 2010.